

# **Evaluation präventiver Praxis gegen Rechtsextremismus**

Positionen, Konzepte und Erfahrungen

Michaela Glaser & Silke Schuster (Hrsg.)



Arbeits- und Forschungsstelle  
Rechtsextremismus  
und Fremdenfeindlichkeit

Die Arbeits- und Forschungsstelle Rechtsextremismus und Fremdenfeindlichkeit wird durch das Bundesministerium für Familie, Senioren, Frauen und Jugend (BMFSFJ) finanziell gefördert.

## **Danksagung**

Diese Publikation wäre ohne die Beiträge und die Unterstützung verschiedener Kolleginnen und Kollegen nicht möglich gewesen. Wir möchten uns insbesondere bei allen Autorinnen und Autoren für ihre Kooperationsbereitschaft und ihr Engagement bedanken. Peter Rieker verdanken wir Anregungen zur Konzeption des Bandes und viele hilfreiche inhaltliche Rückmeldungen; Gabi Elverich war uns durch aufmerksames Korrekturlesen und organisatorische Unterstützung eine große Hilfe. Unser besonderer Dank gilt Renate Schulze für das gründliche Lektorat und für die souveräne Bewältigung aller anfallenden organisatorischen Herausforderungen.

Wir möchten ferner allen Kolleginnen und Kollegen danken, die uns im Rahmen von Interviews und durch ihre Teilnahme an einem Workshop Einblicke in ihre Arbeit gewährten, uns an ihren Erfahrungen teilhaben ließen und mit uns zusammen Fragen und Probleme diskutierten und der DJI-Studie zur Evaluation in der Rechtsextremismusprävention auf diese Weise eine reichhaltige Grundlage boten.

© 2007 Deutsches Jugendinstitut e.V.  
Außenstelle Halle  
Abteilung Jugend und Jugendhilfe  
Arbeits- und Forschungsstelle Rechtsextremismus  
und Fremdenfeindlichkeit  
Franckeplatz 1 – Haus 12/13  
06110 Halle  
Telefon: +49 345 68178-42  
Telefax: +49 345 68178-47  
Internet: <http://www.dji.de>

Layout / Gestaltung / Gesamtherstellung  
omniphon GmbH Leipzig

ISBN: 978-3-935701-21-1

# Inhaltsverzeichnis

## **Einleitung**

*Michaela Glaser/Silke Schuster*

Evaluation präventiver Praxis gegen Rechtsextremismus –  
Besonderheiten und Herausforderungen 5

## **Reflexionen aus der Evaluationspraxis**

*Kerstin Palloks*

„Große Erwartungen“ – zur Wirkungsfrage bei der Evaluation  
von Modellprogrammen 14

*Susanne Klingelhöfer*

Das Programm „Entimon“: Spezifika, Potenziale und Herausforderungen  
einer induktiv-rekonstruierenden Evaluation anhand Logischer Modelle 32

*Miryam Eser Davolio/Matthias Drilling*

Vergleichende Evaluation von Interventionen gegen Rechtsextremismus:  
Perspektiven professionsorientierter wissengenerierender Evaluationen 53

*Michaela Glaser*

Offenheit und Multiperspektivität – Erfahrungen aus der wissenschaftlichen  
Begleitung von Projekten Interkulturellen Lernens 69

*Stephan Schack/Erich Schäfer*

Evaluation in der Demokratie- und Toleranzerziehung – Wirkungsanalyse  
zu Alltagskonflikten von Jugendlichen 86

*Bettina Fritzsche/Ralf Bohnsack*

„Leben und Arbeiten in Vielfalt?“ – Zur dokumentarischen  
Evaluation eines Xenos-Projekts 98

## **Theoretische und methodologische Reflexionen**

*Helmut Kromrey*

Begleitforschung und Evaluation – fast das Gleiche,  
und doch etwas Anderes!

113

*Christian Lüders/Karin Haubrich*

Evaluation in wenig formalisierten pädagogischen Settings

136

*Peter Rieker*

Der soziale Faktor – Ergebnisse einer explorativen Untersuchung  
zu sozialen Beziehungen in der Evaluation von Maßnahmen  
gegen Rechtsextremismus

150

*Silke Schuster*

Improvisation, Partizipation und die Frage der Wirkungen –  
Eine Untersuchung zu Evaluationen der pädagogischen Arbeit  
gegen Rechtsextremismus

169

**Autorinnen- und Autorenverzeichnis**

182

## Evaluation präventiver Praxis gegen Rechtsextremismus – Besonderheiten und Herausforderungen

Mit dem dramatischen Anstieg fremdenfeindlicher und rechtsextrem motivierter Übergriffe in den 1990er Jahren ging auch eine gesteigerte gesamtgesellschaftliche Aufmerksamkeit für aktuelle Erscheinungsformen von Rechtsextremismus und Fremdenfeindlichkeit in der bundesdeutschen Gesellschaft einher. Die Politik reagierte mit Sonderprogrammen, die pädagogische und präventive Arbeit fördern sollten (z. B. das Aktionsprogramm gegen Aggression und Gewalt von 1992 bis 1996). Seitdem wird in der pädagogischen Fachöffentlichkeit die Frage diskutiert, welchen Beitrag Bildungs- und Sozialarbeit leisten kann, um präventiv gegen fremdenfeindliche und rechtsextreme Tendenzen zu wirken.

Mit dem im Jahr 2001 von der Bundesregierung aufgelegten und Ende 2006 ausgelaufenen Aktionsprogramm „Jugend für Toleranz und Demokratie gegen Rechtsextremismus, Fremdenfeindlichkeit und Antisemitismus“ und dessen Teilprogrammen „Entimon – gemeinsam gegen Gewalt und Rechtsextremismus“, „Xenos – Leben und Arbeiten in Vielfalt“ und „Civitas – initiativ gegen Rechtsextremismus in den neuen Bundesländern“ setzten die politisch Verantwortlichen neue inhaltliche Schwerpunkte in der präventiven Arbeit gegen Rechtsextremismus und Fremdenfeindlichkeit. Mit Bildungs- und Sensibilisierungsmaßnahmen, mit Aktivitäten zur Stärkung demokratischer Kultur sowie durch eine Verknüpfung von Ausbildungs- und Arbeitsmarktaktivitäten mit Angeboten gegen Rassismus und Diskriminierung wurden neue Wege in der Arbeit gegen Rechtsextremismus und Fremdenfeindlichkeit eingeschlagen. So hat sich in den letzten Jahren eine breit gefächerte Palette von Ansätzen entwickelt, die den Anspruch, gegen Fremdenfeindlichkeit, Rechtsextremismus und Antisemitismus zu wirken, mit unterschiedlichen thematischen und methodischen Zugängen zu realisieren suchen.

Gleichzeitig zeichnet sich eine weitere Entwicklung ab: Wie auch in anderen Praxisfeldern nimmt auch in der pädagogischen Praxis gegen Rechtsextremismus und Fremdenfeindlichkeit die Forderung nach systematischer, forschungsbasierter Überprüfung und Bewertung der geleisteten Arbeit beständig zu; im Falle öffentlicher Finanzierung sind Evaluationen sogar meist unumgänglich. So kam es durch die Programme des Bundes auch zu einem deutlichen Anstieg von Evaluationen in diesem Feld, weshalb inzwischen Erfahrungen aus einer Reihe von Programm- und Projektevaluationen vorliegen (Klingelhöfer u. a. 2006; Lynen von Berg u. a. 2007; Xenos Zwischenbericht 2005; Schack u. a. 2006; Tonks 2004; Germin 2005; Müller 2004). Der fachliche Austausch zu Voraussetzungen, Möglichkeiten und Schwierigkeiten von Evaluation in diesem Bereich steht jedoch noch am Anfang.

Hier schlägt sich auch eine generelle Schwäche der bundesdeutschen Evaluationsdiskussion nieder, in der – trotz eines beträchtlichen Evaluationsbooms in den letzten Jahren – nach wie vor „fachübergreifende Sammelbände, in denen evaluatorisches Wissen

gebündelt und integriert wird (...) [weitgehend] fehlen“ (Stockmann 2004, S. 34). Dieses Defizit kennzeichnet umso mehr das Evaluationsfeld „Rechtsextremismusprävention“, im dem es bisher keine Tradition der Evaluation gab (vgl. Pingel/Rieker 2003). So wurden zu dieser Thematik auch erst zwei Sammelbände realisiert, denen der Verdienst zukommt, erste Praxiserfahrungen und Analysen versammelt zu haben (Uhl et al. 2004; Lynen v. Berg/Roth 2003). Diese decken jedoch zwangsläufig nur ein kleines Spektrum der inzwischen vorhandenen fachlichen Erfahrungen ab: Während sich der von Uhl et al. herausgegebene Band auf die Evaluation von drei im Rahmen des Projekts „Erziehung zu Demokratie und Toleranz“ umgesetzten politischen Bildungsprogrammen konzentriert, stellt der andere Sammelband erste Erfahrungen mit den Evaluationen der bis 2006 umgesetzten Bundesmodellprogramme vor. Nicht zuletzt der Abschluss dieser Programme und der damit verbundenen Evaluationsvorhaben lässt es deshalb lohnenswert erscheinen, den aktuellen Stand der Evaluationspraxis im Feld der Rechtsextremismus- und Fremdenfeindlichkeitsprävention zu beleuchten und zu diskutieren.

Die vorliegende Publikation will einen Beitrag hierzu leisten, indem sie Praxiserfahrungen und theoretische Reflexionen zu den Anforderungen und Möglichkeiten in diesem Evaluationsfeld versammelt. Zu diesem Zweck haben wir einschlägige Wissenschaftlerinnen und Wissenschaftler aus der Evaluationspraxis aufgefordert, ihre Erfahrungen mit der Evaluation präventiver Ansätze gegen Rechtsextremismus und Fremdenfeindlichkeit, sich hierbei als geeignet erweisende Konzepte und Methoden, aber auch dabei auftretende Schwierigkeiten vorzustellen. Ein besonderes Augenmerk sollte dabei der Frage gelten, durch welche Spezifika sich Evaluationen im Feld der Prävention von Rechtsextremismus und Fremdenfeindlichkeit auszeichnen und welche besonderen Herausforderungen damit für die Arbeit in diesem Feld verbunden sind. Dabei interessierte uns insbesondere auch wie sich Evaluierende in diesem Kontext zur Frage des Wirkungsnachweises positionieren und welche Antworten und Lösungen sie im Umgang mit dieser Frage finden.

Ergänzt werden diese Reflexionen aus der Evaluationspraxis durch eine Reihe von Beiträgen, die die dort aufgeworfenen Fragen aus einer grundsätzlichen, projektübergreifenden Perspektive diskutieren. Neben Positionen aus der theoretischen Evaluationsdiskussion werden hier auch Ergebnisse einer Studie präsentiert, in der Erfahrungen mit Evaluationen in der Rechtsextremismusprävention erhoben und analysiert wurden.

Das in diesem Band vertretene Spektrum von Programm<sup>1</sup>- und Einzelevaluationen sowie von praxisorientierten und theoretischen Reflexionen präsentiert eine große Spannweite von Perspektiven auf dieses Forschungs- und Arbeitsfeld. Bei aller Verschiedenheit der hier vertretenen wissenschaftlichen Traditionen, der praktischen Zugänge und der dabei gemachten Erfahrungen, werden in den Beiträgen auch wesentliche Gemeinsamkeiten erkennbar, die die Evaluation präventiver Angebote gegen Rechtsextremismus und Fremdenfeindlichkeit kennzeichnen:

---

1 Wir verwenden diesen Begriff zur Bezeichnung von Evaluationen öffentlich geförderter Großprogramme und nicht in Anlehnung an die angloamerikanische Fachdebatte zur „program evaluation“.

So sehen sich Evaluationen in diesem Feld einerseits mit besonderen Anforderungen konfrontiert, die aus der inhaltlichen Spezifik ihres Gegenstandes resultieren: Das Thema „Rechtsextremismus und Fremdenfeindlichkeit“ bzw. die Förderung demokratischer und toleranter Orientierungen ist in der Bildungsarbeit kein Thema wie jedes andere. Vielmehr handelt es sich um ein emotional und normativ hoch aufgeladenes Lernfeld, in dem sich auch Evaluierende mit ausgeprägten Sensibilitäten, Polarisierungen und Abwehrhaltungen konfrontiert sehen (vgl. Rieker i. d. Bd.) und das auch mit besonderen normativen Ansprüchen an die Gestaltung von Evaluationen verknüpft ist (vgl. Schack/Schäfer i. d. Bd.). Im Fall staatlich finanzierter Großprogramme gegen Rechtsextremismus kommt hinzu, dass Notwendigkeit, Zielstellung und Strategien solcher Programme stets auch Bestandteil politischer Positionierungen und dementsprechend umstritten sind. Solche Programme stehen häufig im Focus der medialen und politischen Aufmerksamkeit, was auch die Evaluation dieser Aktivitäten maßgeblich prägt. Aus diesen feldspezifischen Voraussetzungen und Gegebenheiten resultieren nicht nur erhöhte Anforderungen an Qualifikation und Kompetenzen von Evaluierenden (vgl. Schuster und Rieker i. d. Bd); sie gehen auch mit einem besonderen „Erfolgsdruck“ hinsichtlich der Ergebnisse dieser Arbeit einher. So sind in diesem Themenfeld die Erwartungen von Auftraggeberseite und interessierter Öffentlichkeit, belastbare Daten zur Wirksamkeit der Arbeit zu erhalten, häufig besonders hoch (vgl. Palloks i. d. Bd.).

Gleichzeitig ist dieses Evaluationsfeld aber auch durch Herausforderungen gekennzeichnet, die vor allem der (pädagogischen) Form seines Gegenstandes geschuldet sind. So haben wir es in diesem Handlungsfeld überwiegend mit sogenannten „persuasiven Programmen“ (Stöss 2003) zu tun, d.h. mit – zumeist im Kontext der Jugend und Sozialarbeit angesiedelten – Bildungs-, Sensibilisierungs- und Beratungsmaßnahmen, welche auf die Veränderung von Einstellungen und Verhaltensweisen ihrer Zielgruppen gerichtet sind. Der Verlauf wie auch die Resultate solcher Angebote hängen, wie stets bei pädagogischen Angeboten, in hohem Maße von der Kooperation und den jeweiligen Voraussetzungen ihrer Adressatinnen und Adressaten ab (vgl. Glaser und Fritzsche/Bohnsack i. d. Bd.); zudem handelt es sich hier um pädagogische Formate, in denen Elemente informellen Lernens eine große Rolle spielen und die vergleichsweise gering strukturiert sind (vgl. Lüders/Haubrich i. d. Bd.). Solche Konstellationen sind zwangsläufig mit großen Unwägbarkeiten verbunden, was die Feststellung von Wirkungen – ein in sozialen Kontexten ohnehin sehr anspruchsvolles Unterfangen – zusätzlich erschwert (vgl. Ahlheim 2005; Haubrich et al. 2005; Uhl et al. 2004).

Die in dieser Publikation diskutierten Evaluationsansätze und -methoden dokumentieren unterschiedliche Möglichkeiten, diesen Spezifika des Feldes und den damit verbundenen, z.T. auch divergierenden Anforderungen Rechnung zu tragen. Dabei fällt auf, dass vor allem qualitative Designs gewählt werden bzw. dort, wo quantitative Verfahren zum Einsatz kommen, diese gezielt durch qualitative Erhebungen ergänzt werden. Insofern liefern die hier versammelten Beiträge Hinweise dafür, dass sich im Evaluationsfeld „Rechtsextremismusprävention“ qualitative Verfahren der Sozialforschung als besonders gegenstandsadäquat erweisen.

Dass es in diesem Feld keine endgültige Wirkungsgewissheit geben kann, darin sind sich die meisten hier vertretenen Autorinnen und Autoren einig. Die Frage, in welcher Hinsicht und mit welcher Reichweite Aussagen zu Wirkungen präventiver Angebote gegen Rechtsextremismus und Fremdenfeindlichkeit möglich sind, wird dabei jedoch durchaus unterschiedlich beurteilt. Allerdings wird auch von den Vertreterinnen und Vertretern einer diesbezüglich eher skeptischen Sichtweise die „Wirkungsfrage“ keineswegs ausgeklammert. Bei allen erkenntnistheoretischen, methodischen und feldbedingten Grenzen und Schwierigkeiten, die in den Beiträgen thematisiert werden, wird vielmehr ein Bestreben erkennbar, sich der Frage der Wirkungen in diesem Feld zumindest anzunähern und hierbei mögliche und hilfreiche Verfahren und Vorgehensweisen auszuloten.

### **Zu den Beiträgen im Einzelnen:**

Eingeleitet wird der Band durch zwei Beiträge, die Erfahrungen aus der Evaluation öffentlich geförderter Modellprogramme gegen Rechtsextremismus thematisieren. Die Autorinnen, die jeweils an einer dieser Programmevaluationen beteiligt waren, stellen zwei unterschiedliche Ansätze zur Evaluation solcher modellhafter Großprogramme vor.

Im Zentrum der Ausführungen von *Kerstin Palloks* steht das häufig konflikträchtige Verhältnis zwischen Erkenntnis- und Legitimationsanspruch von Evaluation. Am Beispiel des Bundesprogramms „Civitas – initiativ gegen Rechtsextremismus in den neuen Bundesländern“ schildert sie den hohen Legitimationsdruck, unter dem staatlich initiierte Programme in diesem Themenfeld stehen und zeigt auf, welche Schwierigkeiten sich daraus für Evaluationen ergeben können. Sodann stellt sie das vom Civitas-Team entwickelte Modell der „kommunalen Kontextanalysen“ vor, mit dem die Forscherinnen und Forscher den divergierenden Erwartungen nach (kritischem) Erkenntnisgewinn, Optimierungs- und Legitimationsfunktion Rechnung zu tragen suchten. Projektbezogene Analysen zur Qualität von Arbeitsprozessen wurden mit einer „soziologischen Tiefenbohrung“ verbunden, die ausgewählte Projekte in ihren spezifischen sozialen Kontexten untersuchte und auf diese Weise die Bedingungen des Handelns gemeinwesenbezogener Präventionsprojekte in den Blick nahm.

Der folgende Beitrag von *Susanne Klingelhöfer*, die an der Wissenschaftlichen Begleitung des Programms „Entimon – gemeinsam gegen Gewalt und Rechtsextremismus“ mitwirkte, diskutiert die Potenziale und Herausforderungen einer induktiv-rekonstruierenden Evaluation mithilfe Logischer Modelle. Die Ausführungen von S. Klingelhöfer stellen die praktische Umsetzung dieses Evaluationsprojekts und die dabei gewonnenen Erfahrungen in den Mittelpunkt und leiten daraus Empfehlungen für zukünftige Evaluationen in diesem Feld ab. In der Entimon-Evaluation wurde ein zweifacher Nutzen des Instruments „Logische Modelle“ deutlich: Zum einen ermöglichte es, Programmstrategien zu identifizieren, deren Potenziale und Grenzen zu beschreiben und diese Resultate für die Übertragung in andere Kontexte fruchtbar zu machen. Durch die Beteiligung der Projekte bei der Erstellung der Logischen Modelle konnten diese zudem bei der Erreichung ihrer Ziele unterstützt werden, indem Selbstreflexions- und Klärungsprozesse initiiert und strukturierend begleitet wurden.

Im Anschluss daran werden Design und Methodik von zwei Untersuchungen vorgestellt, in denen jeweils mehrere unterschiedliche Maßnahmen zur Prävention von Rechtsextremismus bzw. Fremdenfeindlichkeit wissenschaftlich begleitet wurden. Während sich im ersten Fall der Blick stärker auf Wirkungsweisen einzelner Maßnahmen richtete, galt das Erkenntnisinteresse der im zweiten Beitrag diskutierten Studie vor allem den Praxisbedingungen von Projekten und den Voraussetzungen ihrer Wirkfähigkeit.

*Miryam Eser Davolio* und *Mathias Drilling* schildern Möglichkeiten der professionsorientierten Wissensgenerierung am Beispiel einer Studie zu sieben Interventionen der Gemeinwesenarbeit gegen manifesten Rechtsextremismus in der Schweiz. Ein Kriterium bei der Auswahl der untersuchten Interventionen war deren Begleitung durch externe Rechtsextremismusexpertinnen und -experten. Diese wurden neben den lokalen Akteuren befragt, um Informationen zu Wirkungsweisen der Maßnahmen zu erhalten und auf dieser Basis Empfehlungen zur Konzipierung und Durchführung zukünftiger Interventionen für die Fachpraxis zu erarbeiten. Die Autorin und der Autor teilen die verbreitete Skepsis hinsichtlich der Möglichkeit von Wirkungsbeobachtungen nicht; sie kommen vielmehr zu dem Schluss, dass Feststellungen von Wirkungen und Nachhaltigkeit dann möglich sind, wenn Interventionen über einen längeren Zeitraum umfassend und vergleichend untersucht werden.

*Michaela Glaser* schildert in ihrem Beitrag Forschungserfahrungen aus einer qualitativen Untersuchung zu Angeboten Interkulturellen Lernens, in denen Umsetzungsbedingungen sowie Voraussetzungen und Elemente gelingender Praxis in diesem Handlungsfeld am Beispiel von 22 Projekten untersucht wurden. Die Forscherinnen und Forscher wählten für ihre Studie einen offenen, in der Theorietradition der Grounded Theory stehenden multi-methodischen Forschungsansatz, dessen Stärken und Schwächen die Autorin bilanzierend diskutiert. Schwerpunkt ihrer Ausführungen bilden dabei die Möglichkeiten, aber auch erkenntnistheoretischen Grenzen, die sich aus der Triangulation subjektiver Perspektiven für die Analyse von Projektprozessen ergeben. Diese als „Annäherung an komplexe Projektrealitäten“ begreifend, plädiert sie für die verstärkte Realisierung von Evaluationsdesigns, die ein multiperspektivisches Vorgehen und insbesondere die verstärkte Einbeziehung der Adressatenebene ermöglichen.

Den Abschluss der Reflexionen aus der Evaluationspraxis machen zwei Beiträge, in denen jeweils Forschungserfahrungen mit der Evaluation eines einzelnen Projektvorhabens reflektiert werden. In beiden Fällen handelte es sich um Bildungsmaßnahmen, die auf die Förderung toleranter und demokratischer Einstellungen und Verhaltensweisen bei jungen Menschen zielten.

*Stephan Schack* und *Erich Schäfer* stellen ein Evaluationsprojekt vor, in dem ein mehrjähriges, an verschiedenen ostdeutschen Bildungseinrichtungen umgesetztes Trainingsprogramm zur Demokratie- und Toleranzerziehung begleitend untersucht wurde. Um die Wirkungen des Programms zu analysieren, untersuchte das Forscherteam die während des Seminarzeitraums auftretenden Veränderungen in Konfliktwahrnehmung und -verhalten der Teilnehmer/innen. Als besondere Stärke des Designs erwies sich hierbei die

Kombination aus quantitativen und qualitativen Daten, die eine differenzierte, kontextualisierte Interpretation von Veränderungen ermöglichte. Ausgehend von der Überzeugung, dass gegenstandsadäquate Evaluation in diesem Themenfeld auch einen demokratischen Umgang mit den Beteiligten erfordert, bemühten sich die Forscher zudem um einen umfassend partizipativen Ansatz. Dieser Anspruch ließ sich allerdings nur bedingt realisieren; vor dem Hintergrund ihrer Erfahrungen argumentieren die Autoren deshalb für ein differenzierteres Verständnis von Partizipation, das den Restriktionen des Feldes wie auch den realen Partizipationsinteressen der Akteure Rechnung trägt.

*Bettina Fritzsche* und *Ralf Bohnsack* diskutieren in ihrem Projektbeitrag, inwiefern die dokumentarische Evaluation ein Mittel zur Bewertung von pädagogischen Maßnahmen darstellt. Anhand einer mehrjährigen Begleitung eines Projekts zur Toleranzförderung mit lernbehinderten und benachteiligten Jugendlichen zeigen sie auf, wie die Rekonstruktion des handlungsleitenden Erfahrungswissens und der geteilten Orientierungen der Teilnehmer/innen zu Bewertungen über und zu Empfehlungen für die pädagogische Praxis führen kann. Diesem Verfahren sprechen die Verfasser/innen eine zweifache fachliche Bedeutung zu: Zum einen werden nicht die subjektiven Theorien der Beteiligten, sondern deren handlungspraktisches, vorreflexives Wissen zum Ausgangspunkt der Beurteilung und Bewertung gemacht. Damit wird ein methodologischer Versuch unternommen, Evaluation als eigenständigen Forschungstyp zu entwickeln. Zum anderen wird die Perspektive der Adressatinnen und Adressaten zum zentralen Ausgangspunkt der Evaluation gemacht und somit deren Position im Evaluationsdiskurs gestärkt.

Die theoretischen und methodischen Reflexionen im zweiten Teil des Bandes verhandeln zentrale Fragestellungen, die sich im Kontext von Evaluationen präventiver Praxis gegen Rechtsextremismus stellen. In den ersten beiden Beiträgen kommen Expertinnen und Experten aus der wissenschaftlichen Evaluationsdebatte zu Wort, die einige grundsätzliche Herausforderungen an Evaluationen in diesem Arbeitsfeld thematisieren. Dabei handelt es sich um Voraussetzungen und Spezifika, mit denen sich Evaluationen sozialer Sachverhalte bzw. im Kontext der Jugendarbeit und Sozialarbeit generell konfrontiert sehen und die somit auch das Evaluationsfeld „Rechtsextremismusprävention“ maßgeblich prägen.

Aus einer dezidiert wissenschaftstheoretischen Perspektive diskutiert *Helmut Kromrey* Ähnlichkeiten und Verschiedenheiten von Begleitforschung und Evaluation. Gemeinsamkeiten sieht er vor allem in Abgrenzung zur „verwertungsneutralen“ Grundlagenforschung, resultierend aus der Anwendungsorientierung beider Tätigkeiten, die den Forschenden u.a. eine Vereinfachung der Untersuchungssituation zu Forschungszwecken verbietet. Als entscheidenden Unterschied beider Aufgaben arbeitet er den für Evaluationen kennzeichnenden Vorgang des Bewertens heraus, der mit einer dem Paradigma der Werturteilsfreiheit verpflichteten Forschung nicht zu vereinbaren sei. Kromrey kommt zu dem Schluss, dass das „Werturteilsdilemma“ in einer Verbindung der beiden Ansätze letztlich nicht aufzuheben ist. Daraus ergibt sich für den Autor die Notwendigkeit, Forschung – als Erhebung empirisch abgesicherter Informationen – und Evaluation – als Formulierung darauf gestützter Qualitätsaussagen – voneinander zu trennen.

Aus einer ganz anderen, nämlich gegenstandsorientierten Perspektive nähern sich *Christian Lüders* und *Karin Haubrich* dem Gegenstand „Evaluationsforschung“ an. Ihr Ausgangspunkt ist die Feststellung, dass bestimmte pädagogische Formate, die sich auch häufig in der Präventionsarbeit gegen Rechtsextremismus finden, deutlich weniger formalisiert sind als z. B. Formen schulischen Lernens. Damit sind zusätzliche Herausforderungen an Evaluationen verbunden; u. a. zeichnen sich diese Settings dadurch aus, dass es hier bisher kaum fundiertes Wissen über erfolgsversprechende Prozess- und Strukturmerkmale gibt. Um dennoch zu einer Theorie des evaluierten Gegenstandes – als Voraussetzung seiner Evaluierbarkeit – zu gelangen, bietet sich deren empirische Rekonstruktion mithilfe Logischer Modelle an. Eine zukünftige Herausforderung sehen die Verfasser/innen u. a. in der Erweiterung solcher Verfahren durch die Entwicklung qualitativer Zugänge, die eine Ausweitung der bisher primär auf professionellen Selbstauskünften basierenden Datenbasis ermöglichen.

Den Band schließen zwei Beiträge ab, in denen Ergebnisse einer am Deutschen Jugendinstitut durchgeführten Studie zu Evaluationen im Feld der Prävention von Rechtsextremismus und Fremdenfeindlichkeit vorgestellt werden. In einem dreistufigen Verfahren, das eine Dokumentenanalyse, Interviews und einen Expertenworkshop umfasste, wurden 15 Evaluationen auf angewandte Evaluationsansätze, deren Voraussetzungen und die damit gewonnenen Praxiserfahrungen untersucht.

*Peter Rieker* analysiert in seinem Beitrag die Bedeutung sozialer Beziehungen im Rahmen von Evaluationen. Auf der Basis von Interviewauskünften Evaluierender unternimmt er eine systematische Differenzierung möglicher Beziehungskonstellationen in Evaluationskontexten. Er arbeitet verschiedene Faktoren heraus, die die Qualität dieser Beziehungen beeinflussen können und leitet daraus Empfehlungen für die Ausgestaltung von Evaluationen ab. Riekers Ausführungen verdeutlichen, dass soziale Beziehungen in der Evaluation eine entscheidende Rolle für deren Gelingen bzw. Misslingen spielen. Besondere Brisanz kommt diesen sozialen Einflüssen und sensiblen Beziehungsnetzen im heiklen Feld der Rechtsextremismusprävention zu, weshalb sich hier besonders hohe Anforderungen an kommunikative und soziale Kompetenzen von Evaluierenden stellen.

Im Anschluss daran berichtet *Silke Schuster* über Ergebnisse zu konzeptionellen Aspekten, Formen der Wirkungsbeobachtung sowie partizipativen Ansätzen der untersuchten Evaluationen. Die Studie macht deutlich, dass Evaluierende in diesem Themenfeld sich nicht so sehr an bestehenden Evaluationsmodellen orientieren, sondern eigene Ansätze zusammenstellen und dabei vorrangig auf qualitative Methoden setzen. Ein weiteres Kennzeichen des Feldes ist die häufig intransparente Darstellung von Bewertungsprozessen in Evaluationsstudien, die eine Einschätzung zur Qualität und Reichweite von Evaluationsergebnissen erschwert. Angesichts häufig widersprüchlicher Anforderungen an Evaluierende, die sich nicht selten in unrealistischen, „überladenen“ Evaluationsdesigns niederschlagen, plädiert die Autorin für die Etablierung wissenschaftlicher Standards zur Umsetzung von Evaluationen in diesem Feld. Da Evaluation präventiver Arbeit gegen Rechtsextremismus nach wie vor ein Bereich des Ausprobierens und Suchens ist, bedarf es zudem eines ver-

stärkten wissenschaftlichen Austauschs, um eine „Kultur der Evaluation“ zu fördern. Mit den hier versammelten praxisorientierten und theoretischen Reflexionen hoffen wir, einen Beitrag zu diesem fachlichen Austausch zu leisten und Anregungen für die weitere Entwicklung des Forschungs- und Arbeitsfeldes „Evaluation präventiver Arbeit gegen Rechtsextremismus und Fremdenfeindlichkeit“ zu geben.

## Literatur:

- Ahlheim, Klaus (2005): Wirkungsoptimismus und Messbarkeits skepsis – Evaluation und Wirkungsforschung in der politischen Erwachsenenbildung. In: Praxis Politische Bildung, H. 1, S. 17–24
- Bundesministerium für Wirtschaft und Arbeit (2005): Evaluation des Bundesprogramms Xenos – Zwischenbericht 2005, [www.xenos-de.de/Xenos/Redaktion/Medien/Anlagen/2006-06-22-zwischenbericht-2005](http://www.xenos-de.de/Xenos/Redaktion/Medien/Anlagen/2006-06-22-zwischenbericht-2005) (13.11.2007)
- Germin, Cathrin (2005): Evaluation des Xenos-Projektes Verständnis der Kulturen (Schuhfabrik Ahlen), [www.xenos-ahlen.de](http://www.xenos-ahlen.de) (13.11.2007)
- Haubrich, Karin/Holthusen, Bernd/Struhkamp, Gerlinde (2005): Evaluation – einige Sortierungen zu einem schillernden Begriff. In: DJI Bulletin Sonderteil, H. 72, S. 1–4
- Klingelhöfer, Susanne/Schmidt, Mareike/Schuster, Silke/Brüggemann, Ulrich (2006): Ergebnisbericht zu den qualitativen Analysen des Programms „Entimon – gemeinsam gegen Gewalt und Rechtsextremismus“, 2002 bis 2006. Unveröffentlichter Sachstandsbericht. Halle
- Lynen von Berg, Heinz/Roth, Roland (Hrsg.) (2003): Maßnahmen und Programme gegen Rechtsextremismus wissenschaftlich begleitet. Aufgaben, Konzepte und Erfahrungen. Opladen: Leske + Budrich
- Lynen von Berg, Heinz/Palloks, Kerstin/Steil, Armin (2007): Interventionsfeld Gemeinwesen. Evaluation zivilgesellschaftlicher Strategien gegen Rechtsextremismus. Weinheim/München: Juventa
- Müller, Stefan (2004): Evaluation des Projektes der IG-Metall und der Otto-Benner Stiftung – Das Herz in die Hand nehmen, Courage zeigen. Ein gewerkschaftliches Projekt gegen Fremdenfeindlichkeit und für Interkulturalität. [download.herz-und-courage.de/evaluierung/Evaluationsbericht-Endfassung.pdf](http://download.herz-und-courage.de/evaluierung/Evaluationsbericht-Endfassung.pdf), (13.11.2007)
- Pingel, Andrea/Rieker, Peter (2003): Evaluation pädagogischer Praxis gegen Rechtsextremismus: Potenziale und Chancen. In: Lynen von Berg, Heinz/Roth, Roland (Hrsg.) Maßnahmen und Programme gegen Rechtsextremismus wissenschaftlich begleitet. Aufgaben, Konzepte und Erfahrungen. Opladen: Leske + Budrich, S. 103–117
- Schäfer, Erich/Schack, Stephan/Rahn, Peter/Uhl, Sandra (2006): „Wer sich selbst versteht, versteht auch andere besser.“ Eine Längsschnittstudie zu Wirkungen eines Projektes der politischen Jugendbildung zum Demokratie-Lernen. Jena: Edition Paideia

- Stockmann, Reinhardt (2004): Evaluationsforschung in Deutschland. In: ders. Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder. Leske + Budrich, S. 13–45
- Stöss, Richard (2003): Möglichkeiten und Grenzen der Evaluierung von persuasiven Programmen. In: Lynen von Berg, Heinz/Roth, Roland (Hg.) (2003): Maßnahmen und Programme gegen Rechtsextremismus wissenschaftlich begleitet. Aufgaben, Konzepte und Erfahrungen. Opladen: Leske + Budrich, S. 95–102
- Uhl, Katrin/Ulrich, Susanne/Wenzel, Florian M. (Hrsg.) (2004): Evaluation politischer Bildung. Ist Wirkung messbar? Gütersloh: Verlag Bertelsmann Stiftung
- Uhl, Katrin (2004): Evaluation von Demokratie- und Toleranzerziehungsprogrammen – die Erfahrungen aus dem Projekt „Erziehung zu Demokratie und Toleranz“. In: Uhl, Katrin et al. (Hrsg.): Evaluation politischer Bildung. Ist Wirkung messbar? Gütersloh: Verlag Bertelsmann Stiftung, S. 51–55
- Tonks, Iris (2004): Abschlussbericht der Evaluation KICK Im KOPF – Vielfalt in Deutschland – EU-Projekt der Fachhochschule Düsseldorf, Forschungsschwerpunkt Rechtsextremismus und Neonazismus. Duisburger Institut für Sprach- und Sozialforschung (DISS), [www.diss-duisburg.de/Arbeitsbereiche/Kick%20im%20Kopf%20Abschlussbericht.pdf](http://www.diss-duisburg.de/Arbeitsbereiche/Kick%20im%20Kopf%20Abschlussbericht.pdf), (25.11.2007)

## „Große Erwartungen“ – zur Wirkungsfrage bei der Evaluation von Modellprogrammen

Der vorliegende Artikel widmet sich dem Problem des Wirkungsnachweises bei der Evaluation von Modellprogrammen. Am Beispiel der wissenschaftlichen Begleitung des Sonderprogramms „CIVITAS – initiativ gegen Rechtsextremismus in den neuen Bundesländern“<sup>1</sup> werden sowohl die Bedarfe von Auftraggebern als auch die forschungsstrategischen Schwierigkeiten der Evaluation beleuchtet, die sich mit der Erwartung von Wirksamkeitsnachweisen verbinden können. Im Fokus der Betrachtungen steht dabei das Verhältnis zwischen Erkenntnis- und Legitimationsfunktion von Evaluation. Im zweiten Teil des Artikels wird die Evaluationsstrategie der „Kommunalen Kontextanalysen“ als Modell zur Einschätzung von Projekt- und Programmaktivitäten im komplexen Feld der Förderung kommunaler Demokratisierungsprozesse vorgestellt<sup>2</sup>. Der Beitrag schließt mit einer resümierenden Einschätzung der Potenziale und Grenzen dieser neuen Evaluationsstrategie und mit Überlegungen zur Anlage von Evaluationen bei experimentierenden Großprogrammen.

### Evaluation im Spannungsfeld der Erwartungen – das Beispiel des CIVITAS-Programms

In der öffentlichen Wahrnehmung galt das CIVITAS-Programm, mit dem die Bundesregierung in den Jahren 2001 bis 2006 eine Vielzahl unterschiedlicher Projekte in den neuen Bundesländern förderte, als konzeptionelles Gegenstück zu den bis dahin erprobten Interventionsstrategien, der Sozialarbeit mit rechtsextrem orientierten Jugendlichen im Rahmen des Aktionsprogramms gegen Aggression und Gewalt (AgAG). Die Leitziele des nachfolgenden CIVITAS-Programms gaben demgegenüber der gesellschaftlichen Auseinandersetzung mit rechtsextremen und fremdenfeindlichen Erscheinungen Vorrang. So förderte CIVITAS vor allem junge Initiativen, Kreativ- und Bildungsangebote unterschiedlichster Ansätze und Laufzeiten, mit dem Ziel, Impulse zu geben, lokales Engagement gegen Rechtsextremismus für Demokratie zu unterstützen und Alternativen für Jugendliche zu entwickeln. CIVITAS war mit einem Programmvolumen von 52,1 Millionen Euro ausgestattet und förderte insgesamt 1.580 vor allem lokale Projekte (BMFSFJ 2006). Das Kernstück des Programms waren die langfristig geförderten Strukturprojekte (Beratung von Opfern rechtsextremer Gewalt, Mobile Beratungsteams und Netzwerkstellen). Ihre

- 1 Das CIVITAS-Programm wurde von Herbst 2002 bis Frühjahr 2007 durch eine Forschungsgruppe des Instituts für interdisziplinäre Konflikt- und Gewaltforschung (IKG) der Universität Bielefeld extern evaluiert. Auftraggeber der Untersuchung war das Bundesministerium für Familie, Senioren, Frauen und Jugend, das auch Programmträger war. Von Juli 2001 bis Dezember 2001 war das CIVITAS-Programm bereits von der Alice-Salomon-Fachhochschule in Berlin evaluiert worden. Zu den Forschungsergebnissen vgl. Rommelspacher/Polat/Wilpert (2003).
- 2 Ausführliche Beschreibungen des Forschungsdesigns und aller Ergebnisse der Programmevaluierung finden sich in: Lynen von Berg/Palloks/Steil (2007).

Aufgabe war es, durch Beratung und Vernetzung lokaler Akteure dauerhafte Strukturen im Gemeinwesen zu entwickeln, also zunächst wichtige Voraussetzungen für den Aufbau einer demokratischen Zivilgesellschaft zu schaffen. Damit sollten die lokalen Akteure selbst befähigt werden, sich souverän mit Demokratie gefährdenden Phänomenen im Gemeinwesen auseinanderzusetzen. Die Entwicklung zivilgesellschaftlicher Strukturen wurde darum zum allgemeinen Leitziel der Strukturprojekte wie auch des ganzen CIVITAS-Programms erklärt.

CIVITAS war ein Modellprogramm, bei dem bewusst auch konzeptionell unausgereifte Projektansätze erprobt werden sollten<sup>3</sup>. Evaluationen spielen bei Modellprogrammen eine besondere Rolle. Die geförderten Ansätze sind hinsichtlich ihrer Erfolgspotenziale und ihrer Übertragbarkeit in andere Kontexte einzuschätzen, bevor eine mögliche Überführung in die Regelförderung erwogen wird. Von Evaluationen wird also zunächst erwartet, Aussagen zur Umsetzung der Modellprojekte zu erarbeiten. Dann geht es darum, den Programmplanern im laufenden Förderprozess eine Grundlage zur weiteren Programmgestaltung zu erarbeiten und schließlich auch darum, im politischen und öffentlichen Raum eine Argumentationsgrundlage – z. B. für die Weiterführung eines Programms oder von bestimmten Ansätzen – vorzuhalten. Besonders die letztgenannte Anforderung wiederum ist eng an die Frage nach der Wirksamkeit des Programms gekoppelt. Gerade bei teuren Großprogrammen, die politisch unter Legitimationsdruck stehen, werden dabei hohe Erwartungen an die Aussagekraft von Evaluationen gerichtet. Jene Erwartungen bündeln sich in einer bestimmten Vorstellung von Erfolg. Aus der Sicht von Politik und Verwaltung besteht dieser in hoher, wissenschaftlich nachgewiesener Wirkung der Maßnahmen. Der Bedarf an Wirkungsanalysen resultiert also nicht nur aus dem erwähnten Modellcharakter der Programme. Wirkungsnachweise gelten bei Programmen, die in der politischen Öffentlichkeit umstritten sind, als Teil der erforderlichen Legitimationsbeschaffung; sie sind damit generell nicht ohne Brisanz.

Das CIVITAS-Programm war ein solches Programm. Es stand, mehr als die parallel laufenden „Schwesterprogramme“ Entimon und Xenos, im Interesse der Öffentlichkeit. Der programmatische zivilgesellschaftliche Ansatz wurde von unterschiedlichen Akteuren nicht einfach als ein weiterer Zugang zum Rechtsextremismusproblem betrachtet, sondern im Rahmen des ausgerufenen „Paradigmenwechsels“ in der Auseinandersetzung mit Rechtsextremismus und Fremdenfeindlichkeit mit erheblichen Vorschusslorbeeren versehen. Seine Implementierung erfolgte unmittelbar nach den in den Medien hoch emotional geführten Debatten um rechtsextreme und fremdenfeindliche Gewalt, die im Sommer 2000 in den vom damaligen Bundeskanzler ausgerufenen „Aufstand der Anständigen“ mündeten. Damit wurde das Programm in der öffentlichen Wahrnehmung vor allem mit den politischen Konzepten der damaligen Regierungskoalition verbunden. Nicht zuletzt weil der zivilgesellschaftliche Ansatz auch über die Mobilisierungspotenziale Auskunft zu geben vermochte, die über den Leitsatz „Kampf gegen Rechts“ aufzustellen waren, wurde das Programm häufig zum

---

<sup>3</sup> Zum Terminus „Modellprogramm“ und den damit verbundenen Anforderungen an Projekte und Evaluationen vgl. Haubrich u.a. (2007).

Gegenstand parlamentarischer wie außerparlamentarischer Debatten. Der Leitsatz ließ sich nicht nur programmatisch auf das Problemfeld Rechtsextremismus beziehen, sondern transportierte auch die politischen Gegnerschaften des allgemeinen politischen Rechts-Links-Schemas. Die Wirkungsannahmen in Bezug auf das CIVITAS-Programm waren somit auch ein Medium der üblichen parlamentarischen Positionskämpfe<sup>4</sup>. Die in diesem Zusammenhang ausgesprochenen Erfolgs- oder Misserfolgsunterstellungen wurden weit über das bei anderen staatlichen Programmen bekannte Maß hinaus von der Presse aufgegriffen und weitergeführt. Es gab eine kontinuierliche Erfolgswachsamkeit in der Politik, der Öffentlichkeit und der Fachwissenschaft, denn die Frage nach der Wirksamkeit dieses staatlichen Interventionspaketes reichte über das Interesse an den tatsächlichen Programmeffekten hinaus. Sie galt ebenso dem zivilgesellschaftlichen Ansatz wie seinen politischen Vertretern.

Diese Konstellation bildete die Rahmung der Programmumsetzung wie auch der Programmevaluation. Sie war mit ausschlaggebend für den Auftrag des Wirkungsnachweises, aus dem sich für das Evaluationsteam interessante Möglichkeiten, aber auch nicht wenige Schwierigkeiten ergaben. Einige der für die Evaluationsstrategie relevanten Aspekte von Konstellation und Auftrag sollen im Folgenden dargestellt werden.

### **Wirkungsannahmen beim CIVITAS-Programm in der politischen Debatte**

Die Zielvorgaben und damit die unterstellten Wirkungsdimensionen bei politisch ungesicherten Programmen sind umkämpft. Die Arenen liegen häufig außerhalb der Einflussbereiche der Wissenschaftlerinnen und Wissenschaftler und auch jener der Programmakteure. Am Beispiel des CIVITAS-Programms ließ sich erfahren, dass auch die politische Öffentlichkeit zu einem gewissen Teil bestimmte, was als Leitziel und damit als Maßstab für die Erfolgsbeurteilung der Programmaktivitäten zu gelten habe. Dies war etwa der Fall, als zur aktuellen Fragestunde des Bundestages zu den Programmen CIVITAS, Entimon und Xenos festgestellt wurde: „[...] dass sich trotz zusätzlicher eigener Landesprogramme die Zahl der rechtsextremistischen Straftaten im letzten Jahr sogar verdoppelt habe“ (Deutscher Bundestag 2003, S. 2628) und damit nahe gelegt wurde, dass der gültige Erfolgsindikator für die geförderten Maßnahmen im nachweisbaren Rückgang rechtsextremer Straftaten zu sehen ist. Das ist für ein Programm, das sich vor allem der Förderung von Demokratisierungsprozessen widmet, ein eher unrealistisches und darum unangemessenes Erfolgskriterium. Der Bundesrechnungshof stellte ebenfalls eigene Erfolgskriterien auf. Nach Einsichtnahme in offizielle Programmdokumente gelangte man zu folgendem Urteil: „Nicht effizient genug – Rechnungshof fordert Stopp der Programme gegen Neonazis“ titelte u. a. die Tageszeitung „Der Tagesspiegel“ (09.03.2003). Begründet wurde diese Forderung mit

---

4 „Die heutige hohe diskursive Bedeutung des Themas Rechtsextremismus verdankt sich neben den aktuellen Ereignissen dem kulturellen Gedächtnis an den Nationalsozialismus. Wie alle Themen, deren hohe Relevanz in der öffentlichen Kommunikation selbst unbefragt bleibt, ist auch der Rechtsextremismus einer Art Metanutzung unterworfen. Das Thema Rechtsextremismus ist auch Medium für themenun-spezifische Diskurse und Ort von Konfliktaustragungen“ (Klärner/Kohlstruck 2006, S. 9).

Zweifeln an einem „vernünftigen Verhältnis zwischen Kosten und Nutzen, mangelnder Modellhaftigkeit und dem Ausbleiben überregionaler Wirkungen“ der Programme.

Diese von Öffentlichkeit, Politik und Verwaltung postulierten Erfolgs- oder vielmehr Misserfolgskriterien beanspruchen eigene Gültigkeit, auch ohne wissenschaftliche Fundierung. Für die Programme und die Programmverantwortlichen bleiben sie nicht ohne Folgen. Diese müssen auf die ausgesprochenen Urteile reagieren, denn für die politischen Gegner bieten solche Bewertungen eine Gelegenheit, den laufenden Programmen die Existenzberechtigung abzusprechen. Im Falle des CIVITAS-Programms war in der Presse zu lesen, die Maßnahmen der Bundesregierung seien wirkungslos, da die NPD dennoch den Einzug in den Schweriner Landtag geschafft habe. Anhand solcher Beispiele lässt sich nachvollziehen, warum positive Wirkungsnachweise so gefragt sind. Sie deuten gleichzeitig an, vor welche Probleme sich Evaluationen angesichts dieser Erwartungshaltung gestellt sehen.

### **Programmevaluation und Legitimationsbedarfe**

Die Bestandsgefährdungen von „Legislatur-Programmen“ können dazu führen, dass der Argumentationsdruck der politisch Verantwortlichen wächst. Für Evaluationen wiederum kann sich daraus das Problem eines funktionellen Widerspruchs ergeben: Evaluationen haben stets mehrere Funktionen. Legitimationsbeschaffung gilt als eine von ihnen. Daneben stehen jedoch ihre Erkenntnis- und Optimierungsaufgaben. Damit sind auch kritische Befunde möglich, auf deren Basis eben Optimierung laufender Praxis erfolgen kann. Vor dem skizzierten Hintergrund politisch umstrittener Programme geraten diese Funktionen fast zwangsläufig in Konflikt miteinander. In der Situation eines grundsätzlichen Legitimationsdefizits werden Evaluationen mit der Erwartung konfrontiert, die Erkenntnis- und Optimierungsfunktion zugunsten der Legitimationsbeschaffung zumindest zeitweise in den Hintergrund zu stellen (vgl. Dietzel/Troschke 1988, S. 87).

Beim CIVITAS-Programm gestaltete sich dieses Spannungsfeld folgendermaßen: Das von der rot-grünen Bundesregierung aufgelegte Programm war besonders in seiner Anfangszeit in der Öffentlichkeit umstritten. In Bezug auf die Evaluation gab es darum die Befürchtung, dass kritische Befunde als ein Nachweis fehlender oder unerwünschter Wirkungen interpretiert werden könnten und dass das Urteil der Evaluation aus der Hand genommen werden könnte. Dass diese Gefahr tatsächlich bestand, machten die Reaktionen auf eine Studie über die drei Bundesprogramme deutlich, die im Auftrag der Friedrich-Ebert-Stiftung durchgeführt und im gleichen Zeitraum veröffentlicht wurde (vgl. Roth/Benack 2003). Die eher auf Optimierung einer für gut befundenen Programmidee abzielende kritische Studie führte zu einer überhitzten politischen und von den Medien aufgegriffenen Debatte über einen frühzeitigen Abbruch der Bundesprogramme aufgrund unterstellter Wirkungslosigkeit. Diese negativen Erfahrungen mit den Folgen der Programmkritik verfestigten sich zeitweise, so dass sich auch das Evaluationsteam CIVITAS mit entsprechenden Befürchtungen konfrontiert sah, die sich in erhöhten Ansprüchen an den erwarteten Legitimationsgehalt der Evaluation kanalisierten.

Wirkungsannahmen sind also immer auch Munition im umkämpften Terrain, auch unabhängig davon, wer sie produziert. Dieser Umstand kann wiederum mit Folgen für Evaluationen verbunden sein. Sie müssen einen Umgang mit den widersprüchlichen Anforderungen finden. Sie sollen dem Evaluationsauftrag entsprechen, der beides vorsieht: Erkenntnisse – auch kritische – generieren und gleichzeitig das Programm durch Erfolgsmeldungen absichern. Hier besteht die Gefahr, dass „[...] Evaluation instrumentalisiert und für nicht pädagogische Zwecke vereinnahmt wird. In diesem Falle verliert das Projekt Evaluation seine eigentliche aufklärerische Funktion und wird zum bloßen Ritual, das im Gewande der Wissenschaftlichkeit und der Objektivität auftritt oder vorgeschoben wird“ (vgl. Grohmann 1997, S. 266). Diese Überlegung lässt sich als Problem von Auftragsforschung und speziell von Evaluationsforschung verallgemeinern, wenn es in einen gesamtgesellschaftlichen Entwicklungszusammenhang gestellt wird.

### **Exkurs: Evaluation im Kontext von Modernisierung**

Die Erwartungen an Programme und Programmevaluationen haben sich in der Bundesrepublik seit der Zeit der groß angelegten Reformprogramme der 1960er und 1970er-Jahre verändert. Seit Anfang der 1980er-Jahre haben sich Evaluationen im Rahmen eines sich wandelnden Politikverständnisses stärker mit der Frage der Legitimation staatlicher Interventionen und weniger mit Planungs- und Optimierungsaufgaben auseinander zu setzen (vgl. Grohmann 1997, S. 185; Haubrich u.a. 2007). Die Möglichkeit einer illegitimen Funktionalisierung lässt sich auch als ein Entwicklungsaspekt von Modernisierung verstehen. Evaluation im „Kontext der Risikogesellschaft“ (Grohmann 1997, S. 286) kann nicht mehr mit den vermeintlichen Gewissheiten der Reformära rechnen. In der von Beck (1986) entworfenen Gesellschaftsdiagnose werden soziale Probleme nicht mehr als Ordnungsprobleme, sondern als „Risikoprobleme“ verstanden, was bedeutet, dass weder die Ursachen noch die Lösungen gesellschaftlicher Probleme eindeutig zu bestimmen sind. Die Wissensbestände über Wesen, Verbreitung und Lösungsmöglichkeiten sozialer Probleme sind mehrdeutig und sogar ambivalent. Man hat es eher mit Wahrscheinlichkeiten als mit Gewissheiten zu tun, Produktionsinstanzen entscheidungsrelevanten Wissens multiplizieren sich und Deutungen treten in Konkurrenz zueinander. Beide Subsysteme, Politik und Wissenschaft, finden sich in den „Zwängen des gesellschaftlichen Risikomanagements“ (Beck 1986, S. 273) wieder. „Wo die Wirklichkeit als Korrektiv sich in Entscheidungen und Konventionen verabschiedet, beginnt die gesellschaftliche Verwendung (mit)zubestimmen, was als Erkenntnis gilt und was nicht. Der Ort der Kontrolle und die Art der Kriterien verschieben sich: von innen nach außen, von der Methodologie zur Politik, von der Theorie zur gesellschaftlichen Akzeptanz“ (ebd.). Für das Evaluationsthema impliziert dies zweierlei:

1. In Bezug auf das Legitimationsproblem: Evaluation intervenierender Praxis unter den Bedingungen der Risikogesellschaft gerät zunehmend in den Blick des öffentlichen und politischen Interesses (vgl. ebd., S. 289). Der relevante Beitrag und damit auch

die forschungsstrategische Ausrichtung von Evaluation staatlicher Interventionspraxis verschieben sich von der Optimierung der Interventionspraxis hin zur Absicherung laufender Programme. Benötigt wird eine Vergewisserung über die Gültigkeit der zugrunde liegenden Handlungskonzepte durch die wissenschaftliche Referenz ihrer Wirksamkeit (vgl. ebd., 276).

2. In Bezug auf diese Vergewisserung durch Wissenschaft und die damit verbundene Frage nach Wirkungsnachweisen: Die konkurrierenden Konstruktionen sozialer Probleme und daraus resultierende Unsicherheiten haben zur Folge, dass eindeutige Zuordnungen von sozialen Problemen und deren Bedeutung, sozialen Problemen und deren Ursachen sowie sozialen Problemen und denkbaren Lösungsansätzen sehr viel weniger konsensuell akzeptiert werden als in der „einfachen Moderne“. Gerade bei Gegenstandsbe-  
reichen wie Rechtsextremismus oder Fremdenfeindlichkeit wird dies besonders deutlich (vgl. Grohmann 1997, S. 289; Kliche 1996, S. 57ff.). Grohmann spricht diesbezüglich von dem Problem schwacher Kausalitäten (vgl. Grohmann 1997, S. 289). Wenn einfache Kausalitäten nicht mehr zeitgemäß sind, verändert sich das Evaluationsverständnis. Evaluationen haben nun die Aufgabe, in der Diffusität Orientierung zu geben, Ambivalenzen aushaltbar zu machen, die gestiegene Komplexität einerseits zu reduzieren und ihr andererseits mit angemessenen Evaluationsdesigns zu entsprechen, ohne mögliche Wirkungen oder gar die Bewertung gänzlich aus dem Blick zu verlieren.

Damit ist der Auftrag des Wirkungsnachweises schließlich unter methodologischen Gesichtspunkten zu beleuchten – an seine Ausführung sind Bedingungen geknüpft, die mit der Legitimationsbeschaffung verknüpft sein können.

### **Methodische Hürden des Wirkungsnachweises beim CIVITAS-Programm**

Streng genommen bezeichnet Wirkung stets ein direktes Verhältnis zwischen einer strategisch geplanten Intervention und dem daraus resultierenden Ertrag. Um dieses Verhältnis exakt zu beschreiben, bedarf es gewisser Voraussetzungen: „Wenn Evaluation im Sinne eines Soll-Ist-Vergleichs erfolgen soll, dann sind die Ziele, die Realität werden sollen, die Handlungskonzepte, mittels derer diese Ziele erreicht werden sollen und die Kriterien für die Wertung von Erfolg oder Misserfolg präzise zu benennen. Nur dann können Wirkungen, können Zweck-Mittel-Relationen – nach Möglichkeit auch quantitativ – festgemacht, kann Effektivität und Effizienz beurteilt werden“ (Heil 2001, S. 24). Wie oben beschrieben, ist eine direkte kausale Verkettung von Zielen und Ergebnissen bzw. Ursachen und Wirkungen im Bereich (sozial)pädagogischer Handlungsansätze der Demokratieentwicklung problematisch und wird in der Fachdebatte auch entsprechend kritisch gesehen<sup>5</sup>.

---

5 Vgl. u.a. Stöss (2003) zu Evaluationen von „Überzeugungsprogrammen“, Haubrich/Lüders (2007) zur Evaluation experimentierender Modellprogramme, Kelle (2006) zum Kausalitätsparadigma in der qualitativen Evaluationsforschung und Frindte/Preiser (2007) zu den Bedingungen von Wirkungsanalysen bei Maßnahmen in der Prävention von Rechtsextremismus und Fremdenfeindlichkeit.

Der Versuch einer Wirkungsevaluation des CIVITAS-Programms, vor allem im experimentellen oder quasi-experimentellen Design, wäre zum Scheitern verurteilt gewesen. Wie ließe sich der gewünschte Wirksamkeitsnachweis erbringen, wenn es um hochkomplexe, offene und kaum steuerbare Prozesse (Beratung, Vernetzung, Kooperationsweisen) in vielen verschiedenen Gemeinwesen geht? Die Qualität von Arbeitsprozessen zu kennen reicht nicht aus, um deren Wirkungen zu verallgemeinern. Auch gute Projektarbeit kann zu unerwünschten Effekten beitragen; die externen Einflüsse auf die Ergebnisse der Projektarbeit sind zu komplex und selbst nicht zu isolieren. Ein verbindliches theoriegeleitetes Wirkungsmodell (vgl. Kromrey 1988), mit dem gemessene Wirkungen – z. B. lokale Veränderungsprozesse – tatsächlich einer bestimmten Intervention zugeordnet werden könnten, liegt für den Bereich kommunaler Demokratisierungsprozesse nicht vor.

Beim CIVITAS-Programm zeigte sich bspw., dass wesentliche Kontingenzannahmen, von denen das Programm ausging, weder empirisch ohne weiteres begründbar schienen, noch intern widerspruchsfrei waren: Bedeutet z. B. mehr Zivilgesellschaft wirklich auch weniger Rechtsextremismus bzw. mehr Zivilgesellschaft auch gleich mehr Demokratie? Welches Ziel hat Vorrang – die Zurückdrängung rechtsextremer Phänomene oder der Aufbau zivilgesellschaftlicher Strukturen? Wie verhält sich ein politischer Bekämpfungsauftrag zum Auftrag professioneller Beratung kommunaler Akteure? An dieser Stelle verkoppelt sich das Methoden- mit dem Legitimitätsproblem: Das Programm selbst hatte während der Programmaufzeit mit dem Leitziel „Zivilgesellschaft stärken“ bereits einen Operationalisierungsschritt vollzogen, in dem es die unmittelbaren Programm- und Projektziele weitgehend vom Rechtsextremismus entkoppelt hatte. Der programmatische Begriff „Zivilgesellschaft“ war allerdings ebenso wenig angemessen operationalisiert wie die Problemfelder Rechtsextremismus und Fremdenfeindlichkeit (vgl. Lynen von Berg u.a. 2007, S. 325). Mit dieser Unsicherheit in Bezug auf den programmatischen Bezugspunkt war CIVITAS zusätzlich angreifbar. Durch unterschiedliche Deutungen zur Art des Problemzusammenhangs und der zu erreichenden Mittler- und Handlungsziele bis zu der Frage nach der Angemessenheit von Gegenstrategien ergaben sich Kontroversen, die die verschiedenen Ebenen der Programmverwaltung bis zu den Projektakteuren durchzogen. Die verschiedenen Beteiligtengruppen (Programmverantwortliche, Projektmitarbeitende und Wissenschaftlerinnen und Wissenschaftler), aber auch Fachöffentlichkeit und Medien konkurrierten um Deutungs- und Definitions Kompetenzen, die überdies mit spezifischen Eigeninteressen verknüpft waren. Was also wäre der gültige Indikator für den Programmerfolg, wenn sich wesentliche Zielstellungen schon kaum definieren lassen?

Vor dem Hintergrund der verschiedenen Erläuterungsebenen ist die Aufforderung zum Wirksamkeitsnachweis für Evaluationen „exemplarischer Feldexperimente“ (Haubrich u.a. 2007, S. 184) somit doppelt brisant: Sie ist methodisch und politisch heikel. Anspruch und Machbarkeit klaffen deutlich auseinander. Evaluatorinnen und Evaluatoren haben nun verschiedene Möglichkeiten, sich zu diesem Dilemma zu verhalten: Man kann die Diskussion um den Wirkungsbegriff ignorieren und dem Problem auf der semantischen Ebene begegnen, indem man Wirkungen nennt, was sich als zeitlich spätere Ereignisse in einem losen

Zusammenhang mit einer Intervention darstellen lässt<sup>6</sup>. Eine andere Möglichkeit besteht darin, den Auftraggeber und seine Bedarfe zu ignorieren und zu evaluieren, was sich zur Evaluation anbietet. Dies wäre fraglos unredlich und darum nicht zu empfehlen. Oder aber man verlegt sich auf das Aushandeln von Alternativen und entwickelt ein Evaluationsdesign, mit dem das wissenschaftlich Vertretbare dem politisch Gewünschten zumindest angenähert werden kann.

### **Evaluationsansätze beim CIVITAS-Programm**

Die Strategien der CIVITAS-Evaluation stehen für die dritte genannte Alternative. Im Prozess ihrer Planung stellte sich – wie stets bei multizentrischen Programmen – zunächst das Problem der hohen Fallzahl und der schwierigen Vergleichbarkeit der Fälle. Der Evaluation eröffnete sich ein unübersichtliches und äußerst weites Feld (über 1000 Projekte während der Programmlaufzeit) unterschiedlichster Projektarten, Strategien und Ziele, meist kaum operationalisiert und mit einem stark überdehnten Zielgruppenspektrum (die „Zivilgesellschaft“, kommunalpolitische Akteure, professionelle Akteure aus Jugend- und Sozialarbeit). Wie der großen Unterschiedlichkeit Herr werden?

In dieser Situation erfolgte eine Präzisierung des Untersuchungsgegenstandes durch den Auftraggeber: Zunächst sollten Aussagen zum Ende der Modellphase der oben genannten Strukturprojekte erarbeitet werden. Gegenstand der Forschung war der Stand ihrer Professionalisierung. Die Opferberatungen, Mobilen Beratungsteams und Netzwerkstellen (zum Zeitpunkt der Erhebung im Jahr 2003 insgesamt 40 Projekte) wurden deshalb durch eine Totalerhebung mit qualitativen Interviews untersucht. Analysiert wurden die Selbstausskünfte der Projektmitarbeiter/innen – ihre Erfolgsdefinitionen, Strategien und ihre Einschätzungen von Projekterfolgen. Aus Ressourcengründen musste die Ebene der direkten Beobachtung und der situativen Kontexte in dieser Untersuchungsphase noch weitgehend ausgeblendet werden. Mit dieser Untersuchung gelang es aber, die Kerntätigkeiten der Strukturprojekte näher zu beschreiben und die eigenen Erfolgs- und Misserfolgskonzepte der Projektakteure zu erforschen, um auf dieser Basis der Praxis angemessene Kriterien zur Einschätzung von Prozessqualitäten zu erarbeiten.

Ein anderer Analyseteil war die jährliche Dokumentation (quantitative Analyse) von Projektansätzen und Zielgruppen auf der Grundlage der CIVITAS-Projektdatenbank, in der die antragsrelevanten Daten der geförderten Projekte verzeichnet waren. Sie blieb auf eine Systematisierung von Absichtserklärungen beschränkt. Dennoch wurden hier aber die

---

6 Roland Roth plädiert bspw. für einen solchen pragmatischen Umgang mit Kausalitätsproblemen: „Niemand im Feld glaubt an strikte Kausalität, aber doch an mögliche und wahrscheinliche Wirkungszusammenhänge, die vielleicht mit den nächsten Forschungsergebnissen revidiert werden müssen. (...) Evaluationsergebnisse müssen m. E. nicht in einer höheren Spielklasse erzielt werden, sondern es kann nützlich sein, in einigen Dimensionen gezielte Rückmeldungen aus dem Feld zu organisieren, um zu prüfen, ob sich die vermuteten und erhofften Wirkungen eingestellt haben“ (Roth 2004, S. 8). In diesem Verständnis kann von Wirkungen also bereits dann gesprochen werden, wenn die Befragten glauben, dass etwas wirksam ist, somit über die Zuschreibung von Kausalität eine Wirksamkeit unterstellt wird.

wichtigsten steuerungsrelevanten Informationen in der Breite gewonnen (Welche Zielgruppen werden anvisiert, welche Projektformate werden in welchem Umfang gefördert und inwieweit sind diese geeignet, welche Zielsetzungen des Programms umzusetzen?).

Mit der Beschreibung von Prozessqualitäten (Strukturprojekte) und der Analyse der Förderpraxis (Datenbankanalyse) konnten erste wichtige Erkenntnisse über Ansätze und deren Qualitäten auf der Projektebene und über Optimierungsmöglichkeiten bei der Förderstrategie gewonnen werden. Es fehlte jedoch immer noch der bilanzierende Analyseteil, der 2004, zwei Jahre vor dem absehbaren Ende des Programms, erwartet wurde. Nicht nur Prozess-, sondern auch Ergebnisqualitäten sollten bestimmt werden. Nur so schien das Programm politisch legitimierbar. Dem Forscherteam wie auch den Auftraggebern war bewusst, dass das „Summative“ der Programm-Evaluierung nicht über die Breite aller Projekte zu generieren war. So erhielt die Evaluation die Chance, sich an einer soziologischen Tiefenbohrung zu versuchen. Es entstand das Konzept der „Kommunalen Kontextanalysen“, durch die eine längerfristige Begleitung und genaue Untersuchung von Projektaktivitäten in ihren spezifischen Kontexten möglich wurde.

### **Evaluation als Feldforschung: Die Kommunalen Kontextanalysen**

Hinsichtlich der oben beschriebenen Anforderungen an Evaluationen im „Risiko-Kontext“ lässt sich das Projekt der Kommunalen Kontextanalysen einführen als ein Versuch, der Diffusität und Komplexität der Untersuchungsgegenstände gerecht zu werden und dennoch Orientierung durch Bewertungen zu ermöglichen. Um Prozesse und ihre Bedingungen verstehen und einordnen zu können, muss Projektarbeit im spezifischen sozialen Kontext erfahren werden, auf den sie sich bezieht. Denn der Erfolg gemeinwesenorientierter Projektarbeit lässt sich letztlich nur im Gemeinwesen beurteilen, das Gegenstand der Intervention und gleichzeitig Handlungsfeld der Akteure ist. Über die Fragen nach der Qualität von Projektansätzen hinaus sollte untersucht werden, was genau vor Ort passiert, warum es passiert und wie das Geschehene mit den handelnden Projektakteuren zusammenhing bzw. wie es überhaupt einzuschätzen war. Für die Untersuchungen wurden drei Kommunen im ländlichkleinstädtischen, peripheren Raum ausgewählt, die für zwei Jahre zum Forschungsfeld der Evaluation wurden<sup>7</sup>. Die Kommunalen Kontextanalysen bargen die Chance, herauszufinden, was einzelne Projektformate mit hoher Konzept- und Prozessqualität in diesem Interventionsfeld erreichen können. Das Ziel bestand darin, Erfolgchancen verschiedener Formate in ihrem sozialen Umfeld zu eruieren. Damit verlagerte sich das Forschungsinteresse ein Stück weit von den Projekten und ihren Konzeptionen hin

---

<sup>7</sup> Der Auswahl ging eine umfangreiche Explorationsphase voraus. Aus 14 Kommunen wurden drei Kleinstädte in zwei Bundesländern ausgewählt. Auswahlkriterien waren u.a. die Struktur der Projektlandschaft, die Qualität von Projektansätzen (anhand der in der ersten Evaluationsphase erarbeiteten Maßstäbe), die absehbare Kontinuität der Projektarbeit und die Größe der Kommune, nur Gemeinden unter 20.000 Einwohnerinnen/Einwohnern kamen für die Untersuchung infrage, da sonst das Forschungsfeld zu unübersichtlich geworden wäre.

zu den Bedingungen ihres Handlungsfeldes. In dieser Perspektive wird die Kommune als soziales System betrachtet, das die Projektarbeit in zweifacher Weise rahmt: Durch ihre objektiven Bedingungen und durch die Deutungen der kommunalen Akteure. Den objektiven Bedingungsrahmen kennzeichnen materielle Rahmenbedingungen, Sozialstruktur, politische Kräfteverhältnisse und regionale Tradierungen. Dem Deutungsrahmen entspricht das Bild der Kommune, das die Projektakteure und kommunale Akteure entwerfen. Sie deuten ihre Interaktionen im interpretativen Rückschluss auf die Bedingungen des kommunalen Raums, legen darauf abgestimmte Ziele und Erwartungen fest, bestimmen Erfolgs- und Misserfolgskriterien und nehmen strategische Änderungen vor. Die Projektakteure interpretieren wiederum die Motive, Erwartungen und Erwartungserwartungen der kommunalen Akteure und planen so ihr Handeln.

Die Rahmung des Projekthandelns wurde auf mehreren Ebenen untersucht: Die Akteursanalyse beschrieb die Projektarbeit aus der Perspektive strategisch handelnder Akteure, also der Projektmitarbeiterinnen und -mitarbeiter). Die Interaktionsanalyse beschrieb die Interaktionen zwischen den Projektakteuren und ihren Adressaten bzw. anderen Schlüsselakteuren in der Kommune. Die Strukturanalyse interpretierte die Interaktionen und deren Resultate im Rahmen der kommunalen Strukturmerkmale, die zu einem Sozialprofil verdichtet wurden. Aus den Strukturmerkmalen der Sozialprofile ergaben sich wiederum evaluierungsrelevante Fragen: „In welcher Weise werden die Interaktionen und ihre Deutungen durch objektive Faktoren beeinflusst – Größe und Struktur der Gemeinde, sozialstrukturelle und demographische Entwicklungen, Mentalitäten und regionale Tradierungen? Welche (projekt- und programmrelevanten) Problemstrukturen lassen sich beobachten? Sind Projekte und Programm situationsgerecht darauf bezogen? Welche Bedarfe bleiben offen?“ (Lynen von Berg u.a. 2007, S. 357).

Die Analysen hatten den Interaktionsradius der Projekte zum Gegenstand (Wer gehört zum Kreis der Kooperationspartner?), das Interaktionsniveau (Unter welchen Bedingungen sind Kooperationen gelungen bzw. nicht gelungen? Wie sind Konflikte zu begreifen und zu bewerten?) sowie die Resonanzen der Projekte bzw. der Projektaktivitäten, an denen CIVITAS-Projekte beteiligt waren (die Institutionalisierungsformen bürgerschaftlichen Engagements, z.B. Runde Tische, Fachnetzwerke etc.). Die vorgestellten drei Analyseebenen von Akteurs-, Interaktions- und Strukturanalyse wurden nicht als separat abzuhandelnde Untersuchungsschritte vollzogen. In der Datensammlung verschränkten sich vielmehr die Ebenen des „Faktischen“ (objektive Daten) und des „Subjektiven“ (Deutungen der Akteure) in der Interpretation der Forschenden. Das methodische Instrumentarium umfasste Befragungen, Beobachtungen, rekonstruktive Inhaltsanalysen und die Analyse objektivierter Daten. Interaktionssequenzen und ihre Resultate, also die Wechselwirkung zwischen den Projektakteuren und ihren Adressaten ließen sich entweder direkt beobachten oder durch Befragungen und Dokumentenanalysen aus unterschiedlichen Perspektiven rekonstruieren.

Die Befragungen begannen bei den Projektmitarbeitenden und wurden sukzessive auf die kommunalen Schlüsselakteure ausgeweitet. Von Kooperationspartnern und weiteren lokalen Akteuren wurden Referenzen eingeholt und mit den Interpretationen der Projekt-

mitarbeiterinnen und -mitarbeiter ins Verhältnis gesetzt. Schließlich wurden auch solche lokalen Akteure einbezogen, die zwar nicht in Kooperationen mit den CIVITAS-Projekten eingebunden waren, aber aufgrund ihrer Relevanz für die zu bewältigenden Aufgaben bedeutsam waren (z. B. Jugendarbeiterinnen und -arbeiter oder kommunalpolitische Akteure). Ihre Einschätzungen über die lokale Situation, vor allem aber die Gründe für nicht gelungene Kontaktaufnahme oder Kooperationen bzw. ihre Bedarfseinschätzungen boten wichtige Anhaltspunkte für die Einordnung der Projektstrategien. Die Spiegelungen verschiedener Akteursperspektiven ergaben Anhaltspunkte für die im jeweiligen Fall bedeutsamen Strukturbedingungen und damit begründete Einschätzungen über die Angemessenheit der Interventionen im Verhältnis zu den lokalen Problem- und Bedarfslagen. Hier stellte sich jedoch ein Problem: Welches Urteil ist zulässig, wenn die Einschätzungen der Interaktionspartner bzw. der Projektbeteiligten über die Art des Problemzusammenhangs, die Projekterfolge, eine Projektstrategie oder Konflikte nicht übereinstimmen? Für diesen Fall erwies sich eine Klassifizierung der Interaktionsbeteiligten anhand ihrer strategischen Ziele und der habituellen Dispositionen<sup>8</sup> als hilfreich. Unterscheiden ließen sich Interaktionspartner auf der Basis

- geteilter Ziele und habitueller Übereinstimmung,
- divergierender Ziele und habitueller Übereinstimmung,
- geteilter Ziele und habitueller Differenz,
- unterschiedlicher Ziele und habitueller Differenz.

Auf der Basis dieser Klassifizierung ließen sich nun sowohl die Handlungsstrategien als auch die Interaktionen konkreter beschreiben; es war auch genauer zu bestimmen, wo und wann bestimmte Vorgehensweisen (etwa konflikt- oder konsensorientierte Strategien) im Rahmen des jeweiligen Ansatzes sinnvoll und erfolgreich sein können. Ein Beispiel: Das negative Urteil, das ein Sozialdezernent über ein lokales Netzwerkprojekt aussprach, ließ sich relativieren, nachdem rekonstruiert werden konnte, dass die Interessen der Verwaltung (in Bezug auf die Funktion des Netzwerkes als unabhängige zivilgesellschaftliche Assoziation) faktisch nicht mit denen der Netzwerk-Aktiven in Einklang waren (divergierende Ziele bei habitueller Übereinstimmung). In einem anderen Fall ließ sich rekonstruieren, dass vermeintliche inhaltliche Konflikte in Bezug auf unterschiedliche Ziele bei der Unterstützung von Opfern rechtsextremer Gewalt zwischen einer Opferberatung und anderen zivilgesellschaftlichen Akteuren keineswegs auf divergierende Ziele, sondern vielmehr auf habituelle Differenz (Milieufixierung der Berater/innen) und einem damit verbundenen Unwillen zur Kooperation gründeten (geteilte Ziele und habituelle Differenz). Der von lokalen Akteuren attestierte Misserfolg der Beratung ließ sich damit auf einen nicht problemadäquaten Arbeitsansatz zurückführen.

<sup>8</sup> Habituelle Gemeinsamkeiten zeigen sich dort, wo Akteure Ähnlichkeiten im Lebensstil, in weltanschaulichen und politischen Überzeugungen als kommunikativ relevant betrachten. Habituelle Differenzen zeigen sich dort, wo die Kommunikationen durch Unterschiede solcher Art strukturiert werden.

Zusammenfassend lässt sich das Wechselspiel der Analyseebenen bei der Kommunalen Kontextanalyse wie folgt beschreiben: Das strategische Handeln der Projektakteure wird auf die beiden Analyseebenen „Interaktionen mit lokalen Schlüsselakteuren“ und „Strukturmerkmale der Kommune bzw. des ländlichen Raums“ bezogen. Untersucht werden die spezifischen Möglichkeiten, Problemfelder und Bedarfe sowie Interaktions- und Institutionalisierungsformen der Kommunen im ländlichen Raum. Die Analyse der Struktur wird wiederum auf die Beobachtungen zum Projekthandeln zurückbezogen. Am einzelnen Projekt wird somit Symptomatisches erkennbar, das wiederum exemplarische Aussagen zu den Realisierungsbedingungen und -chancen des CIVITAS-Programms in diesem für die Programmintention zentralen Handlungsfeld der Förderung zivilgesellschaftlicher Strukturen zulässt.

### **Einschätzung der Evaluationsstrategie „Kommunale Kontextanalyse“**

Was lässt sich mit dem Evaluationsansatz der Kommunalen Kontextanalysen erreichen und in welchem Verhältnis stehen die Erkenntnisse dieser Forschungsstrategie und des allgemeinen Evaluationsdesigns zu dem beschriebenen Wirkungs- bzw. „Bilanzierungsproblem“? Das Konzept der kommunalen Kontextanalysen ermöglicht zunächst die Bestimmung der Erfolge und Misserfolge von Projekten im spezifischen Fall. Es bezieht diese Wertungen zugleich auf den Bedingungsrahmen der Kommune zurück, so dass ein Horizont von Möglichkeiten umrissen wird, der Chancen und Grenzen der Projektarbeit erkennbar macht. Erst im lokalen Kontext lassen sich taugliche Bewertungsinstrumente entwickeln, an denen der Erfolg von Projekten ablesbar ist. Die Entwicklung von Erfolgsmaßstäben und Erfolgsbewertungen darf allerdings nicht mit einer Wirkungsanalyse verwechselt werden. Das mechanistische Schema von Einfluss und Wirkung ist mit der Untersuchungsperspektive einer kommunalen Kontextanalyse nicht vereinbar. „Wirkungen“ und „Folgen“ unterscheiden sich danach wie streng sie in eindeutig-lineare Kausalitätsrelationen eingebunden werden. Mit „Folgen“ beschreibt man ein Geschehen, das nicht unmittelbar auf strategisches Handeln rückführbar ist. Die untersuchten Projekte haben einen beschreibbaren Anteil an diesem Geschehen, ihr strategisches Handeln ist jedoch nicht als allein ursächlich zu begreifen.

Als Deutungs-, Beziehungs- und Strukturanalysen ermöglichen die Kommunalen Kontextanalysen die genaue Beschreibung und das Einordnen von Projektaktivitäten in einem Gemeinwesen: Erfolge und Scheitern lassen sich aus einer lokalen Bedingungskonstellation heraus verstehen. Damit verbunden ist aber auch die Erkenntnis: Projekterfolge oder -misserfolge sind von vielen Kräften abhängig. Ob eine Projektstrategie zum erwarteten oder auch zu einem unerwarteten Erfolg führt, hängt also zu einem großen Teil von den strukturellen und situativen Bedingungen, nicht absehbaren Konstellationen und Dynamiken ab. So können schlichte und wenig durchdachte Strategien zu guten Ergebnissen führen, ausgereifte Strategien und qualifizierte Arbeit auch im Sande verlaufen. Die Analogie „Operation gelungen, Patient tot“ (Heiner 1992, S. 124, zitiert nach Grohmann 1996,

S. 232) bringt das ungesicherte Verhältnis zwischen Qualität und Erfolg lakonisch auf den Punkt. Projektfolgen sind darum in diesem Feld nicht unbedingt als zentrale Indikatoren für erfolgreiche Programmstrategien anzusehen und sollten darum letztlich für die Evaluation zweitrangig bleiben.

### **In welchem Verhältnis stehen Einsatz und Ertrag?**

Die Antwort auf diese Frage hängt wiederum eng mit dem Nutzen bzw. dem Erkenntnisinteresse der Evaluation zusammen. Die Kommunalen Kontextanalysen sind ein spannendes, aber äußerst zeit- und arbeitsintensives Unternehmen<sup>9</sup>. Mit Auswertung und Berichtslegung standen hierfür zwei Jahre Forschungstätigkeit zur Verfügung. Sicher ist weiterhin, dass die Beobachtungen damit einen von außen definierten Endpunkt hatten – hier muss die erste Einschränkung vorgenommen werden: Die Entwicklungen in den drei Kommunen konnten zwar intensiv, aber eben nur episodenhaft untersucht werden, die wirklichen Endpunkte der Entwicklung hingegen sind auch heute nicht abzusehen, so dass auch unter diesem Gesichtspunkt der Indikator „Projekterfolg“ fraglich bleibt.

Rückblickend lohnt sich der hohe Aufwand von Kommunalen Kontextanalysen nicht, wenn das oberste Erkenntnisziel allein die summative Bewertung eines Programms ist. Die Stärken dieser Evaluationsstrategie bestehen vielmehr darin, Prognosen über Erfolgchancen von bestimmten Projektformaten in bestimmten Sozialräumen zu formulieren und die Anlage eines Sonderprogramms aus der Maulwurfsperspektive des Einzelfalls und seines strukturellen und situativen Kontextes zu begutachten. Vor allem aber besteht ihre Stärke darin, die Förderbedingungen eines spezifischen Sozialraumes zu erschließen – Bedarfe, Potenziale, Barrieren wie auch förderliche Bedingungen von Projektarbeit im ländlichen Raum.

### **Welche Ergebnisse wurden verwertet?**

Diese Frage betrifft nicht nur die Ergebnisse der Kommunalen Kontextanalysen, sondern alle Elemente der Programmevaluation. Über die Frage der Nützlichkeit der Evaluation lässt sich nur mutmaßen, denn natürlich ist kein Auftraggeber einer Evaluation verpflichtet, die anfallenden Erkenntnisse zu berücksichtigen oder sich in dieser Frage überhaupt zu befertigen. Formative Einflüsse auf das ehemalige CIVITAS-Programm lassen sich am ehesten bei der Modifikation von Förderstrategien vermuten. Die jährlichen Datenbankanalysen zeigten eine kontinuierliche positive Entwicklung der Förderentscheidungen in Richtung

<sup>9</sup> Insgesamt wurden von März 2004 bis Oktober 2005 129 Interviews mit 133 Personen sowie viele Projektbesuche und teilnehmende Beobachtungen bei öffentlichen Veranstaltungen, Sitzungen und Seminaren durchgeführt. Neben diversen Interviews wurden verschiedene Medien (Lokalzeitschriften, Veröffentlichungen und informelle Notizen der Projekte, offizielle Dokumente der Städte) ausgewertet, relevante Ereignisse und öffentliche Diskussionen in den Städten einbezogen. Von den drei ausgewählten Kommunen wurde zudem ein Set von Strukturdaten erhoben und ausgewertet, welches in die Erstellung des Sozialprofils der Kommunen einging.

der formulierten Empfehlungen. Gesichert ist, dass die umfangreichen Ergebnisse aus den Kontextanalysen für die Projektpraxis fruchtbar gemacht wurden, denn die Programmverantwortlichen förderten explizit einen solchen Wissenschaft-Praxis-Transfer. Als dessen Ergebnis entstand ein Arbeitsbuch für die Praxis, das Reflexionen und Anregungen für zukünftige Projektakteure im ländlichen Raum enthält (vgl. Palloks/Steil, i.E.). Ob und in welcher Weise die Erkenntnisse aus den kritischen Programmreflexionen und den Informationen über die Förderbedingungen im ländlichen Raum nützlich waren, kann schwer eingeschätzt werden. Erkennbar ist, dass wesentliche Evaluationsergebnisse aus dem CIVITAS-Programm offenbar kaum in das seit Januar 2007 laufende neue Bundesmodellprogramm eingeflossen sind (vgl. BMFSFJ 2006). Die Evaluationsergebnisse wurden – zumindest bei ihrem vielleicht chancenreichsten Unterfangen, der Vorbereitung neuer Programmkonzepte auf der Grundlage der erarbeiteten Bedarfe und Förderbedingungen – damit weniger fruchtbar gemacht.

„Der Begriff Wirkung (impact) einer Evaluation bezieht sich darauf, welchen Einfluss die Evaluation auf die Entscheidungen und das weitere Vorgehen der Adressaten hat“ (Joint Committee/Sanders 2006, S. 87). Der Einfluss von Evaluationen zeigt sich in der späteren Anwendungspraxis und für diese gibt es naturgemäß kein einheitliches Prozedere. Ob und ggf. welche Ergebnisse oder Anregungen von Evaluationen in welcher Weise wozu verwendet werden, ist von vielen Faktoren abhängig und von Fall zu Fall verschieden. So könnte eine mögliche Ursache wenig produktiver Verwertung auch in der Art der Evaluationsergebnisse und deren Vermittlung liegen. Die Deutsche Gesellschaft für Evaluation empfiehlt für die Formulierung von Evaluationsberichten knappe Texte mit deutlichen Urteilen. Eine solche Reduktion der Komplexität von Programm- und Projektpraxis zu verdichteten Thesen über den Programmerfolg ist eine sinnvolle Aufgabe von Evaluationen, wenngleich sie nicht in jedem Fall leistbar bzw. angemessen scheint. So kommen im unerschlossenen Feld experimentierender Programme offene Untersuchungsdesigns und qualitative Methoden zum Einsatz, weil ein theoretischer Rahmen des relevanten Geschehens erst entwickelt werden muss. In der Regel bedeutet dies, dass sich die Komplexität des Untersuchungsgegenstandes kaum reduzieren lässt. Die Kommunalen Kontextanalysen haben durch den multiperspektivischen Zugang unterschiedliche Deutungen bestimmter Phänomene und damit ihre soziale Konstruktion nachvollzogen. Sie führten damit nicht zu einer verknappten Beschreibung der Problemzusammenhänge, sondern förderten die volle Komplexität ihrer Strukturbedingungen und damit mögliche Wirkungspotenziale von Interventionen erst zu Tage. Mit aller Vorsicht ließen sich mögliche „kausale Pfade“ (Kelle 2006, S. 129f.) benennen, ohne Aussagen über ihre Verbreitung oder Verallgemeinerbarkeit machen zu können.

Die Darstellung der Ergebnisse solcher Forschungen erfordern viel Aufwand (vgl. Klawe 2006, S. 139). Die Art der Ergebnisse lassen nur vorsichtige Verallgemeinerungen zu. Aussagen können nur getroffen werden, wenn gleichzeitig die Bedingungen, unter denen sie gelten, mitgenannt werden. Als Grundlage für strategische Argumentationen oder Mediadarstellungen sind solche komplexen und tentativen Aussagen weniger geeignet, denn es

mangelt ihnen an Deutlichkeit und Schärfe. Anders als bei klassischen Wirkungsstudien ist der Legitimationsgehalt bei Studien wie der CIVITAS-Evaluation deshalb eher gering.

## Resümee und Anregung

„Gelegentlich kam auch die Erwartung zum Ausdruck, dass die Beurteilung des schlecht Gemachten auf die moralische Unanfechtbarkeit des gut Gemeinten Rücksicht zu nehmen habe. Dass die Selbstdegradierung wissenschaftlicher Beobachtung zu werbewirksamer Öffentlichkeitsarbeit nur zur Diskreditierung des Programms und der Arbeit der Projekte hätte führen müssen, blieb dabei unbedacht“ (Lynen von Berg u.a. 2007, S. 14). Das dem Evaluationsbericht entlehnte Zitat spiegelt in der dortigen Verwendung die Perspektive einiger vom „Projekt Evaluation“ betroffener Projektakteure wieder. Es verdichtet die Ausführungen dieses Artikels aber zu einem wesentlichen Gedanken. Evaluationen, vor allem solche, die sich um politisch umkämpfte Programme bemühen, stehen mitunter vor der Aufgabe, ihren Anspruch wissenschaftlicher Autonomie unter den Bedingungen der spezifischen Programmrealität und der daraus resultierenden Bedarfe der Programmverantwortlichen durchzusetzen bzw. immer wieder neu zu verhandeln. Bei solchen Programmen wird von externen wissenschaftlichen Evaluationen vor allem Legitimationsbeschaffung erwartet. Das ist in Anbetracht des öffentlichen Drucks nachvollziehbar. Es sind damit andererseits Erwartungen verbunden, die Wissenschaft nicht immer einlösen kann. Sie ist ihrerseits den Standards wissenschaftlicher Rationalität verpflichtet, die mit einer Legitimationsbeschaffung eben nicht in jedem Fall zu vereinbaren sind. Hier treffen unterschiedliche Systemlogiken aufeinander und eine Vermittlung ist mitunter schwierig. Ein hoffnungsloser Gegensatz?

Die Zwänge von Programmgestaltern sind real und anzuerkennen, dennoch kann Wissenschaft hier nur begrenzt einlenken. Bei nicht ausgereiften, komplexen, persuasiven Programmen ist sie mit echten Wirkungsnachweisen methodisch überfordert, für die dadurch erhoffte Legitimationsbeschaffung praktisch nicht zuständig. Evaluationen werden dadurch in Widersprüche verstrickt, die sich nur schwer auflösen lassen. Es sind verschiedene Wege denkbar, um dieser Legitimationsfalle zu entgehen. So bestünde die Möglichkeit, Legitimität für neue Programme von vornherein erheblich zu erhöhen, wenn die wichtigen Fragen bereits bei der Programmplanung gestellt werden, die sich ansonsten postfestum an bilanzierende Untersuchungen richten: Welche Problemanalysen gibt es, welche theoretischen Annahmen liegen dem Programm zugrunde, von welchen Kausalitätsvermutungen kann auf dieser Grundlage ausgegangen werden? Ist die Qualität und Fachlichkeit von Projektkonzepten gesichert? Wie sind die Rahmenbedingungen gestaltet und was sind realistische Erfolgsmaßstäbe? Auf welche Erfahrungen kann ggf. zurückgegriffen werden? Solche klärenden Reflexionen würden einen längeren und gründlicheren Vorlauf erfordern, der für die Planung umfangreicher und vor allem teurer Programme aber auch angemessen scheint.

Gleiches gilt natürlich auch für die Anlage der Evaluationen. Grohmann stellt die Frage nach der generellen Evaluierbarkeit eines Programms an den Beginn der Planungsarbeiten:

„Als positiv scheint sich dagegen auszuwirken, wenn innerhalb der Institution ausreichende Freiräume für die Konzeptionierung einer Evaluation vorhanden sind und gleichzeitig ein legitimatorisches Eigeninteresse besteht, das nicht die Form eines Legitimationsdrucks annimmt. (...) Es ist in der Vorphase zu klären, inwieweit auf der Grundlage der vorhandenen institutionellen Bedingungen und der Spielräume, die diese bieten, Evaluation mit einem inhaltlichen sozialpädagogischen Zuschnitt möglich ist, inwieweit rein funktional-organisatorische Analysen angebracht sind oder ob nicht andere Formen der wissenschaftlichen Begleitung geeignet wären“ (Grohmann 1997, S. 268). Bei experimentierenden Modellprogrammen im „offenen Feld“ ist die Erwartung eines Wirkungsnachweises fehl am Platz. Präsentationen von Projektfolgen sind in diesem Feld weniger aussagekräftig als die Gewissheit, für Konzept- und Projektqualität gesorgt, also bestmögliche Bedingungen für gute Projektarbeit geschaffen zu haben. Gelungene Planung von Evaluationen beinhaltet demnach neben der Definition ihrer Zuständigkeit auch die Verständigung über die Grenzen ihrer Aussagekraft. Dass Evaluationen von Modellprogrammen generell andere Aufträge als den Wirksamkeitsnachweis haben könnten, zeigt ein Blick auf die internationale Fachdebatte. In der amerikanischen Evaluationsforschung werden Wirksamkeitsanalysen nur noch bei „ausgereiften Programmen“ als sinnvoll erachtet. Die Stärken der Evaluation von Programmen im experimentellen Stadium liegen demgegenüber auf vorgelagerten Ebenen, z. B. bei der Programmzielbeschreibung oder der Erarbeitung von Erfolgsindikatoren (vgl. Haubrich u.a. 2007, S. 188). Die Kommunalen Kontextanalysen der CIVITAS-Evaluation haben in diesem Feld ihren Beitrag geleistet, indem sie Bedarfsanalysen vorhalten und die Bedingungen der spezifischen Handlungsfelder von Projektarbeit beschreiben. Das erprobte Instrumentarium ermöglicht es, kontroverse Einschätzungen von Projektleistungen durch die Rekonstruktion von Deutungsmustern einzuordnen – und so die Entwicklung realistischer Erfolgsmaßstäbe voranzutreiben. Der wesentliche Erkenntnisgewinn dieser Forschungsstrategie ist demnach auf der planerischen Ebene zu verorten.

Wünschenswert ist ein souveräner Umgang mit formativen Evaluationen, deren Anliegen die Optimierung von laufenden und die Planung künftiger Programme auf der Basis anfallenden Erfahrungswissens sind. In Bezug auf das Wirkungsproblem bei Großprogrammen bleibt zu hoffen, dass es zukünftig nicht nur in Hinblick auf seine methodischen Voraussetzungen diskutiert, sondern auch vermehrt als Gegenstand kritischer wissenschaftssoziologischer Debatten entdeckt wird.

## Literaturverzeichnis

- Beck, Ulrich (1986): Risikogesellschaft. Auf dem Weg in eine andere Moderne. Frankfurt am Main: edition suhrkamp
- BMFSFJ (2006): Abschlussbericht zur Umsetzung des Aktionsprogramms „Jugend für Toleranz und Demokratie – gegen Rechtsextremismus, Fremdenfeindlichkeit und Antisemitismus“. Berlin, [www.jugendstiftung-civitas.org/downloads/Abschlussbericht\\_zum\\_Aktionsprogramm.pdf](http://www.jugendstiftung-civitas.org/downloads/Abschlussbericht_zum_Aktionsprogramm.pdf) (22.03.07)

- Deutscher Bundestag (2003): Plenarprotokoll 15/33; Stenographischer Bericht, 33. Sitzung (BT-Drs. 15–33 Haushaltsdebatte 2003 am 18.03.2003\_Erz17.pdf)
- Dietzel, Gottfried T. W./Troschke, Jürgen v. (1988): Begleitforschung bei staatlich geförderten Modellprojekten – strukturelle und methodische Probleme. Schriftenreihe des Bundesministers für Jugend, Familie, Frauen und Gesundheit. Berlin/Köln/Mainz: Verlag W. Kohlhammer
- Frindte, Wolfgang/Preiser, Siegfried (2007): Präventionsansätze gegen Rechtsextremismus. In: Aus Politik und Zeitgeschichte. Beilage zur Wochenzeitschrift Das Parlament 11/2007, S. 32–38
- Grohmann, Romano (1997): Das Problem der Evaluation in der Sozialpädagogik. Frankfurt am Main: Europäischer Verlag der Wissenschaften
- Haubrich, Karin/Christian Lüders/Gerlinde Struhkamp (2007): Wirksamkeit, Nützlichkeit, Nachhaltigkeit. Was Evaluationen von Modellprogrammen realistischer Weise leisten können. In: Schröder, Ute B./Streblow, Claudia (Hrsg.): Evaluation konkret. Fremd- und Selbstevaluationsansätze anhand von Beispielen aus Jugendarbeit und Schule. Opladen/Farmington Hills: Verlag Barbara Budrich, S. 183–202
- Heil, Karolus (2001): Ziele, Probleme und systematisch-methodische Gesichtspunkte der Evaluation sozialer Arbeit. In: Heil, K./Heiner, M./Feldmann, U. (Hrsg.): Evaluation Sozialer Arbeit. Eine Arbeitshilfe mit Beispielen zur Evaluation und Selbstevaluation. Frankfurt am Main: Eigenverlag des Deutschen Vereins für öffentliche und private Fürsorge, S. 17–26
- Joint Committee on Standards for Educational Evaluation/Sanders, James R. (Hrsg.) (2006): Handbuch der Evaluationsstandards. (3. bearb. Aufl.). Wiesbaden: Verlag für Sozialwissenschaften
- Kelle, Udo (2006): Qualitative Evaluationsforschung und das Kausalitätsparadigma. In: Flick, Uwe (Hrsg.): Qualitative Evaluationsforschung. Konzepte Methoden Umsetzungen. Reinbeck bei Hamburg: Rohwolt Taschenbuch Verlag, S. 117–134
- Klawe, Willy (2006): Multiperspektivische Evaluationsforschung als Prozess- Wirkungsrekonstruktion aus Sicht der Beteiligten. In: Projekt „EXE“ (Hrsg.): Wirkungsevaluation in der Kinder- und Jugendhilfe. Einblicke in die Evaluationspraxis. München: DJI, S. 125–140
- Klärner, Andreas/Kohlstruck, Michael (2006): Rechtsextremismus – Thema der Öffentlichkeit und Gegenstand der Forschung. In: Klärner, A./Kohlstruck, M. (Hrsg.): Moderner Rechtsextremismus in Deutschland. Hamburg: Hamburger Edition
- Kliche, Thomas (1996): Interventionen, Evaluationsmaßstäbe und Artefaktbildung. Zehn Thesen zur Konstruktion von Rechtsextremismus. In: Heiland, Hans-Günther/Lüdemann, Christian (Hrsg.): Soziologische Dimensionen des Rechtsextremismus. Opladen: Westdeutscher Verlag, S. 57–84
- Kromrey, Helmut (1988): Akzeptanz- und Begleitforschung. Methodische Ansätze, Möglichkeiten und Grenzen. In: Massacommunicatie, Tilburg. H. 3, S. 221–242

- Lynen von Berg, Heinz/Palloks, Kerstin (2004): Evaluierung eines Programms zur Stärkung der Zivilgesellschaft – Teil 1: Anlage, Konzeption und Vorgehensweisen der wissenschaftlichen Begleitforschung des CIVITAS-Programms. In: Journal für Konflikt- und Gewaltforschung, 6. Jg., H. 1, S. 41–69
- Lynen von Berg, Heinz/Palloks, Kerstin/Steil, Armin (2007): Interventionsfeld Gemeinwesen. Evaluation zivilgesellschaftlicher Strategien gegen Rechtsextremismus. Weinheim: Juventa
- Palloks, Kerstin/Steil, Armin (i. E.): Von Blockaden und Bündnissen. Praxismaterialien zur Auseinandersetzung mit Rechtsextremismus im Gemeinwesen. Weinheim: Juventa
- Rommelspacher, Birgit/Polat, Ülger/Wilpert, Czarina (2003): Die Evaluation des CIVITAS-Programms. In: Lynen von Berg, Heinz/Roth, Roland (Hrsg.): Maßnahmen und Programme gegen Rechtsextremismus wissenschaftlich begleitet. Aufgaben, Konzepte und Erfahrungen. Opladen: Leske + Budrich, S. 51–76
- Roth, Roland (2004): Reden Sie mit dem Pferd – Bedingungen, Möglichkeiten und Grenzen der Evaluationskultur in Deutschland. In: Sozial Extra, H. 6/2004, S. 6–9
- Roth, Roland/Benack, Anke (2003): Bürgernetzwerke gegen Rechts. Evaluierung von Aktionsprogrammen und Maßnahmen gegen Rechtsextremismus und Fremdenfeindlichkeit. Erstellt für den Arbeitskreis „Bürgergesellschaft und Aktivierender Staat“ der Friedrich-Ebert-Stiftung, Bonn
- Stöss, Richard (2003): Möglichkeiten und Grenzen der Evaluierung von persuasiven Programmen. In: Lynen von Berg, Heinz/Roth, Roland (Hrsg.): Maßnahmen und Programme gegen Rechtsextremismus wissenschaftlich begleitet. Konzepte und Erfahrungen. Opladen, S. 95–102

*„Evaluation’s most important purpose is not to prove, but to improve“  
(Stufflebeam 2003, zit. nach Beywl 2006, S. 44)*

## Das Programm „Entimon“: Spezifika, Potenziale und Herausforderungen einer induktiv-rekonstruierenden Evaluation anhand Logischer Modelle

Das Programm „Entimon – gemeinsam Gewalt und Rechtsextremismus“ wurde in den Jahren 2002–2006 unter dem Dach des Aktionsprogramms „Jugend für Toleranz und Demokratie – gegen Rechtsextremismus, Fremdenfeindlichkeit und Antisemitismus“ bundesweit umgesetzt. Die Initiatorinnen und Initiatoren<sup>1</sup> setzten sich zum Ziel, durch die Förderung pädagogischer – und insbesondere politisch-bildnerischer – Arbeit sowie der Unterstützung von Vernetzungsprozessen Zivilgesellschaft zu stärken und einen Beitrag zur Förderung von Demokratie und Toleranz zu leisten.

Betrachten wir als Evaluationsteam vergleichend das Vorgängerprogramm „Aktionsprogramm gegen Aggression und Gewalt“ („AgAG“), das in den Jahren 1992 bis 1996 durch die Bundesregierung implementiert worden war, so zeigt sich in unserer Wahrnehmung ein wichtiger Paradigmenwechsel bezüglich der inhaltlichen Ausrichtung des Aktionsprogramms „Jugend für Toleranz und Demokratie“: Im Fokus seiner Teilprogramme „Entimon“, „CIVITAS“ und „XENOS“ stand nunmehr nicht – wie noch bei „AgAG“ – die Arbeit mit rechts-extremen und gewaltbereiten Jugendlichen, sondern vorrangig das Empowerment derjenigen gesellschaftlichen Kräfte, die sich gegen Rechtsextremismus und Fremdenfeindlichkeit sowie für die Förderung zivilgesellschaftlicher Strukturen und Prozesse engagieren und in dieser Arbeit unterstützt, gestärkt und (weiter-)qualifiziert werden sollten. In das Programm einbezogen wurden eine Vielzahl von Standorten sowie ein breites Spektrum an Zielgruppen, das von Kindern, Jugendlichen und jungen Erwachsenen bis hin zu Multiplikatorinnen und Multiplikatoren, Eltern sowie lokalen, einflussreichen Akteuren reichte.

Das Programm „Entimon“ wurde während seiner gesamten Laufzeit durch das Deutsche Jugendinstitut e.V. (DJI) wissenschaftlich begleitet<sup>2</sup>. Es hatte durch das Bundesministerium für Familie, Senioren, Frauen und Jugend (BMFSFJ) den Auftrag erhalten, im Rahmen der Evaluation:

- 
- 1 Das Programm „Entimon – gemeinsam gegen Gewalt und Rechtsextremismus“ wurde durch das Bundesministerium für Familie, Senioren, Frauen und Jugend (BMFSFJ) finanziell gefördert.
  - 2 Zu dem Evaluationsteam gehörten Ulrich Brüggemann (seit 2001), Mareike Schmidt (seit 2003), Silke Schuster (2005 bis 2006) sowie die Autorin (seit 2001). Bedanken möchte ich mich bei meinen Kollegen und Kolleginnen für ihre wertvollen Anregungen zum vorliegenden Beitrag.

- die Umsetzungsphase des Programms wissenschaftlich zu begleiten,
- eine dichte Beschreibung modellhafter Praxis im Programm „Entimon“ zu liefern und
- empirisch gesicherte Aussagen über übertragbare Methoden und Konzeptionen der pädagogisch-bildnerischen Arbeit zur Stärkung von Demokratie und Toleranz sowie zur Prävention und Bekämpfung von Rechtsextremismus zu treffen (vgl. BMFSFJ 2006, S. 3).

Im Rahmen dieses Beitrags wird das gewählte Evaluationsdesign und -vorgehen vorgestellt, begründet und – basierend auf den gewonnenen Erfahrungen – retrospektiv in Bezug auf seine Potenziale und Herausforderungen reflektiert. Fokussiert werden dabei

- der gewählte Ansatz der Programmevaluation, der auf die Zielsetzung und Charakteristika des Programms „Entimon“ abgestimmt wurde – entwickelt wurde eine Variante der „Cluster Evaluation“, die „induktiv-rekonstruierende Cluster-Evaluation“ –,
- das Evaluationsinstrument „Logisches Modell“, das als heuristisches Modell den Evaluationsverlauf maßgeblich mitsteuerte.

### **Die Wahl und Adaption des Evaluationsansatzes: Übertragung der Heuristik der „Cluster-Evaluation“<sup>3</sup>**

Das Evaluationsdesign – das verdeutlicht auch die aktuelle Fachliteratur<sup>4</sup> – muss den Charakteristika des Programms ebenso angepasst werden wie der Evaluationsfragestellung und dem Nutzungsinteresse im Hinblick auf die Evaluationsergebnisse.

Es ist demnach sinnvoll, zunächst nach dem Charakter, der Struktur und der Zielsetzung des Programms „Entimon“ zu fragen. Was macht seine Besonderheiten aus – und welche Anforderungen an das Evaluationsdesign lassen sich hieraus ableiten?

Als Programm mit stark modellhaften Komponenten verfolgte „Entimon – gemeinsam gegen Rechtsextremismus und Gewalt“ im Sinne der Anregungsfunktion des Bundes vorrangig das Ziel, „Real-“ bzw. „Feldexperimente“ zur Weiterentwicklung der Kinder- und Jugendhilfe im Feld der Förderung von Toleranz und Demokratie sowie der Prävention von Rechtsextremismus und Gewalt durchzuführen. Angestrebt war, im Anschluss an das Programm einen verbesserten und übertragbaren Erfahrungsschatz an Konzepten, Strategien und Methoden in den Handlungsfeldern des Programms zur Hand zu haben und diesen für die weitere Praxis nutzbar zu machen (vgl. BMFSFJ 2006, S. 3). Dieser Implementationsmotivation lag die Wahrnehmung zugrunde, dass im Bereich der pädagogisch-bildnerischen und vernetzenden Arbeit gegen Rechtsextremismus und Gewalt sowie der Förderung

3 Die folgende Darstellung basiert auf den konzeptionellen Überlegungen, die im Rahmen mehrerer interner Workshops mit Karin Haubrich und Dr. Christian Lüders ausgetauscht wurden. Ihnen gebührt mein herzlicher Dank für ihre wertvollen Anregungen und Hinweise.

4 Vergleiche hierzu die Reflexionen zu der Frage, welche „Entwicklungsstufe“ bzw. welchen „Reifegrad“ des Programms Wirkungsevaluationen voraussetzen (vgl. u.a. Lüders/Haubrich 2006, S. 16; Haubrich 2006, S. 105f.; Rossi u.a. 2004, S. 39).

von Demokratie und Toleranz bisher wenig ausgereiftes und empirisch überprüftes Wissen über „wirkfähige“ und übertragbare Konzepte und Ansätze vorliegt. Analog dieser Zielsetzung von Programmen mit modellhaftem Charakter nahm die Evaluation eine primär wissensgenerierende Funktion und Perspektive ein, denn im Rahmen wissensgenerierender Evaluationen „soll ein fachliches Wissen darüber gewonnen werden, warum unter gegebenen Voraussetzungen und Rahmenbedingungen bestimmte Arbeitsformen erfolgreiche Ansätze zur Lösung eines gesellschaftlichen Problems darstellen oder auch nicht“ (Haubrich u.a. 2005, S. 2). Die wissensgenerierende Evaluation dient demnach dem Ziel, Fachpraxis bzw. Fachlichkeit weiterzuentwickeln (vgl. ebd.).

Als Modellprogramm mit innovativem Charakter<sup>5</sup> verfügte „Entimon“ über wenig explizierte und operationalisierte Programmziele, sondern eröffnete – ganz im Sinne der Zielsetzung und Logik von modellhaften, innovativen Programmen – ein breites Experimentierfeld für Konzeptübertragungen, -entwicklungen und -erprobungen. Die Programmleitlinien sowie die formulierten Förderschwerpunkte („Interkulturelles Lernen“, „Politische Bildungsarbeit“ sowie „Lokale Netzwerke“) limitieren und orientieren diesen Experimentierraum zwar, indem sich jedes Projekt bei der Antragstellung einem Förderschwerpunkt zuordnen musste, es zeigte sich z. B. jedoch bei ersten Analysen, dass das Programm in seiner Umsetzung z. B. durchaus Experimente über die Förderschwerpunktgrenzen hinweg zuließ (z. B. in Form der Umsetzung interkultureller Netzwerke oder dem Versuch der Integration von Konzepten der interkulturellen und politischen Bildung). Es wurde demnach durch das Programm ein genügend breites Experimentierfeld für Innovationen eröffnet.

Mit dieser experimentorientierten Ergebnis- und Veränderungsoffenheit des Programms ist jedoch in Hinblick auf die Evaluation die Einschränkung verbunden, dass für das Programm „Entimon“ nicht – wie bei reifen Programmen, bei denen schon ein überprüftes Fach- und Praxiswissen bezüglich der Beschaffenheit und Wirkfähigkeit der geförderten Konzepte vorliegt – explizite Wirkungsannahmen formuliert werden konnten. Mit anderen Worten, die Rekonstruktion einer Programmtheorie – d.h. die Explikation der dem Programm zugrunde liegenden logischen Verbindungen zwischen Aktivitäten/ Interventionen und Resultaten/ Effekten (vgl. Haubrich 2006, S. 105f.; Beywl 2006, S. 33–38) die eine wirkungsorientierte Evaluation steuert und strukturiert, war zu Beginn der Evaluation nicht möglich. Dieses Erkenntnis hatte weit reichende Auswirkungen auf das Evaluationsvorhaben, auf die weiter unten eingegangen werden soll. Zunächst lässt sich festhalten: Ein deduktives Vorgehen im Rahmen der Programmevaluation – nämlich die Orientierung an einer Programmtheorie – war nicht möglich, da das Programmsetting zu wenig formalisiert war. Aus dem gleichen Grund waren die Möglichkeiten, Wirkungen des Programms zu messen, begrenzt.

---

5 Innovative modellhafte Programme stellen einen spezifischen Programmtypus dar, der sich durch Ergebnis- und Veränderungsoffenheit kennzeichnen lässt. Im Rahmen modellhafter innovativer Programme sollen Ansätze und Konzepte entwickelt und erprobt werden, die bisher noch nicht Eingang in die entsprechende Regelpraxis gefunden haben. Hierfür wird ein Experimentierraum zur Verfügung gestellt, der – da ja noch wenig Wissen über das zu entwickelnde Handlungsfeld vorliegt – mit wenig konkretisierten Zielsetzungen verbunden ist (vgl. Haubrich 2006, S. 104).

Das Programm „Entimon“ wurde bundesweit durch ein breites Spektrum an Trägern und an zahlreichen unterschiedlichen Standorten (d.h. im ländlichen, klein- und großstädtischen Bereich) umgesetzt. Es handelte sich somit um ein sogenanntes multizentrisches Programm (vgl. Haubrich 2003, S. 214). Da die Projektmitarbeitenden vor Ort auf Grundlage der lokalen Problemlagen und Bedarfe entschieden, welche Konzepte sie mit welchen Zielgruppen umsetzen, entstand im Programmkontext eine gewünschte Vielfalt höchst heterogener Aktivitäten. Insgesamt wurden über 1000 Maßnahmen gefördert.

Der Eindruck von Heterogenität erhöht sich weiter, wenn man die Projektlaufzeiten in die Analyse mit einbezieht: Die geförderten Projekte wiesen unterschiedliche Laufzeiten von jeweils bis zu drei Jahren auf und setzten in den einzelnen Förderjahren jeweils differente, zum Teil aufeinander aufbauende Projektmodule um. Auch der Beginn der Projektarbeit war unterschiedlich, da das Programm durch mehrere Bewilligungszeitpunkte geprägt war. Im Evaluationsdesign musste der Größe und Heterogenität des Programms u.a. durch die Einplanung unterschiedlicher Befragungszeitpunkte und -module im Rahmen der quantitativen Analysen sowie die Bildung eines qualitativen Untersuchungssamples Rechnung getragen werden.

Die drei genannten Charakteristika des Programms – die Modellhaftigkeit, die Ergebnis- und Veränderungsoffenheit sowie die Größe und Heterogenität – bestimmten maßgeblich die Wahl des Evaluationsansatzes und -vorgehens.

Einen erprobten Rahmen für die Evaluation modellhafter, innovativer und multizentrischer Programme bietet der Ansatz der Cluster-Evaluation (vgl. Haubrich 2006, S. 101), der – ursprünglich aus dem nordamerikanischen Raum stammend<sup>6</sup> – bereits in unterschiedlichen Kontexten adaptiert wurde<sup>7</sup>. Den Darstellungen und Ausarbeitungen von Karin Haubrich (vgl. ebd. 2006, S. 101–122) folgend, betrachten Cluster-Evaluationen die – nicht-standardisiert erfolgenden – Implementationen der Projekte im Rahmen innovativer und multizentrischer Programme als Variationen einer Programmidee in unterschiedlichen Kontexten – und somit als Möglichkeit, aus den Erfahrungen der Projekte mehr über das zu entwickelnde Handlungsfeld sowie Bedingungen für das Gelingen oder Mislingen von Ansätzen zu lernen. Im Mittelpunkt steht dabei nicht nur die Frage, welche Resultate die Programmumsetzungen zeitigen, sondern vor allen Dingen, warum sich bestimmte Ansätze kontextübergreifend als erfolgversprechend erweisen oder aber scheitern. In die Evaluation einbezogen werden dabei ausschließlich die Projektbereiche, die einen Beitrag zur Umsetzung der Programmidee leisten – Projektaktivitäten, die keine relevante Dimension der Programmidee berühren, werden demnach aus der Analyse ausgeblendet.

---

6 Die Cluster-Evaluation wurde in den 1990er Jahren durch die W.K. Kellogg Foundation entwickelt (vgl. [www.wkkf.org](http://www.wkkf.org)).

7 Die erste Adaption in Deutschland erfolgte durch Karin Haubrich/Kerstin Frank im Kontext der Evaluation des Modellprogramms „Mobile Jugendsozialarbeit für junge Menschen ausländischer Herkunft“ (vgl. Haubrich/Frank 2000). Für weitere Adaptionen vgl. Haubrich/Vossler 2001, Vossler/Obermaier 2003.

Weitere wichtige Prämissen der Cluster-Evaluationen sind das kooperative Vorgehen aller Beteiligten sowie die Vertraulichkeit der Evaluation, die Anonymität bzw. die Präsentation der Ergebnisse in aggregierter, d.h. zusammenfassender und projektübergreifender Form sichert. Die Aufarbeitung von Lernerfahrungen aus der Praxis in ansatz- bzw. strategiebezogener Weise verdeutlicht den Projektmitarbeitenden gleichzeitig, dass auch das Scheitern von Projektumsetzungen entscheidend zu den „lessons learned“ beitragen kann und daher ungestraft an die Evaluatorinnen/Evaluatoren zurückgekoppelt werden darf.

In seiner klassischen Form wird im Rahmen dieses Evaluationsansatzes zu Beginn der Untersuchungen ein Cluster von Projekten gebildet, das alle Maßnahmen des Programms umfasst, welche als heterogene Implementationen einer gemeinsamen projektübergreifenden Programmidee verstanden werden (vgl. ausführlich Haubrich (i. E.)). Idealerweise übersteigt die Anzahl der Projekte im Cluster dabei nicht 25 bis 30 Maßnahmen (vgl. Parsons/Haubrich 2002). Da die Anzahl der geförderten Projekte im Programm „Entimon“ diese Zahl weit überstieg und die Projekte sich überdies nicht einer rekonstruierbaren Programmidee zuordnen ließen, war der Ansatz der Cluster-Evaluation in seiner klassischen Form nicht auf das Programm „Entimon“ anwendbar. Übertragen werden konnte jedoch das heuristische Modell der Cluster-Evaluation und die dargestellte Perspektive auf modellhafte, innovative Programme. Die Adaptionleistung der wissenschaftlichen Begleitung lag in der Anpassung des konkreten Evaluationsvorgehens an die Charakteristika des Programms „Entimon“ sowie einer empirischen Neubestimmung der Anzahl der Untersuchungscluster. Aufgrund der Maßgabe, dass es nicht möglich war, eine – die Evaluation orientierende und strukturierende – Programmtheorie zu rekonstruieren und die Evaluation deduktiv an ihr auszurichten, musste die Evaluation induktiv angelegt werden. Ausgehend von 24 Projekten, die nach dem Kriterium der „Innovativität“<sup>8</sup> ausgewählt wurden, und ihrer vergleichenden und kontrastierenden qualitativen Gegenüberstellung haben wir die wesentlichen Dimensionen des Programms<sup>9</sup> im Evaluationsverlauf erst induktiv rekonstruierend herausgearbeitet. Hierbei konnten wir – ausgehend von unserem qualitativen Sample – vier Kerndimensionen des Programms identifizieren. In Form der vier Programmstrategien „Bildung & Sensibilisierung“, „Begegnung“, „Beratung“ und „Vernetzung“ können diese mit Abschluss der Evaluation nun beschrieben werden (vgl. Klingelhöfer u.a. 2007 (i. E.)). Jede dieser Programmstrategien wird verstanden als ein induktiv rekonstruiertes Programmcluster.

---

8 Das Auswahlkriterium „Innovativität“ erachteten wir im Kontext der Evaluation eines Modellprogramms als sinnvoll, da es im Erkenntnisinteresse des Auftraggebers lag, aus dem Programm mehr über neue Handlungsansätze zu lernen. „Innovativität“ wurde durch uns wie folgt operationalisiert: „Übertragung einer bewährten Methode in ein neues Feld“, „Übertragung einer bewährten Methode in ein neues Setting“, „Übertragung und Weiterentwicklung einer bewährten Methode in neue regionale Bezüge, „Innovatives Engagement in Bereichen, die sich in den Auswertungen der Vorjahre als zentral, aber auch besonders problematisch herauskristallisiert haben“ (z. B. geschlechts-spezifische Jungenarbeit im Feld interkulturellen und antirassistischen Lernens oder der politischen Bildung) und „Entwicklung einer neuen Methode“.

9 Mit dem Begriff der „Programmdimension“ lehnen wir uns an K. Haubrich an (Haubrich 2006, S. 104, 115).

Die Analyse der Strategien richtete den Blick auf die Umsetzung befördernder und/oder behindernder Rahmenbedingungen und Ressourcen, strategietypische Aktivitäten sowie auf Zielerreichungsprozesse. Weiterhin konnten Bedingungen in Hinblick auf die Übertragung in andere Kontexte extrahiert werden. Die strategiebezogene Auswertungs- und Darstellungsform gewährleistet einerseits den Anspruch der Cluster-Evaluation, Evaluationsergebnisse nur in aggregierter Form zur Verfügung zu stellen. Darüber hinaus macht sie im Hinblick auf die Modellhaftigkeit des Programms Sinn, denn: Zielsetzung der Evaluation von modellhaften Programmen ist es, Aussagen über die Übertragbarkeit von erprobten Projektansätzen zu liefern. Übertragen werden können jedoch nur Strategien, nicht aber Einzelprojekte.

Das konkrete Vorgehen der Evaluation sowie die Erhebungs- und Auswertungsschritte wurden ganz wesentlich durch das eingesetzte Evaluationsinstrument „Logisches Modell“ strukturiert, das bereits in mehreren Fällen – und in je unterschiedlicher Weise – in Kombination mit dem Programmevaluationsansatz „Cluster-Evaluation“ angewandt wurde (vgl. Haubrich/Frank 2000; Haubrich/Vossler 2001; Vossler/Obermaier 2003).

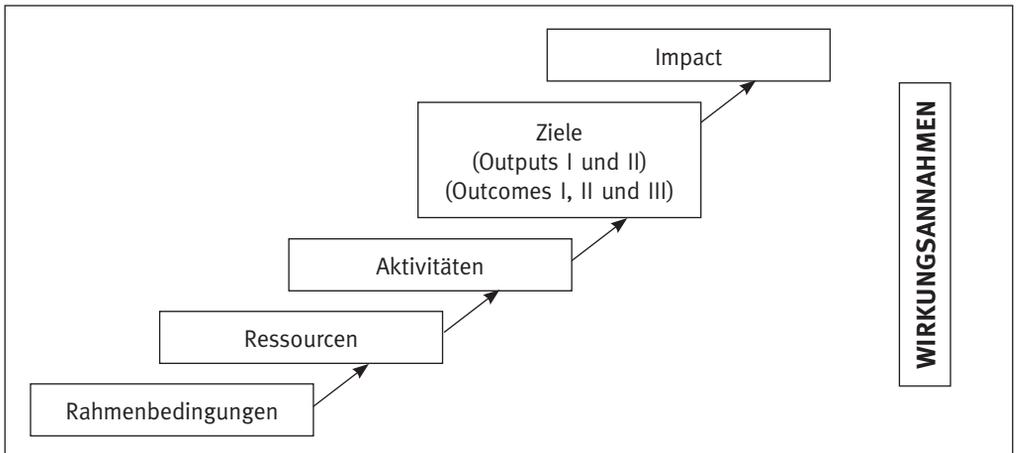
### **Die Rekonstruktion und Explikation von Wirkungsannahmen: Die Arbeit mit dem Evaluationsinstrument „Logisches Modell“**

Im Rahmen der Programmevaluation wurde – wie im vorangegangenen Abschnitt erläutert – ein stark induktiv orientiertes Vorgehen gewählt. Im Rahmen dieses „bottom-up“-Verfahrens kam der Analyse der Projektebene eine besondere Bedeutung zu. Auf dieser Ebene setzte das Bemühen der wissenschaftlichen Begleitung an, Praxis „sichtbar“ zu machen, ihre innere Logik nachzuvollziehen und – im Sinne einer „Erhellung der black box“ – die Frage zu stellen, was warum im Projektalltag geschieht und zu welchen Ergebnissen dies führt. Damit verbunden war die Prüfung, welche Bedingungen sich als relevant für das Scheitern oder Gelingen von Projektarbeit erweisen. Gleichzeitig fokussierten die Analysen die Projektebene, um über ein induktives Vorgehen die relevanten Programmdimensionen zu erschließen und herauszuarbeiten. Während es auf der Programmebene aus den oben genannten Gründen nicht möglich war, *eine* Programmtheorie und damit Wirkungsannahmen zu rekonstruieren, zeigt sich auf der Ebene der (Einzel-)Projekte ein anderes Bild: Für das Einzelprojekt sowie für Gruppen vergleichbarer Projekte war der Nachvollzug der eigenen Wirkungsannahmen möglich und erwies sich gleichzeitig als zentral, um den oben formulierten Fragen nachzugehen.

Die wissenschaftliche Begleitung wählte als unterstützendes und strukturierendes Evaluationsinstrument das sogenannte „Logische Modell“. Logische Modelle bieten eine Heuristik an, die es erlaubt, die „implizite Handlungslogik“ (und somit auch Wirkungsannahmen) von Projekten zu rekonstruieren und zu visualisieren. Sie stellen einfach strukturierte „Abbilder“ der Projektlogik dar und machen letztere somit einer Überprüfung zugänglich. Inzwischen wurden elaborierte Konzepte Logischer Modelle entworfen und erprobt (vgl. Beywl 2006, S. 34ff.).

Logische Modelle basieren auf der Annahme, dass Programme bzw. Projekte auf der Grundlage bestimmter beschriebener Rahmenbedingungen und Ressourcen sowie auf der Grundlage der auf dieser Basis geplanten Aktivitäten bestimmte Zielsetzungen verfolgen (vgl. Lüders/Haubrich i. d. Bd.). Im Modell selbst werden die einzelnen Bereiche in Form von „Wenn-dann-Beziehungen“ verknüpft ausgewiesen. Aus diesem Grund werden Logische Modelle auch als „chain of events“ – d.h. als „Kette von Ereignissen“ bezeichnet, als Schritte, die vollzogen werden müssen, um die erwünschten Ziele zu erreichen (vgl. Weiss 1996, S. 174). Die Verknüpfungen zwischen den Bereichen des Logischen Modells sind nicht als „harte“ Kausalbeziehungen zu verstehen, sondern bilden zunächst die Annahmen der Projektmitarbeitenden über die Wirksamkeiten ihrer Arbeit ab. Die Rekonstruktion dieser Annahmen machen diese erst einer Überprüfung zugänglich (z. B. „Kann ich mit den von mir eingesetzten Ressourcen tatsächlich die geplanten Aktivitäten umsetzen?“, „Erreiche ich mit meinen Aktivitäten die angestrebte Zielgruppe?“). Das Nachzeichnen des Projektverlaufs anhand der Aktualisierung der Logischen Modelle in regelmäßigen Zeitabständen erlaubt es gleichzeitig, den Einfluss förderlicher und hinderlicher Rahmenbedingungen sowie förderlicher und fehlender Ressourcen auf die Projektaktivitäten genauer herauszuarbeiten.

**Abb. 1:** Aufbau eines einfachen Logischen Modells



In der Darstellung wird zwischen den folgenden Bereichen unterschieden:

**Rahmenbedingungen**, die durch das Projekt zunächst nicht beeinflussbar sind; Ressourcen, die die finanziellen, personellen, zeitlichen, räumlichen etc. Mittel abbilden, die dem Projekt zur Umsetzung seiner Aktivitäten zur Verfügung stehen. Zu den Ressourcen zählten wir auch die Potenziale und Kompetenzen der Teilnehmenden (vgl. Beywl u.a. 2004, S. 27);

**Aktivitäten**, die die geplanten Handlungsschritte bezeichnen;

**Outputs**, die im Bereich der formulierten Ziele die strukturellen und produktorientierten Ziele abbilden. Unterschieden wird zwischen der Output-Ebene I, die die produktorientierten Zielsetzungen, z. B. die Bildung eines Netzwerks, die Entwicklung eines Bildungsmoduls etc. umfasst und der Output-Ebene II, die Ziele in Bezug auf die angestrebte Teilnehmerstruktur (z. B. Teilnehmerkreis, -anzahl, Intensität der Teilnahme) detailliert (vgl. Beywl 2006, S. 37);

**Outcomes**, die die Ziele auf Adressatenebene bezeichnen. Die Outcomes I beschreiben angestrebte Veränderungen oder Stabilisierungen auf der Ebene des Wissens und der Emotionen. Zielstellungen im Bereich des Handelns und Verhaltens bilden die Outcomes II, während Ziele, die die Lebenslage und den Status der Zielgruppe betreffen, als „Outcome III“ bezeichnet werden;

**Impact**, der angestrebte, über einen Zeitraum von mehreren Jahren erreichbare dauerhafte Veränderungen beschreibt.

Logische Modelle in der oben dargestellten Form wurden durch uns als Evaluations- und Kommunikationsinstrument eingesetzt. Sie strukturierten die Evaluation (und somit die qualitativen und quantitativen Erhebungen)<sup>10</sup>, die Kommunikation mit den Projektmitarbeitenden sowie die Auswertung und den Aufbau der Evaluationsberichte.

### **Im Evaluationsverlauf wurden Logische Modelle auf drei Ebenen entwickelt:**

**Logische Modelle auf der Projektebene**, die zunächst durch die Projektmitarbeitenden selbst entwickelt und im Projektverlauf immer wieder aktualisiert wurden, um die Realisierung des jeweiligen Projektvorhabens nachzuzeichnen. Am Abschluss dieser Phase stand die Entwicklung eines „Gesamtmodells“ – eines die Gesamtentwicklung der Maßnahme nachvollziehenden Logischen Modells – durch die wissenschaftliche Begleitung. Dieses Gesamtmodell stellte eine erste Abstraktionsebene im Rahmen der Auswertung auf der Einzelprojektebene dar.

---

<sup>10</sup> Im Rahmen der Evaluation wurden sowohl qualitative als auch quantitative Analysen durchgeführt. Das quantitativ orientierte Programm-Monitoring umfasste Fragebogenerhebungen mit integrierten Follow-up-Modulen unter allen geförderten Projekten. Die Fragekomplexe im Fragebogen orientierten sich gleichfalls an der Struktur Logischer Modelle (vgl. Klingelhöfer u.a. 2007 (i.E)). Die im Rahmen des Monitorings erhobenen Projektbeschreibungen wurden in der eigens entwickelten Datenbank „Entimon“ öffentlich zugänglich gemacht. vgl.: [www.dji.de/entimon](http://www.dji.de/entimon) (30.05.2007).

*Logische Modelle auf der Ebene der Umsetzungsformen der Strategien*, die durch die wissenschaftliche Begleitung nachgezeichnet wurden und die eine zweite Abstraktionsebene zur Erschließung der Programmdimensionen darstellten. Über die Kontrastierung und die vergleichende Analyse der einzelnen Logischen Modelle der Projekte wurden Strategieformen (z. B. „Beratung von Eltern rechtsextremer Kinder“, „Beratung von Kommunen“) herausgearbeitet, deren Gemeinsamkeiten im Bereich des methodischen Heran-gehens und der angesprochenen Zielgruppe/n lagen.

*Logische Modelle auf der Strategieebene*, die eine dritte Abstraktionsebene herstellen. Diese bezeichnet die *Programmcluster*, die durch die Evaluation induktiv rekonstruiert wurden. Beispielsweise die Strategie „Beratung“ bildet ein Programmcluster, das mehrere Strategieumsetzungsformen (z. B. „Beratung von Eltern rechtsextremer Kinder“, „Beratung von Kommunen“) umfasst, die wiederum jeweils mehrere Projekte bzw. Projektanteile in-kludieren (z. B. „telefonische Einmal- und Sachberatung von Eltern rechtsextremer Kinder“, „psychosoziale Beratung von Eltern rechtsextremer Kinder im Rahmen von face-to-face“-Settings“).

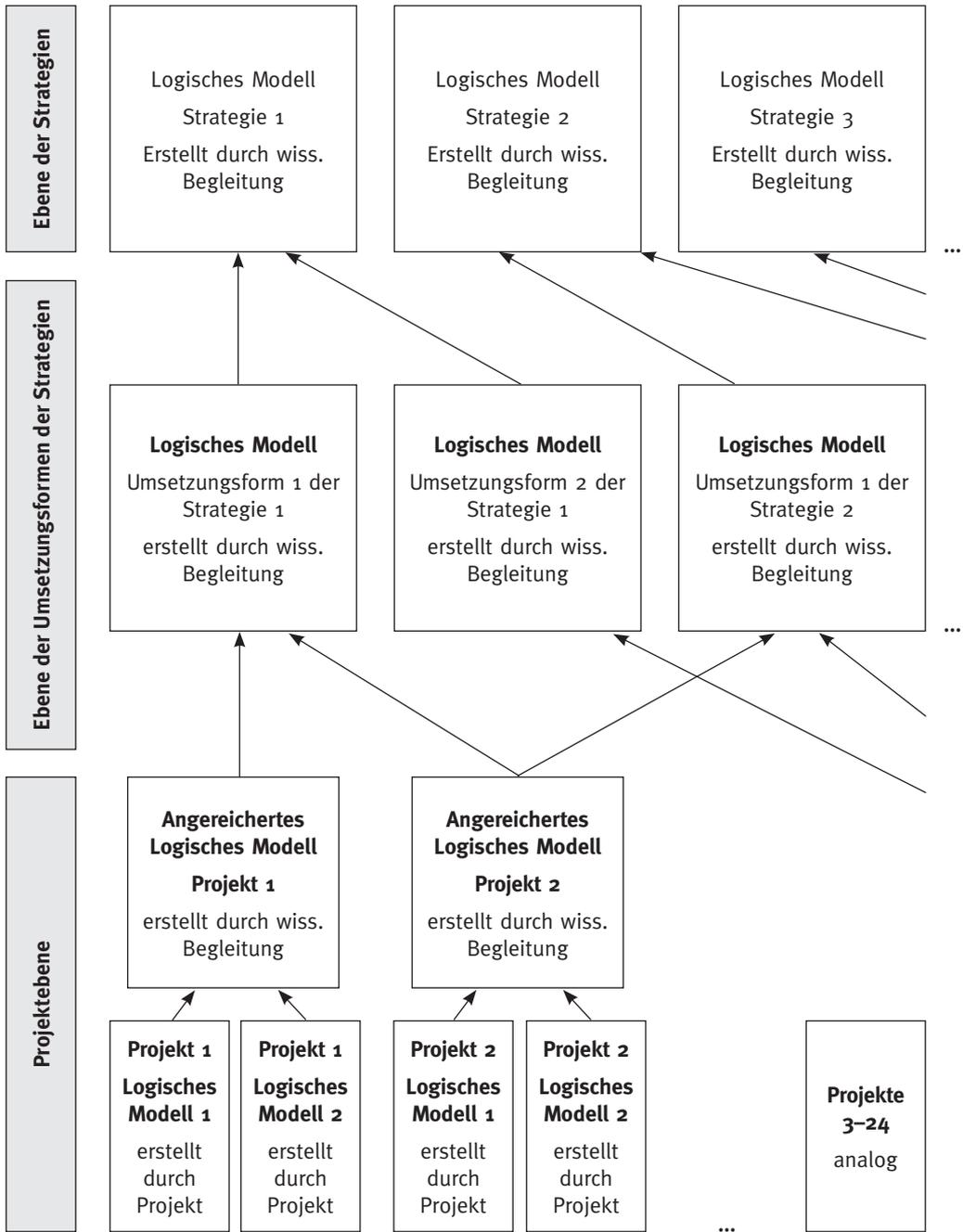
Die Programmcluster bilden im Sinne der Clusterevaluation jeweils eine Programmidee ab (z. B. „Beratung“), die auf Projektebene je nach Kontext, Ressourcen und Zielgruppe unterschiedlich variiert wurde. Aus den unterschiedlichen Variationen der Programmidee sowie den mit ihnen verbundenen Lernerfahrungen konnte im Rahmen der wissenschaftlichen Begleitung Wissen über entwickelte Konzepte und Angebote (z. B. im Bereich der „Beratung“), förderliche und hinderliche Einflussfaktoren in Bezug auf deren Umsetzung und Umsetzbarkeit sowie Kriterien für die Übertragung der Strategien in andere Kontexte generiert werden (wissensgenerierende Funktion und Perspektive der Evaluation). In der folgenden Abbildung werden die drei Abstraktionsebenen sowie die Systematik des Analyseprozesses veranschaulicht.

Die Logischen Modelle auf den drei Ebenen stehen gleichzeitig für drei klar unterscheidbare Phasen der Evaluation:

- Erhebungen und Auswertungen auf Projektebene,
- Auswertungen auf der Ebene der Umsetzungsformen der Strategien,
- Auswertungen auf Strategieebene.

Im Folgenden sollen diese näher vorgestellt werden. Dem kritischen Blick auf Potenziale und Herausforderungen und auch Grenzen des durch uns gewählten Vorgehens soll dabei ein besonders Gewicht zukommen: Was hat sich aus welchen Gründen bewährt? An welchen Stellen zeigten sich Schwierigkeiten? Wie musste das Vorgehen deshalb angepasst und verändert werden? Und welche Lernerfahrungen können wir hieraus generieren, die ggf. für weitere Evaluationen nutzbar gemacht werden können?

**Abb. 2:** Phasen der Evaluation: Einsatz Logischer Modelle auf drei Abstraktionsebenen



## **Erste Phase: Erhebungen und Auswertungen auf Projektebene**

Wie bereits beschrieben, wurden die Logischen Modelle auf Projektebene durch die Projektmitarbeitenden selbst entwickelt. Soweit dem Evaluationsteam bekannt, wurde dieser Zugang durch uns in Deutschland erstmals erprobt. Mit diesem Vorgehen waren drei wesentliche Ziele verbunden: Die Logischen Modelle sollten so weitgehend wie möglich die individuelle Handlungslogik und authentisch den geplanten Projektverlauf abbilden. Die Evaluation sollte im Rahmen der Möglichkeiten partizipations- und empowermentorientiert verlaufen und den Projektmitarbeitenden sollte mit dem „Logischen Modell“ gleichzeitig ein Instrument zur Selbstevaluation an die Hand gegeben werden, das sie in Bezug auf seine Einsatzmöglichkeiten und Potenziale einschätzen und in der Zukunft selbstbestimmt einsetzen können.

Mit dem gewählten Vorgehen war für uns als Evaluierende das positive Ergebnis verbunden, authentische Logische Modelle auf Projektebene zu erhalten, die die Handlungslogik der Projekte ohne zweiten Filter nachzeichnen.

Gleichzeitig stellte uns dieses Vorgehen aber vor zwei Herausforderungen, die den Evaluationsverlauf maßgeblich beeinflussten: Zum einen mussten die Projektmitarbeitenden in das Konzept Logische Modelle eingeführt werden, zum anderen hatten wir es als Evaluatorinnen/Evaluatoren mit individuellen Logischen Modellen zu tun, die z.T. wenig vergleichbar waren, da die jeweiligen chains of events – je nach Zuständigkeitsbereich des beteiligten Mitarbeiters/der beteiligten Mitarbeiterin – unterschiedliche Projektausschnitte nachzeichneten bzw. unterschiedlich ausgearbeitet und differenziert waren. Dies erforderte von uns zusätzliche Zeitressourcen, da z. B. Nacharbeiten auf der Basis von Telefoninterviews vorgenommen werden mussten. Eine erste, auf unseren Lernerfahrungen basierende Empfehlung für ähnlich angelegte Evaluationen lautet daher, entsprechend umfangreiche Ressourcen für die Kommunikation mit Projektmitarbeitenden von Anfang an einzuplanen.

### ***Die vorbereitenden Workshops***

Zur Einarbeitung der Projektmitarbeitenden in die Erstellung von Logischen Modellen entwickelten wir ein eigenes Workshop-Konzept. Diese Form der Zusammenarbeit ermöglichte es den beteiligten Praxisprojekten und uns, sich in einer lockeren Arbeitsatmosphäre gegenseitig kennen zu lernen, Transparenz über das gesamte Evaluationsvorhaben herzustellen und Vertrauen zu entwickeln. Insgesamt führten wir im gesamten Evaluationsverlauf drei Einführungs- und einen Auswertungsworkshop mit den am Evaluationsprozess beteiligten Projekten durch. Eingeladen wurden jeweils die Mitarbeitenden, die an der direkten Umsetzung des Projekts beteiligt waren und einen unmittelbaren Einfluss auf dessen Verlauf nehmen konnten. Hintergrund dieser Entscheidung war, dass Logische Modelle nur für den eigenen Arbeits- und Verantwortungsbereich erstellt werden können und es uns wichtig war, die konkrete methodisch-didaktische Umsetzung des Projekts in die

Analyse einzubeziehen. Trotz einer gezielt vorgenommenen Ansprache der entsprechenden Projektmitarbeitenden kam es in wenigen Fällen vor, dass

- a) sich Zuständigkeitsbereiche von Projektmitarbeitenden änderten,
- b) die konkrete Umsetzung des Vorhabens vorrangig durch Honorarkräfte übernommen wurde, an den Workshops aber die hauptamtlichen Konzeptentwickler/innen teilnahmen oder
- c) Mitarbeitende mit einem formal vergleichsweise schwachen Standing im Projekt (z. B. Praktikant/in)<sup>11</sup> zum Workshop anreisten.

Auf die mit diesen Fällen verbundenen Schwierigkeiten soll in den Ausführungen zum Bereich „Auswertung“ näher eingegangen werden. Aus unseren Erfahrungen lässt sich die folgende zweite Empfehlung ableiten: Die Auswahl der Projektmitarbeitenden, die Logische Modelle erstellen, sollte im Rahmen intensiver Vorab-Befragungen noch stringenter angelegt bzw. ggf. gezielt mehrere Mitarbeitende beteiligt werden, um einen genügend großen „Projektausschnitt“ beschreiben zu können.

Nahezu alle in das Fremdevaluationskonzept des DJI eingebundenen Projektmitarbeiter/innen blieben über die Projektlaufzeit konstant als Ansprechpartner/innen für uns erhalten<sup>12</sup>. In Bezug auf wichtige zeitliche Eckpfeiler möchten wir hervorheben, dass wir anstrebten, die ausgewählten Projekte möglichst direkt nach der Bewilligung anzusprechen und zeitnah Workshops zur Entwicklung Logischer Modelle anzubieten. Dennoch ließ es sich nicht vermeiden, dass das erste Logische Modell je nach Zeitpunkt der erfolgten Bewilligung sowie je nach Entwicklungsvoraussetzungen, -stand und -verlauf der Projekte zu unterschiedlichen Zeitpunkten erstellt wurde (in einem Fall noch vor Beginn des Projekts, in anderen Fällen mehrere Monate nach Projektbeginn). Auch wenn wir in den letztgenannten Fällen die Projektmitarbeitenden baten, die Anfangssituation im Rahmen der Erstellung ihres Logischen Modells zu rekonstruieren, nahm diese Unterschiedlichkeit einerseits einen hinderlichen Einfluss auf „Vergleichbarkeiten“ im Rahmen der kontrastierenden Auswertung und andererseits auf den Erkenntnisgewinn bei der Analyse der Logischen Modelle, da die nachträgliche Erinnerbarkeit von Anfangssituationen zumeist eingeschränkt ist. Unsere dritte Empfehlung lautet daher, sehr frühzeitig mit Programminitiator/innen und Projektverantwortlichen geeignete Termine für vorbereitende Workshops abzusprechen und festzulegen.

Die Workshops selbst waren im Regelfall durch eine offene, interessierte und engagierte Arbeitsstimmung geprägt. Sie dienten dazu, für die Projektmitarbeitenden eine transparente Verortung der Arbeit mit Logischen Modellen in unserem Evaluationsdesign

---

11 In dem genannten Fall musste das Evaluationsvorhaben abgebrochen werden, da die Ressourcen des Praktikanten nicht ausreichten, um die Evaluation in das Projekt zu implementieren und das Logische Modell zu erstellen, abzustimmen und zu aktualisieren.

12 Ich bedanke mich bei allen Mitwirkenden für die kontinuierliche, engagierte, motivierte und kritisch-reflektierte Zusammenarbeit.

vorzunehmen und die Projekte in die Erstellung von Logischen Modellen und die Arbeit mit ihnen einzuführen. Ein besonderer Schwerpunkt wurde von uns dabei auf die Ebene der Zielformulierung und -operationalisierung gelegt. Diese Fokussierung erwies sich als richtig, da es den Projektmitarbeitenden in der Regel vergleichsweise leicht fiel, die Rahmenbedingungen und Ressourcen ihrer Projektarbeit detailliert und reflektiert zu beschreiben. Schwierigkeiten zeigten sich jedoch insbesondere dort, wo es um die systematische, zwischen den Ebenen Aktivitäten und Ziele unterscheidende und Wenn-dann-Beziehungen konkretisierende Beschreibung ihrer Handlungslogik ging. Es fiel ihnen schwer, im Einzelfall Aktivitäten und Zielsetzungen zu unterscheiden und Zielsetzungen so zu operationalisieren, dass ihre Erreichung überprüfbar ist (Indikatorenentwicklung). Eine weitere Schwierigkeit bildete die Differenzierung der unterschiedlichen Zielebenen (Wissen und Emotionen, Handeln und Verhalten, Lebenslage und Status) und die Darstellung der vermuteten Kausalzusammenhänge z. B. zwischen Einstellungen und Verhaltensweisen im Logischen Modell<sup>13</sup>. Die Handlungslogik des eigenen Gesamtprojekts nachvollziehbar und widerspruchsfrei abzubilden, stellte somit keine einfache Aufgabenstellung dar.

Für den Bereich Zieloperationalisierung boten wir daher unterschiedliche Übungen, z. B. Einstufungen von fremden Zieloperationalisierungen nach Bewertungskriterien, eigene, methodisch unterstützte Operationalisierungen und moderierte Zielfindungsklärungen an. Als Qualitäts- und Bewertungskriterien für Zielformulierungen wurden die „s.m.a.r.t.“-Kriterien<sup>14</sup> herangezogen.

Zum Abschluss des Workshops wurden die Grundzüge der entwickelten Logischen Modelle im Plenum vorgestellt. Auf diese Weise konnte im Rahmen kollegialer Feedbackrunden eine gemeinsame Verständigung darüber erfolgen, ob die eigene Handlungslogik und die Zielsetzungen der Projektarbeit für Dritte nachvollziehbar dargestellt wurden. Zusätzlich entwickelten wir eine Arbeitshilfe für die Erstellung Logischer Modelle, die die Projektmitarbeitenden bei der Erstellung und in regelmäßigen Zeitabständen vorzunehmenden Aktualisierung der Modelle unterstützen sollte.

Von einigen Projektmitarbeitenden wurde das Konzept Logisches Modell anfangs als etwas starr empfunden. Es ist richtig, dass im Logischen Modell zunächst Komplexität vereinfacht dargestellt werden muss. Bei der Analyse der Workshopverläufe wurde allerdings deutlich, dass gerade durch die Fokussierung auf die eigene Handlungslogik erhebliche und tiefgehende Selbstreflexions- und Klärungsprozesse bei und unter den Projektmitarbeitenden auf der Ebene von Werten, Zielen und Handlungskonzepten angestoßen

---

13 Diese Schwierigkeiten sind nachvollziehbar und begründbar. Die Fragestellung wie Bildungs- und Wissensinhalte verarbeitet und verknüpft werden bzw. wie kognitives, emotionales sowie auf das Handeln bezogenes Lernen zusammenwirkt, ist hochkomplex und nicht nur im Rahmen wirkungsevaluativer Fragestellungen in Bezug auf den Bildungsbereich sehr aktuell. Forscher/innen aus der Sozialpsychologie kamen z. B. zu dem Ergebnis, dass Änderungen im Verhalten nicht wie vielfach angenommen zwangsläufig Einstellungsänderungen voraussetzen und Einstellungsveränderungen nicht automatisch Änderungen im Handeln und Verhalten nach sich ziehen (vgl. Eagly/Chaiken 1993).

14 Sie stehen für „spezifische“, „messbare“, „akzeptable“, „realistische“ und „terminierte“ Ziele (vgl. Heiner 1996).

wurden. Diese machten nicht nur die Projektarbeit einer (Fremd- und Selbst-)Evaluation und somit auch einer Überprüfung der Wirksamkeitsannahmen überhaupt erst zugänglich, sondern bewirkten zum Teil auch Änderungen von Projektkonzepten. Hervorgehoben werden muss: Die Rekonstruktion der impliziten Handlungslogik von Projekten bewirkte, dass Werte (z. B. unterschiedliche, emotional unterlegte Verständnisse von Integration), Ziele und Annahmen über Zielerreichungsprozesse geklärt und offen gelegt wurden/werden mussten. In den Fällen, in denen mehrere Mitarbeitende an der Erstellung des Logischen Modells beteiligt waren, kamen Aushandlungsprozesse hinzu, da Werte, Ziele und Umsetzungsvorstellungen nicht automatisch geteilt werden.

Die Erstellung Logischer Modelle und die Fremdevaluation durch das DJI veränderten somit den Projektverlauf. In diesem Sinn kann von einer Evaluation mit formativen Anteilen gesprochen werden. Für uns leitend war die Prämisse, dass die „Wirkungsmodellierung“<sup>15</sup> im Rahmen der Erstellung Logischer Modelle die Wahrscheinlichkeit der Zielerreichung erhöht. In diesem Zusammenhang ist auch das dem Beitrag vorangestellte Zitat von Stuffelbeam zu verstehen: Unsere Evaluation verfolgte, in der Arbeit mit Logischen Modellen, vorrangig das Ziel des Verbesserns, nicht des Beweisens.

### *Aktualisierung der Logischen Modelle: Projektbesuche und leitfadengestützte Interviews*

Im weiteren Verlauf der Evaluation wurden die Projekte aufgefordert, in regelmäßigen Zeitabständen ihr Logisches Modell zu aktualisieren, um den konkreten Projektverlauf sowie vorgenommene Änderungen nachvollziehbar zu machen. Zu diesem Zweck besuchten wir die Projekte vor Ort, um im Rahmen von Interviews die aktuelle Projektsituation zu beschreiben sowie Änderungen im Hinblick auf die Rahmenbedingungen, Ressourcen, Aktivitäten und Zielsetzungen diskursiv nachzuvollziehen und zu begründen. Die Interviews wurden mit den Projektmitarbeitenden auf Grundlage eines Interviewleitfadens geführt, der sich stark an der Struktur des „Logischen Modells“ orientierte (Abfrage der das Logische Modell konstituierenden Bereiche „Rahmenbedingungen“, „Ressourcen“, „Aktivitäten“ und „Ziele“).

Es handelte sich somit um eine spezifische Form des Interviews, das stärker noch als z. B. das problemzentrierte Interview (vgl. Witzel 2000) den Interviewgegenstand fokussierte und die Kommunikationssituation strukturierte.

In Bezug auf die Interviewsituationen ist hervorzuheben, dass im Rahmen der Interviews durch die gemeinsame Rekonstruktion und Überprüfung der Handlungslogik der Projektmitarbeitenden sowie der Zielerreichung (z. B. „Habe ich meine ursprüngliche Zielgruppen erreicht?“, „Wenn nicht, warum nicht?“, „Welche Rahmenbedingungen, Ressourcen und/oder Aktivitäten wirkten sich förderlich oder hinderlich auf die Zielgruppenerreichung

---

15 Vergleiche Beywl u.a. 2004. Die Autorinnen und Autoren verstehen „Wirkungsmodellierung“ als Ansatz, „Wirkungen von Programmen in theoretische/logische Argumentationsmuster einzubetten (Programmtheorie/Logisches Modell) und so argumentativ plausibel zu machen, dass die gemessenen Resultate auf die dokumentierte Programmdurchführung rückführbar sind“ (ebd., S. 19).

aus?) intensive Selbstreflexions- und Klärungsprozesse angestoßen wurden. Dies galt umso stärker, je mehr Projektmitarbeitende an dem Interview und somit an der Darstellung und Interpretation der „Ist-Situation“ sowie dem gewünschten Zustand beteiligt waren<sup>16</sup> und je höher die Bereitschaft und das Vertrauen der Interviewten war, sich in Bezug auf unsere Evaluation zu öffnen. Der stark strukturierende und fokussierende Charakter der Interviews führte somit einerseits zu einer Fülle an dichtem, stark selbstreflexiv geprägten Datenmaterial, zum anderen begrenzte das Interview Komplexität, indem es bestimmte Projektbereiche fokussierte und andere – nicht programmgeförderte – Bereiche ausblendete. Weiterhin wurden durch die Visualisierungen und Rekonstruktionen im Rahmen der Erstellung Logischer Modelle Vereinfachungen vorgenommen, indem modellhaft sichtbar gemacht wurde, welche Zielsetzungen mit welchen Handlungsschritten auf der Grundlage welcher Rahmenbedingungen und Ressourcen erreicht werden sollten. Mit anderen Worten: In den Interviews konnte nicht der gesamten Komplexität des Projektalltags Rechnung getragen werden, sondern das Logische Modell lieferte einen stark strukturierenden und gegenstandsbestimmenden Rahmen. Es durfte daher sowohl bei der Erhebung als auch bei der Auswertung und Darstellung der Evaluationsergebnisse nicht aus dem Blick verloren werden, dass „Modelle die Wirklichkeit nicht umfassend abbilden können“ (Haubrich 2006, S. 119), sondern vereinfachend den Blick auf bestimmte, vermutete bzw. zu überprüfende Zusammenhänge richten.

Ergänzend wurden in den Fällen, in denen es das Projekt-Setting, die Ressourcen der Projektmitarbeitenden sowie die Ressourcen der wissenschaftlichen Begleitung ermöglichten, punktuell teilnehmende Beobachtungen durchgeführt.

### *Auswertung des Datenmaterials*

Die Auswertung des Datenmaterials wurde von mehreren Herausforderungen geprägt. Zum einen war in Bezug auf jedes einzelne Projekt eine Fülle an Datenmaterial entstanden/erhoben worden, das Logische Modelle, Interviews, Netzwerkkarten, Fragebögen mit Follow-up-Modulen, Projektberichte, Projektmaterialien und Ergebnisberichte für den Fördergeber umfasste. Das Datenmaterial war darüber hinaus gezielt zu unterschiedlichen Zeitpunkten erfasst worden und zeichnete im Längsschnitt einen Projektverlauf nach.

Zum anderen war das erhobene Datenmaterial stark heterogen: Schon beim Einzelprojekt waren einzelne Datenquellen unterschiedlich stark differenziert und umfangreich – bei dem interprojektbezogenen Vergleich der Datenmaterialien wurde darüber hinaus deutlich, dass auch die einzelnen Logischen Modelle unterschiedlich detailliert waren und zudem differente Projektausschnitte beschrieben. So hatte ein Geschäftsführer, der allein für mehrere Projekte und Aktivitäten zuständig war, aus seiner Perspektive ein sehr umfassendes Bild seiner Projektaktivitäten nachgezeichnet, also ein „weites Objektiv“

---

<sup>16</sup> Die Anzahl der am Interview teilnehmenden Mitarbeiter/innen reichte je nach Projektgröße und -anlage von einer Einzelperson bis zu vier Personen.

gewählt. Ein Mitarbeiter eines anderen Projekts konnte in seinem Logischen Modell nur den Projektausschnitt nachzeichnen, den er selbst in der Verantwortung hatte: Die Durchführung eines internationalen Austauschs im Inland. Er hatte demnach im Vergleich zu dem Geschäftsführer ein engeres Objektiv ausgewählt/auswählen müssen, zeichnete aber ein stark differenziertes und detailliertes Bild der geplanten Aktivitäten. Um es zugespitzt zu formulieren: Durch die sorgfältig vorbereitete Auswahl der an der Evaluation beteiligten Projektmitarbeiter/-innen nach ihrem Verantwortungsbereich, das strukturierende Element der Logischen Modelle sowie die Orientierung an Projektleitfäden gelang es zwar, die dem Untersuchungsgegenstand „Programm Entimon“ inhärente Heterogenität zu begrenzen, nicht aber, sie gänzlich aufzulösen. Das in anderen Zusammenhängen gewählte Vorgehen, die Logischen Modelle durch die Evaluierenden in standardisierter Form zu entwickeln, hätte die Heterogenität ggf. weiter verringert, dies aber um den Preis der Individualität und Authentizität der einzelnen Logischen Modelle und nachgezeichneten Handlungslogiken.

Mit den Herausforderungen gingen wir wie folgt um: In Bezug auf die zwischen den Projekten differierende Detailliertheit der Logischen Modelle nahmen wir einige Nacherhebungen bei Einzelprojekten vor, um die Datengrundlagen homogener zu gestalten. Differenzen in Bezug auf die Größe der dargestellten Projektausschnitte berücksichtigten wir darüber hinaus in der vergleichenden Auswertung der Projekte.

Die Datenfülle zum Einzelprojekt brachte darüber hinaus die Notwendigkeit mit sich, das Datenmaterial im Rahmen von Einzelfallanalysen zu integrieren und zu strukturieren. Als integrierendes Moment bot sich wiederum die Folie des Logischen Modells an. Für die Einzelfallanalysen wurden jeweils die gesamten projektbezogenen Materialien in das Textanalyseprogramm MAXqda eingelesen und codiert, wobei der Codebaum sich an der Struktur und den unterschiedlichen Analysebereichen Logischer Modelle (vgl. Abb. 1) orientierte.

Die leitenden Fragen für die Einzelfallanalysen fokussierten den Einfluss förderlicher und hinderlicher Rahmenbedingungen und Ressourcen auf die Umsetzung der einzelnen Handlungsschritte, die Lernerfahrungen der Projektmitarbeitenden in Bezug auf die inhaltliche und methodische Umsetzung des Projekts sowie die Zielerreichungsindikatoren und -prozesse. Basis unserer Auswertungen waren somit in allen Fällen die Selbstauskünfte der Projektmitarbeitenden.

Im Rahmen der Auswertung entstanden auf diese Weise für jedes Einzelprojekt ein um alle Datenquellen angereichertes Logisches Modell in Fließtextform, das im Rahmen eines Gesamtmodells zugleich den zeitlichen Verlauf der Projektentwicklung nachvollzog, d.h. zum Beispiel der Frage nachging, welche Aktivitäten im Projektverlauf warum und in welcher Form adaptiert und verändert wurden. Da sich die Interpretation der Textmaterialien auf die Selbstauskünfte der Projektmitarbeitenden stützte, haben wir die unterschiedlichen Datenquellen bei den Integrations- und Auswertungsprozessen zur Erhöhung der Validität priorisiert: Den Projektinterviews wurde grundsätzlich ein stärkeres Gewicht zugesprochen als z. B. den Ergebnisberichten, da wir vor dem Hintergrund unserer Evaluationserfahrungen davon ausgingen, dass im Regelfall die unter Zusicherung von

Anonymität und im persönlichen Gespräch vor Ort gegebenen Auskünfte selbstreflexiver und weniger stark vom Einflussfaktor antizipierte soziale Erwünschtheit geprägt waren als Ergebnis- und Sachberichte.

Zusätzlich triangulierten wir die auf Selbstauskünften basierenden Datenquellen mit Erkenntnissen aus durchgeführten teilnehmenden Beobachtungen. Widersprüche zwischen einzelnen Datenquellen, im Logischen Modell selbst oder zwischen den Aussagen von Projektmitarbeitenden aus dem gleichen Projekt wurden unter den Evaluatorinnen und Evaluatoren diskutiert, zum Gegenstand von Interpretationen gemacht und diese in mehreren Fällen mit den entsprechenden Mitarbeitenden im Rahmen angekündigter und transparent angelegter Telefoninterviews kommunikativ validiert. Auf der Basis des bereits angelegten Vertrauensverhältnisses verliefen die Telefoninterviews reflektiert und offen. Wir können daher begründet davon ausgehen, dass sie einen wichtigen Beitrag zur Validierung unserer Auswertung leisteten.

Die angereicherten Logischen Modelle stellten als Einzelfallanalysen die Grundlage für die weiteren Analyseschritte zur Erschließung der Programmdimensionen dar (vgl. Abb. 2).

### **Zweite und dritte Phase: Auswertungen auf der Ebene der Umsetzungsformen der Strategien und der Ebene der Strategien**

Ziel der Evaluation war es, Programmpraxis sichtbar zu machen und in Bezug auf ihre Potenziale zur Weiterentwicklung der Fachpraxis zu beschreiben. Programmpraxis und nicht „nur“ Projektpraxis sichtbar zu machen, hieße gleichzeitig, die Programmdimensionen herauszuarbeiten.

In einem zweiten abstrahierenden Schritt wurden daher die Einzelfallanalysen vergleichend und kontrastierend betrachtet. Ziel der vergleichenden Auswertung war es, Gruppen von Projekten empirisch herauszuarbeiten und zu beschreiben, die starke Ähnlichkeiten in Bezug auf die umgesetzten Ansätze, die angesprochenen Zielgruppen sowie die formulierten Zielsetzungen aufwiesen.

Dabei zeigte sich als zusätzliche Schwierigkeit, dass weniger die Einzelprojekte in ihrer Gesamtheit und Komplexität vergleichbar waren – diese setzten komplexe Kombinationen von abgrenzbaren Einzelmaßnahmen bzw. -aktivitäten (z. B. einen internationalen Austausch im Ausland und Seminare politischer Bildung in Deutschland, die sich jeweils an andere Zielgruppen wandten) um – sondern dass es vielmehr Projektbereiche – und hier Ansätze wie z. B. Beratung oder Begegnung – waren, die sich in mehreren Projekten wiederfanden und konsistent in Bezug auf ein Zusammenwirken von Rahmenbedingungen, Ressourcen, Aktivitäten und Ziele beschrieben werden konnten. Dieses Zusammenwirken beschrieben wir als „Strategie“.

Dieses Vorgehen löste auch ansatzweise die beschriebenen Differenzen in Bezug auf die unterschiedliche Größe des im Logischen Modell beschriebenen Projektausschnitts auf, da nunmehr Projektbereiche – und nicht das Gesamtprojekt – in den Fokus der Analyse genommen wurden. Deutlich hervorgehoben werden soll: Die Identifizierung von „programm-

förmigen“ Projektbereichen konnte in unserem Fall nicht wie bei der klassischen Cluster-Evaluation zu Beginn der Evaluation vorgenommen werden, sondern erfolgte induktiv im Verlauf der Evaluation.

Auf der zweiten Abstraktionsebene konnten wir dabei kleinere Gruppen von Projektaktivitäten typologisieren, die sich mit einem ähnlichen Ansatz an ähnliche Zielgruppen wandten (z. B. „Beratung von Eltern rechtsextremer Kinder“, „Beratung von Kommunen“ oder „Internationale Begegnung im Ausland unter Einbezug interreligiöser Fragestellungen“, „Begegnung im Inland unter Einbezug statusungleicher Gruppen“). Diese bezeichneten wir als „Umsetzungsformen der Strategien“.

Auf einer dritten Abstraktionsebene konnten wir diese Gruppen empirisch weiter fassen und beschreiben. Auf diese Weise arbeiteten wir die Programmstrategien (z. B. „Beratung“ und „Begegnung“) heraus, die als Programmcluster jeweils unterschiedliche Programmideen (z. B. „Prävention von Rechtsextremismus und Förderung von Demokratie durch Beratung“, „Prävention von Rechtsextremismus und Förderung von Demokratie durch Begegnung“) abbilden. Im wenig formalisierten Setting des Programms „Entimon“ setzten die Projekte verschiedene als Programmstrategien zu bezeichnende Ansätze um, die durch die Umsetzenden des Programms als relevante pädagogische, bildnerische bzw. vernetzende Strategien in der Prävention von Rechtsextremismus und Gewalt bzw. der Förderung von Toleranz und Demokratie definiert worden sind. Um diese Strategien im Hinblick auf ihre Potenziale und Grenzen näher zu beschreiben, beantworteten wir die gleichen Auswertungsfragen wie auf der Projektebene auch für die typologisierten Gruppen von Projekten und erstellten Logische Modelle, die das jeweilige Zusammenwirken von Rahmenbedingungen, Ressourcen, Aktivitäten und Zielerreichungen beschreiben. Aus dem Vergleich der Lernerfahrungen der Projekte konnten gleichzeitig Voraussetzungen für eine Übertragung der Strategien in andere Kontexte abgeleitet werden, indem auf diese Weise u.a. notwendige Umsetzungsressourcen herausgearbeitet wurden.

Die Evaluationsergebnisse wurden den Projektmitarbeitenden auf einer Zwischenstufe der Auswertung vorgestellt und rückgekoppelt. Auch an dieser Stelle bot ein gemeinsamer Workshop Raum für Fragen, Validierungsprozesse und Diskussionen.

Da im Gegensatz zur „klassischen Form“ der „Cluster-Evaluation“ nicht alle geförderten Projekte in das Untersuchungscluster einbezogen werden konnten, wurde in einem letzten Auswertungsschritt eine zusätzliche qualitative Analyse aller geförderten Projekte auf der Basis der in der Datenbank „Entimon“<sup>17</sup> gespeicherten Projektdaten durchgeführt. Ziel war eine Verortung des qualitativen Untersuchungssamples und der auf dieser Basis erzielten Evaluationsergebnisse. Es zeigte sich, dass alle Projekte den herausgearbeiteten Programmstrategien zugeordnet werden konnten.

---

17 Vergleiche [www.dji.de/entimon](http://www.dji.de/entimon).

## Fazit

Das Instrument Logisches Modell erfüllte für uns im Evaluationsverlauf vielfältige Funktionen, die wir im Folgenden zusammenfassen: Es ermöglichte den Nachvollzug der Handlungslogik und der Wirkungsannahmen auf der Projekt-, Strategieumsetzungsform- und Strategieebene im Rahmen eines induktiv-rekonstruierenden Verfahrens, diente als strukturierendes Element der Kommunikation mit den Projektmitarbeitenden und ermöglichte weiterhin die Erschließung der inhaltlichen Dimensionen des Programms Entimon durch die empirische Herausarbeitung der Programmstrategien. Es bildete weiterhin das strukturierende Element der Darstellung und Präsentation der Evaluationsergebnisse.

Den Projektmitarbeitenden wurde mit dem Logischen Modell zugleich ein Instrument in die Hand gegeben, das im Rahmen von Selbstevaluationsprozessen nutzbar gemacht werden kann. Wie unsere Befragungen zeigten, schätzten die Fachkräfte der Praxisprojekte das Logische Modell größtenteils als hilfreich zur Überprüfung und Bewertung ihrer eigenen Praxis ein. Sie sahen allerdings die Notwendigkeit, das Logische Modell in der Selbstevaluation zu adaptieren und Projektausschnitte und Handlungsschritte kleinteiliger in den Blick zu nehmen, um der Komplexität des Projektalltags und ihrer Projektperspektive gerecht zu werden. Sie nutzen das Logische Modell weiterhin, um die Projektkonzeption und Projektergebnisse öffentlich vorzustellen.

Für uns bleibt festzuhalten: Das Logische Modell ermöglichte es uns, unserer Evaluationsfragestellung fokussiert und strukturiert nachzugehen. Es trug sowohl auf der Seite der Projektmitarbeitenden als auch auf unserer Seite grundlegend zur Klärung der Projekt- und Programmdimensionen bei. Zugleich beschnitt es an einigen Stellen notwendigerweise Komplexität, um die Programmperspektive in der Evaluation wahren zu können. Wie auch die Projektmitarbeitenden deutlich machten, verlangt es kleinteilige Adaptionen, die stärkere Beachtung zirkulär verlaufender Prozesse etc., um sich auf der Projektebene der Frage nach Wirkungen der Projektarbeit anzunähern. Auf der Ebene der Programmevaluation erwies sich die Arbeit mit Logischen Modellen sowohl in der Vorbereitung und Durchführung der Erhebungen als auch in der Auswertung als sehr zeitaufwendig. Die Notwendigkeit der Erarbeitung von Workshop-Konzepten, die Durchführung von telefonischer Beratung zur Entwicklung von Logischen Modellen, Rückkopplungen und Nachbefragungen beschnitten die Ressourcen der Evaluierenden an anderen Stellen. So konnten wir im Rahmen unserer Evaluation die umgesetzten Konzepte und Ansätze detailliert beschreiben – die zeitlichen Ressourcen reichten aber nicht aus, um regelmäßige teilnehmende Beobachtungen und/oder Adressatenbefragungen durchzuführen (Einbezug der Adressatenebene). Diese ergänzend und systematisch durchzuführen, sehen wir im Sinne einer erwünschten Triangulation der Daten für weitere Evaluationen als fruchtbar an.

## Literaturverzeichnis

- Beywl, Wolfgang (2006): Demokratie braucht wirkungsorientierte Evaluation – Entwicklungspfade im Kontext der Kinder- und Jugendhilfe. In: Projekt eXe (Hrsg.): Wirkungsevaluation in der Kinder- und Jugendhilfe: Einblicke in die Evaluationspraxis. München: DJI, S.25–46
- Beywl, Wolfgang/Speer, Sandra/Kehr, Jochen (2004): Wirkungsorientierte Evaluation im Rahmen der Armut- und Reichtumsberichtserstattung. Perspektivstudie. Online unter: [www.bmg.bund.de/nn\\_603382/SharedDocs/Publikationen/Forschungsprojekte-Lebenslagen/a-323-10006,templateld=raw,property=publicationFile.pdf/a-323-10006.pdf](http://www.bmg.bund.de/nn_603382/SharedDocs/Publikationen/Forschungsprojekte-Lebenslagen/a-323-10006,templateld=raw,property=publicationFile.pdf/a-323-10006.pdf) (29.06.2007)
- BMFSFJ (2006): Leitlinien zur Umsetzung des Programms „Entimon – gemeinsam gegen Gewalt und Rechtsextremismus“ – [www.entimon.de/content/e28/e34/index\\_ger.html](http://www.entimon.de/content/e28/e34/index_ger.html) (30.05.2007)
- Eagly, Alice/Chaiken, Shelly (1993): The Psychology of Attitudes. Fort Worth, TX: Harcourt Brace Jovanovich
- Haubrich, Karin (i. E.): Evaluation innovativer multizentrischer Programme: Möglichkeiten und Grenzen programmtheoriebasierter Cluster-Evaluation am Beispiel von Bundesmodellprogrammen in der Kinder- und Jugendhilfe. Dissertation an der Freien Universität Berlin
- Haubrich, Karin (2006): Wirkungsannahmen sichtbar machen: Cluster-Evaluation innovativer multizentrischer Programme. In: Projekt eXe (Hrsg.): Wirkungsevaluation in der Kinder- und Jugendhilfe: Einblicke in die Evaluationspraxis. München: DJI, S. 101–122
- Haubrich, Karin (2003): Die Evaluation von Modellprogrammen im Spannungsfeld zwischen Wirkungs- und Nutzungsorientierung. In: Loidl-Keil, Rainer/Viechtbaur, Karin (Hrsg.): Evaluation in der außerschulischen Kinder- und Jugendarbeit in Österreich – Berichte aus der Praxis. Linz: Trauner, S. 205–222
- Haubrich, Karin/Frank, Kerstin (2000): Vom Aufsuchen zur beruflichen Integration. Evaluationsstudie zum Bundesmodellprogramm „Mobile Jugendsozialarbeit für junge Menschen ausländischer Herkunft“. DJI-Arbeitspapier Nr. 1–156. München: DJI
- Haubrich, Karin/Vossler, Andreas (2001): Das Modellprogramm „Interkulturelles Netzwerk der Jugendsozialarbeit im Sozialraum“ – Erste Ergebnisse und Erfahrungen. In: BAG JAW 8 (Hrsg.): Jugend-Beruf-Gesellschaft, Arbeitsdruck 39. Sozialanalyse. Bonn, S. 22–33
- Haubrich, Karin/Holthusen, Bernd/Struhkamp, Gerlinde (2005): Evaluation – einige Sortierungen zu einem schillernden Begriff. In: DJI Bulletin Sonderteil, 2005, Heft 72, S. 1–4
- Heiner, Maja (Hrsg.) (1996): Qualitätsentwicklung durch Evaluation. Freiburg: Lambertus-Verlag

- Klingelhöfer, Susanne/Schmidt, Mareike/Schuster, Silke/Brüggemann, Ulrich (2007): Abschlussbericht der wissenschaftlichen Begleitung des Programms „Entimon – gemeinsam gegen Gewalt und Rechtsextremismus“ 2002 bis 2006, Halle: DJI
- Lüders, Christian/Haubrich, Karin (2006): Wirkungsevaluation in der Kinder- und Jugendhilfe: Über hohe Erwartungen, fachliche Erfordernisse und konzeptionelle Antworten. In: Projekt eXe (Hrsg.): Wirkungsevaluation in der Kinder- und Jugendhilfe: Einblicke in die Evaluationspraxis. München: DJI, S. 5–22
- Parsons, Beverly A./Haubrich, Karin (2002): Features of Cluster Evaluation, An Emerging Methodology. Paper presented at the annual meeting of the American Evaluation Association, Washington
- Rossi, Peter H./Lipsey, Mark W./Freeman, Howard E.: Evaluation (2004): A Systematic Approach, 7th Edition. Thousand Oaks/London/New Delhi
- Vossler, Andreas/Obermaier, Andrea (2003): Netze knüpfen – Integration fördern. Evaluationsstudie zum Bundesmodellprogramm „Interkulturelles Netzwerk der Jugendsozialarbeit im Sozialraum“. München
- Weiss, Carol: Excerpts from Evaluation Research (1996): Methods of Assessing Program Effectiveness 1. In: American Journal of Evaluation. Vol. 17, S. 173–175
- Witzel, Andreas (2000): Das problemzentrierte Interview. In: Forum Qualitative Sozialforschung/Forum: Qualitative Research [Online Journal], 1(1). [www.qualitative-research.net/fqs-texte/1-00/1/1-00witzel-d.htm](http://www.qualitative-research.net/fqs-texte/1-00/1/1-00witzel-d.htm) (30.05.2007)

## Vergleichende Evaluation von Interventionen gegen Rechts- extremismus: Perspektiven professionsorientierter wissens- generierender Evaluationen

### Zur Einordnung von Evaluationen in der Sozialen Arbeit

Der vor allem in Deutschland geführten Debatte um Qualität, Qualitätsstandards, Rahmenrichtlinien, Leitlinien oder Guidelines ist es zu verdanken, dass sich die Evaluationstätigkeit in der Sozialen Arbeit von einem reinen Untersuchungsauftrag zu einem mehrdimensionalen Ansatz mit Anspruch, Evaluationsforschung zu sein, verbreitert hat. Heute liegt aus dem deutschsprachigen Raum eine „nicht mehr überschaubare Anzahl an Hand- und Grundlagenbüchern“ (Speck 2006, S. 38) vor, die u.a. eine Broschürenreihe (im Rahmen der Bundesinitiative „Qualitätssicherung in der Kinder- und Jugendhilfe“ entstanden) und Themenhefte von Zeitschriften umfasst. In der Schweiz, wo die im Folgenden dargestellte Studie durchgeführt wurde, sind sowohl der unmittelbare Zusammenhang zwischen Steuerung und wirkungsorientierter Evaluation als auch die damit verbundene Fachdiskussion inklusive Fachliteratur erst in Teilen diffundiert.

Seit dem Jahr 2000 ist zwar in Art. 170 der Schweizer Bundesverfassung verankert, dass die Maßnahmen des Bundes auf ihre Wirksamkeit hin zu überprüfen sind, doch kann derzeit nicht von einer systematischen Evaluationstätigkeit in der Schweiz gesprochen werden (vgl. Bussmann 2005). Außerdem differieren die Evaluationsansätze, -vorgehensweisen oder -praktiken in den Kantonen stark und es besteht die Tendenz, die Aktivitäten anderer Kantone zu ignorieren (vgl. Kaufmann 2006). Dies betrifft auch Programme und Projekte der Kantone und Gemeinden im Bereich der Sozialen Arbeit. Sie werden eher sporadisch fallweise oder exemplarisch auf ihre Wirkung hin betrachtet.

Eine Datenbank für Evaluationen in der Schweiz fehlt ebenso wie ein Diskurs über Evaluationsforschung in der Sozialen Arbeit. Anhaltspunkte hierzu gibt eine kürzlich durchgeführte Sichtung von wissenschaftlichen Evaluationsstudien in der Sozialen Arbeit, die sich auf Recherchen in verschiedenen Forschungsdatenbanken (vor allem Schweizerisches Informations- und Datenarchiv SIDOS sowie Schweizerischer Nationalfonds) stützte und versuchte, aus wissens- und professionsorientierter Sicht die Evaluationsstudien zu typisieren (vgl. Drilling 2007). Insbesondere sollte dabei herausgearbeitet werden, inwieweit es sich um „wissensgenerierende“ Evaluationen (Haubrich u.a. 2005) handelt, die eine über den direkten Nutzen der Auftraggeber hinausreichende Entwicklung von Fachlichkeit anstreben. Dies zu betrachten ist aus professionstheoretischer Perspektive interessant, da sich die in den 1970er Jahren gepflegte Idee eines direkten Transfers von wissenschaftlichem Wissen in berufliches Handeln als weitgehend illusorisch erwiesen hat. Diese expertokratische Vorstellung der Wissensverwendung ist einem Modell gewichen, das den konstitutiven Differenzen zwischen wissenschaftlichem Wissen, Erfahrungswissen und

professionellem Wissen Rechnung trägt. Wissenschaftliches Wissen wird hier mit einem Fallverständnis relationiert oder anders ausgedrückt: Das Verhältnis von sozialwissenschaftlicher Theorie und professioneller Handlungspraxis erfährt eine Überbrückung durch eine auf Wissensgenerierung abzielende Evaluationstätigkeit.

Demnach lassen sich drei Typen von wissenschaftlichen Evaluationen in der Sozialen Arbeit unterscheiden (vgl. Drilling 2007):

- Auftragsorientierte wissensverwertende Evaluationen,
- auftragsorientierte wissensvermittelnde Evaluationen,
- professionsorientierte wissensgenerierende Evaluationen.

Während sich die beiden ersten Typen streng an den Fragestellungen der Auftraggebenden orientieren, eine theoretische Auseinandersetzung eher intern unter Evaluierenden stattfindet, ist bei der professionsorientierten wissensgenerierenden Evaluation die Anreicherung der evaluativen Fragestellung mit einer berufsbezogenen Perspektive der markanteste Unterschied. Die hier dargestellte Evaluation fragt „Was kann Soziale Arbeit in Bezug auf Rechtsextremismus ausrichten? Unter welchen Bedingungen? Mit welchen Ansätzen?“ und versucht aus den Evaluationsergebnissen „Guidelines“ zu entwickeln, welche die in der Sozialen Arbeit tätigen Personen im Umgang mit rechtsextremen Vorfällen anleiten.

Zudem werden beim Typ professionsorientiert wissensgenerierende Evaluation im Hinblick auf die Übertragbarkeit der Ergebnisse Modelle entwickelt bzw. anhand der Evaluationsergebnisse präzisiert und diskutiert. So entsteht eine zweite Forschungsebene, die zur Evaluation in eigener Sache hinzutritt, wobei die Fragestellung, die durch die Forschenden generiert wird, eine Passung zwischen Evaluation und Forschungsinteresse der Sozialen Arbeit darstellt. Dieses so erworbene Wissen wird „offen“ generiert, d.h. diskursiv im Bericht unter Rückgriff auf die Literatur dargelegt. Auch die im Folgenden dargestellte Evaluation, die verschiedene Interventionen der Sozialen Arbeit gegen Rechtsextremismus und Fremdenfeindlichkeit analysierte, vereint so drei Zieldimensionen:

- Die Interventionsanbietenden erhalten Antworten über die Wirkung ihrer Angebote anhand der Evaluationsergebnisse,
- die Interventionsanbietenden sowie die am Fachdiskurs interessierten Fachpersonen erhalten Einblicke in eine aktuelle fachwissenschaftliche Debatte anhand konkreter Beispiele (Wissenstransfer),
- die im Arbeitsfeld Rechtsextremismus und Fremdenfeindlichkeit tätigen Fachpersonen erhalten eine theoriebasierte Argumentationshilfe, die sowohl internationale Debatten berücksichtigt als auch Überlegungen für weitere Aktivitäten umfasst (Professionalisierung).

## Evaluationsstudie „Soziale Arbeit und Rechtsextremismus – Evaluation von Interventionsansätzen und Entwicklung von Guidelines“

Unsere Evaluationsstudie mit dem oben aufgeführten Titel war Teil des Forschungsprogramms „Rechtsextremismus: Ursachen und Gegenmaßnahmen“ des Schweizer Nationalfonds (Nationales Forschungsprogramm 40+, [www.nfp40plus.ch](http://www.nfp40plus.ch)), dessen Ziele Grundlagenforschung, Wissenstransfer und Umsetzung der Forschungsergebnisse sind. Unser Interesse galt innovativen Interventionsansätzen in lokal eingegrenzten Kontexten, wofür wir eine wissenschaftsgenerierende professionsorientierte Evaluationsstudie konzipierten.

Zu Beginn orientierten wir uns an einem quasi-experimentellen Design, auf dessen Grundlage wir erfolgreich antirassistische Bildungsprojekte evaluiert hatten (vgl. Eser Davolio 2000; Eckmann/Eser Davolio 2003). Dabei waren wir uns der Schwierigkeiten der Adaptierbarkeit bewusst, ist doch die Komplexität der Wirkungsanalyse von Interventionen der Sozialen Arbeit, insbesondere wenn es sich um Gemeinwesenarbeit handelt, erheblich größer. Ausgehend vom Anspruch, Interventionen über ihren ganzen Zeitraum von Beginn bis Abschluss zu begleiten, sahen wir eine wissenschaftliche Begleitung der Interventionen mit Vor- und Nachbefragungen der beteiligten Akteure vor, um deren Wirkung und Nachhaltigkeit zu überprüfen. Dieses Forschungsdesign stellte sich in der Folge als wenig realitätsnah heraus, da die Interventionsanbieter keine Interventionen gegen Rechtsextremismus auf Lager hatten, die in unser vorgegebenes Zeitfenster gepasst hätten und zudem eine hohe Abbruchquote bestand. Bevor weiter auf die Evaluationsbedingungen eingegangen werden soll, hier eine Beschreibung unserer Studie, die sich in zwei Forschungsetappen gliedert:

- Vorstudie: Befragung von Akteuren (Schule, Polizei, Jugendarbeit) in Gemeinden mit manifesten Rechtsextremismusproblemen,
- Hauptstudie: Evaluation von sieben Interventionen gegen Rechtsextremismus, welche von auf Rechtsextremismus spezialisierten Expertinnen/Experten der Sozialen Arbeit begleitet wurden.

Der Grund für diese Zweigliederung war die Absicht, mit der Vorstudie Gemeinden bezüglich ihrer Wahrnehmung und Reaktionen auf die vorhandenen Rechtsextremismusprobleme zu untersuchen, wofür wir in den neun untersuchten Gemeinden lokale Akteure aus den Bereichen Polizei, Sozialdienst und Schule sowie Aktivistinnen, Aktivistinnen und Politiker/innen (N=33) interviewten. Mit der Hauptstudie hingegen wollten wir von Expertinnen/Experten begleitete Interventionen evaluieren. Unsere Erwartung bezüglich der Vorstudie, dass nämlich nicht alle konfrontierten Gemeinden auch auf die vorhandenen Rechtsextremismusprobleme reagieren würden, erwies sich als unbegründet. Das ausgelöste Medienecho und der damit verbundene drohende Imageverlust hatten in den

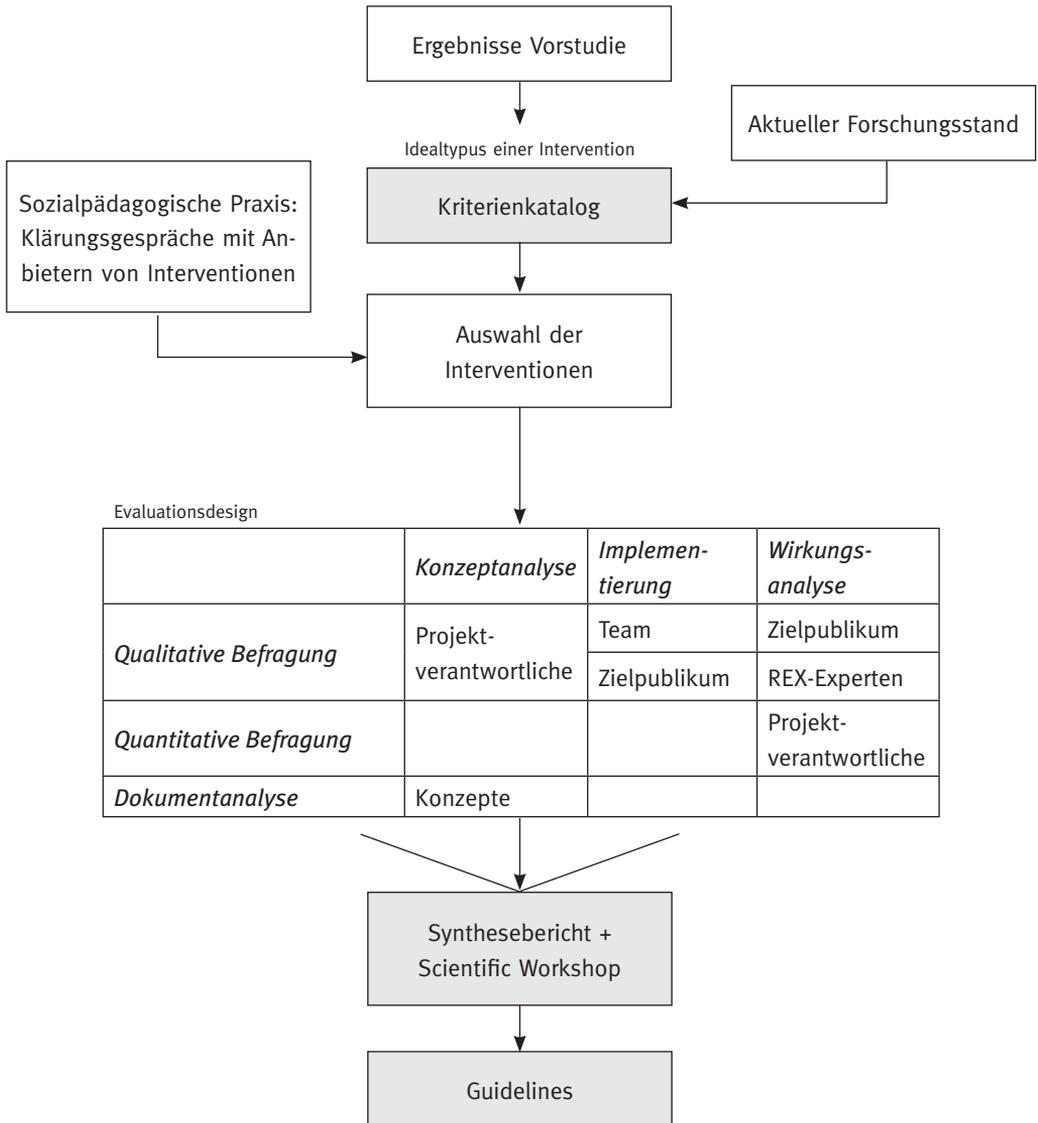
betroffenen Gemeinden einen politischen Handlungsdruck ausgelöst, der zu einer Mobilisierung der lokalen Akteure geführt hatte. Die Tatsache, dass wir auf keine indifferenten Gemeinden gestoßen sind, steht in Zusammenhang mit der Medienbezogenheit des ersten unserer drei Auswahlkriterien:

- Die Häufigkeit von Medienberichterstattungen über rechtsextremistische Vorfälle im Zeitraum von 2000 bis 2003,
- das Vorhandensein einer lokal ansässigen rechtsextremen Szene,
- die Bereitschaft der Gemeinde sich für unsere Evaluationsstudie zur Verfügung zu stellen.

Für die Hauptstudie wählten wir insgesamt sieben von Expertinnen und Experten begleitete Interventionen aus. Die Auswahl der Interventionen erfolgte danach, ob sich die Intervention ausdrücklich gegen Rechtsextremismus richtete, mindestens sechs Monate dauerte und den Kontext bzw. das Umfeld als auch verschiedene Gemeindeakteure mit einbezog.

Bei der Planung der Vorstudie bestand unser Erkenntnisinteresse in erster Linie darin, zu untersuchen, ob die betroffenen Gemeinden Rechtsextremismus als Problem wahrnehmen und Handlungsbedarf sehen. Durch das Auswahlkriterium der erfolgten Berichterstattung über rechtsextremistische Vorfälle erfassten wir jedoch ausschließlich Gemeinden, in denen aufgrund des Medienechos ein politischer Handlungsdruck entstanden war, weshalb in allen untersuchten Gemeinden auf die Probleme reagiert worden war – meist ohne Expertenbegleitung. Das heißt jedoch nicht, dass es in der Schweiz keine Gemeinden mit fehlendem Problembewusstsein trotz rechtsextremistischer Vorkommnisse gibt, wurden uns doch solche Fälle von einigen Interventionsanbietenden geschildert. Dadurch wurde unsere Vorstudie zu einer Art Kontrollgruppenvergleich, indem wir unbegleitete Interventionen solchen mit Expertenbegleitung gegenüberstellen konnten.

Folgendes Schema gibt Auskunft über den Aufbau unserer Evaluation und die angewandten Instrumente der Hauptstudie:



**Tab. 1:** Übersicht über die Forschungsphasen der Hauptstudie (REX steht für Rechtsextremismus)

## Modelle und Instrumente der Evaluation

Zur Bestimmung manifester rechtsextremistischer Phänomene haben wir angelehnt an das Modell zur Mobilisierungstheorie von Pierre Gentile et al. (1995), wonach rechtsextreme Phänomene den drei Kategorien „demonstrativ“, „konfrontativ legal“ und „konfrontativ illegal“ zugeordnet werden, die Kategorie „demonstrativ“ zusätzlich unterteilt und so folgendes erweitertes 4-Stufen-Modell entwickelt:

- Demonstrativ legale Phänomene (z. B. Petitionen, Unterschriftensammlungen, bewilligte Versammlungen, Propaganda, Feste, Konzerte, Leser- und Protestbriefe etc.),
- demonstrativ illegale Phänomene (Unbewilligte Versammlungen und Konzerte, Sprayereien und Schmierereien mit rassistischem/nazistischem Bedeutungsgehalt etc.),
- konfrontativ legale Phänomene (z. B. Ehrverletzungsklagen, massives Auftreten in Gemeindeversammlungen, zu Einbürgerungen oder anderen Anlässen etc.),
- konfrontativ illegale Phänomene (z. B. physische und psychische Gewaltakte, Flugblätter oder andere Meinungskundgebungen mit starken Drohungen, Sachbeschädigungen, anonyme Telefonanrufe, öffentliches Auftreten mit Drohcharakter, Störungen von Gemeindeversammlungen oder anderen Anlässen etc.).

Dieses Modell schien uns für unsere Zwecke besonders geeignet, da es legale Phänomene von Rechtsextremismus mit einbezieht wie etwa Bürgermobilisierung zur Verhinderung von Einbürgerungen oder aggressive Leserbriefe, welche von Erwachsenen mit rechtsextremen Haltungen bevorzugt werden, während Jugendliche die mit illegalen Aktionen verbundenen Risiken weniger scheuen. Mithilfe dieses Modells ermittelten wir im Rahmen unserer Befragungen der lokalen Akteure rechtsextreme Phänomene sowohl auf der Ebene rechtsextrem orientierter Jugendlicher als auch indirektere Formen auf der Erwachsenenenebene. Durch die retrospektive Befragung der Gemeindeakteure bezüglich rechtsextremer Phänomene vor und nach der Intervention wollten wir deren quantitativen Umfang zu diesen zwei Zeitpunkten in Erfahrung bringen. Daraus konnten wir ableiten, welche unerwünschten Erscheinungsformen minimiert werden und welche Zielgruppen und Prozesse durch die Intervention gestärkt werden konnten (vgl. Stöss 2003).

Im Rahmen unserer Befragungen stellte sich dieses Modell zur Erfassung manifester Rechtsextremismusphänomene nicht in allen Gemeinden als ergiebig heraus, ganz in Abhängigkeit der lokalen Ausprägung und Erscheinungsform von Rechtsextremismus. Insbesondere bei schwacher Ausprägung (sporadische Vorfälle, Gruppierungen mit patriotischer Ausrichtung etc.) handelte es sich in erster Linie um latente und schwer fassbare Phänomene. Zudem scheuten sich einzelne Gesprächspartner/innen solche Phänomene – wie etwa Sprayereien – als rechtsextrem motiviert einzuordnen, weil z. B. die Urheberschaft nicht geklärt oder der Zusammenhang mit rechtsextremen Gruppierungen nicht klar ersichtlich war. Diese Schwierigkeit der Feststellung manifester Phänomene muss auch in Zusammenhang damit gesehen werden, dass in der Schweiz rechtsextrem orientierte

Jugendliche meist sozial und beruflich integriert sind, weshalb illegale Ausdrucksformen eher vermieden werden.

Rechtsextremismus und Fremdenfeindlichkeit sind in ihrer latenten Form ein schwer fassbares Phänomen, weshalb es auch für unsere Interviewpartner/innen oft nicht einfach war, diese zu benennen und zu quantifizieren. Äußerungen wie „das spürt man hier“, „das ist ein hartes Pflaster“ oder „das ist nicht der Boden, wo Rechtsextremismus gedeihen kann“ bleiben vage und mussten durch Nachfragen konkretisiert werden. So waren z. B. Sympathien für die rechtsextreme Szene konkreter in Leserbriefen als in Stammtischgesprächen oder hinter vorgehaltener Hand gemachten Äußerungen zu orten.

Solche Deutungsschwierigkeiten stellten auch Christian Hirschi und Thomas Widmer (2006) in ihrer Metaevaluation fest, wonach latente Phänomene im Gegensatz zu manifesten nicht beobachtbar seien und aufgrund ihrer Gradualität das Problem ihrer Zuordnung zum Rechtsextremismusbegriff diskutabel wird. Doch auch manifeste Phänomene können z. B. durch die erfolgte Medienberichterstattung und dadurch ausgelöste Emotionalisierung in ihrer Wahrnehmung beeinflusst und verfälscht werden (vgl. ebd.). Das Problem der Gradualität konnten wir nicht durch „Härtekriterien“ lösen, lag uns doch viel an der Erfassung solcher „weicher“ Kontextvariablen (vgl. Tab. 2), welche den Boden für manifesten Rechtsextremismus legen können.

Die Frage nach der Eingrenzung respektive Elastizität des Rechtsextremismusbegriffs stellte sich schon bei der Auswahl der zu evaluierenden Interventionen. So bezogen wir z. B. eine Intervention in die Evaluation mit ein, die für das Problem „Jugendgewalt“ von der Gemeinde erfragt und vom Interventionsanbietenden konzipiert wurde, in deren Verlauf sich aber unter anderem auch Probleme mit nationalistischen und gewaltbereiten Jugendlichen herauskristallisierten. Eine weitere Intervention hatten wir einbezogen, obwohl sie sich in erster Linie gegen das fremdenfeindliche und ausgrenzende Klima in der Bevölkerung richtete; das Vordringen von Rechtsextremismus wurde aufgrund der Betroffenheit umliegender Gemeinden zwar befürchtet, doch waren derart manifeste Phänomene in der Gemeinde nicht auszumachen. Diese Ausdehnung der Auswahlkriterien hatte auf der einen Seite Folgen für unsere Befragungsinstrumente, die nun eben gerade auch solche „weichen“ Formen erfassbar machen mussten. Und auf der anderen Seite galt es bei der Darstellung der Ergebnisse darauf zu achten, bezüglich des Rechtsextremismusbegriffs zu differenzieren, um eine mehrheitlich präventive und auf die Reduktion von Fremdenfeindlichkeit ausgerichtete Intervention nicht als eigentliche Intervention gegen Rechtsextremismus erscheinen zu lassen.

Zur Typologisierung der unterschiedlichen Interventionsstrategien entwickelten wir ausgehend von Krafelds (1992) Interventionsmodell ein Kontext-Täter-Schema (vgl. Eckmann/Eser Davolio 2003). Dieses Vierfelderschema ist bestimmt durch die beiden Achsen Kontext- und Tätervariable und nimmt auf die Verfestigungsgrade von Fremdenfeindlichkeit, Rassismus und Gewaltaffinität Bezug. So erfasst die Kontextvariable das Klima in der Bevölkerung bezüglich Minderheiten und Rechtsextremismus und reicht von einer mehrheitlich toleranten und von Rechtsextremismus distanzierter Haltung der Bevölke-

rung bis zu einem fremdenfeindlichen und rassistischen Klima mit Sympathien für rechts-extreme Anliegen, während die Tätervariable von nationalistischen, wenig gewaltbereiten Jugendlichen und jungen Erwachsenen bis zu verfestigten rechtsextremen Ideologien und Gewaltausübung reicht. Die beiden Achsen sind als Kontinuum zu verstehen, deren grobe Unterteilung in ein Vierfelder-Schema uns zur Verortung der angetroffenen Interventionsstrategien diene, da jedes Interventionsmilieu eine methodische Schwerpunktsetzung erfordert:

| Kontextvariable  | <i>Tätervariable</i>                            |   |
|--|---|---|
|  | <i>Jugendliche mit rechtsextremen Tendenzen</i> | <i>Jugendliche mit ausgeprägtem Rechtsextremismus</i> |
| <i>Bevölkerung ist offen und tolerant, distanziert sich von Rechtsextremismus</i>  | I. Prävention und Aufklärung                    | II. Mediation und Konfrontation                       |
| <i>Bevölkerung grenzt sich kaum von REX ab, Fremdenfeindlichkeit und Rassismus</i> | III. Minderheiten im Kontext stärken            | IV. Deeskalation und Opferhilfe                       |

**Tab. 2:** Vierfelderschema zu Interventionsmilieus und Interventionsstrategien (REX steht für Rechtsextremismus)

In den meisten der untersuchten Gemeinden zeichnete sich das Klima in der Bevölkerung zu Beginn der Intervention durch latente Fremdenfeindlichkeit und Kontaktvermeidung aus. Das Klima veränderte sich aufgrund der einzelnen Interventionen, indem sich die Bevölkerung zunehmend von Rechtsextremismus distanzierte. Wurden in der Folge auch die Minderheiten in die Intervention einbezogen und gestärkt, fand zunehmend Austausch zwischen Zugewanderten und Einheimischen statt, was sich auf das Ausmaß der Fremdenfeindlichkeit positiv auswirkte. Dies zeigt, dass die beiden Variablen Kontext und Täter nicht nur als Kontinuum zu verstehen sind, sondern sich auch im Verlauf der Intervention verändern können.

Dies gilt auch für die Strategien des Vierfelderschemas: Für die meisten Interventionen stellte das Feld IV. mit Deeskalation, Repression und Opferhilfe den Ausgangspunkt der Intervention dar. Nach der eigentlichen Krisenintervention war es möglich, mit weiteren Strategien, die den Feldern I, II oder III zugeordnet werden können, zu arbeiten. Somit gilt es, das Schema als dynamisches System zu verstehen, da die Intervention Prozesse auslöst und aufgrund sich verändernder Umstände und Konstellationen ständig angepasst und neu ausgerichtet werden muss.

### Evaluation mit Schwerpunkt Wirkungsanalyse

Für die Hauptstudie sahen wir ein dreistufiges Vorgehen vor: Konzeptanalyse, Implementations-evaluation und Wirkungsanalyse. Als Forschungsmethode dienten uns vorwie-

gend qualitative Verfahren sowie eine quantitative Fragebogenerhebung. Alle sieben Evaluationen waren bezüglich der Forschungsfragen und der angewandten Methoden in vergleichbarer Weise angelegt, um die Wirkungsweise der unterschiedlichen Interventionsansätze abschätzen zu können. Dabei galt es, die strukturellen Projektvoraussetzungen und -bedingungen sowie eine differenzierte Analyse des Interventionsprozesses zu berücksichtigen, um die Evaluationsergebnisse für die Praxis effektiv nutzbar zu machen (vgl. Sanders 2000).

Anders als bei der CIVITAS-Programmevaluation (vgl. Lynen von Berg et al. 2003; Rommelspacher et al. 2003), bei welcher die Prozessqualität im Vordergrund stand, ging es bei unserer Studie in erster Linie um die Ergebnisqualität heterogener Interventionen ohne einheitliche Projektaufträge und Rahmenbedingungen. Deshalb wandten wir eine Kombination von präadaptiver Vorgehensweise (vgl. Grohmann 1997), bei der vor der Erhebung bestimmte Urteilskriterien festgelegt werden, und interpretativer Vorgehensweise an. Bei letzterer werden die Kriterien aus dem gesammelten Datenmaterial hergeleitet; die Erfolgsbeurteilung wird somit aus der Perspektive externer Akteure – in unserem Fall die an der Intervention beteiligten Personengruppen in den einzelnen Gemeinden – vorgenommen.

Unser Evaluationsdesign sah die Befragung von direkt und partiell involvierten Gemeindeakteuren vor. Die Auswahl und die Kontakte zu den direkt involvierten Akteuren stellten wir über die Interventionsanbietenden her, die uns Auskunft über die Zusammensetzung der lokalen Arbeitsgruppen gaben. Teils aufgrund von Informationen der Anbietenden, teils aufgrund von Aussagen der interviewten Gemeindeakteure erhielten wir Hinweise auf weitere Personen, die Kenntnis von der Intervention hatten oder über einen beschränkten Zeitraum daran mitgearbeitet hatten. Von diesen Akteuren erhofften wir uns eine distanziertere und kritischere Beurteilung der Intervention, was sich bei der Auswertung der Interviews auch bestätigte. Für die Wirkungsanalyse unserer Evaluation stützten wir uns ausschließlich auf die Urteile der involvierten und distanziierten Gemeindeakteure sowie auf die Einschätzung der Rechtsextremismusexpertinnen und -experten mit lokalen Kenntnissen der Szene, während wir die Bewertungen der Interventionsanbietenden bezüglich der Wirkung und Nachhaltigkeit ihrer Interventionen nur zum Vergleich heranzogen. Wir gingen davon aus, dass die Fremdurteile eine objektivere und kritischere Sichtweise zur Beurteilung der Wirkung und Nachhaltigkeit der einzelnen Interventionen garantieren würden. Dies bestätigte sich insofern, dass die Intervenierenden bezüglich der Einschätzung des Interventionserfolges durchgängig optimistischer waren als die beteiligten Gemeindeakteure.

Insbesondere was die Einschätzung der Wahrnehmung und Resonanz der Intervention in der Bevölkerung betraf, gingen die distanziierten Akteure von einem geringeren erreichten Sensibilisierungsgrad aus als die direkt involvierten Gemeindeakteure. Dies lässt sich damit erklären, dass Letztere ihr Umfeld über die Intervention informierten und zum Teil auch darin involviert waren, weshalb sie über weniger Kontakten zu Bevölkerungskreisen ohne Kenntnisnahme der Intervention verfügten. Somit fehlte den involvierten Akteuren zum Teil

diese Außenwahrnehmung der Intervention. Doch muss man nahezu allen befragten Gemeindeakteuren zugute halten, dass sie sich bezüglich des erreichten Rechtsextremismusrückgangs sehr vorsichtig äußerten und ein jederzeitiges Wiederaufflammen der Probleme für möglich hielten. So waren sie sich in der Regel der großen Komplexität und soziogeografischen Reichweite des Rechtsextremismusphänomens und damit des beschränkten Wirkungsradius der Intervention bewusst.

Als besonders wichtig stellten sich im Verlauf der Interviews die emotionalen Aspekte, welche mit der Intervention assoziiert wurden, heraus. Wir stellten fest, dass beinahe alle der angefragten Gemeindeakteure kooperationsbereit sowie großzügig und flexibel bezüglich der für die Interviews aufzuwendenden Zeit reagierten. Diese Bereitschaft schien damit zusammenzuhängen, dass sie die Interventionen meist als wirkungsvoll und persönlich bereichernd erlebt hatten. Im Verlaufe der Interviews zeigte sich, dass etwa bei Fragen nach ihrer persönlichen Betroffenheit durch Rechtsextremismus, eventuellen Widerständen gegenüber der Intervention oder ihrem persönlichen Gewinn aus dem Engagement, stets auch gefühlsmäßige Anteile mitschwangen wie etwa vorübergehende Enttäuschungen, Empörung, Enthusiasmus oder Stolz. Da sich diese emotionale Dimension für die Motivation zum weiteren Engagement sowie zur Aufrechterhaltung der Intervention und somit für deren Nachhaltigkeit als relevant herausstellte, halten wir den Einbezug von Fragekategorien, die auf solche emotionalen Dimensionen abzielen, für empfehlenswert.

Für die Wirkungsanalyse untersuchten wir zwei Dinge: Zum einen die Aspekte der erfolgreichen Stärkung zivilgesellschaftlicher Kräfte und zum anderen den Rückgang von Rechtsextremismus. Die befragten Gemeindeakteure schilderten uns, dass zur breiten Mobilisierung zuerst ein Problembewusstsein bezüglich Rechtsextremismus und eine öffentliche Auseinandersetzung mit Informationsveranstaltungen, Podiumsdiskussionen, Öffentlichkeitsarbeit und Kundgebungen geschaffen werden mussten. Hierfür sei es von großer Wichtigkeit gewesen, dass sich Schlüsselpersonen in der Gemeinde über alle politischen Grenzen hinweg gemeinsam gegen Rechtsextremismus, Gewalt und Ausgrenzung öffentlich geäußert hätten. Dies habe eine Signalwirkung erzeugt und das Meinungsklima in der Bevölkerung sensibilisiert.

Bei der Bewertung der Wirkung von Interventionen wurden von den interviewten Gemeindeakteuren folgende Zusammenhänge für den Rückgang von Rechtsextremismus aufgeführt: Aufgrund der Krisenintervention und verstärkter Repression sei das Terrain für Rechtsextreme unattraktiv geworden und eine Verringerung der Vorfälle und Anlässe (Veranstaltungen, Treffen) sei die Folge. Durch den sozialen Druck, der von Eltern, Lehrkräften, Lehrmeistern und Arbeitgebern ausgegangen sei, hätten sich rechtsextreme Jugendliche und potenzielle Mitläufer/innen von der Szene distanziert. Des Weiteren wurde eine Beruhigung bezüglich Jugendgewalt ins Feld geführt, da etwa durch Konfrontation (mit „Dorfautoritäten“) und Mediation Gewalteskalationen zwischen polarisierten Szenen wie etwa links vs. rechts, Schweizer/innen vs. Ausländer/innen vermieden worden seien. Ebenso habe die gestiegene öffentliche Aufmerksamkeit die Signalisierungs- und Anzeigebereitschaft erhöht und die Vernetzung zwischen Polizei und Jugendarbeit ein frühzeitiges

oder präventives Eingreifen bei Konfrontationen zwischen Jugendgruppierungen ermöglicht. Denn das Interventionsziel „Rückgang von Rechtsextremismus“ gilt es differenziert zu betrachten, da neben den intendierten auch nicht intendierte Wirkungen wie etwa lokale Verlagerung oder Vermeidung von Sichtbarkeit, Verschiebung von manifesten zu latenten Erscheinungsformen, mitberücksichtigt werden müssen. Solche Phänomene spielen sich auch außerhalb des Erfahrungsbereichs der befragten Gemeindeakteure ab, weshalb wir zu deren Beurteilung externe Rechtsextremismusexpertinnen und -experten (Kantonale Polizeistellen, Bundespolizei, NGO's, Journalistinnen/Journalisten) interviewt hatten. Diese schätzten die Wirkung der Interventionen meist zurückhaltender und geringer ein als die befragten Gemeindeakteure, da sie in einzelnen Fällen (insbesondere bei Interventionen, die weniger auf Nachhaltigkeit ausgerichtet waren) ein Ausweichen auf andere Gemeinden und Regionen nach dem Aktivwerden der betroffenen Gemeinden vermuten respektive feststellen konnten.

Diesbezüglich muss angemerkt werden, dass die meisten Interventionen nicht den Anspruch oder die Möglichkeit hatten, das Problem Rechtsextremismus großräumig bearbeiten und lösen zu können. Andererseits müsste der Prozess der Mobilisierung der zivilgesellschaftlichen Kräfte auch andernorts, insbesondere in den Gemeinden, wo sich rechtsextremistische Gruppierungen neu formieren und manifestieren, stattfinden, um ein Ausweichen zu verhindern.

Was die Diskussion um die Wirkungsanalyse von Rechtsextremismusinterventionen betrifft, so halten wir die Beurteilung der Wirkung und Nachhaltigkeit einer Intervention für machbar und vertretbar und teilen die Sichtweise von Lynen von Berg u.a. (2003) und Scherr (2003) nicht, dass aufgrund der hohen Komplexität des Interventionssettings keine direkten Kausalwirkungen solcher Programme und ihrer Konzepte festgestellt werden könnten. Aufgrund unserer Forschungserfahrungen sind wir der Ansicht, dass die Wirkungsqualität einer Intervention über einen längeren Zeitraum und im Zusammenspiel mit weiteren Faktoren beobachtet werden muss, was wir aber in unserer Untersuchung bei den meisten evaluierten Interventionen auch leisten konnten. Was die Zuschreibung der Kausalwirkungen betrifft, so haben wir in der Tat bei einzelnen Interventionen festgestellt, dass die intendierte Wirkung wie z. B. Opferhilfe mittels Helpline, nicht erreicht werden konnte, doch die Bevölkerung mit einer erhöhten Signalisierungsbereitschaft reagierte, was ebenfalls im Sinne der Intervention ist, während Bumerangeffekte oder kontraproduktive Wirkungen die Ausnahme darstellten. Natürlich standen auch wir vor dem Problem der direkten Zuordnung von Maßnahmen und Effekten, doch denken wir durch eine umfassende, vergleichende Sichtweise auf die untersuchten Interventionen und ihr jeweiliges Interventionsmilieu Aussagen zu deren Wirkung machen zu können, indem wir einzelne Findings durch die Gegenüberstellung zu verdichten und zu erhärten vermögen. Die pessimistische Beurteilung der Durchführbarkeit einer Wirkungsanalyse in der Studie von Lynen von Berg u.a. (2007), hat zum einen mit einem engen Begriffsverständnis nach Stöss (2003) zu tun, wonach nur direkte Einwirkungs-Effekt-Zusammenhänge als Wirkung bezeichnet werden dürfen und zum anderen mit dem Kontext der untersuchten Programme

und Interventionen. Hier vermuten wir, dass die schwierige Ausgangssituation in den jeweiligen Einsatzorten der neuen Bundesländer, bei welchen die Intervenierenden zum Teil demokratisches Basiswissen vermitteln mussten, bis die eigentlichen Interventionen gegen Rechtsextremismus beginnen konnten, mit dazu führten, dass es wenig an direkten Wirkungen aus den ursprünglichen Interventionszielen nach zu verfolgen gab. Hier trafen wir in den untersuchten Schweizer Gemeinden auf ungleich günstigere Ausgangslagen mit einem zum Teil sehr ausgeprägten Grad an Mobilisierung zivilgesellschaftlicher Kräfte – ohne damit zu sagen, dass dies auf alle Schweizer Gemeinden generalisiert werden könne. Keine einzige der untersuchten Gemeinden wies von ihrem Kontext her ein ausgesprochen fremdenfeindliches und rassistisches Klima auf oder unterstützte die jungen Rechtsextremen offen oder latent. Diese Ausgangssituation, welche sich ganz anders als bei den von Lynen von Berg u.a. evaluierten Interventionen in den neuen Bundesländern darstellt, machte es uns möglich, Wirkungszusammenhänge festzustellen und durch die Beurteilung externer Rechtsextremismusexpertinnen und -experten mit Kenntnissen der regionalen Szene zu validieren. Trotzdem kann die Einwirkung weiterer äußerer Faktoren nicht völlig ausgeschlossen werden, die mit zu diesem positiven Ergebnis beigetragen haben wie etwa eine mögliche modebedingte Abnahme des Rechtsextremismusphänomens.

Ein wichtiges Anliegen stellte bei der Wirkungsanalyse die Überprüfung der Nachhaltigkeit der Interventionen dar. Da die Befragungen der involvierten/distanzierten Akteure in den Gemeinden in einem zeitlichen Abstand von einem halben Jahr bis vier Jahre nach Abschluss der Intervention respektive nach eigentlicher Krisenintervention durchgeführt wurden, stellten wir Fragen zur Einschätzung der längerfristigen Wirkung der einzelnen Interventionen. Es wurde somit eine retrospektive und eine prospektive Beurteilung der Interventionsnachhaltigkeit erfragt, indem sowohl rückblickend die Reichweite der erzielten Wirkung als auch deren zukünftige Potenziale abgeschätzt werden mussten. Dabei erfragten wir Faktoren, welche eine längerfristige Wirkung begünstigen wie etwa die Aufrechterhaltung der Vernetzung oder die Weiterführung von Aktivitäten. Im Speziellen ging es hier auch um den Multiplikationsaspekt als ein mögliches Nachhaltigkeitsprinzip, nämlich die Verselbständigung der Intervention, indem die lokalen Akteure befähigt wurden, selbst Interventionen gegen Rechtsextremismus zu konzipieren und durchzuführen sowie ihre Erfahrungen anderen zugänglich zu machen.

Die Darstellung der aus dem umfangreichen Datenmaterial gewonnenen Aussagen und Ergebnisse stellte aufgrund der Komplexität des Forschungsgegenstandes keine einfache Aufgabe dar. Um den vier wichtigsten Evaluationsstandards: Nützlichkeit, Durchführbarkeit, Korrektheit und Genauigkeit (vgl. Widmer 2004) Genüge zu leisten, hielten wir ein Vorgehen für angebracht, welches sich auf die ausführliche Wiedergabe von Aussagen der interviewten Gemeindeakteure stützt. So hofften wir die Einschätzung der Intervention und deren Wirkung für die Leser/innen nachvollziehbar zu machen und subjektive Beurteilungskriterien und verallgemeinernde Zusammenfassungen von uns Forschenden so weit wie möglich zu vermeiden. Auch wenn die Auswahl der Interviewausschnitte und die Durchführung der Interviews stets von uns Forschenden gesteuert sind, so halten wir die

detaillierte Wiedergabe von Original-Interviewaussagen für valider als eine zusammenfassende Darstellung, zumal in den Interviewaussagen wichtige Zusatzinformationen über Kontextbedingungen sowie relativierende und differenzierende Inhalte enthalten sind. Dadurch versprochen wir uns eine bessere Transparenz bezüglich der Reliabilität unserer Schlussfolgerungen und Guidelines, indem deren Begründungen von Leser/innen im ausführlichen Abschlussbericht nachvollzogen und überprüft werden können.

### **Professionsorientierte Wissensgenerierung für unterschiedliche Zielgruppen**

Die ausgewählten Interventionen zeichnen sich durch eine große Heterogenität aus, sowohl was ihre Interventionskonzepte als auch angewandte Vorgehensweisen betrifft. Unter den einbezogenen Interventionsanbietenden finden sich sowohl private als auch öffentliche sowie halböffentliche Stellen, was unterschiedliche Legitimierungen für Interventionen impliziert. Auch was den professionellen Hintergrund der Intervenierenden betrifft, gab es große Differenzen. Dieser reichte von Ausbildungsabschlüssen in Sozialer Arbeit über Psychologie bis hin zu Jurisprudenz, was sich auch in den Interventionsstrategien und ihrer Gewichtung widerspiegelte. So lag etwa der Arbeitsschwerpunkt des Rechtsextremismusexperten mit juristischem Hintergrund auf der Beratung der Gemeinde bezüglich des rechtlichen Vorgehens gegen einen Rechtsextremistentreffpunkt. Doch regte er gleichzeitig Informationsveranstaltungen und Verhandlungen mit Wirten bezüglich der Verweigerung von Vermietungen an, führte Beratungsgespräche mit rechtsextremen Jugendlichen und deren Eltern etc., so dass sein Vorgehen nicht auf den juristischen Weg reduziert werden kann. Die Wahl mittels baurechtlicher Vorschriften die Schließung des Treffpunkts herbeizuführen erwies sich in der Folge zwar als rasche Lösung des eskalierenden Problems, doch konnte in dieser Gemeinde kein weiterführender Prozess ausgelöst werden, welcher zu einer Auseinandersetzung mit den Kontextfaktoren (Klima in der Bevölkerung etc.) geführt hätte. Dass mit dem eindimensionalen Vorgehen eine Chance für einen Bewusstwerdungsprozess vertan wurde, konnte durch den Vergleich mit anderen Interventionsstrategien in Gemeinden mit vergleichbarer Ausgangslage gezeigt werden. Somit stellte diese Heterogenität zwar hohe Anforderungen an die Operationalisierung eines einheitlichen Evaluationsdesigns, doch stellte sie sich nachträglich als Gewinn für die Formulierung der Guidelines heraus, da vielfältiges und umfangreiches Datenmaterial aus dem Vergleich so unterschiedlicher Interventionsansätze gewonnen werden konnte.

Die Quintessenz unserer Ergebnisse fand Eingang in die Guidelines, welche Empfehlungen zur Konzipierung und Durchführung von Interventionen geben, auf mögliche Hindernisse und Fallen hinweisen und den Prozesscharakter und die langfristige Ausrichtung einer Intervention für die Entfaltung einer nachhaltigen Wirkung betonen. Insbesondere enthalten sie Anleitungen zur Formierung einer Arbeitsgruppe mit Hinweisen zu deren politischer Abstützung, zu Zielformulierungen und Namensgebung, zum Einbezug der Medien, zur Vernetzung, zum Einbezug von Experten und Expertinnen, zur Strategieentwicklung und zur Umsetzung der Maßnahmen sowie zur Überprüfung der erreichten Ziele. Mit den

Guidelines und der Evaluationsstudie als Ganzes ist Wissen für folgende Zielgruppen generiert worden:

1. Für Interventionsanbietende und Rechtsextremismusexpertinnen und -experten sind zur Überprüfung und Erweiterung ihrer Interventionskonzepte fachliches Wissen und professionelle Standards produziert worden.
2. Für Gemeindeakteure sind Informationen zur Abschätzung der Möglichkeiten einer selbständigen Intervention respektive zum Einbezug auswärtiger Expert/innen sowie Anleitungen zu deren Durchführung erarbeitet worden (vgl. Eser Davolio/Drilling 2007).
3. Für die Praxis der Sozialen Arbeit, insbesondere Gemeinwesenarbeit und Jugendarbeit ist Wissen über verschiedene Interventionsstrategien, ihre Implementierung und Wirkungsmöglichkeiten, sowie den Umgang mit Rechtsextremismus generiert worden.
4. Für Evaluierende und Sozialwissenschaftlerinnen und Sozialwissenschaftler wurden Forschungsinstrumente, Modelle und Referenzsysteme zur Evaluation unterschiedlicher Interventionsstrategien gegen Rechtsextremismus entwickelt.

Für die Wissensgenerierung erwies sich unser eintägiger Scientific Workshop mit in- und ausländischen Rechtsextremismusexpertinnen/-experten (Prof. B. Rommelspacher, Prof. F.-J. Krafeld, Prof. F. Hamburger, Prof. K. Weiss, Prof. U. Wolff-Jontofsohn, Dr. P. Rieker, Prof. A. Scherr, Prof. M. Eckmann) und der bilaterale Austausch mit C. Mude (B), Y. Carlsson (N) und R. Pugh (GB) als sehr fruchtbar. Insbesondere was die Übertragbarkeit der Ergebnisse im internationalen Kontext betrifft, zeigte sich, dass sich einzelne Interventionsansätze nur bedingt auf andere Interventionsmilieus übertragen lassen, wenn z. B. die Polizei nicht über die Möglichkeit verfügt, sich lokalen Vernetzungsgruppen anzuschließen und auch präventive Arbeit zu leisten oder die Gemeindeautonomie, welche für die untersuchten Interventionen grundlegend war, aber nur in föderalistischen Staatssystemen vorausgesetzt werden kann. Die Guidelines und Empfehlungen für zukünftige Interventionen wurden kritisch diskutiert, ergänzt und gegebenenfalls relativiert. Bezüglich der vergleichenden Evaluation wurde uns von einer Bewertung der einzelnen Interventionen abgeraten, da dies eine Quantifizierbarkeit der Wirkung und Nachhaltigkeit einer Intervention suggerieren würde. Erstens seien diese Dimensionen nur schwer abschätzbar und zweitens müssten die Qualitäten respektive Defizite einer Intervention stets vor dem Hintergrund der spezifischen Problemkonstellation gesehen werden. Das Ziel unserer vergleichenden Evaluation war schließlich auch nicht das Erstellen einer Rangliste unterschiedlich erfolgreicher Interventionen, sondern die professionsorientierte Wissensgenerierung, um die Strategien zukünftiger Interventionen zu optimieren.

Der Wissenstransfer sowie die Verbreitung der erworbenen Erkenntnisse sind beim Schweizerischen Nationalfonds wichtige Aspekte, weshalb schon bei der Eingabe von Forschungsprojekten spätere Umsetzungsaktivitäten geplant werden müssen, wie etwa Transfermöglichkeiten, Austausch mit in- und ausländischen Forschenden, Publikationen und Vorträge etc. Somit waren von Beginn an die Ressourcen für Zeit und Finanzierung

dieser Aktivitäten eingeplant und gewährleistet. Eine unserer Wissenstransferleistungen bestand in der Rückmeldung der Evaluationsergebnisse an die untersuchten Interventionsanbietenden. Ihren Einbezug in den Scientific Workshop ermöglichte den „wettbewerbsfreien“ Austausch innerhalb der Profession, schuf Austauschmöglichkeiten mit in- und ausländischen Rechtsextremismusexpertinnen/-experten und stellte gemäß Aussagen der Teilnehmenden eine fokussierte Weiterbildung dar. Zudem fanden innerhalb des Nationalfondsprogramms mehrere Austauschrunden der beteiligten Forschungsteams statt, welche unterschiedliche Facetten des Rechtsextremismusphänomens beleuchteten, die ebenfalls zur Wissensgenerierung innerhalb der Schweizer Forschungslandschaft beitrugen. Ebenso gilt es den Austausch mit dem Eidgenössischen Fonds für Rassismusbekämpfung zu erwähnen, welcher Interventionen gegen Rechtsextremismus in der gesamten Schweiz finanziell unterstützt und die vorliegenden Ergebnisse in Form einer Broschüre als Orientierungshilfe für Antragsstellende verbreiten wird. Somit kann von einer professionsorientierten Wissensgenerierung auf Ebene der Praktiker/innen und Forschenden gesprochen werden, welche wir mit Referaten auf internationalen Tagungen und diversen Publikationen auch versuchen, über den nationalen Rahmen hinaus zu tragen.

## Literaturverzeichnis

- Busmann, Werner (2005): Evaluation in der Europäischen Gemeinschaft. Aktuelle Praxis und zukünftige Trends. In: Bulletin der Schweizerischen Evaluationsgesellschaft, Nr. 24, S. 2–4
- Drilling, Matthias (i. E.): Zum Stand der Evaluation in der Sozialen Arbeit: Länderbeitrag Schweiz. In: Widmer, Thomas/Beywl, Wolfgang/Fabian, Carlo (Hrsg.): Kompendium der Evaluation. Wiesbaden: VS-Verlag
- Eckmann, Monique/Eser Davolio, Miryam (2003): Evaluation du projet „A l'école de la solidarité“ de l'OSAR. Genève: Editions ies
- Eser Davolio, Miryam (2000): Fremdenfeindlichkeit, Rassismus und Gewalt – Festgefahrenes durch Projektunterricht verändern. Bern: Haupt Verlag
- Eser Davolio, Miryam/Drilling, Matthias (2007): Wie können Gemeinden gegen Rechtsextremismus vorgehen? In: Fachstelle für Rassismusbekämpfung (Hrsg.): Rechtsextremismus bekämpfen. Wirksame Maßnahmen und griffige Arbeitsinstrumente für Gemeinden. Bern: EDI, S. 5–56
- Eser Davolio, Miryam/Eckmann, Monique/Drilling, Matthias (2004): Rechtsextremismus und Soziale Arbeit in der Schweiz. In: Drilling, M./Eser Davolio, M. (Hrsg.): Rechtsextremismus und Soziale Arbeit. Forschungsberichte des Moduls „Lernen und Forschen“. Fachhochschule für Soziale Arbeit beider Basel, S. 6–10. [www.forschen.ch](http://www.forschen.ch)
- Eser Davolio, Miryam/Gerber, Brigitta/Eckmann, Monique/Drilling, Matthias (2006): The Special Case of Switzerland: Research Findings and Thoughts from a Context-Oriented Perspective: In: Rieker, Peter/Glaser, Michaela/Schuster, Silke (Hrsg.): Prevention of Right-Wing-Extremism, Xenophobia and Racism in European Perspective, Halle: Deutsches Jugendinstitut, S. 30–47

- Gentile, Pierre/Jegen, Maya/Kriesi, Hanspeter/Marquis, Lionel (1995): Die rechtsradikale Radikalisierung: Eine Fallstudie. In: Altermatt, Urs/Kriesi, H. (Hrsg.): Rechtsextremismus in der Schweiz. Zürich: Verlag Neue Zürcher Zeitung, S. 157–225
- Grohmann, Romano (1997): Das Problem der Evaluation in der Sozialpädagogik: Bezugspunkte zur Weiterentwicklung der evaluationstheoretischen Reflexion. Frankfurt am Main/Bern: Lang Verlag
- Haubrich, Karin/Holthusen, Bernd/Struhkamp, Gerlinde (2005): Evaluation – einige Sortierungen zu einem schillernden Begriff. In: DJI-Bulletin 72plus. [www.dji.de/evaluation/Evaluation\\_Begriff.pdf](http://www.dji.de/evaluation/Evaluation_Begriff.pdf) (30.3.2007)
- Hirschi, Christian/Widmer, Thomas (2006): Evaluating Measures against Right-wing Extremism in Switzerland: Approaches and Challenges for Evaluation. Paper presented at the Joint International Conference of the European Evaluation Society and the United Kingdom Evaluation Society, Queen Elisabeth II Conference Centre, London, UK, 2–4, October 2006
- Kaufmann, Philippe (2006): Wissensmanagement: Fehlendes Glied zwischen Politik und Evaluation? In: Bulletin der Schweizerischen Evaluationsgesellschaft, H. 26, S. 2–4
- Krafeld, Franz-Josef (1992): „Eskalation der Gewalt gegen Ausländer – und was tun?“ In: deutsche Jugend, H. 11, Weinheim: Juventa Verlag, S. 500–502
- Lynen von Berg, Heinz/Pallocks, Kerstin/Steil, Armin (2007): Interventionsfeld Gemeinwesen. Evaluation zivilgesellschaftlicher Strategien gegen Rechtsextremismus. Weinheim; München: Juventa Verlag
- Lynen von Berg, Heinz/Pallocks, Kerstin/Vossen, Johannes (2003): Bericht der wissenschaftlichen Begleitforschung über die Modellphase der Strukturprojekte des Programms CIVITAS – initiativ gegen Rechtsextremismus in den neuen Bundesländern. Berlin
- Rommelspacher, Birgit/Polat, Ülger/Wilpert, Czarina (2003): Die Evaluation des CIVITAS-Programms. Berlin: Alice Salomon Fachhochschule
- Sanders, James R. (2000): Handbuch der Evaluationsstandards. Opladen: Leske + Budrich
- Scherr, Albert (2003): Pädagogische Konzepte gegen Rechts – was hat sich bewährt, was ist umstritten, was sollte vermieden werden? In: Lynen von Berg, Heinz/Roth, Roland (Hrsg.): Maßnahmen und Programme gegen Rechtsextremismus wissenschaftlich begleitet. Opladen: Leske + Budrich, S. 249–264
- Speck, Karsten (2006): Qualität und Evaluation in der Schulsozialarbeit. Konzepte, Rahmenbedingungen und Wirkungen, Wiesbaden: VS-Verlag
- Stöss, Richard (2003): Möglichkeiten und Grenzen der Evaluierung von persuasiven Programmen. In: Lynen von Berg, Heinz/Roth, Roland (Hrsg.): Maßnahmen und Programme gegen Rechtsextremismus wissenschaftlich begleitet. Opladen: Leske + Budrich, S. 95–102
- Widmer, Thomas (2004): Qualität der Evaluation – Wenn Wissenschaft zur praktischen Kunst wird. In: Stockmann, Richard (Hrsg.): Evaluationsforschung. Opladen: Leske + Budrich, S. 83–110

# Offenheit und Multiperspektivität – Erfahrungen aus der wissenschaftlichen Begleitung von Projekten Interkulturellen Lernens

## Einleitung

Dieser Beitrag dokumentiert Erfahrungen mit der Evaluation präventiver Bildungsmaßnahmen, die im Rahmen einer dreijährigen Untersuchung<sup>1</sup> zu Angeboten Interkulturellen Lernens in Jugendbildung und Jugendarbeit gesammelt wurden. Das Untersuchungsinteresse galt dabei speziell solchen Angeboten, die als Präventionsmaßnahme gegen Fremdenfeindlichkeit umgesetzt werden, d.h. mit Ansätzen Interkulturellen Lernens einen Beitrag gegen fremdenfeindliche Orientierungen und Verhaltensweisen bei jungen Menschen leisten wollen.

Anders als bei vielen Evaluationen in diesem Feld stand zu Beginn unserer Erhebung nicht der Auftrag eines Fördergebers, der die Überprüfung und Bewertung eines bestimmten von ihm geförderten Projekts oder Programms wünschte. Anlass war vielmehr ein Maßnahmen übergreifendes, wissenschaftliches Interesse an diesem Praxisfeld: Die Studie sollte zum einen Einblicke in die existierende Projektlandschaft Interkulturellen Lernens gegen Fremdenfeindlichkeit und Rahmenbedingungen, Umsetzungsweisen und Entwicklungsstand dieser Arbeit dokumentieren. Zum anderen sollte der Frage nachgegangen werden, welche Möglichkeiten verschiedene Ansätze Interkulturellen Lernens für die Prävention von Fremdenfeindlichkeit bergen und mit welchen Schwierigkeiten sich diese Ansätze in der Praxis konfrontiert sehen. Ein besonderer Schwerpunkt wurde dabei auf Projekte gelegt, die mit sogenannten „schwierigen Zielgruppen“ – Jugendlichen mit geringem Bildungskapital sowie fremdenfeindlich orientierten Adressatinnen und Adressaten – arbeiten. Ziel war es, Bedingungen und Faktoren gelingender Praxis Interkulturellen Lernens in der Prävention von Fremdenfeindlichkeit zu identifizieren sowie Herausforderungen für die Weiterentwicklung dieses pädagogischen Feldes zu bestimmen.

Das Nichtgebundensein an einen projektbezogenen Evaluationsauftrag erleichterte uns nicht nur in einigen Fällen den Feldzugang, da wir von den Projekten als „unabhängiger“ wahrgenommen wurden als durch den Fördergeber initiierte Evaluationen; es ermöglichte uns auch, mit unterschiedlichen Methoden bei der Erhebung zu experimentieren und uns nicht auf ein bestimmtes, einheitliches Erhebungsverfahren festzulegen. Unter diesen Bedingungen konnten wir ein offenes, multiperspektivisches Untersuchungsdesign

---

<sup>1</sup> Die Studie wurde in den Jahren 2004 bis 2006 von der Arbeitsstelle Rechtsextremismus und Fremdenfeindlichkeit des DJI durchgeführt und aus Mitteln des BMFSFJ finanziert. An der Untersuchung waren die Autorin und Peter Rieker beteiligt, dem auch dieser Beitrag wichtige Überlegungen und Anregungen verdankt. Bei der Recherche unterstützte uns Nicole Schneider im Rahmen eines Praktikums, auch ihr sei an dieser Stelle herzlich gedankt. Zu den Ergebnissen siehe den veröffentlichten Forschungsbericht zur Untersuchung (Glaser/Rieker 2006).

umsetzen, das sich als gut geeignet erwies, um den Spezifika dieses Feldes Rechnung zu tragen und das es uns ermöglichte, Einblicke in die Bedingungen gelingender Praxis Interkulturellen Lernens gegen Fremdenfeindlichkeit zu erhalten. Die Erfahrungen, die wir hierbei gemacht haben, werden in diesem Beitrag reflektiert. Zunächst werden die konzeptionelle Anlage der Untersuchung sowie die Überlegungen, die diesem Design zugrunde lagen, dargelegt. Daran anschließend werden ausgewählte Aspekte unseres Erhebungs- und Auswertungsverfahrens detaillierter vorgestellt, wobei insbesondere die Potenziale eines multimethodischen und multiperspektivischen Vorgehens ausgeleuchtet werden: Die Chancen, die sich aus der Einbeziehung und Gegenüberstellung unterschiedlicher Perspektiven für die Analyse komplexer Projektrealitäten in diesem Themenfeld ergeben, werden anhand von zwei Fallbeispielen dargelegt und analysiert. Abschließend werden der Gewinn, aber auch die Begrenzungen und Schwierigkeiten dieses Vorgehens bilanzierend diskutiert. Dabei werden auch die Möglichkeiten diskutiert, sich im Rahmen eines solchen Designs der Wirkungsfrage anzunähern und zu objektiven Aussagen über das untersuchte Projektgeschehen zu gelangen.

### **Das Untersuchungsdesign: Die Bestimmung von Wirkungsvoraussetzungen in einem qualitativen, offenen Verfahren**

Für unsere Untersuchung wählten wir ein ausschließlich auf qualitativen Erhebungs- und Auswertungsverfahren basierendes Design. Auf quantitativen Methoden basierende, experimentelle Verfahren gelten zwar häufig als „Königsweg“ (Lüders/Haubrich 2006) bzw. „Goldstandard“ (Kromrey i. d. Bd.) der Evaluation, da sie statistisch abgesicherte Zielerreichungs-Messungen ermöglichen (vgl. Beywl 2006, S. 39). Ihre Eignung zur Anwendung in pädagogischen Settings ist in der Fachwelt jedoch umstritten. So lassen sich valide Messdaten nur mittels (quasi)experimenteller Kontroll- und Vergleichsgruppensdesigns erheben (vgl. Frindte/Preisner 2007, S. 36), die in pädagogischen Kontexten nur sehr schwer realisierbar sind (vgl. Lüders/Haubrich 2006, S. 19). Zum anderen wird bezweifelt, dass sich Wirkungen von Humandienstleistungen und insbesondere von „Überzeugungsprogrammen“ durch solche Messdaten überhaupt sinnvoll erfassen lassen (vgl. Ahlheim 2005, Haubrich u.a. 2005, Hirsland u.a. 2004, Uhl 2004). Da es uns bei unserer Erhebung auch weniger um die Erfassung sogenannter „Outputs“, d.h. von quantifizierbaren Produkten eines Projekts wie Teilnehmerdaten, Anzahl der Seminare etc. ging, als um das Nachzeichnen und Verstehen inhaltlicher Prozesse und Zusammenhänge, erschien uns ein qualitatives Design sowohl im Hinblick auf dieses spezifische Untersuchungsfeld als auch auf unser konkretes Erkenntnisinteresse als das geeignetere Verfahren.

Mit Blick auf den Fokus der Erhebung wurde das Untersuchungsdesign außerdem bewusst sehr offen angelegt. Da unser Interesse nicht nur konzeptionellen Aspekten, sondern auch den tatsächlichen Gegebenheiten im Feld galt, war uns daran gelegen, diese möglichst realitätsnah und „unverfälscht“ in den Blick zu bekommen. Wir bemühten uns deshalb um eine unvoreingenommene Perspektive auf das Untersuchungsfeld, indem wir

auf eine Vorab-Festlegung untersuchungsrelevanter (und damit die Ausklammerung „nicht relevanter“) Zusammenhänge und Dimensionen der von uns untersuchten Projektrealität so weit als möglich verzichteten<sup>2</sup>. Damit trugen wir gleichzeitig der Tatsache Rechnung, dass es zu den meisten von uns untersuchten Angebotsformen bisher nur wenig wissenschaftlich fundiertes Wissen gibt, auf dessen Basis eine entsprechende Einschränkung und Formulierung theoriegeleiteter Hypothesen hätte vorgenommen werden können: Insbesondere zu Ansätzen, die mit schwierigen Zielgruppen in diesem Feld arbeiten, aber auch zum Interkulturellen Lernen in der offenen Jugendarbeit oder in kultur- und medienpädagogischen Lernarrangements liegen bisher kaum systematische Untersuchungen vor (vgl. auch Glaser 2005, S. 49f.), auf die wir uns hätten stützen können.<sup>3</sup>

Der „Preis“ für diese offene Haltung bestand darin, dass es in vielen Fällen nicht möglich war, klare Gegenstandsbeschreibungen der zu untersuchenden Projekte zu erhalten. Denn häufig waren die Projekte nicht oder nur sehr begrenzt dazu in der Lage, Zielstellungen, darauf ausgerichtete Handlungsschritte und Kausalitätsannahmen in Bezug auf ihre Arbeit in systematischer Weise zu definieren und zu benennen.<sup>4</sup> In solchen Fällen gehen manche Evaluationen den Weg, zunächst gemeinsam mit den Projektbeteiligten die „interne Logik“ (Lüders/Haubrich 2006, S. 15f.) oder „Programmtheorie“ der zu untersuchenden Projekte zu rekonstruieren (vgl. Klingelhöfer i. d. Bd.). Diese Alternative kam für uns jedoch nicht in Frage, da eine solche gemeinsame „Modellierung“ (Beywl 2006, S. 34) des Untersuchungsgegenstands mit den Projekten zwangsläufig auch stark formative Effekte gehabt hätte, wir jedoch an der existierenden Praxis Interkulturellen Lernens mit ihren de facto gegebenen Bedingungen, Einschränkungen und auch Defiziten interessiert waren. Ein weiteres Gegenargument war der Zeitfaktor: Projekte Interkulturellen Lernens

---

2 Mit diesem Vorgehen orientierten wir uns an dem von Glaser und Strauss entwickelten Modell einer „gegenstandsbezogenen Theorie“: Um eine „materiale, der empirischen Situation treue“ Theorie zu generieren, kann und sollte man Glaser/Strauss zufolge „den fraglichen Bereich auch ohne irgendeine vorgefasste Theorie untersuchen, die der eigentlichen Forschung vorweg die ‚relevanten‘ Konzepte und Hypothesen bestimmt“ (ebd. 1998, S. 43). Andernfalls bestehe die Gefahr, dass „möglicherweise auftauchende relevante Konzepte und Hypothesen vernachlässigt“ werden (ebd.). Dabei war uns bewusst, dass ein 100%ig naiver Blick, der ohne jegliche das Feld vorstrukturierende Kenntnisse und Vorannahmen auskommt, nicht möglich und sinnvoll ist. Solche Vorannahmen flossen in unserem Fall etwa infolge der Rezeption pädagogischer Fachliteratur oder von Expertengesprächen (siehe folgender Abschnitt) ein.

3 Eine Ausnahme stellen Internationale Begegnungen und Jugendaustausche dar, die in der Bundesrepublik bereits seit den 1950er Jahren gefördert werden und vergleichsweise umfassend dokumentiert und erforscht sind (vgl. Breitenbach 1979; Thomas 1988 und 1996). Aber auch hier gilt, dass zur Arbeit mit Bildungsfernen und Fremdenfeinden bisher kaum systematisierte Erfahrungswerte vorliegen.

4 Diese Schwierigkeit scheint ein generelles Kennzeichen pädagogischen Handelns zu sein. Ein Grund hierfür ist sicherlich im „Eigensinn der Adressaten/innen sozialer Arbeit sowie [der] Unvorhersehbarkeit alltäglicher Lebensvollzüge und -ereignisse“ (Ader 2007) zu suchen, die (sozial)pädagogisches Handeln schwer kalkulierbar machen; in unser Untersuchung gewannen wir aber auch den Eindruck, dass den Pädagoginnen und Pädagogen z.T. auch das methodische „Rüstzeug“ (Hansbauer 2002, S. 835, zit. nach ebd.) fehlt, um Arbeitsprozesse in einer Weise zu strukturieren, die „das eigene Vorgehen nach außen kommunizierbar“ (Ader, a. a. O.) machen.

werden oft nur über einen begrenzten Zeitraum gefördert, wodurch auch unser Feldzugang in vielen Fällen zeitlich sehr begrenzt war. Da Verfahren der Programmtheoriekonstruktion in der Regel sehr zeitaufwendig sind, hätte dies bei vielen Projekten die Beschränkung der Begleitung auf eine derartige Rekonstruktionsleistung und den Verzicht auf darüber hinausgehende, empirische Felderhebungen bedeutet. (Hier deutet sich eine grundsätzliche Problematik für Evaluationen in diesem Feld an: Befristete Laufzeiten und die unsichere Finanzierungssituation vieler Projekte begrenzen auch die Möglichkeiten längerfristig angelegter Erhebungen. Selbst wenn Maßnahmen in einer neuen Förderrunde fortgeführt werden, findet dies häufig mit neuem Personal statt, da die bisherigen Mitarbeitenden sich in der Zwischenzeit neue, z.T. auch themenfremde Beschäftigungsfelder suchen (müssen), was vergleichende Untersuchungen erschwert.)

Unser Untersuchungsinteresse und unsere methodische Orientierung an den im Feld gegebenen Realitäten beinhaltete somit auch den Verzicht auf eine präzise Bestimmung des Untersuchungsgegenstands im Vorfeld, die wir unseren Erhebungen hätten zugrunde legen und z. B. zum Ausgangspunkt von Wirkungsüberprüfungen hätten machen können. Stattdessen bestimmten wir die relevanten, d.h. im Verlauf der Erhebungen und bei der anschließenden Auswertung ins Zentrum rückenden Dimensionen der von uns untersuchten Projektabläufe während des Untersuchungsprozesses selbst – wir generierten diese Dimensionen sozusagen unmittelbar aus dem Feld heraus, indem wir sie „aus den Daten selbst hervortreten“ (Glaser/Strauss 1998, S. 83) ließen.

Der Gewinn dieser Vorgehensweise bestand zum einen darin, die realen Voraussetzungen und Rahmenbedingungen der Arbeit (z. B. die vorhandene Kompetenz der Projekte, klar benennbare und nachvollziehbare Ziele und Zielerreichungsschritte ihrer Arbeit zu definieren) erfassen zu können. Außerdem ermöglichte sie uns, solche Aspekte der Projektarbeit zu identifizieren und in den Mittelpunkt der Analyse zu stellen, die sich in der praktischen Umsetzung von Konzepten, im Projektalltag, als relevant für den Verlauf eines Angebots erwiesen. Einschränkung ist hier allerdings anzumerken, dass Erhebungsmöglichkeiten und Feldzugang insofern eine Vorauswahl „relevanter“ Aspekte beinhalteten, als sie uns während einer Maßnahme oder unmittelbar nach ihrer Beendigung beobachtbare und erfragbare Prozesse beschränkten. Dabei handelte es sich um insgesamt vier Dimensionen der Projektumsetzung, wobei sich je nach Projekttypus unterschiedliche Aspekte als (vorrangig) bedeutsam erwiesen. Die folgende Auflistung gibt einen Überblick über diese von uns als relevant herausgearbeiteten Untersuchungsebenen:

### *Das Erreichen und die Einbindung der projektrelevanten Zielgruppen*

Hier ging es um die Frage, ob für eine Maßnahme auch tatsächlich diejenigen Zielgruppen zur (aktiven) Teilnahme gewonnen werden konnten, an die sich das Projekt primär richtete (z. B. fremdenfeindliche Jugendliche) bzw. die für die Realisierung interkultureller Lerninhalte von wesentlicher Bedeutung waren (z. B. Teilnehmerinnen/Teilnehmer unterschiedlicher ethnischer Herkunft in Begegnungsprojekten). Dieses Kriterium erwies sich als

bedeutsam – und als eine zentrale Schwierigkeit – insbesondere bei Maßnahmen, deren Teilnehmerschaft sich nicht aus einer festen, bereits bestehenden Gruppe konstituierte.

### *Die Verfolgung und Realisierung interkulturell orientierter Lerneinheiten*

Auf dieser Ebene untersuchten wir, ob Elemente interkulturellen Lernens im Rahmen einer Maßnahme überhaupt zum Einsatz kamen: Verfolgten die umsetzenden Pädagoginnen und Pädagogen interkulturelle Zielstellungen oder verbanden sie mit ihrer Arbeit primär andere Ziele? Konnten vorgesehene interkulturelle Lerneinheiten im Projekt auch tatsächlich umgesetzt werden oder traten, aufgrund von Gruppendynamik, Teilnehmerzusammensetzung etc., andere Themen bzw. Aufgaben – spezifische Teilnehmerinteressen und -bedürfnisse, andere Aspekte sozialen Lernens – in den Vordergrund?

### *Die Akzeptanz von Maßnahmen seitens der Teilnehmenden*

Hier interessierte uns, wie das Angebot bei den teilnehmenden Jugendlichen „ankam“, ob sie eine positive Haltung, Mitwirkungsbereitschaft und Interesse an den Aktivitäten des Projekts zeigten. Die Mitwirkungsbereitschaft der Zielgruppen kann generell als zentrale Voraussetzung für die Wirkfähigkeit pädagogischer Maßnahmen gelten, da erzielte Wirkungen im pädagogischen Kontext immer das Ergebnis einer Koproduktion von Pädagoginnen/Pädagogen und Adressatinnen/Adressaten sind (vgl. Haubrich u.a. 2005, S. 2; Beywl 2006, S. 31). Diese Bereitschaft bei den teilnehmenden Jugendlichen zu erwirken bzw. zu erhalten, erwies sich insbesondere bei Projekten, die fremdenfeindliche oder rechtsextreme Einstellungen direkt adressierten, als „Schlüsselfrage“ für den Verlauf von Angeboten.

### *Die Etablierung (Umfang und Qualität) interkultureller Beziehungen*

Hier erhoben wir nicht nur, ob Angebote zur Entstehung von Kontakten zwischen Jugendlichen unterschiedlicher Herkunft führten, wir versuchten auch, soweit im begrenzten Erhebungszeitraum möglich, Informationen über Dauer und Intensität solcher Kontakte einzuholen. Dieses Kriterium legten wir ausschließlich bei solchen Ansätzen an, die mit Begegnungselementen arbeiteten – in denen also die Entwicklung solcher Kontakte ein zentrales Element der Maßnahme bildete. Während die zuvor genannten Kriterien sich nicht immer mit eigenen Maßstäben der Projekte deckten, handelte es sich hier um einen Punkt, den die entsprechenden Projekte selbst als ein Beurteilungskriterium für ihre Arbeit formulierten.

Bei der Bestimmung dieser Indikatoren für gelingende Projektverläufe im Bereich Interkulturellen Lernens handelte es sich genaugenommen um erste und fortlaufende Interpretationen des Materials im Verlauf des Erhebungsprozesses, die dann wiederum die weiteren Erhebungen wie auch die Auswertung strukturierten (vgl. Glaser/Strauss 1979,

S. 92–96). In einem zweiten Schritt wurden dann Faktoren herausgearbeitet, die sich im Hinblick auf einen gelingenden Verlauf dieser unterschiedlichen Projektdimensionen als förderlich bzw. hinderlich erwiesen. Wie die Auflistung zeigt, bezogen sich die auf diese Weise definierten Untersuchungsfelder (mit Ausnahme des Punktes „interkulturelle Beziehungen“), schwerpunktmäßig auf den Projektverlauf und weniger auf die im Rahmen eines Projekts erzielten Effekte. Wenn das Material Anhaltspunkte für Effekte wie Wissens-, Einstellungs- oder Verhaltensänderungen lieferte, wurden diese ebenfalls in die Auswertung mit einbezogen. Aufgrund der eingangs dargestellten methodischen und feldspezifischen Begrenzungen ließen sich diese allerdings nicht systematisch erheben – so dass wir hierzu auch weniger verlässliche Hinweise als zu den umsetzungsbezogenen Dimensionen fanden. Der Schwerpunkt unserer Untersuchung lag somit weniger auf der Ergebnis- als auf der Prozessqualität der untersuchten Projekte. Dabei handelte es sich zum einen um Untersuchungsdimensionen, die einer qualitativen, mit den Restriktionen dieses Feldes konfrontierten Erhebung gut zugänglich waren. Gleichzeitig trug dieses Vorgehen der Heterogenität der untersuchten Projekte Rechnung, für die sich in der Umsetzung unterschiedliche Aspekte als unterschiedlich relevant erwiesen.

### *Exploration des Feldes und Feldzugang*

In die Erhebung einbezogen wurden 22 Projekte, die wir im Zuge einer Recherche per Internet, Datenbanken und Anfragen bei Jugendämtern etc. identifiziert hatten. Bei der Auswahl achteten wir auf eine möglichst große Repräsentativität der vertretenen Angebote (Ansätze und Methoden; Stadt und Land; Ost und West). Die Recherche diente hierbei nicht nur zur Identifikation potenzieller, interessanter „Untersuchungsobjekte“ für unsere Erhebung; sie ermöglichte uns auch eine erste Bestimmung unterschiedlicher „Lernarrangements“ Interkulturellen Lernens, die wir anhand von Kriterien wie zentrale Methoden, zeitlicher und räumlicher Rahmen u. Ä. vornahmen (vgl. Glaser/Rieker 2006, S. 15f.) Diese Unterteilung, die wir im Verlauf der Untersuchung noch weiter ausdifferenzierten, bildete die Basis für eine möglichst repräsentative Zusammenstellung des Untersuchungssamples; sie diente uns außerdem als wesentliches Strukturierungselement bei der Auswertung und Darstellung der Ergebnisse.

Im Vorfeld wurden außerdem leitfadengestützte Experteninterviews mit Fachleuten aus Wissenschaft und Praxis Interkulturellen Lernens geführt, die der Eingrenzung des Feldes und der Identifikation interessanter Fragestellungen bzw. von besonderen Forschungsbedarfs dienten. Diese Fragen bzw. Forschungsinteressen – z. B. Interkulturelles Lernen in Ostdeutschland, Angebote für „schwierige“ Zielgruppen – wurden bei der Projektauswahl mit berücksichtigt. Expertengespräche und Recherchen lieferten uns darüber hinaus Einblicke in Praxis und Umsetzungsbedingungen Interkulturellen Lernens, die als ergänzende Informationen in die Auswertung mit einfließen.

Den Projekten wurde Vertraulichkeit und Anonymisierung der Ergebnisse zugesichert. Als Gegenleistung für ihre Mitwirkung wurde ihnen die Möglichkeit einer ebenfalls ver-

traulichen, mündlichen oder schriftlichen Auswertung angeboten (dieses Angebot wurde nicht der Projektleitung oder dem Fördergeber, sondern den unmittelbar an der Umsetzung Beteiligten unterbreitet); außerdem wurden insgesamt vier Workshops durchgeführt, die allen teilnehmenden Projekten offenstanden und neben themenspezifischen Weiterbildungen einen Rahmen für Erfahrungsaustausch und Reflexion boten.

### *Die Datenerhebung: flexibel und multiperspektivisch*

Bei der Wahl der Erhebungsmethoden legten wir besonderes Augenmerk darauf, eine Vielzahl von Perspektiven auf das Geschehen zu erfassen. Pädagogische Prozesse – wie soziale Prozesse generell – zeichnen sich durch eine im hohen Maße „subjektive“ Realität aus. Sie entfalten ihre Wirkung erst durch den „Sinn, den Akteure ihren Erfahrungen, Lernprozessen etc. geben“ (Klawe 2006, S. 126), welcher sich wiederum vor dem Hintergrund je individueller Wertsysteme und -präferenzen der beteiligten Akteure konstituiert (vgl. Beywl 2006, S. 31). Das gilt sowohl für die Adressatinnen und Adressaten von Maßnahmen als „Koproduzenten sozialer und pädagogischer Prozesse“ (Klawe 2006, S. 125) als auch für alle anderen an einer Maßnahme Beteiligten. Indem wir unterschiedliche Sichtweisen der verschiedenen Beteiligten zu erheben suchten, erhofften wir uns eine Annäherung an diese vielschichtige „Projektrealität“ bzw. die unterschiedlichen Realitäten unterschiedlicher Beteiligter.

Zu diesem Zweck setzten wir eine Reihe unterschiedlicher Erhebungsformen ein. Dazu gehörten:

- Dokumentenanalysen (Konzeptpapiere, Projektberichte),
- Interviews mit Projektfachkräften (Projektleitung, Teamerinnen/Teamer),
- teilnehmende Beobachtungen,
- Gruppeninterviews mit Teilnehmenden,
- informelle Gespräche (Projektfachkräfte, Teilnehmende, Kooperationspartner),
- Workshops mit Projektmitarbeitenden.

Die einzelnen Projekte wurden unterschiedlich intensiv begleitet – sowohl den zeitlichen Umfang als auch die Vielzahl eingesetzter Methoden betreffend<sup>5</sup>. Diese unterschiedliche Intensität ergab sich vor allem aus den Laufzeiten der Projekte und den Möglichkeiten des Feldzugangs, es spielten aber auch inhaltliche Fragen eine Rolle: Um eine besonders intensive Begleitung bemühten wir uns bei solchen Projekten, die uns im Hinblick auf unsere spezifischen Untersuchungsinteressen, z. B. die Arbeit mit fremdenfeindlichen Zielgruppen, besonders vielversprechend erschienen.

---

5 Mit acht Projekten wurde ein einmaliges, zwei- bis dreistündiges Interview geführt; bei zehn Projekten fanden zwei bis drei Vor-Ort-Termine statt, die Interviews und teilnehmende Beobachtungen umfassten; vier Projekte wurden mit einer Intensität von sechs bis neun Besuchen begleitet (vgl. Glaser/Rieker 2006, S.18) Bei allen mehr als einmal besuchten Projekten wurde die Befragung von Projektmitarbeitenden durch die Einbeziehung anderer Perspektiven ergänzt.

Im Vorfeld jeder Begleitung analysierten wir – sofern diese vorlagen und uns zugänglich waren – schriftliche Dokumente wie Projektkonzeptionen und Projektberichte. Dieses Material erwies sich allerdings nur als begrenzt ertragreich; es war vor allem insofern aufschlussreich, als es eine häufig anzutreffende Diskrepanz zwischen berichtsförmiger Außendarstellung und beobachteter bzw. im Gespräch reflektierter Projektrealität verdeutlichte. Den Auftakt der Erhebungen vor Ort bildete dann stets ein Interview mit Projektfachkräften, das grundlegende Informationen zu Rahmenbedingungen und Inhalten des Projektes lieferte; auch wurden Einschätzungen der Akteure zu förderlichen und hinderlichen Faktoren und zu den Erträgen ihrer Arbeit erfragt. Dieses erste Interview mit Projektvertreterinnen und -vertretern bildete zudem die Entscheidungsgrundlage für oder gegen eine weitere Begleitung des Projektes. Sofern wir dieses Gespräch mit nicht unmittelbar an der Umsetzung beteiligten Projektakteuren führten, bemühten wir uns bei weiteren Interviews gezielt um Gespräche mit den direkt in die pädagogische Arbeit involvierten Personen.

Bei allen mehrfach besuchten Projekten führten wir zudem teilnehmende Beobachtungen von Seminareinheiten bzw. Projektsequenzen durch. Diese ermöglichten nicht nur einen vergleichsweise distanzierten Blick „von außen“ auf das Projektgeschehen, es zeigte sich auch verschiedentlich, dass wir als nicht unmittelbar in die Projektgestaltung Involvierte bestimmte Prozesse – Teilnehmerreaktionen, Stimmungen in der Gruppe – beobachten konnten, die den Projektmitarbeitenden entgingen oder auch gar nicht in ihrem Aufmerksamkeitsfokus standen. Unsere Anwesenheit – deren Funktion wir den Beteiligten stets offen legten – eröffnete uns darüber hinaus vielfältige Gelegenheiten für informelle Gespräche mit Mitarbeitenden, Teilnehmenden und Kooperationspartnern. Dabei zeigte sich auch mehrfach, dass uns als Außenstehenden und „neutralen“ Wissenschaftlerinnen und Wissenschaftlern andere Informationen und Einschätzungen genannt wurden, als dies gegenüber Projektvertreterinnen und -vertretern der Fall war. In einem ausgewählten Teil von Projekten fanden überdies Gruppeninterviews mit den Teilnehmerinnen und Teilnehmern statt, die wir nach Projektende, bei mehrteiligen Projekten auch nach Abschluss einer Seminareinheit führten und in denen wir, in einer Kombination aus offenem Gespräch und gezielten Nachfragen, die Sicht der Jugendlichen auf das Projekt erhoben. Auch für diese Gespräche erwies es sich als sehr förderlich, wenn wir den Jugendlichen durch die vorherige Teilnahme an Projektsequenzen bereits vertraut waren.

Als ein unerwartet gewinnbringendes Erhebungsinstrument entwickelten sich die insgesamt vier Workshops mit den Projekten bzw. die dort geführten Diskussionen und formulierten Erfahrungsberichte: In der vertrauten Atmosphäre, die im Rahmen der zweitägigen Seminare wie auch durch die Kontinuität des Angebots möglich wurde und im Austausch mit Kolleginnen und Kollegen, die sich mit ähnlichen Anforderungen und Schwierigkeiten konfrontiert sahen, ermöglichten Projektmitwirkende vielfach weiterreichende Einblicke in ihre Arbeit und entwickelten auch andere Perspektiven und Einsichten auf diese, als dies in der Interviewsituation der Fall war. (In einigen Fällen konnte aber auch der umgekehrte Effekt beobachtet werden: Schwierigkeiten und wenig befriedigende Projektverläufe, die in der unmittelbaren Projektsituation sehr selbstkritisch beurteilt und primär auf eigene

Fehler zurückgeführt wurden, wurden nun – mit zeitlichem Abstand und vor der professionellen „Konkurrenz“ – durch „externe“ Faktoren begründet, z.T. sogar in Erfolge umgedeutet.)

Mit Projektfachkräften, in einzelnen Fällen auch mit Teilnehmenden, führten wir darüber hinaus auch Interviews und Gespräche zu verschiedenen Zeitpunkten der Projektumsetzung, so dass wir auch in zeitlicher Hinsicht unterschiedliche Perspektiven auf das Geschehen erhoben.

Ein Vorteil dieses breiten Spektrums an Erhebungsmethoden bestand in der so gewonnenen Flexibilität: Es erlaubte uns, auf die jeweils in einer bestimmten Situation am besten geeigneten bzw. in dieser Situation überhaupt realisierbaren Methoden zurückzugreifen. Ein solcher flexibler, situationsbezogener Methodeneinsatz erwies sich insbesondere in hierarchischen, belasteten Kontexten (z. B. bei Projekten, die in einer Jugendstrafvollzugsanstalt umgesetzt wurden) als hilfreich, aber auch in anderen Kontexten konnte es vorkommen, dass z. B. eine zuvor abgesprochene teilnehmende Beobachtung vor Ort doch nicht möglich war, so dass stattdessen verstärkt Gespräche mit Beteiligten und Kooperationspartnern geführt wurden.

### *Erträge einer multiperspektivischen Vorgehensweise*

Die besondere Stärke dieses Methodenmix lag jedoch in der Möglichkeit, eine Vielzahl von Perspektiven auf das Geschehen zu erfassen und aufeinander zu beziehen. Diese Triangulation unterschiedlicher subjektiver Perspektiven<sup>6</sup> erwies sich im Rahmen unserer Untersuchung in verschiedener Hinsicht als ertragreich:

#### – *Ergänzende Perspektiven*

Zum einen konnten auf diesem Wege Informationen gewonnen werden, die sich in ihrem je spezifischen Zugang zur untersuchten Projektrealität ergänzten und in ihrer Verbindung ein vertieftes Verständnis von Ereignissen und Gegebenheiten ermöglichen.

#### Fallbeispiel:

Ein außerschulisches Bildungsprojekt führt ein mehrtägiges interkulturelles Bildungsseminar mit einer 9. Hauptschulklasse durch. Von Anfang an reagieren die Jugendlichen, von einzelnen Ausnahmen abgesehen, mit Desinteresse oder gar deutlicher Ablehnung auf das Projekt; es herrscht große Unruhe, nur

---

6 Mit Triangulation wird in der Sozialforschung die „Betrachtung eines Forschungsgegenstandes von mindestens zwei Punkten“ (Seipel/Rieker 2003, S. 224) bezeichnet, wobei zwischen verschiedenen Formen der Triangulation unterschieden wird. Folgt man der in der Fachliteratur gängigen Unterteilung nach Denzin (ders., zit. nach Flick 2000, S. 310; vgl. auch Seipel/Rieker 2003, S. 225f.), dann stellt unser Vorgehen eine Kombination aus Daten-Triangulation (Erhebung bei unterschiedlichen Akteursgruppen und zu verschiedenen Zeitpunkten) und methodologischer Triangulation (Verwendung verschiedener Methoden) dar.

wenige lassen sich überhaupt auf Gesprächsangebote ein. Lärm und offene Gesprächsverweigerungen nehmen in den Folgestunden so sehr zu, dass das Seminar am 3. Tag nur nach einer Intervention des Direktors und der Ankündigung disziplinarischer Maßnahmen zu Ende geführt werden kann. Alle Beteiligten sind sich im Nachhinein einig, dass das Projekt ein ziemliches Fiasko war.

Zur Erklärung des konflikthaften Verlaufs wurden von den Projektmitarbeitenden und vom Lehrpersonal der Schule folgende Deutungen ins Spiel gebracht: Es handelte sich um eine besonders schwierige Klasse, in der viele Schüler/innen Lernprobleme haben, einige ausgeprägt fremdenfeindliche Einstellungen vertreten und das Klassenklima insgesamt nicht gut ist, da viele so kurz vor dem Abschluss immer noch keine Aussicht auf eine Arbeit oder Lehrstelle haben. Bei unserer teilnehmenden Beobachtung wurde außerdem erkennbar, dass die Teamerinnen und Teamer in manchen Situationen ungeschickt agierten und auf diese Weise ihrerseits zu einer Eskalation der Situation beitrugen. Jede dieser Informationen trug einen Teil zur Erklärung des Geschehens bei; dennoch konnte damit nicht befriedigend geklärt werden, warum die Klasse bereits von Anfang an mit so starker Abwehr auf das Projekt reagierte – zumal das Angebot bereits in anderen Schulklassen mit vergleichbaren Rahmenbedingungen durchgeführt wurde, die Teamerinnen und Teamer eine ähnliche Situation nach eigenem Bekunden jedoch noch nicht erlebt hatten. Erst in den nach Projektende durchgeführten Gruppeninterviews mit den Schülerinnen und Schülern wurde ein weiterer Aspekt sichtbar, der sich aus den anderen Perspektiven nicht erschloss: Wie die Jugendlichen in den Interviews übereinstimmend berichteten, wurde ihnen seitens der Schule deutlich vermittelt, dass sie aufgrund ihrer Einordnung als „Problemklasse“ für die Teilnahme an dem Seminar ausgewählt worden waren. „Da gibt es so ein Projekt, das euch vielleicht bei ein paar eurer Probleme helfen kann“ – so erinnert sich ein Schüler an die diesbezüglichen Erläuterungen seines Lehrers. In den Ausführungen der Jugendlichen wird zudem erkennbar, dass sie das Seminar vor diesem Hintergrund als stigmatisierende Sonderbehandlung wahrnahmen – was bei ihnen verständlicherweise eine negative Grundeinstellung und Abwehrreaktionen gegenüber dem Projekt erzeugte.

Ursachen für das Scheitern des Seminars waren nicht nur auf Seiten der Zielgruppen zu suchen (oder auf ungeschicktes Agieren der Projektmitarbeiterinnen/-mitarbeiter zurückzuführen); dazu beigetragen hatte auch eine mangelnde bzw. unter den Kooperationspartnern unzureichend abgestimmte Vorbereitung der Maßnahme im Vorfeld, die dem Projekt in einer ohnehin schwierigen Situation extrem ungünstige „Startbedingungen“ verschaffte. Während in diesem Fall durch die Rekonstruktion unterschiedlicher Perspektiven Gründe für das Nichterreichen angestrebter Lernziele herausgearbeitet

wurden, konnten in anderen Fällen Elemente gelingender Praxis sichtbar gemacht werden, die sich aus der Perspektive einzelner Projektakteure alleine nicht erschlossen hätten. Wie im vorgestellten Beispiel war es dabei in vielen Fällen insbesondere die Einbeziehung der Adressatenperspektive, die solche weitergehenden Einblicke eröffnete. In anderen Fällen waren es aber auch Gespräche mit Kooperationspartnern oder informelle Gesprächssituationen mit einzelnen Projektakteuren – z. B. während einer gemeinsamen Heimfahrt nach einem Seminar – die eine für das Verständnis bestimmter Entwicklungen wesentliche Zusatzinformation lieferten.

– Divergierende Perspektiven

In anderen Fällen erbrachte die Erhebung unterschiedlicher Perspektiven aber auch divergierende, sich z.T. auch offen widersprechende Informationen, die alternative Deutungen des Geschehens zuließen.

Fallbeispiel:

In einer Schule mit hohem Migrantenanteil wird ein interkulturelles Schülercafé betrieben – ein unter Beteiligung von Schülerinnen und Schülern organisierter Treffpunkt mit flankierenden Angeboten wie Hausaufgabenbetreuung, Bewerbungstraining, Tanz-AG etc. Der Projektleiter charakterisiert das Projekt im Interview als ein interkulturelles Begegnungsangebot für Schülerinnen und Schülern unterschiedlicher Herkunft; als wesentliche Ziele nennt er ein besseres gegenseitiges Verständnis und den Abbau von Vorurteilen zwischen deutschstämmigen Schülerinnen und Schülern und Migrantenkindern; gefragt nach den Resultaten der Arbeit formuliert er die Überzeugung, dass das Projekt erfolgreich entsprechende Lernimpulse zu setzen vermag.

Bei zwei Vor-Ort-Besuchen in der Einrichtung machen wir die Beobachtung, dass der Treffpunkt überwiegend von Migrantenkindern genutzt wird; die wenigen Deutschen bleiben – von einzelnen Ausnahmen abgesehen – unter sich bzw. nutzen punktuell spezifische Angebote, ohne sich auf das kontinuierliche Begegnungsangebot einzulassen. Insgesamt gewinnen wir den Eindruck, dass das Angebot kaum dazu beiträgt, dass es zu Begegnungen zwischen den beiden Gruppen kommt. Es werden für uns aber auch keine Bestrebungen der anwesenden Pädagoginnen und Pädagogen erkennbar, solche Begegnungen gezielt zu fördern.

Wie sich im Gespräch mit dem vor Ort zuständigen Pädagogen zeigt, ist diesem die Diskrepanz zwischen von der Projektleitung formulierter Zielstellung und Projektrealität allerdings überhaupt nicht bewusst; mehr noch, sie ist für ihn auch nicht relevant. Im weiteren Gesprächsverlauf wird vielmehr erkennbar, dass dieser Pädagoge die Funktion des Projekts auf ganz anderer Ebene, nämlich in der sozialpädagogischen Hilfestellung für sozial benachteiligte

Jugendliche, verortet – ein Verständnis, das sowohl seine Interpretation des Projektgeschehens als auch sein eigenes Handeln in diesem Kontext prägt. Die Einbeziehung weiterer Perspektiven relativierte somit nicht nur die Aussage der Projektleitung zu den vom Projekt verfolgten Zielen und den dabei erreichten Erfolgen; ihre Gegenüberstellung ermöglichte auch die Erschließung von Faktoren, die die beobachtete Diskrepanz zwischen offiziellem Anspruch und de facto fehlender interkultureller Orientierung zu erklären vermochten: Es wurde erkennbar, dass bei Projektleitung und mit der Umsetzung betreutem Pädagogen ganz unterschiedliche Vorstellungen über Projektinhalte und -ziele existierten, die möglicherweise aus unterschiedlichen Zuständigkeiten und damit verbundenen professionellen Interessen – hier die erfolgreiche Akquise von Fördergeldern, dort das situationsbezogene Reagieren auf konkrete Bedarfslagen – resultierten und offensichtlich zwischen den Beteiligten unzureichend kommuniziert wurden.

Für die Interpretation des Projektgeschehens konnten solche divergierenden Perspektiven in zweifacher Hinsicht informativ sein:

- Zum einen erwiesen sie sich als hilfreich, um den Stellenwert einzelner Aussagen besser beurteilen zu können, d.h. sie halfen bei der Bewertung von Informationen. Eine Schwierigkeit, die sich hier stellen konnte, war die Frage der Gewichtung der einzelnen Perspektiven. Wo sich diese nicht aus der Gegenüberstellung selbst ergab (indem z. B. objektive Kontextinformationen/Fakten eine subjektive Einschätzung widerlegten) bemühten wir uns, durch ergänzende Erhebungen – weitere Gespräche, gezielte Nachfragen, eigene Beobachtungen – weitere Hinweise zu erlangen, die die eine oder andere Position untermauerten. So war im obigen Beispiel das Gespräch mit dem Projektmitarbeiter gezielt gesucht worden, da das Interview mit der Projektleitung und der eigene Eindruck beim Projektbesuch deutlich auseinander lagen. Wie in diesem Fall reichte häufig bereits die Einbeziehung einer dritten Perspektive, um das Bild zu verdeutlichen. Es konnte aber auch vorkommen, dass jede weitere Perspektive eine alternative Interpretation des Geschehens und seiner Bedeutung erbrachte und die unterschiedlichen Informationen ein so widersprüchliches Bild zeichneten, dass eine Situation letztlich nicht interpretierbar war.
- Zum anderen konnten solche unterschiedlichen Perspektiven auch an sich aufschlussreich sein, indem sie etwa eine mangelnde inhaltliche Übereinstimmung oder den unzureichenden Austausch beteiligter Akteure sichtbar machten, die wiederum bestimmte Entwicklungen (oder auch das Ausbleiben von Entwicklungen) zu erklären vermochten. In dieser Hinsicht erwiesen sich vor allem die Sichtweisen unterschiedlicher Projektbeteiligter (Konzept- und Umsetzungsebene, aber auch Kooperationspartner) als interessant. Erhellend war in manchen Fällen aber auch die Gegenüberstellung von

Projektakteurs- und Adressatenperspektive. So zeigte sich verschiedentlich bei Projekten, die sich nicht in der erhofften Weise entwickelten, dass sich die Perspektiven von Teamenden und Teilnehmenden auf Projekthinhalte, aber auch auf das Projektgeschehen z.T. erheblich unterschieden. In manchen Fällen wurde auf diese Weise deutlich, dass Projekte das primäre Anliegen ihrer Arbeit überhaupt nicht hatten vermitteln können; oder es zeigte sich, dass Pädagoginnen und Pädagogen über die Köpfe der Teilnehmenden hinweg agiert hatten, weil sie deren Interessen und Bedürfnisse falsch eingeschätzt hatten. Erkennbar wurden aber auch Missverständnisse zwischen Beteiligten, die sich belastend auf das Seminarklima auswirkten und dadurch den Projekterfolg beeinflussten.

Wie auch die geschilderten Beispiele zeigen, ermöglichte dieses multiperspektivische Vorgehen im Rahmen unserer Untersuchung vor allem eine Identifikation von Ursachen schwieriger Projektverläufe sowie von Faktoren, die zum Scheitern angestrebter Lernprozesse führten. Über die Bestimmung solcher „Stolpersteine“ konnten dann wiederum – sozusagen im Umkehrschritt – Kriterien gelingender Praxis definiert werden. Verschiedentlich trug dieses Verfahren aber auch dazu bei, erfolgreiche Strategien und Ansätze zu identifizieren, indem es eine „Verdichtung“ von Hinweisen auf Voraussetzungen und Elemente gelingender Praxis ermöglichte.

### *Schlussbetrachtung: Offenheit und Multiperspektivität in der Annäherung an komplexe Projektrealitäten*

In der hier vorgestellten wissenschaftlichen Begleitung von Projekten Interkulturellen Lernens setzten wir ein sehr offenes Untersuchungsdesign um, indem wir auf eine Festlegung des zu untersuchenden Gegenstandes im Vorfeld und auf ein einheitliches Verfahren bei der Erhebung verzichteten. Mit diesem Vorgehen orientierten wir uns zum einen an den Gegebenheiten in diesem pädagogischen Handlungsfeld; zum anderen trugen wir damit einem Untersuchungsinteresse Rechnung, das weniger der Überprüfung auf dem Papier vorliegender Konzepte unter idealtypischen Bedingungen galt, als der existierenden Projektlandschaft und den Bedingungen ihrer Umsetzung im Projektalltag.

Die Entscheidung für ein solches Untersuchungsdesign hatte auch Auswirkungen darauf, inwiefern – für welchen Bereich und mit welcher Reichweite – Aussagen über Wirkungen der untersuchten Projekte getroffen werden konnten. So war es im Rahmen dieses Designs weder möglich, statistisch abgesicherte – in der Fachwelt ohnehin strittige – Wirkungsnachweise in Form valider Messdaten zu erbringen, noch Wirkungsüberprüfungen auf der Basis abgesicherter Hypothesen oder im Vorfeld klar definierter Ursache-Wirkungs-Annahmen vorzunehmen (vgl. Haubrich u.a. 2005, S. 2). Stattdessen näherten wir uns der Wirkungsthematik in anderer Weise an: Indem wir der Frage nachgingen, „Was funktioniert nicht?“ nahmen wir zugleich eine negative Bestimmung der Wirkfähigkeit von Projekten

vor und wählten so den Weg einer „Annäherung durch Ausschluss“<sup>7</sup>. Zum anderen konnten wir auf der Basis unserer Daten – der identifizierten „Stolpersteine“ und Elemente gelingender Praxis – zentrale Wirkungsvoraussetzungen herausarbeiten, die gegeben sein müssen, damit Projekte im Bereich Interkulturelles Lernens überhaupt Wirkung entfalten können.

Dieses Vorgehen ist allerdings auch mit bestimmten Anforderungen und Schwierigkeiten verbunden, die man sich bei der Entscheidung für ein solches Design vergegenwärtigen sollte. So brachte es der weite, zunächst bewusst nicht selektierende Untersuchungsfokus mit sich, dass wir auch einige „Irrwege“ gingen, indem wir uns in Aspekte vertieften, die zunächst vielversprechend erschienen, sich im weiteren Verlauf jedoch als wenig relevant erwiesen. Die Unterschiedlichkeit des Vorgehens bei der Datenerhebung schränkte zudem die Vergleichbarkeit der gewonnenen Daten ein, so dass zu bestimmten Aspekten die Datenbasis im Endeffekt zu schmal war, um sie systematisch auswerten zu können. Letztlich beinhaltete dieses Vorgehen deshalb auch einen erhöhten Ressourcenaufwand oder anders formuliert: Um ein solches Design umzusetzen, bedarf es auch einer bestimmten Kapazität an Ressourcen, die den „Luxus“ eines mit Suchbewegungen, Umwegen und Sackgassen verbundenen Erhebungsprozesses erlauben. Dies und die günstigen Ausgangsvoraussetzungen einer weitgehenden Unabhängigkeit von Auftraggeberinteressen und -vorgaben machen das von uns gewählte Vorgehen sicherlich nicht auf jeden Evaluationskontext übertragbar.

Dennoch bleibt festzuhalten, dass sich gerade die offene Anlage des Designs, das dadurch ermöglichte induktive, gegenstandsbezogene Vorgehen wie auch die damit verbundene Flexibilität als gut geeignet erwiesen, um den Spezifika dieses Arbeitsfeldes Rechnung zu tragen und die Umsetzungsbedingungen und -möglichkeiten von pädagogischer Praxis in diesem Feld in den Blick zu bekommen.

Ein weiteres zentrales Element dieses Designs war die mithilfe verschiedener Erhebungsmethoden realisierte Perspektivenvielfalt, durch die eine Triangulation unterschiedlicher subjektiver Perspektiven möglich wurde.

Trotz der in der Fachliteratur wie auch in Berichten aus der Evaluationspraxis immer wieder konstatierten Begrenztheit von Erklärungsmodellen, die nur auf einer Sorte Daten fußen bzw. sich einseitig auf die Perspektiven einer Akteursgruppe gründen (vgl. etwa Flick 2000; Glaser/Strauss 1998; S. 72ff.; Lüders i. d. Bd. sowie die Evaluationsberichte von Klawe (2006) und Masscanae (2006)), sind Evaluationsdesigns, in denen unterschiedliche Perspektiven systematisch aufeinander bezogen werden, im Feld der Rechtsextremismusprävention und Toleranzerziehung noch vergleichsweise selten. Zwar wird gerade in diesem Feld ein Bemühen erkennbar, mehrere Perspektiven in eine Evaluation zu integrieren (vgl. Schuster i. d. Bd.); häufig beziehen sich diese jedoch auf verschiedene Aspekte eines Projekts oder dienen der Umsetzung unterschiedlicher Evaluationsphasen – indem z. B.

---

7 Mit diesem Argument greife ich eine Überlegung auf, die Christian Lüders auf einem Expertenworkshop der Arbeitsstelle zum Thema „Evaluation in der Rechtsextremismusprävention“ (vgl. Schuster i. d. Bd.) am 09.11.2006 in Halle formulierte.

Fachkräfte im Vorfeld interviewt und auf dieser Basis ein Fragebogen erstellt wird, mit dem Adressatinnen und Adressaten abschließend befragt werden. Insbesondere mangelt es an Ansätzen, in denen die Sichtweisen von Interessierten eine adäquate Einbeziehung erfahren (vgl. auch Rieker i. d. Bd.).

Ein Grund dürfte auch hier der erhöhte Ressourcenaufwand sein, den solche Designs bei der Datengewinnung unzweifelhaft erfordern. Hinzu kommt, dass „die Einbeziehung verschiedener Perspektiven [...] nicht nur mit einem Zugewinn an Informationen, sondern auch mit zusätzlicher Komplexität verbunden“ (Rieker 2005, S. 96) ist, die auch an die Auswertung erhöhte Ansprüche stellen. Die z.T. auch bei unserer Untersuchung aufgetretenen Schwierigkeiten bei der Gewichtung von Daten und die dabei gemachte Erfahrung, dass sich die gewonnene Perspektivenvielfalt manchmal auch einer eindeutigen Interpretation entzog, verweisen dabei auf ein grundsätzliches, erkenntnistheoretisches Problem: Dem Streben nach objektiver „Wahrheit“ ist auch bei diesem Vorgehen durch die Eigenheit sozialer Prozesse Grenzen gesetzt, die ihre Relevanz für die Beteiligten eben immer auch vor dem Hintergrund subjektiv unterschiedlicher Erfahrungen und Wertmuster entfalten. Es ist deshalb nicht auszuschließen, dass unterschiedliche Perspektiven Ausdruck subjektiv unterschiedlicher Wirklichkeiten sind, weshalb die Frage, wessen Sicht nun „objektiver“ ist bzw. die „objektive“ Realität wiedergibt, letztlich nicht beantwortbar ist (vgl. Rieker 2005, S. 97). Insofern können auch multiperspektivische Verfahren immer nur eine Annäherung an komplexe Projektrealitäten darstellen, sie tragen „zu einem vertieften Verständnis des untersuchten Gegenstandes [bei] (...) und weniger zu Validität und Objektivität in der Interpretation“ (Flick 2000, S. 311).

Trotz dieser Einschränkungen verweisen die in unserer Untersuchung gemachten Erfahrungen darauf, dass die Einbeziehung unterschiedlicher subjektiver Perspektiven dazu beitragen kann, die zwangsläufige Selektivität von Methoden und Perspektiven in der Evaluation von Bildungsprozessen zu vermindern und zu einem solchen „vertieften Verständnis“ des Untersuchungsgegenstandes zu gelangen. Die sichtbar gewordenen Diskrepanzen und Begrenzungen der unterschiedlichen Perspektiven bestätigten zum einen den von Lüders und Haubrich formulierten Befund, dass „Evaluationen, wenn sie die Qualität pädagogischen Handelns in den Blick nehmen, ohne Einbezug der Adressatinnen und Adressaten kaum möglich erscheinen“ (Lüders/Haubrich 2003, S. 326). Vor dem Hintergrund der Umsetzungsbedingungen von Projekten in diesem Feld erwies sich aber auch die Triangulation von Perspektiven unterschiedlicher beteiligter Fachkräfte als wichtiger Zugang, um einen differenzierten Blick auf das Projektgeschehen zu erhalten. Angesichts dieser Erfahrungen erscheint es uns für die zukünftige Weiterentwicklung der Evaluationspraxis in pädagogischen Settings wünschenswert, verstärkt Evaluationsdesigns zu entwickeln sowie die hierfür erforderlichen Ressourcen zur Verfügung zu stellen, die eine systematische Einbeziehung unterschiedlicher subjektiver Perspektiven ermöglichen.

Abschließend lässt sich festhalten: Die spezifischen Anforderungen, die unser Untersuchungsgegenstand an uns stellte und auf die wir mit unserem Forschungsdesign reagierten, waren bestimmten Charakteristika dieses pädagogischen Handlungsfeldes geschul-

det. Dazu gehörten eine wenig erforschte und gering operationalisierte Praxis, ein durch kurze Projektlaufzeiten begrenzter Feldzugang sowie die generelle Schwierigkeit, die Vielschichtigkeit und intersubjektiv sich konstituierende „Wirklichkeit“ sozialer Prozesse zu erfassen. Diese Voraussetzungen prägen jedoch nicht nur die pädagogische Praxis im Bereich Interkulturellen Lernens; sie sind auch kennzeichnend für andere Bereiche der präventiven Arbeit gegen Rechtsextremismus und Fremdenfeindlichkeit. Offene, explorative Verfahren, die durch Einbeziehung verschiedener Methoden ein multiperspektivisches Vorgehen ermöglichen, erscheinen uns deshalb generell erforderlich, um vertiefte Kenntnisse über Voraussetzungen und Bedingungen gelingender Praxis in der Prävention von Rechtsextremismus und Fremdenfeindlichkeit zu erhalten.

## Literaturverzeichnis

- Ader, Sabine (2007): „Wissen um die eigene Wirkung“. Von der Bedeutsamkeit professioneller und institutioneller Deutungskompetenz und Reflexivität. In: *Jugendhilfe*, 45. Jg., H. 4, S. 187–196
- Ahlheim, Klaus (2005): Wirkungsoptimismus und Messbarkeitskepsis – Evaluation und Wirkungsforschung in der politischen Erwachsenenbildung. In: *Praxis Politische Bildung*, H. 1, S. 17–24
- Beywl, Wolfgang (2006): Demokratie braucht wirkungsorientierte Evaluation – Entwicklungspfade im Kontext der Kinder- und Jugendhilfe In: Projekt eXe (Hrsg.): *Wirkungsevaluation in der Kinder- und Jugendhilfe. Einblicke in die Evaluationspraxis*. München, DJI, S. 25–46
- Breitenbach, Diether u.a. (1979): Kommunikationsbarrieren in der Internationalen Jugendarbeit, Saarbrücken/Fort Lauderdale
- Flick, Uwe (2000): Triangulation in der qualitativen Forschung. In: Flick, Uwe/Kardorff, Ernst von, Steinke, Ines (Hrsg.), *Qualitative Forschung. Ein Handbuch*, Reinbek bei Hamburg, Rowohlt: S. 309–318
- Frindte, Wolfgang/Preiser, Siegfried (2007): Präventionsansätze gegen Rechtsextremismus. In: *Aus Politik und Zeitgeschichte* H. 11, S. 32–38
- Glaser, Barney/Strauss, Anselm (1979): Die Entdeckung gegenstandsbezogener Theorie: Eine Grundstrategie qualitativer Sozialforschung. In: Hopf, Christel, Weingarten, Elmar (Hrsg.), *Qualitative Sozialforschung*, Stuttgart: Klett-Cotta, S. 91–108
- Glaser, Barney/Strauss, Anselm (1998), *Grounded Theory. Strategien qualitativer Forschung*, Bern/Göttingen/Toronto/Seattle: Huber
- Glaser, Michaela (2005): Von der Einsamkeit der Gipfel und den Mühen der Ebene. Theorie und Praxis interkulturellen Lernens in der Jugendarbeit – eine Landschaftsbeschreibung. In: *kursiv – Journal für politische Bildung*, H. 2, S. 46–56
- Glaser, Michaela/Rieker, Peter (2006): *Interkulturelles Lernen als Prävention von Fremdenfeindlichkeit*. Halle: DJI

- Haubrich, Karin/Holthusen, Bernd/Struhkamp, Gerlinde (2005): Evaluation – einige Sortierungen zu einem schillernden Begriff. In: DJI Bulletin Sonderteil, H. 72, S. 1–4
- Hirsland, Katrin/Cecchini, Michela/Odom, Tanya M. (2004): Evaluation in the Field of Education for Democracy, Human Rights and Tolerance. A Guide for Practitioners. Gütersloh: Bertelsmann
- Klawe, Willy (2006): Multiperspektivische Evaluationsforschung als Prozess-Wirkungskonstruktion aus der Sicht der Beteiligten. In: Projekt eXe (Hrsg.): Wirkungsevaluation in der Kinder- und Jugendhilfe: Einblicke in die Evaluationspraxis. München: DJI, S. 125–140
- Lüders, Christian/Haubrich, Karin (2003): Qualitative Evaluationsforschung. In: Schweppe, Cornelia (Hrsg.): Qualitative Forschung in der Sozialpädagogik. Opladen: Leske + Budrich, S. 305–330
- Lüders, Christian/Haubrich, Karin (2006): Wirkungsevaluation in der Kinder- und Jugendhilfe. Über hohe Erwartungen, fachliche Erfordernisse und konzeptionelle Antworten. In: Projekt eXe (Hrsg.): Wirkungsevaluation in der Kinder- und Jugendhilfe. Einblicke in die Evaluationspraxis. München: DJI, S. 5–23
- Mascenaere, Michael (2006): Wirkungen der Kinder- und Jugendhilfe sind messbar! Methoden, Ergebnisse und Empfehlungen der Jugendhilfe – Effekte – Studie (JES) und weiterer darauf beruhender wirkungsorientierter Evaluationen. In: Projekt eXe (Hrsg.): Wirkungsevaluation in der Kinder- und Jugendhilfe. Einblicke in die Evaluationspraxis. München: DJI, S. 49–78
- Rieker, Peter (2005): Delinquenz aus Sicht von Eltern und Kindern – Herausforderungen und Chancen der Triangulation subjektiver Perspektiven. In: Psychotherapie und Sozialwissenschaft. Zeitschrift für qualitative Forschung und klinische Praxis, 7. Jg., H. 2, S. 95–129
- Seipel, Christian/Rieker, Peter (2003): Integrative Sozialforschung. Konzepte und Methoden der qualitativen und quantitativen empirischen Forschung. Weinheim/München: Juventa
- Thomas, Alexander (1988) (Hrsg.): Interkulturelles Lernen im Schüleraustausch. SSIP-Bulletin, Nr. 58, Saarbrücken/Breitenbach
- Thomas, Alexander (1996): Können interkulturelle Begegnungen Vorurteile verstärken? In: ders. (Hrsg.): Psychologie und multikulturelle Gesellschaft. Göttingen/Bern/Toronto/Saettle: Hogrefe-Verlag, S. 227–237
- Uhl, Katrin (2004): Evaluation von Demokratie- und Toleranzerziehungsprogrammen – die Erfahrungen aus dem Projekt „Erziehung zu Demokratie und Toleranz“. In: Uhl, Katrin (Hrsg.): Evaluation politischer Bildung. Ist Wirkung messbar? Gütersloh: Bertelsmann Stiftung, S. 51–55

## Evaluation in der Demokratie- und Toleranzerziehung – Wirkungsanalyse zu Alltagskonflikten von Jugendlichen

Während im Bereich der Evaluation von politischer Bildung in den letzten Jahren einiges unternommen wurde – verwiesen sei hier u.a. auf die Evaluationsstudien politischer Bildung von Schröder u.a. (2004) und Fritz/Maier (2005) sowie Publikationen in der Zeitschrift „Praxis Politische Bildung“, zuletzt Heft 4/2006 – gibt es in der Demokratie- und Toleranzerziehung noch einen Nachholbedarf. Zum einen ist die Evaluation im Bildungsbereich in Deutschland noch ein relatives Neuland und zum anderen findet der Begriff Evaluation auf recht unterschiedliche Sachverhalte Anwendung. Hier fehlt es bisher an einem umfassenden Konzept der Wirkungsforschung (vgl. Uhl 2004). Unsere Evaluation des Projektes „Schule und Ausbildung für Toleranz und Demokratie (SAFT)“ reiht sich in die bisher wenigen Untersuchungen ein, die zu Programmen dieses pädagogischen Arbeitsfeldes vorliegen. Mit den Wirkungen der auch in diesem Projekt verwendeten Programme haben sich bisher drei andere Evaluationen befasst, deren Ergebnisse die Diskussion zu diesem Themenfeld weiter vorangebracht haben (Hirsland u.a. 2004; Kehr 2004; Bommes u.a. 2004).

Im Folgenden soll es darum gehen, die in unserer Untersuchung (vgl. Schäfer u.a. 2006) gesammelten Erfahrungen zu diskutieren. Im Mittelpunkt unserer Projektevaluation stand die Untersuchung von Wirkungen von Angeboten zum Demokratie lernen im Lebensalltag von Jugendlichen. Der Fokus wurde dabei auf die Einschätzung von Konflikten und das individuelle Konfliktverhalten gerichtet. Nach einer kurzen Darstellung des Forschungsgegenstandes (1) stellen wir den Forschungsansatz (2) dar und berichten von Erfahrungen (3) aus dem Evaluationsvorhaben, die sich zugleich als Konsequenzen für künftige Untersuchungen in diesem Themenfeld diskutieren lassen.

### **Forschungsgegenstand: Trainingskurse im Rahmen des Projekts „Schule und Ausbildung für Toleranz und Demokratie (SAFT)“**

Bis Ende 2004 fand in der Europäischen Jugendbildungs- und Jugendbegegnungsstätte Weimar ein dreijähriges XENOS-gefördertes Projekt zur Demokratie- und Toleranzerziehung statt. Schwerpunkt dieses von uns evaluierten Projektes waren drei fünftägige Trainingskurse mit Schul- und Ausbildungsklassen unterschiedlicher Thüringer Schulen, die innerhalb von etwa zwei Kalenderjahren durchgeführt wurden. Methodische Grundlage stellten dabei Bildungsprogramme dar, die in den vergangenen Jahren mit Unterstützung der Bertelsmann Stiftung durch das Centrum für angewandte Politikforschung (CAP) entwickelt bzw. adaptiert und erprobt wurden.

Im ersten Projektjahr fand zunächst für alle Teilnehmenden ein Einführungsseminar statt, in dem sie mit den Methoden und Lernformen im Rahmen der Demokratie- und Tole-

ranzerziehung vertraut gemacht wurden. Zudem wurden erste Annäherungen an einige im Rahmen des Projektes wesentliche Begriffe wie „Demokratie“ und „Toleranz“ unternommen.

Das darauf folgende Toleranzseminar bestand schwerpunktmäßig aus Übungen des Programms „Achtung (+) Toleranz“ (Ulrich u.a. 2001). Im Mittelpunkt dieses Seminars stand die Vermittlung von Definitionen der Begriffe „Konflikt“, „Gewalt“ und „Toleranz“. Diese sind nötig, um das am CAP entwickelte operationalisierbare Toleranzmodell für die politische Bildung zu verstehen (Bertelsmann Forschungsgruppe Politik 2001). Im weiteren Seminarverlauf wurden Grundlagen der Kommunikation vermittelt und Elemente einer partnerschaftlichen Kommunikation als Möglichkeit für das Erlernen von tolerantem Verhalten im Alltag eingeübt.

Der dritte Kurs, das Miteinanderseminar, wurde mit Übungen aus dem Programm „Miteinander – Erfahrungen mit Betzavta“ (Ulrich u.a. 1997) durchgeführt. Im Mittelpunkt dieses Seminars standen die Auseinandersetzung mit dem Thema „Demokratie“ und die Entwicklung eines qualitativen Demokratieverständnisses. Neben der Thematisierung von Aspekten einer Demokratie als Gesellschaftsform ging es dabei im Kern des Seminars um Fragen einer Demokratie als Lebensform (vgl. Himmelmann 2001).

Die inhaltliche Klammer zwischen diesen drei Seminaren stellte die Entwicklung eines gegenseitigen zwischenmenschlichen Verständnisses für den oder das Andere dar. Auf der Basis einer Anerkennung des gleichen Rechts auf freie Entfaltung sollten tolerantes Verhalten und ein Verständnis für demokratisches Miteinander entwickelt und befördert werden.

Im Anschluss an diese drei Kurse fanden für beteiligte Schulen Abschlussseminare statt, an denen neben einzelnen Schülerinnen und Schülern, die das reguläre Programm des Projektes durchlaufen hatten, auch Lehrerinnen und Lehrer, Eltern sowie andere Schüler/innen teilnehmen konnten.

Neben den Seminaren für die Jugendlichen fanden Fortbildungsveranstaltungen für Lehrpersonen aus den beteiligten Schulen bzw. Ausbildungseinrichtungen statt. Damit wurde eine umfassende Projektstrategie verfolgt, die dem Anspruch, der sich im Titel „Schule und Ausbildung für Toleranz und Demokratie“ widerspiegelt, in der konzeptionellen Ausrichtung entsprechen sollte (vgl. Schäfer u.a. 2006).

Die im Projekt SAFT verwendeten Seminar- und Trainingskonzepte zur Demokratie- und Toleranzerziehung, fußen auf der Überzeugung, dass eine demokratische Gesellschaft auf Toleranz angewiesen ist. Intoleranz – in welcher Form auch immer – schafft zunehmend Konflikte, die im Vorfeld der Rechtsordnung bzw. der Anwendung des staatlichen Gewaltmonopols nur durch Toleranz zu regeln sind. Toleranz wird zur existentiellen Bedingung von Demokratie und Erziehung zur Toleranz eine bildungspolitische und -praktische Zukunftsaufgabe.

„Demokratie lernen“ bedeutet mehr als das Lernen von demokratischen Strukturen und geht über das übliche Verständnis der Demokratie als Staatsform hinaus. Die Frage nach Demokratie und demokratischen Verhaltensweisen stellt sich somit nicht nur in der Politik, sondern auch in der Schule, in der Familie und in jeder alltäglichen Interaktion. Demokratie

als Lebensform zu verstehen bedeutet somit, zu lernen wie alltägliche Konflikte demokratisch und gleichberechtigt geregelt werden können. Dieser Prozess ist kaum allein durch Wissen zu vermitteln, sondern muss im Interesse der Nachhaltigkeit erfahren werden.

### **Forschungsansatz: Evaluation im Feld der Demokratie- und Toleranzerziehung**

Die Evaluation von Methoden der Erziehung zu Toleranz und Demokratie als Projekt der politischen Bildung stellt erhebliche Anforderungen an die wissenschaftliche Begleitung. Zum einen erweist sich die Messung von Toleranz- und Demokratievariablen als komplexe und zum Teil umstrittene Herausforderung (vgl. Ahlheim 2003). Zum anderen begleitete das Projekt der EJBW sowohl Auszubildende als auch Schülerinnen und Schüler in einem mehrjährigen Lernprozess, der den spezifischen politisch sozialisatorischen Kontext der neuen Bundesländer berücksichtigen musste und zugleich den Auftrag des Förderprogramms „XENOS – Leben und Arbeiten in Vielfalt“ zu erfüllen hatte: Toleranz und demokratischen Respekt gegenüber Angehörigen anderer Kulturen und Nationen zu fördern. Eine wissenschaftliche Begleitung, die diese Mehrdimensionalität der Problemlage adäquat abbilden möchte, erfordert einen erheblichen Ressourceneinsatz, der an dieser Stelle aber nur bedingt zur Verfügung stand.

#### *Das Forschungskonzept*

Die Evaluation von Angeboten der politischen Bildung sollte dem Bildungsziel angemessen sein: Eine auf demokratischen und partizipativen Prinzipien beruhende Bildungsarbeit, die Menschen zu selbstverantwortlichem Handeln als politische Subjekte befähigen will, darf diese in der Evaluation nicht zu Objekten machen, sondern sollte sie weitgehend beteiligen. Wenngleich der Anspruch an eine partizipative Evaluation (vgl. Ulrich/Wenzel 2003) prinzipiell aufrechtzuerhalten ist, so konnte er im Rahmen des Projektes nur bedingt realisiert werden. Es wurde aber großer Wert gelegt auf eine enge Kooperation zwischen dem Projekträger der EJBW und dem Evaluationsteam am Fachbereich Sozialwesen der Fachhochschule Jena.

#### *Die Forschungsziele*

Die Ziele der Evaluation wurden durch die Zielstellung des Projektes präjudiziert. Es hatte sich zur Aufgabe gemacht, (a) soziale Schlüsselkompetenzen der Schüler zu befördern, (b) ein demokratisches Miteinander in den Seminaren erlebbar zu machen sowie (c) das demokratische Miteinander an der Schule der Projektteilnehmenden zu stärken. Es ging also um den sowohl kognitiven wie emotionalen Erwerb von Kenntnissen und Fähigkeiten sowie ihren Transfer in den Schulalltag. Damit stand die Frage der Wirkung der Seminare im Zentrum der Evaluation. Wenn hier die Rede von Wirkungen ist, so bedarf dieser Begriff einer Explikation. Mit Wirkung sind nicht objektive Entitäten gemeint, die in einem kausalen Ursache-Wirkungs-

verhältnis eindeutig und zweifelsfrei auf bestimmte Bildungsinhalte zurückzuführen sind. Wirkungen haben u. E. vielmehr einen konstruktivistischen Charakter; sie entstehen in kommunikativen Prozessen als eine gemeinsame Konstruktion sozialer Wirklichkeit, die sich zwar wesentlich aus inneren Deutungsmustern der Individuen speist, sich aber gleichzeitig in ihren Handlungen manifestiert. In diesem Sinne geht es um die Untersuchung der realen und potenziellen Wirkungen der Seminare auf unterschiedlichen Ebenen. Wir haben zwei Ebenen besonders in den Blick genommen: die individuelle Ebene der Persönlichkeitsentwicklung und die institutionelle Ebene, den Kontext der Schule. Um verstehen zu können, was sich auf den einzelnen Ebenen vollzieht, galt es die Impulse der Lernprozesse zu identifizieren. Deshalb wurde der Lernprozess der Projektteilnehmenden durch quantitative und qualitative Instrumente nachgezeichnet. Im Blick auf Wirkungen politischen Lernens sollte weiterhin auch die gesellschaftliche Ebene berücksichtigt werden. An dieser Stelle tut sich jedoch ein Hiatus auf zwischen dem Anspruch nach verlässlichen Aussagen über die (Miss-)Erfolge politischer Bildungsprogramme in der Gesellschaft und dem was eine zeitlichen, materiellen und personellen Friktionen unterliegende Evaluationsforschung zu leisten vermag.

### *Konflikte als zentraler Ansatzpunkt der Evaluation*

Wirkungen politischer Bildung können sich an ganz unterschiedlichen Orten manifestieren; uns ging es darum, sie an einem bestimmten Ereignis der alltäglichen Lebenswelt zu identifizieren, und zwar einem solchen, das in enger Beziehung zu den Bildungszielen der Seminare steht. Eine spezifische Toleranzkompetenz basiert auf der persönlichen Bereitschaft, individuellen Fähigkeiten und entsprechender Sachkenntnis, sich in Konfliktsituationen angemessen, d.h. insbesondere gewaltlos zu verhalten. Aus diesem Grund konzentrierten wir uns auf die von den Teilnehmenden erlebten Konflikte und deren Umgang damit. Die Konflikte wurden auf drei Ebenen beleuchtet:

- auf der emotionalen Ebene ging es um die während des Konflikts wahrgenommenen Gefühle,
- auf der kognitiven Ebene stand die Frage im Mittelpunkt, welche rationalen Schlüsse aus dem Konflikt gezogen und welche Einstellungen gewonnen werden und
- auf der handlungsorientierten Ebene wurden die praktischen Verhaltensweisen in Konflikten analysiert.

Einerseits wurden die genannten drei Ebenen in ihrem wechselseitigen Bezug untersucht und andererseits Entwicklungen während des Projektverlaufes aufgezeigt. Das Forschungsinteresse richtete sich gleichermaßen auf Konflikte in der Schule, im familiären Bereich sowie in der Peergroup als den zentralen Bezugspunkten der Lebenswelt von Auszubildenden und Schülerinnen und Schülern. Des Weiteren differenzierten wir in der Analyse der Daten nach Wirkungsarten, -räumen und -weisen. So entstand ein relativ komplexes Konstrukt der Wirkungsanalyse.

## *Die Forschungsinstrumente*

Im Verlauf des Projektes wurden unterschiedliche quantitative und qualitative Erhebungen durchgeführt. Alle Teilnehmenden wurden mit einem standardisierten Fragebogen zu vier Zeitpunkten befragt. Der Schwerpunkt lag auf der Konfliktbeurteilung und Konfliktregelungskompetenz der Teilnehmenden.

Nach jedem der drei Trainingsseminare wurden die Teilnehmenden mit einem Reflexionsbogen nach ihrer Einschätzung befragt. Ein wichtiger Schwerpunkt war dabei das Verständnis und dessen mögliche Veränderungen zu vier im Rahmen des Projektes zentralen Begriffen: Demokratie; Toleranz; Gewalt; Konflikt.

Der Fragebogen bezog sich als überwiegend summatives Instrument auf die Kernelemente des Toleranzkonzeptes. Der Reflexionsbogen hingegen diente in erster Linie als Instrument der Selbstevaluierung und der formativen Evaluation und damit zum Monitoring des Projektablaufes und der Qualitätsentwicklung des Projektes. Beide Instrumente „Fragebogen“ und „Reflexionsbogen“ sind voneinander weitgehend unabhängig und konnten als jeweils eigenständige Informationsquelle genutzt werden. Die Anlage der standardisierten Fragebogenuntersuchung machte es möglich, echte Paneldaten (Längsschnittstudie) über den gesamten Verlauf des fast dreijährigen Bildungszyklus zu gewinnen. Dies erbringt eine viel höhere Datenqualität als reine „Zeitreihen“-Analysen, die auf der jeweils gleichen Befragung, aber je unterschiedlichen Befragten beruhen. Dieser Ansatz ist mindestens in der Arbeit mit den verwendeten Bildungskonzepten bisher einmalig. Vergleichbare Längsschnittstudien zur Demokratie- und Toleranzerziehung sind uns nicht bekannt.

In zwei Befragungswellen wurden ergänzend zu den quantitativen auch qualitative Methoden der Evaluation eingesetzt. Schwerpunkt dabei war die Durchführung von leitfadengestützten Interviews mit ausgewählten Teilnehmenden. In diesen Interviews wurde auf drei Ebenen gefragt nach: Wirkungsarten, Wirkungsräumen und Wirkungsweisen der Seminarinhalte auf die Teilnehmenden. Ein weiteres Element der qualitativen Untersuchungsmethoden waren Gruppendiskussionen (vgl. Lamnek 1998, S. 29) mit Lehrkräften, Teamerinnen und Teamern sowie den Schulleiterinnen und Schulleitern, in denen es um Außenperspektiven auf die Teilnehmenden ging.

Die Abschlussseminare, die zum Ende des Projektes mit verschiedenen Teilnehmendengruppen der beteiligten Schulen stattgefunden haben, wurden durch das Evaluationsteam einer teilnehmenden Beobachtung unterzogen.

## **Forschungserfahrungen: Erkenntnisse und Konsequenzen zur Evaluation der Demokratie- und Toleranzerziehung**

Wenn hier von Forschungserfahrungen die Rede ist, so wollen wir dies nicht verstanden wissen als inhaltliche Ergebnisse unserer Evaluationsstudie; dieser Aspekt wird hier zugunsten einer Perspektive zurückgestellt, der es ausschließlich um solche Forschungserfahrungen geht, die sich aus dem gewählten Evaluationsdesign ergeben und die der

Diskussion um eine Weiterentwicklung des Evaluationsansatzes im Dreieck zwischen Selbst- und Fremdevaluation sowie Wirkungsforschung dienen sollen.

### *Partizipative versus kooperative Evaluation*

In seinem Aufsatz „Zur gesellschaftlichen Bedeutung und Entwicklung (qualitativer) Evaluationsforschung“ schreibt Ernst von Kardorff: „Evaluation befindet sich (...) in guter Gesellschaft mit Stichworten wie Empowerment, Partizipation oder Community, aber auch mit Mentoring, Aktivierung oder Selbstverantwortung. Die Nähe des Begriffs zu Konzepten der Entwicklung einer demokratischen bürgerschaftlichen Kultur selbst gesteuerter und partizipativer Aktivitäten von unten bezeichnet den einen Pol der mit Evaluation verbundenen Erwartungen und Hoffnungen“ (Kardorff 2006, S. 64). Angesichts der prinzipiellen Wirkungsungleichheiten gewinnt ein Evaluationskonzept an Bedeutung, das der Tatsache Rechnung trägt, dass Bildung stets als ein in permanenter Bewegung befindlicher Prozess der aktiven Aneignung durch die Lernenden zu verstehen ist. Da Bildung nicht vollständig plan- und steuerbar ist, kann es auch eine adäquate Evaluation nicht sein. Mit dem Konzept der „partizipativen Evaluation“ (Ulrich/Wenzel 2004) liegt ein Ansatz vor, der darauf abzielt, die unterschiedlichen Interessen und Perspektiven der Stakeholder zu integrieren. Durch eine partizipative Vorgehensweise, „bei der alle an einer Evaluation Beteiligten aktiv in den Dialog einbezogen werden und ein Höchstmaß an Transparenz gegeben ist, wird einem Großteil der Standards für Evaluation Rechnung getragen“ (Caspari 2006, S. 383). Von diesen Ansprüchen haben wir uns in unserer Evaluation leiten lassen; allerdings gibt es Grenzen in der praktischen Realisation partizipativer Evaluation. Diese liegen zum einen in den personellen und zeitlichen Ressourcen, die hohe Anforderungen an alle Beteiligten stellen und zum anderen kann nicht davon ausgegangen werden, dass alle Stakeholder wirklich an einer derart intensiven Zusammenarbeit interessiert sind. Vor diesem Hintergrund scheint es uns notwendig zu sein, den vielfach schillernden Begriff der Partizipation näher zu bestimmen, seine deskriptiven und normativen Aspekte zu erkunden und unterschiedliche Arten der Partizipation in Abhängigkeit von ihrer Intensität zu unterscheiden, so wie dies Oser und Biedermann (2006) getan haben. Da der Anspruch einer partizipativen Evaluation – trotz eines äußerst intensiven Austausches zwischen der Leitung des Projekts bei der EJBW und dem Evaluationsteam am Fachbereich Sozialwesen der Fachhochschule Jena – unter den gegebenen Bedingungen nicht einlösbar bzw. in der letzten Konsequenz nicht umsetzbar war, sprechen wir lieber von einer kooperativen Evaluation.

### *Gegenstandsadäquatheit*

Die Prämisse der Demokratisierung als Lerngegenstand und Prozessdimension politischen Lernens gilt es auch für die Durchführung der Evaluation geltend zu machen (vgl. Kuper 2005, S. 88). „Qualitative Forschung ist gekennzeichnet durch das Postulat der Gegenstandsangemessenheit von Methoden“ (Flick 2006, S. 19). Insbesondere eine formative,

d.h. prozessbegleitende Evaluation, die Wert auf Qualitätssicherung und -entwicklung legt, erfordert einen intensiven Einsatz. Um den demokratiepolitischen Impulsen von solchen Projekten auch in der wissenschaftlichen Begleitung eine komplementäre Strategie anbieten zu können, lässt sich eine summative Evaluation durch Methoden der Selbstevaluation ergänzen, die die beteiligten Akteure handlungsorientiert bei der Selbstreflexion unterstützt. Strategien der Selbstevaluation erscheinen darüber hinaus auch in soweit methodisch angemessen, dass sie eine langfristige Selbstbeobachtung der mit dem Projekt eingesetzten und weiterentwickelten Instrumentarien erlauben. Sowohl aus prinzipiellen Überlegungen kooperativer Evaluation als auch aus pragmatischen Gründen bietet sich eine Kombination aus Fremd- und Selbstevaluation an. Dabei werden die grundsätzlichen Fragestellungen der Erhebungsinstrumente zwischen den Partnern abgesprochen. Die Methodenentwicklung erfolgte adaptiv in Bezug auf die Bedingungen der Kurse, die Entscheidung für ihren Einsatz regulativ hinsichtlich des Verlaufs der Trainingskurse und die Beurteilung der Prozesse reflexiv im Rahmen der kooperativen Evaluation. Mit diesem Vorgehen wurde das Ziel verfolgt, das Evaluierungsinstrumentarium als Hilfsmittel projektbezogener Planungs-, Monitorings-, Evaluierungs- und Selbstlernprozesse in Kooperation mit den Projektverantwortlichen weiterzuentwickeln und ihnen zur weiteren eigenverantwortlichen Nutzung zur Verfügung zu stellen. Damit ist neben den summativen Ergebnissen des Evaluationsprozesses gleichzeitig auf der formativen Ebene die Möglichkeit gegeben, regulierend in die Planung der einzelnen Seminare einzugreifen. Diesem Zweck dient auch die kommunikative Validierung. Ziel ist nicht nur, die Projektverantwortlichen und -beteiligten über Prozesse und Resultate der Evaluation zu informieren, sondern diese darüber hinaus auch mit ihnen kritisch zu diskutieren. Einerseits wird diesem Personenkreis die Möglichkeit geboten, anhand der vorliegenden Ergebnisse ihre Arbeit zu überprüfen und ggf. zu modifizieren. Andererseits wird die Validität der Daten und ihre Interpretation durch diesen Personenkreis einer kritischen Betrachtung unterzogen, so kann eine ergänzende Sichtweise auf die Untersuchungsbefunde gewonnen werden.

### *Kombination qualitativer und quantitativer Methoden*

Die Zeiten, in denen der Einsatz von qualitativen und quantitativen Methoden als paradigmatischer Streit eines Entweder-Oder geführt wurde, gehören glücklicherweise der Vergangenheit an. Bezug nehmend auf Hammersley unterscheidet Flick drei Formen der Kombination qualitativer und quantitativer Methoden: „die Triangulation beider Ansätze, wobei er den Schwerpunkt auf die wechselseitige Überprüfung der Ergebnisse legt und weniger die wechselseitige Ergänzung der Erkenntnismöglichkeiten sieht. Facilitation betont die unterstützende Funktion des jeweils anderen Ansatzes – so liefert der eine Ansatz Hypothesen und Denkanstöße für die Weiterführung der Analysen mit dem anderen Ansatz. Schließlich können beide Ansätze als komplementäre Forschungsstrategien kombiniert werden“ (Flick 2006, S. 16). In unserer Evaluation lassen sich Elemente aller drei Kombinationsformen erkennen.

In der Evaluationsforschung wird häufig das Stufenmodell nach Kirkpatrick (1998) verwendet. Damit lässt sich gut illustrieren wie sich die von uns eingesetzten Untersuchungsschritte und -methoden verorten. Auf der ersten Stufe steht die Meinung der unmittelbaren Zielgruppen des Angebots im Vordergrund; mit der Reflexionsbogenuntersuchung haben wir genau diesen Aspekt abgedeckt. Wir wollten hier u.a. wissen wie die Teilnehmenden das Angebot beurteilen und wie sie damit zufrieden sind. Die zweite Stufe rückt das Können in den Mittelpunkt. Über das Wissen und Können der Schülerinnen und Schüler haben wir uns zum einen über die quantitativen Erhebungsinstrumente und zum anderen über die teilnehmende Beobachtung Klarheit verschafft. Die dritte Stufe behandelt die Anwendung respektive den Transfererfolg. Erkenntnisse zu dieser Frage strebten wir primär mit den qualitativen Befragungen und Gruppendiskussionen sowohl der Schülerinnen und Schüler bzw. Auszubildenden als auch der Personen aus ihrem sozialen Umfeld an. In der letzten und vierten Stufe geht es schließlich um die schon thematisierten Wirkungen.

Die qualitativen und quantitativen Evaluationsmethoden stehen in einem engen inhaltlichen Bezug zueinander. Durch die Verschränkung von quantitativen und qualitativen Forschungsmethoden entsteht ein Informationsraum, in dem multidimensionale Zugänge zu Sachverhalten, auf die sich das Evaluationsinteresse bezieht, geschaffen werden. Auf diese Weise wird es möglich, die einzelnen Befunde in Beziehung zueinander zu setzen, Aussagen über die Möglichkeiten und Grenzen der Verwendung der drei eingesetzten Programme für spezielle Zielgruppen zu machen und Aussagen über die Wirkung des Projektes zu formulieren.

### *Anforderungen an das Evaluationskonzept*

Eine große pädagogische Herausforderung an die Didaktik und Methodik von Bildungsangeboten besteht darin, produktiv mit jener Spannung umzugehen, die sich ergibt, wenn die Beschäftigung mit den geplanten Seminarinhalten in Konkurrenz tritt zu aktuell auftretenden Konflikten in der Seminargruppe. Die Leistung des Seminarteams ist es, unter der Leitlinie Störungen haben Vorrang, eine Verbindung zwischen beiden Aspekten herzustellen. Auf welche Weise dies geschieht, hängt von den situativen Gegebenheiten sowie den Bedürfnissen und Fähigkeiten der Beteiligten ab. In der Praxis der Bildungsarbeit haben wir es hier mit impliziten Passungsdialogen zu tun. Wenn es gelingt, die Bearbeitung der auftretenden Konflikte in das Seminarkonzept einzubinden, so trägt dies – wie unsere empirischen Befunde zeigen – zu einem besseren Verständnis und auch einer besseren Übernahme der Seminarziele bei. Hierauf hat auch das Evaluationskonzept Rücksicht zu nehmen, indem es diese Gegebenheiten berücksichtigt und flexibel darauf reagiert. Wir haben uns deshalb für eine Prozessevaluation entschieden, die zwar einen relativ fest fixierten Rahmen vorab setzt, die aber durch die Auswahl der Forschungsmethoden offen für Veränderungen bleibt.

## Wirkungsanalyse

Dem vielfältigen und breiten Erfahrungswissen praktischer Bildungsarbeit im Bereich der Erziehung zu Toleranz, Menschenrechten und Demokratie steht kein entsprechend breit fundiertes Wirkungsverständnis gegenüber. Vermehrte Anstrengungen zur Evaluierung und wissenschaftlichen Begleitung dieser Bildungsarbeit sind somit dringend geboten. Anhand des Projektes SAFT sollen im Folgenden erste Antworten auf die Frage am Beispiel unseres Evaluationsansatzes gegeben werden.

In der wissenschaftlichen Diskussion ist unbestritten, dass es weder einen Königsweg der Evaluation gibt, der sich umstandslos von einem Programm auf ein anderes übertragen lässt, noch dass sich Kausalwirkungen weicher Programme wie der politischen Bildung messen lassen (vgl. Lynen von Berg/Hirsland 2004, S. 19ff.). Die Befürchtungen und Kritiken sind vielmehr gegen eine politische Instrumentalisierung von Evaluationsprozessen und deren Ergebnissen gerichtet. So lässt sich auch die Schelte an dem „Marktdienstleistungskompetenzevaluationsstandardisierungskartell“ in der politischen Bildung von Klaus-Peter Hufer (2004) interpretieren. Ahlheim verweist unter Rückgriff auf Hans Tietgens zu Recht darauf, dass der Fokus des Forschungsinteresses nicht im Messbaren aufgeht. Seiner Meinung nach kann es „keiner Wirkungsforschung im Feld politischer Erwachsenenbildung gelingen, eindeutig qualifizierbare, repräsentativ-objektive Erfolgsdaten zu liefern, die sich auch noch rechnen und nachrechnen lassen“ (Ahlheim 2005, S. 18). Er wendet sich deshalb gegen jeden „naiven Glauben an die Macht des Messbaren“ (Ahlheim 2003, S. 5).

Die Studie „Evaluation der politischen Erwachsenenbildung“ der Dresdener Forschungsgruppe hat in ihrem Bericht festgestellt, dass sie keinen Wirkungsnachweis politischer Bildungsmaßnahmen in Form einer Input-Output-Messung erbringen konnte. Wenn es auch nicht möglich ist, eine „Wirkungsgewissheit“ (Schiele 2004, S. 3) politischer Bildung zu reklamieren, so ist sie in ihren Aktivitäten doch permanent darauf ausgerichtet, „dass ihre Leistungen allmähliche und längerfristige Einstellungsänderungen zum Ziel haben“, die sich zukünftig „auch in politischem Engagement und Handeln manifestieren“ (Fritz/Maier 2005, S. 11).

Ein technisches Wirkungsverständnis im Sinne einer klaren Ursache-Wirkungs-Relation ließe die Eigengesetzlichkeiten politischer Prozesse außer Acht. In dem vom Deutschen Institut für Internationale Pädagogische Forschung im Auftrag der Kultusministerkonferenz herausgegebenen „Bildungsbericht für Deutschland“ heißt es folglich: „Was immer als Wirkung von institutionalisierter Bildung angesehen wird, entsteht aus dem Zusammenspiel institutioneller und unterrichtlicher Lernbedingungen mit soziokulturellen Faktoren, familiären Einflüssen sowie den individuellen Voraussetzungen und individuellen Verarbeitungsprozessen der einzelnen Schülerinnen und Schüler. Dieser Sachverhalt ist bei einer Bilanzierung von Effekten und Bildungserträgen stets zu berücksichtigen“ (Avenarius et al. 2003, S. 167). Diese Aussage gilt uneingeschränkt auch für unsere Untersuchungsergebnisse.

Ein Aspekt verdient noch eine besondere Aufmerksamkeit. Wie unsere Ergebnisse zeigen, kam es zum Teil im Verlauf der quantitativen Paneluntersuchung zu einer kritischeren Einschätzung des Schul- bzw. Klassenklimas durch die befragten Schüler/innen. Dies könnte mit einer realen Verschlechterung der das Klima bestimmenden Faktoren zusammenhängen; die qualitativen Interviews deuten allerdings darauf hin, dass sich vielmehr die Wahrnehmung der sozialen Wirklichkeit als Folge der Bildungsarbeit verändert hat, was zu einer Sensibilisierung sowohl der Sichtweisen als auch der Handlungsoptionen und damit zu einer differenzierteren und kritischeren Einschätzung des Schul- bzw. Klassenklimas geführt hat. Insofern können vermeintlich schlechtere Umfragewerte ein Beleg für erfolgreiche Bildungsarbeit sein. Dies macht deutlich wie wichtig es ist, qualitative und quantitative Daten wechselseitig aufeinander zu beziehen und jeweils den Kontext zu betrachten.

Konzepte der Wirkungsforschung dürfen nicht nur danach fragen, ob und inwieweit die vorgegebenen Ziele der Bildungsarbeit erreicht wurden. Unsere Ergebnisse deuten darauf hin, dass die Jugendlichen sehr viel mehr aus den Seminaren für sich mitnehmen konnten, als in den quantitativen Messinstrumenten erfragt wurde. Möglicherweise handelt es sich dabei um Aspekte, die über die beabsichtigte Beförderung von Konflikt-, Toleranz- oder Demokratiekompetenz hinausgehen bzw. sich nicht unter diesen Begriffen fassen lassen. Damit wird die Frage der nicht intendierten Effekte berührt, die bei künftigen Evaluationen noch stärker in den Blick genommen werden sollte.

### *Auswertungsstrategien*

Eine im Kontext der Auswertung qualitativer Daten diskutierte Frage ist die, ob sich die Bewertungen der Forscherinnen und Forscher auf die der Akteure im Feld stützen oder aber unabhängig davon vorgenommen werden sollen (vgl. Lüders 2006). Mit der Entscheidung für eine partizipative respektive kooperative Evaluationsstrategie haben wir die Antwort darauf zugunsten der ersten Alternative – zumindest was die Projektleitung betrifft – gegeben. In einem sehr zeitintensiven Prozess hat das Evaluationsteam mit der Leitung des Projektes sowohl die qualitativen wie quantitativen Daten diskursiv gesichtet, erörtert und interpretiert. Dies war ein äußerst produktiver Prozess in dem vor allem auch die Kontexte sowohl des Seminargeschehens als auch der Datenerhebungen ausgeleuchtet und reflektiert werden konnten. Wer Demokratiekompetenz vermitteln bzw. erwerben will, muss Demokratie leben und lernen. Dies geschieht in einem permanenten Prozess, der Zeit, Erfahrungsräume und Wiederholungen benötigt. Erst das Wissen und das Können zusammen machen die habituelle Kompetenz möglich, d.h. eine demokratische Gesinnung, die in inneren Einstellungen und Werthaltungen wurzelt (vgl. Scherb 2005, S. 272). Auf dieser Basis lassen sich dann neue Ideen für die Fortschreibung unserer Bildungsarbeit, der Evaluationspraxis sowie des Gemeinwesens entwickeln.

## Literaturverzeichnis

- Ahlheim, Klaus (2005): Wirkungsoptimismus und Messbarkeitskepsis – Evaluation und Wirkungsforschung in der politischen Erwachsenenbildung. In: Praxis Politische Bildung, H. 1, S. 17–24
- Ahlheim, Klaus (2003): Vermessene Bildung? Wirkungsforschung in der politischen Erwachsenenbildung. Schwalbach/Ts.: Wochenschau
- Avenarius, Hermann/Ditton, Hartmut/Döbert, Hans/Klemm, Klaus/Klieme, Eckhard/Rürup, Matthias/Tenorth, Heinz-Elmar/Weishaupt, Horst/Weiß, Manfred (2003): Bildungsbericht für Deutschland. Erste Befunde. Opladen: Leske + Budrich
- Bertelsmann Forschungsgruppe Politik (Hrsg.) (2001): Toleranz – Grundlage für ein demokratisches Miteinander. Gütersloh: Bertelsmann Stiftung
- Bommes, Michael/Guter, Jochen/Wolff-Jontofsohn, Ulrike (2004): „Miteinander – Erfahrungen mit Betzavta“. In: Uhl, Katrin/Ulrich, Susanne/Wenzel, Florian (Hrsg.): Evaluation politischer Bildung. Ist Wirkung messbar? Gütersloh: Bertelsmann Stiftung, S. 97–120
- Caspari, Alexandra (2006): Partizipative Evaluationsforschung – zur Entmystifizierung eines Begriffs in der Entwicklungszusammenarbeit. In: Flick, Uwe (Hrsg.): Qualitative Evaluationsforschung. Reinbek: Rowohlt, S. 365–384
- Flick, Uwe (2006): Qualitative Evaluationsforschung zwischen Methodik und Pragmatik. In: Flick, Uwe (Hrsg.) (2006): Qualitative Evaluationsforschung. Reinbek: Rowohlt, S. 9–29
- Fritz, Karsten/Maier, Katharina (2005): Zur Evaluation der politischen Erwachsenenbildung. In: Praxis Politische Bildung, H. 1, S. 5–16
- Himmelman, Gerhard (2001): Demokratie Lernen als Lebens-, Gesellschafts- und Herrschaftsform. Ein Lehr- und Sachbuch. Schwalbach/Ts.: Wochenschau
- Hirsland, Andreas/Dimbath, Oliver/von Hayek, Julia/Schneider, Werner (2004): Evaluation des Programms „Achtung (+) Toleranz“ – ein Praxisbericht. In: Uhl, Katrin/Ulrich, Susanne/Wenzel, Florian (Hrsg.): Evaluation politischer Bildung. Ist Wirkung messbar? Gütersloh: Bertelsmann Stiftung, S. 57–82
- Hufer, Klaus-Peter (2004): Das Marktdienstleistungskompetenzevaluationsstandardisierungskartell in der politischen Bildung. In: Außerschulische Bildung. H. 1, S. 20–23
- Kardorff, Ernst von (2006): Zur gesellschaftlichen Bedeutung und Entwicklung (qualitativer) Evaluationsforschung. In: Flick, Uwe (Hrsg.): Qualitative Evaluationsforschung. Reinbek: Rowohlt, S. 63–91
- Kehr, Jochen (2004): Praxisbericht: Evaluation des Programms „Eine Welt der Vielfalt“. In: Uhl, Katrin/Ulrich, Susanne/Wenzel, Florian (Hrsg.): Evaluation politischer Bildung. Ist Wirkung messbar? Gütersloh: Bertelsmann Stiftung, S. 83–95
- Kirkpatrick, Donald L. (1998): Evaluating Training Programs. The Four Levels. 2. ed. San Francisco: Berret-Koehler
- Kuper, Harm (2005): Evaluation in der Erziehungswissenschaft. Stuttgart: Kohlhammer
- Lamnek, Siegfried (1998). Gruppendiskussion. Theorie und Praxis. Weinheim: UTB

- Lüders, Christian (2006): Qualitative Evaluationsforschung – was heißt hier Forschung? In: Flick, Uwe (Hrsg.): Qualitative Evaluationsforschung. Reinbek: Rowohlt, S. 33–62
- Lynen von Berg, Heinz/Hirsland, Andreas (2004): Zivilgesellschaft und politische Bildung – Zur Evaluation von Programmen und Projekten. In: Uhl, Katrin/Ulrich, Susanne/Wenzel, Florian (Hrsg.): Evaluation politischer Bildung. Ist Wirkung messbar? Gütersloh: Bertelsmann Stiftung, S. 15–26
- Oser, Fritz/Biedermann, Horst (2006): Partizipation – ein Begriff, der ein Meister der Verwirrung ist. In: Quesel, Carsten/Oser, Fritz (Hrsg.): Die Mühen der Freiheit. Probleme und Chancen der Partizipation von Kindern und Jugendlichen. Zürich: Rüegger, S. 17–37
- Schäfer, Erich/Schack, Stephan/Rahn, Peter/Uhl, Sandra (2006): „Wer sich selbst versteht, versteht auch andere besser“. Eine Längsschnittstudie zu Wirkungen eines Projektes der politischen Jugendbildung zum Demokratie-Lernen. Jena: IKS Garamond
- Scherb, Armin (2005): Demokratie-Lernen und reflexive Urteilskompetenz. In: Himmelmann, Gerhard/Lange, Dirk (Hrsg.): Demokratiekompetenz. Wiesbaden: VS, S. 270–285
- Schiele, Siegfried (2004): Ein halbes Jahrhundert staatliche politische Bildung in Deutschland. In: Aus Politik und Zeitgeschichte, B. 7–8, S. 3–6
- Schröder, Achim/Balzter, Nadine/Schroedter, Thomas (2004): Politische Jugendbildung auf dem Prüfstand. Ergebnisse einer bundesweiten Evaluation. Weinheim und München: Juventa
- Uhl, Katrin (2004): Evaluation von Demokratie- und Toleranzerziehungsprogrammen – die Erfahrungen aus dem Projekt „Erziehung zu Demokratie und Toleranz“. In: Uhl, Katrin/Ulrich, Susanne/Wenzel, Florian (Hrsg.): Evaluation politischer Bildung. Ist Wirkung messbar? Gütersloh: Bertelsmann Stiftung, S. 51–55
- Ulrich, Susanne/Henschel, Thomas R./Oswald, Eva (1997): Miteinander – Erfahrungen mit Betzavta. Ein Praxishandbuch auf der Grundlage des Werks „Miteinander“ von Uki Maroshek-Klarman, Adam Institut, Jerusalem. Gütersloh: Bertelsmann Stiftung
- Ulrich, Susanne/Heckel, Jürgen/Oswald, Eva/Rappenglück, Stefan/Wenzel, Florian (2001): Achtung (+) Toleranz. Wege demokratischer Konfliktregelung. Gütersloh: Bertelsmann Stiftung
- Ulrich, Susanne/Wenzel, Florian M. (2004): Partizipative Evaluation. In: Uhl, Katrin/Ulrich, Susanne/Wenzel, Florian (Hrsg.): Evaluation politischer Bildung. Ist Wirkung messbar? Gütersloh: Bertelsmann Stiftung, S. 27–48
- Ulrich, Susanne/Wenzel, Florian M. (2003): Partizipative Evaluation. Ein Konzept für die politische Bildung. Gütersloh: Bertelsmann Stiftung

## „Leben und Arbeiten in Vielfalt?“ – Zur dokumentarischen Evaluation eines Xenos-Projekts

Das im Rahmen des Aktionsprogramms „Jugend für Toleranz und Demokratie – gegen Rechtsextremismus, Fremdenfeindlichkeit und Antisemitismus“ 2001 von der Bundesregierung lancierte Teilprogramm „XENOS – Leben und Arbeiten in Vielfalt“ ermöglichte es zahlreichen pädagogischen Institutionen, Maßnahmen im Bereich der Rechtsextremismusprävention und der Demokratie- und Toleranzförderung durchzuführen. Das Programm wurde als so erfolgreich eingestuft, dass nach seinem Auslaufen 2006 mittlerweile ein Nachfolgeprogramm gestartet wurde. XENOS „zielt darauf ab, Fremdenfeindlichkeit, Rassismus und Diskriminierung in der Gesellschaft nachhaltig entgegenzuwirken“<sup>1</sup>, wobei Aktivitäten gegen Rassismus mit arbeitsmarktpolitischen Maßnahmen verbunden werden sollen.

Im vorliegenden Beitrag möchten wir die dokumentarische Evaluation eines – von 2003 bis 2006 an einem ostdeutschen Ausbildungszentrum für „lernbehinderte und benachteiligte Jugendliche“<sup>2</sup> durchgeführten XENOS-Projekts vorstellen. Dieses bestand in der Fortbildung der Auszubildenden zur Thematik „Akzeptanz und Toleranz“ im Rahmen von Projektwochen, gemeinsam mit Asylbewerberinnen durchgeführten Ausbildungsmodulen und weiteren pädagogischen Maßnahmen.

### Rahmen und Methode der Evaluation

Vorrangiges Ziel des XENOS-Projekts war es laut Konzept, bei den Jugendlichen die Entwicklung einer weltoffenen und antirassistischen Grundhaltung zu fördern. Durch berufliche Qualifikation und den Aufbau sozialer Beziehungen sollte darüber hinaus eine bessere Integration der Zielgruppe in den Arbeitsmarkt und die Gesellschaft erreicht werden. Grundgedanke hierbei war, dass eine realistische Berufswahl, das Akzeptieren des eigenen Qualifikationsprofils und die Übernahme von Verantwortung für die eigene Entwicklung eine Mündigkeit fördern, die Fremdenfeindlichkeit keinen Raum lässt. In diesem Sinne war ein Schwerpunkt des Projekts, einer möglichen Anfälligkeit der Jugendlichen für rechtes Gedankengut präventiv gegenzusteuern. Als weitere Ansatzpunkte zur Förderung von Akzeptanz und Integration werden in der Konzeption des Projekts allgemeine und berufliche Bildung genannt sowie Begegnungen auf der Ebene des Miteinandertuns und die Integration von Asylbewerberinnen und -bewerbern in die Ausbildung.

---

1 Vergleiche [www.xenos-de.de](http://www.xenos-de.de)

2 Zitat aus der Projektkonzeption

In Absprache mit den Auftraggeberinnen und Auftraggebern wurde die Maßnahme in einer Längsschnittstudie vom „Centrum für qualitative Evaluations- und Sozialforschung e.V.“<sup>3</sup> kontinuierlich wissenschaftlich begleitet. In regelmäßigen Abständen wurden insgesamt sieben Gruppendiskussionen mit verschiedenen Gruppen von Auszubildenden sowie zwei Gruppendiskussionen mit den im Ausbildungszentrum tätigen Pädagoginnen und Pädagogen durchgeführt. Im Zentrum des Forschungsinteresses standen die Lern- und Erfahrungsprozesse der Zielgruppe des Projekts. Hierbei ging es zum einen um eine Analyse der Erwartungen gegenüber dem Projekt und zum anderen um eine Rekonstruktion der konkreten Erfahrungen und Erlebnisse während dessen Durchführung. Auf der Grundlage einer Rekonstruktion des handlungsleitenden Erfahrungswissens und der gemeinsam geteilten Orientierungsrahmen im Bereich des alltagspraktischen Handelns der Jugendlichen sollten insofern mögliche durch das XENOS-Projekt ausgelöste Irritationen und Veränderungen ermittelt werden.

Die dokumentarische Methode (vgl. u.a. Bohnsack 2007) eröffnet den methodisch kontrollierten Zugang zu jenen Bereichen des Wissens, die handlungsorientierenden Charakter haben. Karl Mannheim (1964) hat dieses in die Alltagspraxis eingelassene vorreflexive oder implizite Wissen auch als atheoretisches Wissen bezeichnet – im Unterschied eben zum theoretischen Wissen, also zu den Alltags-Theorien oder Common Sense-Theorien der Erforschten. In der theoretischen Diskussion zur Evaluationsforschung in den USA hat auch Schwandt (1997) darauf aufmerksam gemacht, dass erst die genaue Kenntnis dieses handlungspraktischen Wissens einen Zugang zur Logik des Handelns jenseits der Theorien der Erforschten ermöglicht und somit auch die Bedingungen der Möglichkeit für eine Beurteilung und Bewertung dieser Praxis und für eine mögliche Einflussnahme auf diese schafft. Schwandt (2002, S. 47) spricht in diesem Sinne auch von der Notwendigkeit einer „praktischen Hermeneutik“ („practical hermeneutics“). In Auseinandersetzung mit der neueren Diskussion in den Vereinigten Staaten und auch mit den Arbeiten von Robert E. Stake (u.a. 2000), einem der prominentesten Begründer qualitativer Evaluationsforschung, und mit dessen Modell der „responsiven Evaluation“<sup>4</sup> sowie in der Rekonstruktion der eigenen Forschungspraxis haben wir die dokumentarische Evaluationsforschung entwickelt (vgl. Bohnsack 2006; Nentwig-Gesemann 2006).

Indem es im Sinne der dokumentarischen Evaluationsforschung darum gehen muss, jenes die Handlungspraxis orientierende Erfahrungswissen der Evaluationsbeteiligten zur begrifflich-theoretischen Explikation zu bringen, erweist sich Evaluation als ein genuin wissenschaftliches Projekt. Die zentrale wissenschaftlich-hermeneutische Aufgabe besteht in der empirisch fundierten Transzendenz der Oberflächenebene der Common Sense-Theorien in Richtung auf die tiefer liegende Semantik der praktisch-existentiellen Ebene. Vor allem aus diesem Grunde verstehen wir Evaluation als einen Typ sozialwissenschaftlicher

---

3 Für nähere Informationen zum Centrum siehe [www.ces-forschung.de](http://www.ces-forschung.de)

4 Auch Robert E. Stake (1978, S. 6) hat die Bedeutung des impliziten Wissens („tacit knowledge“) für die Orientierungen und Werthaltungen der sog. Stakeholder, der Evaluationsbeteiligten, herausgestellt.

Forschung und – in Übereinstimmung mit Lüders/Haubrich (2003, S. 319) – als einen „eigenständigen Forschungstyp“<sup>5</sup>.

Im vorliegenden Beitrag möchten wir insbesondere verdeutlichen, inwiefern eine Rekonstruktion des Erfahrungswissens der Stakeholder zu Bewertungen einer Maßnahme und konkreten Empfehlungen führen kann. Dies sei im Folgenden anhand einiger ausgewählter Ergebnisse der Studie erläutert.

### **Ausgewählte Ergebnisse der Studie**

Auf der Grundlage der Überlegung, dass ein vertieftes Verständnis der Erfahrungszusammenhänge und Orientierungsmuster der Auszubildenden und ihrer Handlungspraxis Voraussetzung ist, um deren Reaktionen und Perspektiven auf die verschiedenen Maßnahmen im Rahmen des Xenos-Projekts nachvollziehen zu können, möchten wir im Folgenden zunächst auf die Wahrnehmung und Deutung der eigenen alltäglichen Erfahrungen seitens der Klientel des von uns „Bildungszentrum Möllenwerder (BZM)“ genannten Ausbildungszentrums eingehen. Im nächsten Schritt werden wir sowohl positive als auch negative Erfahrungen der Jugendlichen mit dem XENOS-Projekt diskutieren um schließlich einige zentrale Ergebnisse der Evaluation zusammenzufassen.

#### *Akzeptanz und Toleranz im „Bildungszentrum Möllenwerder (BZM)“*

Das BZM als Zentrum für „lernbehinderte und benachteiligte Jugendliche“ ist kein Ausbildungsort, der in der Regel freiwillig und gerne gewählt wird, er stellt vielmehr oftmals die einzige Alternative zur Arbeitslosigkeit dar und die Jugendlichen landen hier aufgrund einer Zuweisung durch das Arbeitsamt. Wie sich in allen im Rahmen der Evaluation durchgeführten Gruppendiskussionen zeigte, ist gleichzeitig der Status als BZM-Auszubildende/r mit einschneidenden Stigmatisierungserfahrungen verbunden, die das Selbstbild der Einzelnen sowie die Beziehungen untereinander und nach außen stark prägen<sup>6</sup>.

---

5 Allerdings gilt es, das Verständnis von „Eigenständigkeit“ genauer zu klären. Wir gehen davon aus, dass die bekannten Verfahren der qualitativen Sozialforschung mit ihren Standards (vgl. Bohnsack 2005) umfassend auf den Bereich der Evaluationsforschung übertragen werden können und müssen. Um den mit einer Evaluation verbundenen Ansprüchen gerecht werden zu können, bedarf es jedoch einer Erweiterung dieser Verfahren vor allem um eine Methodik der Moderation (vgl. Bohnsack 2006).

6 Alle verwendeten Orts- und Personenamen sind maskiert. Zum Transkriptionssystem: ( ): unverständliche Äußerung, je nach Länge; @(.): kurzes Auflachen; @nein@: lachend gesprochen; [: überlappend gesprochene Redebeiträge; (Panik): unsichere Transkription; (.): kurze Pause; (3): Anzahl der Sekunden, die eine Pause dauert; unterstrichen: betont; fett: laut; merkt=s: ineinander übergehend gesprochene Wörter; (...): Auslassung

- Denise:** ick kannte dat BZM ja nich, und hab irgendwann dem ( ) erzählt und sag BZM und dann hat er anjefanzen zu lachen
- ?f:** @(.)@
- Janine:** [@(.)@ is schon peinlich wenn man det sagt, irgendjemand ist in Möllenerwerder=(in), also ick finde dat schon peinlich
- Denise:** danach hatte ick (Panik) klar zu kommen, nachdem mich mehrere ausgelacht haben und jesagt haben, mein Gott da verblödest du
- (...)
- Janine:** also jedenfalls unten in der Küche (.) sage ick ma (.) wie die anderen Bereiche so sind, weef ick nicht (.) also ooch die Leute unten sind noch ziemlich normal, kann man sagen
- ?:** noch
- Janine:** noch, ja (.) also ick et jibt hier wirklich scho:n (.) na ja wie sage ick det jetze? (1) will jetz niemand beleidigen oder so
- Denise:** ziemlich viele Chaoten
- Janine:** ja (.) Chaoten
- Denise:** daran merkt man, dass das ein Ausbildungszentrum für Benachteiligte ist
- Janine:** für Behinderte und Benachteiligte @(.)@

Wie sich in dieser Passage zeigt, wird den Jugendlichen von ihrem Umfeld signalisiert, dass es lächerlich und peinlich sei, gerade im BZM ausgebildet zu werden. Der Hinweis darauf, dass man dort „verblöde“, lässt auf die Annahme eines niedrigen Niveaus schließen, die Auszubildenden selbst sind somit mit dem Stigma behaftet, zu „blöd“ für andere Ausbildungsstätten zu sein. Janine und Denise, die beide in der Küche arbeiten, grenzen ihre positiven Erfahrungen als BZM-Auszubildende auf ihren eigenen Ausbildungsbereich ein, in welchem die Leute „noch ziemlich normal“ seien. Der Status der Normalität, der in allen im Rahmen der Evaluation mit Jugendlichen durchgeführten Gruppendiskussionen eine große Rolle spielt, ist offensichtlich nicht selbstverständlich für die Auszubildenden. In der Beschreibung der beiden Mädchen ist Normalität ein Privileg ausgewählter Bereiche und auch hier ein prekärer Status, der wieder verloren gehen kann. Sie selbst fühlen sich bedroht von dem Stigma, das damit verbunden ist, in einem Ausbildungszentrum „für Behinderte und Benachteiligte“ zu lernen und weisen dieses deshalb Auszubildenden aus anderen Bereichen der Einrichtung zu (den „Chaoten“). In ihrer Bemerkung, sie wolle niemanden beleidigen, unterstreicht Janine einerseits, dass sie sich selbst unzweifelhaft auf der Seite der Normalität verortet und deutet andererseits an, dass sie nicht allen Teilnehmenden der Gruppendiskussion diesen Status zugestehen würde.

Wo genau die Grenze zwischen den „noch Normalen“ im BZM und den nicht mehr Normalen verläuft, ist eine umstrittene Frage. Dadurch wird der eigene Status notwendigerweise für alle Auszubildenden prekär. Ein weiteres Ergebnis der Evaluation ist, dass viele Jugendliche mit der Bedrohung, selbst als nicht normal gelten zu können, in der Weise umgehen, dass sie andere Auszubildende entnormalisieren und degradieren.

**I:** na und was könnte für euch jetzt besser laufen am BZM, also wenn ihr (.) euch jetzt vorstellt (.) ihr wärt jetzt Direktor hier von dem Laden, was würdet ihr alles verändern?

**Lars:** na viel mehr Sachen freizeitmäßig

**Tino:** oh, richtig!

**Uwe?:** erst mal (.)

**Tino:** dass wern Fernseher aufm Zimmer haben dürfen

**Uwe?:** erst mal die janzen (.) Keimlinge aus der Küche, die ganzen fetten Schweine raus

**me:** @(1)@

**Uwe?:** raus, raus (.) also (.) wenn ich da esse ( )

**?:** also wenn ick da esse

**Uwe?:** weil wenn ick da manschmal essen jehe und dann so=n dicket Ding an mir vorbei läuft, dann (.) kann ich

**Lars?:** [( ) zum Beispiel?

**Tino:** uäh

**?:** den Teller gleich wieder vollbrechen

**Tino:** @(. )@

**Christoph?:** ( )

**Uwe?:** das so wat inne Küche arbeitet, und dann keene (.) rejlmäßige Körperpflege

**Lars:** det gloobe ick

**Uwe?** ohh::: wenn ick daran schon denke

**Dennis:** jetzt weiß ick ooch woher ick immer die (Kriegen) kriege

**Uwe?:** wenn die sich anmelden würden bei mir, dann würde ick gleich sagen, erstma gucken von oben bis unten so und dann würd ick sagen wat, nimm erstma ein paar Kilo ab und dann kannst du wiederkommen ((Räuspern))

**me::** @(7)@

**?:** @(der war jut)@

Neben mehr Freizeit und einem Fernseher auf dem Zimmer verbinden die Auszubildenden eine Verbesserung der Verhältnisse im BZM vor allem mit einem Rauswurf der „Keimlinge“ und „fetten Schweine“ aus der Küche. In dieser sehr dichten Passage demonstrieren die beteiligten Jugendlichen übereinstimmend Abneigung und körperlichen Ekel („Teller gleich wieder vollbrechen“) gegenüber den als „dick“ wahrgenommenen Auszubildenden im Hauswirtschaftsbereich, wobei Übergewicht mit mangelnder Hygiene gleichgesetzt wird. Das Verfehlen einer körperlichen Norm wird mit unzureichender Körperpflege assoziiert, wodurch gleichzeitig den betreffenden Personen die Verantwortung für die Nicht-Erfüllung dieser Norm zugewiesen wird<sup>7</sup>. Die Abneigung der Teilnehmenden gegenüber als zu dick empfundenen Anderen geht so weit, dass sie scherzhaft das Szenario entwerfen, als

<sup>7</sup> Da es sich hier um eine Gruppendiskussion mit ausschließlich männlichen Jugendlichen aus dem Technikbereich handelt, die über die weiblichen Auszubildenden aus dem Hauswirtschaftsbereich reden, ist deren Diskriminierung und Beleidigung darüber hinaus geschlechtlich konnotiert.

Direktor des BZM jugendliche Anwärter/innen von vornherein wegen ihres Übergewichtes als Auszubildende abzulehnen.

Wie in zahlreichen anderen Passagen auch zeigt sich hier die offensichtlich gängige Praxis unter den Auszubildenden, das eigene Selbstbewusstsein auf dem Wege der Abwertung Anderer zu retten, wobei als andere die Mitschülerinnen und Mitschüler im BZM gewählt werden. Auch in anderen Gruppendiskussionen zeigt sich, dass Degradierungen der Auszubildenden untereinander insbesondere über eine Orientierung an einer Norm körperlicher Attraktivität verlaufen, die offensichtlich eine große Bedeutung im Umgang miteinander hat. Hierbei wird mangelnde körperliche Attraktivität als Wesensmerkmal bestimmter Personen festgeschrieben und mit mangelndem Leistungswillen, Disziplinlosigkeit, unzureichender Hygiene und sozialem Abstieg in Verbindung gebracht.

Interessant in diesem Kontext ist, dass hingegen eine öffentlich zur Schau gestellte rassistische oder rechtsextreme Einstellung von vielen der befragten Jugendlichen kritisiert wird: In ihrem Bemühen, als „normal“ anerkannt zu werden, suchen die Auszubildenden Positionen zu vermeiden, die sie für extrem oder ungewöhnlich halten. Rassismus und Rechtstextremismus werden dabei nicht als moralisches Problem wahrgenommen, sondern als problematisch für das Ansehen des BZM als Institution und somit auch für das eigene Ansehen<sup>8</sup>. Gleichzeitig dokumentieren sich in Bemerkungen über „die Neger“ oder „die Türken“ in allen durchgeführten Gruppendiskussionen rassistische Positionen.

### *Erfahrungen in der antirassistischen Arbeit mit den Jugendlichen*

In den Rückmeldungen der Jugendlichen zum XENOS-Projekt fällt zunächst auf, dass dieses einhellig als positiv bewertet wird. Die Projektwochen gelten als willkommene Alternative zu einem als arbeitsreich und eintönig empfundenen Alltag und werden als „Erholung“ und „Abwechslung“ gelobt. Worin ein weitergehender Reiz bestehen kann, soll im Folgenden am Beispiel der Erzählungen eines Jugendlichen über seine Teilnahme am im Rahmen der XENOS-Projektwoche angebotenen Trommelkurs erörtert werden.

**Lars:** so letzten Tach denn, uffn Freitag, sollten wir unten denn so=ne: (.) Präsentation machen, mit so=ner Trommel

**Stefan?:** @(.)@

**Lars:** da hat denn eener von de:r Trommelspieler mitgemacht (...) hat er sich eu-(.) hat er eben mitgebracht, ein Freund, dann haben wir mit denen geübt drei Stunden oder vier Stunden lang, nur gebüffelt (.) und denn sind wa runterjeggangen (.) erst wollten wa eigentlich bloß hinstellen die Dinger und vorführn (.) mal (.) aber denn ham wa doch gespielt (1) dann ham sie so viel mal gerufen, noch mal, noch mal, dann mussten wa natürlich noch mal uff de Bühne ruff

<sup>8</sup> Diese Einstellung findet sich auch in den Gruppendiskussionen mit im Ausbildungszentrum tätigen Pädagoginnen und Pädagogen wieder.

?: na jeht doch, is doch okay  
**Lars:** @(.)@  
 ??: @(hm)@  
 ??: so muss et sein ( ) (.)  
**Lars:** dann ham wa draußen noch (.) als se abgehauen sind ham draußen sich noch paar hingestellt, und wir ham natürlich aus Spaß uns draußen hingestellt, Musik gemacht, dann sind se alle uff een Mal wieder gekommen.  
 ??: @(.)@  
**Lars:** wollten alle nich mehr nach Hause gehen, Autos wieder zugeschlossen, hinjekom zu uns (.) dann ham wir draußen Musik gemacht (.) sachma, draußen hat=s viel mehr Spaß gemacht  
 ??: das hat ma jehört  
**Lars:** ja is klar dass des bis oben hinjehört hast, des hat=s (.) des ham wir uns schon gedacht (.) unser Meester hat sich da draußen vor der Halle schon hingesetzt, mit d= Stuhl (.) und hat jehorcht  
 ??: @(.)@ tja  
**Lars:** dann ham wir die ganzen Dinger hochgebracht, und denn war ja schon Feierabend (.) n=paar sind nach Hause jejangen; er und der Musikspieler sind da geblieben, die ham den nen Sixpack usm Auto geholt, und denn sind wer eenen, uff den Feierabend (schönen) trinken (.) war cool

Am Ende der Projektwoche bereiteten die Teilnehmenden sich intensiv auf die für alle Kurse vorgesehene Präsentation der Ergebnisse vor, wobei sie zunächst dachten, es ginge vorrangig darum, die Trommeln vorzuzeigen. In der Situation ergab es sich schließlich wie von selbst, dass sie auf den Trommeln spielten und somit weniger diese, sondern vor allem sich selbst als Musiker präsentierten. Die Vorbereitung durch die Experten und die Bestätigung durch das Publikum brachte die Jugendlichen in die Situation, die eigenen musikalischen Fähigkeiten demonstrieren zu können und sich somit selbst als kompetent in einem völlig neuen Feld zu erleben.

Ihr Erfolg bei diesem Auftritt zeigte sich insbesondere nach Beendigung des offiziellen Programms: Da sie Spaß am Trommeln gefunden hatten, spielten die Jugendlichen jenseits der institutionell vorstrukturierten Projektwochenpräsentation noch weiter und das Publikum, das bereits auf dem Weg nach Hause gewesen war, kehrte um und sogar der Meister hat sich Zeit genommen und in Ruhe zugehört. Lars' Betonung, dass das Spielen „draußen“ viel mehr Spaß gemacht habe, verdeutlicht, dass die im pädagogischen Setting der offiziellen Präsentation erlebte Anerkennung durch dessen Arrangiertheit beeinträchtigt wird: Hier lauschte das Publikum, weil es von ihm erwartet wurde, während es beim Spielen „draußen“ aus echtem Interesse blieb. Nachdem Teile des Publikums gegangen waren, verbrachten die Jugendlichen noch gemeinsam mit ihrem Teamer und dessen Freund ihren Feierabend, indem sie zusammen Bier tranken. Die Hierarchie zwischen Kursleiterinnen/Kursleitern und Teilnehmenden wurde durch diese gemeinschaftliche Praxis aufgebrochen und Lars erlebte diese

gleichberechtigte Situation als „cool“. Die affirmativen Reaktionen der anderen Teilnehmenden auf seine Erzählung („so muss et sein“, „das hat ma jehört“, mehrmaliges Lachen), die im Kontrast zum ansonsten eher rauen Umgangston untereinander stehen, lassen darauf schließen, dass die Trommelvorführung als gemeinsames positives Erlebnis erinnert wird.

Diese Passage verweist unseres Erachtens in diesem Sinne auf eine mit den Xenos-Projektwochen verbundene zentrale Chance: Diese erlauben es den Auszubildenden teilweise, eine ungewohnte Erfahrung der Selbstwirksamkeit (Bandura 1993) zu machen<sup>9</sup>. Ein respektvolles und anerkennendes Verhalten der externen Teamer/innen und das Erstellen von prestigeträchtigen Produkten erlebten die Jugendlichen als Aufwertung ihrer Person, was insbesondere vor dem Hintergrund ihrer Stigmatisierungserfahrungen bedeutsam für sie ist. Das Erproben neuer Fertigkeiten im Rahmen eines alternativen Lern-Settings während der Projektwochen erlaubte es ihnen, ungeahnte Kompetenzen zu entwickeln, für die sie auch innerhalb des BZM und der eigenen Peergroup Anerkennung bekamen.

In der didaktischen Konzeption der Projektwoche sollte der Trommelkurs auch mit einer Begegnung „mit Anderen/Fremden“<sup>10</sup> einhergehen und wurde von einem afrikanischen Asylbewerber geleitet. Dieser wurde von den Jugendlichen jedoch offensichtlich gerade nicht als „Fremder“ oder „Ausländer“ sondern als „Trommelspieler“ wahrgenommen. In diesem Zusammenhang lässt sich davon ausgehen, dass Erfahrungen der Selbstwirksamkeit und der eigenen Aufwertung während der Projektwochen eine gute Voraussetzung dafür schaffen konnten, dass in deren Rahmen stattfindende Begegnungen auf einer respektvollen Ebene abliefen und Exotisierungen vermieden wurden.

In einer anderen Passage zeigt sich, dass arrangierte „Begegnungen mit Anderen/Fremden“ andererseits auch die Gefahr bergen, die Jugendlichen zu überfordern und bestehende Vorurteile eher zu bestätigen als zu hinterfragen. Nadine berichtet hier über den im Rahmen einer Projektwoche unternommenen Besuch in einem Obdachlosenheim:

- Nadine:** det (.) Asylding=da ham wir ooch noch ma (.) sollten wir ooch noch Mittag essen (.) also (.) nee danke @(.)@
- me:** @(.)@
- I:** des (.) weil es so dreckig war?
- Nadine:** ja, also ick meine ick hätte da nicht jeessen (.) und die Frau, mit die wer uns dann halt da unterhalten haben, die des allet jezeichnet hat, die hat je ooch die Frau die dann halt da kocht ooch total runtergezogen (...) ick hab mich bei ner Freundin uffn Schoß gesetzt also ick hab mir=ick hab mich nich oof die Stühle setzen können, dit war ja so widerlich ey
- ?:** °@(.)@°

9 Bandura zufolge können Überzeugungen der eigenen Wirksamkeit (self efficacy beliefs) das Handeln und die Leistungen der Individuen völlig unabhängig von deren individueller Ausstattung mit anderen Eigenschaften wie Intelligenz positiv beeinflussen (vgl. auch Edelstein 1995, S. 13).

10 Im Konzept werden als Bestandteil des Projekts Begegnungen „mit Anderen/Fremden“ auf den Ebenen „des miteinander Tuns, des voneinander Lernens, des miteinander Redens“ beschrieben.

- Nadine:** und denn (3) ne (.) also Mittag essen könn ich mir da nich vorstellen  
 ?.: °@(.)@°
- Nadine:** man ist da rinjekommen, da waren irgendwelche Kunden jewesen, Alter, die keen bisschen Kohle hatten, nischt (.) ick meine gegen solche Leute habe ick ja nischt (.) aber dann sind wir da rin, und denn (3) ach, kam ein blöder Kommentar, oder ham sie hinterher gepffiffen und allet, also ick bin mir da janz schön dämlich vorjekommen weil (.) wat die für Sachen anhatten und dann (.) steht hier so eene drinne, die hat (.) frisch jefärbte Haare @(.)@ frisch jewaschene Haare, denn saubere Sachen (.) und wat weiß ick für teure Sachen und dann kommt man sich ooch ein bisschen dämlich vor, wenn Leute da drin, sind die kaum wat haben
- Mario:** mh
- Patrick?:** mh
- I:** mhm
- Nadine:** [ mh (.) det war schon (.) heftig (.) also noch mal möchte ich da nich hin

Ihren Aufenthalt im Obdachlosenheim bringt Nadine mit einer Konfrontation mit Schmutz und starken Ekelgefühlen in Verbindung, auch kamen „blöde Kommentare“ von den Bewohnerinnen und Bewohnern. Sie erlebte sich als privilegiert, indem sie sich der Sauberkeit ihres Äußeren und des Wertes ihrer Frisur und ihrer Kleidung bewusst wurde. Dieser Teil der Projektwoche löste selbstreflexive Überlegungen bei Nadine aus und konnte sie offensichtlich nachhaltig beeindrucken. Da sie die Kontakte im Obdachlosenheim jedoch vorrangig als feindselig empfand, bewertet sie diese Erfahrung letztlich als negativ und möchte sie nicht wiederholen.

Die Erzählung Nadines zeigt, dass Begegnungen mit als anders wahrgenommenen Gruppierungen wie Obdachlosen Erfahrungen der Selbstreflexion ermöglichen können. Jedoch liefert sie auch deutliche Hinweise auf die Grenzen solcher pädagogisch inszenierter Begegnungen: Diese können auch bereits verfestigte Stereotype etwa über die „Ekeligkeit“ von Obdachlosen bestätigen.

Weiterhin ergab die Evaluation, dass die größte Schwierigkeit mit den Projektwochen für die Auszubildenden darin lag, durch diese in die Situation eines engeren Kontaktes untereinander zu kommen. Wie bereits erläutert, ist eine zentrale Erfahrung der Jugendlichen ihre kontinuierliche Stigmatisierung und Degradierung als BZM-Auszubildende und eine verbreitete Reaktionsweise auf diese Erfahrung besteht in der Projektion dieses Stigmas auf Andere, verbunden mit Strategien der Distinktion von diesen. Vor diesem Hintergrund müssen Kontakte innerhalb des BZM, die über die einmal gefundene Peergroup hinausgehen, notwendig mit gravierenden Irritationen und Herausforderungen einhergehen. In diesem Sinne empfanden einige Jugendliche die gemeinsame Unterbringung während der Projektwoche in einer Bildungsstätte als unerträglich und ein gemeinsames Pizzaessen mit Asylbewerberinnen und -bewerbern wurde als negativ bewertet, weil die Pizza von den verachteten „fetten Weibern“ aus dem Hauswirtschaftsbereich gebacken wurde.

## Konsequenzen der Evaluationsergebnisse

Die Forschungsergebnisse wurden im Laufe der prozessbegleitenden Evaluation mehrmals für die Verantwortlichen des XENOS-Projekts transparent gemacht und gemeinsam diskutiert. Diesen wurden nach jeder durchgeführten Erhebung mündliche Rückmeldungen gegeben, weiterhin wurden ein Zwischenbericht und ein Abschlussbericht angefertigt, die beide in mehrstündigen Feedback-Runden gemeinsam mit dem Xenos-Team erörtert wurden. Die Ergebnisse dieses kommunikativen Verständigungs- und Aushandlungsprozesses flossen in den Abschlussbericht ein. In diesem Sinne handelte es sich um eine formative nutzungsorientierte Evaluation, die an der Ausgestaltung des Projekts mitwirkte und an der Optimierung der Praxis orientiert war. Im Folgenden seien einige der Bewertungen einzelner Angebote im Rahmen des XENOS-Projekts benannt. Wie dies dem „Gesprächscharakter der Evaluation“ (vgl. Bohnsack 2006) im Sinne der dokumentarischen Evaluationsforschung entspricht, sind sie im Dialog mit den Stakeholdern, hier den Auftraggebern, über die Evaluationsergebnisse entstanden.

- Vor dem Hintergrund der oben erläuterten Interpretationsergebnisse vermuteten wir, dass vor allem all jene Angebote sich als sinnvoll erweisen könnten, die den Jugendlichen die Chance geben, ein Gefühl eigener Verantwortung, Kompetenz und Selbstwirksamkeit zu entwickeln. Als Voraussetzungen hierfür erscheinen uns eine respektvolle und anerkennende Haltung der Kursleiter/innen sowie eine egalitäre Lernatmosphäre, die es den Einzelnen ermöglicht, ihre persönlichen Fähigkeiten einzubringen. Diese Einschätzung korrespondierte mit den Bewertungen einzelner Projektwochen durch die Jugendlichen, die das Xenos-Team im Rahmen einer Selbstevaluation ermittelt hatte.
- Die Evaluationsergebnisse verweisen darauf, dass Angebote, die die Auszubildenden überforderten, kontraproduktiv waren, da sie deren kontinuierliche Erfahrung der Degradierung wiederholten und deshalb Abwehrhaltungen und Trotzreaktionen provozierten. Auch in diesem Punkt stimmten die Verantwortlichen uns zu. Auf dieser Grundlage ließ sich in der gemeinsamen Betrachtung einzelner Projektwochenangebote entscheiden, welche fortgesetzt werden sollten und welche nicht.
- Die Auswertungen der Gruppendiskussionen mit den Jugendlichen zeigen deutlich, dass arrangierte Begegnungen mit als different wahrgenommenen Gruppierungen zur Förderung einer „weltoffenen und antirassistischen Grundhaltung“ nicht ausreichen<sup>11</sup>. Wichtig erscheinen ergänzende Maßnahmen, die den Jugendlichen vor dem Hintergrund eigener Erfahrungen begreifbar machen, wie Diskriminierungen funktionieren.

---

<sup>11</sup> Auch Pingel/Rieker (2003, S. 107f.) weisen in ihren Überlegungen zur Evaluation pädagogischer Praxis gegen Rechtsextremismus darauf hin, dass arrangierte Begegnungen mit Fremden die Gefahr bergen, dass die entsprechende Zielgruppe diese letztlich nur wegen bestimmter Fähigkeiten respektiert, beziehungsweise dass sie als positive Ausnahme betrachtet und andere Angehörige derselben Gruppe weiterhin missachtet werden.

Die Evaluierenden empfahlen in diesem Zusammenhang, auf Ansätze zurückzugreifen, die eine Verbindung von persönlicher Erfahrung und gesellschaftlichen Dimensionen von Diskriminierung anstreben wie dies etwa im Anti-Bias-Training der Fall ist (vgl. Derman-Sparks 1989). Die Xenos-Verantwortlichen vermuteten allerdings, ihre Klientel sei nicht offen für derartige Formen der Selbstreflexion. Sie favorisierten stattdessen Angebote im Bereich der Selbstverteidigung.

- Vor dem Hintergrund der Annahme, dass die angestrebte „antirassistische Grundhaltung“ in einer toleranten Einstellung gegenüber der eigenen Peergroup verankert sein müsste, wurde von Seiten der Durchführenden der Evaluation die Empfehlung formuliert, die alltäglichen Diskriminierungen der Auszubildenden untereinander zum zentralen Thema des Projekts zu machen. Gemeinsam mit dem Xenos-Team wurde überlegt, welche Art von Angeboten es den Jugendlichen ermöglichen könnten, eine größere Toleranz untereinander zu entwickeln, beziehungsweise auch, sich selbst gegen die Diskriminierungen anderer zur Wehr setzen zu können.

Die Ansprechpartnerinnen und -partner der Evaluierenden waren drei im Rahmen des dreijährigen XENOS-Projekts angestellte Pädagoginnen und Pädagogen, die an der Entwicklung dieser Empfehlungen beteiligt waren und diese zumindest in Teilen auch während der Projektlaufzeit umsetzten. Da wir mit den Verantwortlichen des Ausbildungszentrums während der Durchführung der Evaluation kaum Kontakt hatten, bestand nach Abgabe des Abschlussberichts, der auch Empfehlungen in Bezug auf günstige Voraussetzungen für eine Nachhaltigkeit des Projekts enthielt, zunächst die Befürchtung, dass dieser nicht oder kaum im BZM zur Kenntnis genommen würde. Mittlerweile wurde das „Centrum für qualitative Evaluations- und Sozialforschung e.V.“ jedoch von der Geschäftsführung des Ausbildungszentrum beauftragt, ein neues XENOS-Projekt zu evaluieren, bei dessen Konzeption auf zentrale Ergebnisse des vorliegenden Evaluationsberichtes zurückgegriffen wurde<sup>12</sup>.

### **Fazit: Dokumentarische Evaluationsforschung als Mittel zur Bewertung von Maßnahmen im Bereich der Rechtsextremismusprävention**

Wie Ulrike Hormel und Albert Scherr (2004) in ihrer Auseinandersetzung mit verschiedenen Konzepten der Bildung für die Einwanderungsgesellschaft konstatieren, liegen mittlerweile einige Studien im Bereich der qualitativen Jugendforschung vor, die belegen, dass fremdenfeindliche, rassistische und rechtsextreme Artikulationen nicht einfach als Übernahme gesellschaftlich zirkulierender Ideologien erklärt werden können:

---

<sup>12</sup> Eine der Empfehlungen des Abschlussberichtes bezog sich auf eine notwendig erscheinende verstärkte Sensibilisierung der im Ausbildungszentrum tätigen Pädagoginnen und Pädagogen im Bereich Rassismus und Rechtsextremismusprävention. Das nun neu beantragte XENOS-Projekt soll Fort- und Weiterbildungen für die im BZM tätigen Berufsschullehrer/innen in diesem Bereich ermöglichen.

„Sie sind vielmehr als Formen der Auseinandersetzung mit je spezifischen eigenen Lebensbedingungen und darin eingelassenen Erfahrungen sinnverstehend zu rekonstruieren. Daraus folgt, dass pädagogische Interventionen sich nicht auf eine aufklärungsorientierte Kritik von Vorurteilen, Feindbildern und Ideologien beschränken können, die deren innere Widersprüche und sachlichen Fehlleistungen aufdeckt. Vielmehr ist es erforderlich, in die Auseinandersetzung mit den jeweiligen Sichtweisen der eigenen Lebenssituation, Probleme und Interessen einzutreten, die sich in der Verwendung fremdenfeindlicher, rassistischer und rechtsextremer Symbole, in Diskursen und Handlungen artikulieren“ (ebd., S. 29).

Die dokumentarische Evaluation des im BZM durchgeführten XENOS-Projekts ermöglichte Rückschlüsse auf die eigene Lebenssituation der Auszubildenden, die sich als durchaus hilfreich für den weiteren Verlauf des Projekts erwiesen.

Die dokumentarische Evaluationsforschung bzw. die ihr zugrunde liegende wissenssoziologische Analyse eröffnet nicht nur einen anspruchsvollen methodischen Zugang zum handlungsleitenden Erfahrungswissen der Evaluationsbeteiligten, der Stakeholder, sondern darüber hinaus eröffnet sie auf dem Wege der Explikation impliziten Wissens in optimaler Weise die Möglichkeit, der von Seiten der Evaluationsforscher/innen immer wieder geäußerten – Forderung nachzukommen, allen am Evaluationsprozess beteiligten Parteien die Chance zu eröffnen, ihre Orientierungen und Werthaltungen zur Artikulation zu bringen<sup>13</sup>. Eine derartige Artikulationshilfe ist nun aber im Kontext der dokumentarischen Evaluationsforschung nicht eine normative oder ethische Forderung, die zur Methodologie noch hinzutritt, wie uns dies unter dem Begriff des „empowerment“ bei Fetterman (1994) und daran anknüpfend bei Guba/Lincoln (1989, S. 203) begegnet. Vielmehr ist sie integraler und elementarer Bestandteil der Methodologie selbst. Eine Explikation impliziten Wissens erhöht – indem sie das praktische Wissen der unterschiedlichen Milieus in einen einheitlichen ‚Code‘ übersetzt – die Chance, gehört zu werden, gerade für diejenigen, die vom dominanten ‚Code‘ weiter entfernt sind, also für Angehörige von Randmilieus, aber auch für die in der institutionellen Hierarchie weiter unten angesiedelten Gruppen (also beispielsweise für Schüler in der Institution Schule allgemein und für Schüler oder Auszubildende aus Randmilieus noch einmal insbesondere).

In diesem Sinne ist es dann auch integrales Element einer dokumentarischen Evaluationsforschung, der Perspektive jener Stakeholder zur Artikulation zu verhelfen, die in der Regel – im Unterschied zu den anderen Evaluationsbeteiligten – ihre Interessen nicht direkt in den Diskurs einbringen können (vgl. Bohnsack 2006).

Die Rekonstruktion des Erfahrungswissens und der Orientierungen der Jugendlichen führten zu der Erkenntnis, dass sich die Zielgruppe der angestrebten Förderung von

---

<sup>13</sup> Beispielsweise betonen Guba/Lincoln (1989, S. 203): „If audiences are not sufficiently knowledgeable or sophisticated at the start, it is the evaluator’s moral obligation to take steps necessary to bring them to whatever minimal level is required to full participation“.

Akzeptanz und Integration selbst in hohem Maße von mangelnder Akzeptanz und einer unzureichenden Integration in die Gesellschaft bedroht sieht. Diese eigene Erfahrung hat offensichtlich auch großen Einfluss auf die Interaktionen der Jugendlichen untereinander sowie ihren Umgang mit anderen.

Die BZM-Auszubildenden verfügen vor diesem Hintergrund über Orientierungsmuster, an die rechtsextreme Ideologien der Ungleichwertigkeit, der hierarchischen Ordnung sozialer Beziehungen und des Rechts des Stärkeren gut anknüpfen können. In diesem Sinne konstatieren auch Rommelspacher u.a. (2003) unter Bezug auf verschiedene empirische Studien:

„Entscheidend ist bei rechtsextremen Jugendlichen eine Sozialisation, die ihnen vermittelt hat, dass sie ihre Probleme vor allem durch die Abwertung anderer „lösen“ können und dass die Welt primär als eine feindselige zu verstehen ist, der mit Misstrauen begegnet werden muss und in der jedes Mittel recht ist, um ans Ziel zu kommen. Es geht also um die Ideologie der Ungleichwertigkeit und ein starkes Konkurrenzmotiv“ (ebd., 54).

Insbesondere aufgrund der Tatsache, dass ‚bildungsferne‘ Jugendliche in den seit Beginn der 1990er-Jahre in der BRD durchgeführten Antirassismusmaßnahmen über lange Zeit signifikant vernachlässigt wurden (vgl. Roth u.a. 2003, S. 16; Klingelhöfer/Brüggemann 2003, S. 87), erscheint es außerordentlich sinnvoll, für diese Zielgruppe ein breitgefächertes Angebot politischer Bildung im Bereich der Rechtsextremismusprävention und der Demokratie- und Toleranzförderung bereitzustellen.

Wenngleich wir der Ansicht sind, dass die spezifische Problematik der BZM-Auszubildenden sie zu einer besonders sinnvollen Zielgruppe für ein derartiges Angebot qualifiziert, so möchten wir jedoch gleichzeitig davor warnen, diese per se als defizitär einzustufen. Wie sich aus den Gruppendiskussionen eindeutig ergibt, sind nicht mangelnde Bildung oder Intelligenz die Voraussetzung für bestimmte Orientierungen der Jugendlichen, sondern ihre Erfahrung der gesellschaftlichen Stigmatisierung als behindert oder benachteiligt und insofern als anormal. Die wiederum von Rommelspacher u.a. (2003) formulierte folgende Einschätzung: „Die Frage ist also nicht in erster Linie, ob sich rechtsextreme Jugendliche in besonderen Problemlagen befinden, sondern wie sie diese interpretieren“ (ebd., S. 55) ließe sich insofern wie folgt ergänzen: und wie diese von ihrer Umgebung interpretiert werden. Objektivistische Zugänge, bei denen die Forscherinnen und Forscher Problemlagen als objektive, jenseits der Interpretationen und Erfahrungen sowie des handlungsleitenden Erfahrungswissens der Jugendlichen zu identifizieren suchen (vgl. Bohnsack 2000), erweisen sich somit nicht nur als erkenntnistheoretisch fragwürdig, sondern auch als unfruchtbar hinsichtlich ihres Erkenntnisgewinns.

## Literaturverzeichnis

- Bandura, Albert (1993): Perceived self-efficacy in cognitive development and functioning. In: *Educational Psychologist*, No. 28, S. 117–14
- Bohnsack, Ralf (2000): Jugendliche als Opfer und Täter. Das Fehlen der Jugend in der Forschung zur Jugendkriminalität. In: Sander, Uwe/Vollbrecht, Ralf (Hrsg.): *Jugend im 20. Jahrhundert*. Neuwied u. Berlin: Luchterhand, S. 316–336
- Bohnsack, Ralf (2005): Standards nicht-standardisierter Forschung in den Erziehungs- und Sozialwissenschaften. In: *Zeitschrift für Erziehungswissenschaft (ZfE)*, 8. Jg., Beiheft Nr. 4. Ingrid Gogolin, Heinz-Hermann Krüger, Dieter Lenzen, Thomas Rauschenbach (Hrsg.) *Standards und Standardisierungen in der Erziehungswissenschaft*. Wiesbaden: VS-Verlag für Sozialwissenschaften, S. 63–81
- Bohnsack, Ralf (2006): *Qualitative Evaluation und Handlungspraxis. Grundlagen dokumentarischer Evaluationsforschung*. In: Flick, Uwe (Hrsg.): *Qualitative Evaluationsforschung*. Reinbek bei Hamburg: Rowohlt, S. 135–155
- Bohnsack, Ralf (2007): *Rekonstruktive Sozialforschung – Einführung in qualitative Methoden*. (6. Auflage). Opladen/Farmington Hills: Verlag Barbara Budrich
- Derman-Sparks, Louise (1989): *Anti-Bias-Curriculum – Tools for Empowering Young Children*. Washington: ERIC
- Edelstein, Wolfgang (1995): *Krise der Jugend – Ohnmacht der Institutionen. Eine Einleitung im Anschluss an Emil Durkheims Theorie*. In: Edelstein, W. (Hrsg.): *Entwicklungskrisen kompetent meistern. Der Beitrag der Selbstwirksamkeitstheorie von Albert Bandura zum pädagogischen Handeln*. Heidelberg: Asanger, S. 13–24
- Fetterman, David M. (1994): *Empowerment Evaluation*. In: *Evaluation Practice*, Vol. 15, No.1, S. 1–15
- Guba, Egon G./Lincoln, Yvonna S. (1989): *Fourth Generation Evaluation*. Newbury Park, CA.: Sage
- Hormel, Ulrike/Scherr, Albert (2004): *Bildung für die Einwanderungsgesellschaft. Perspektiven der Auseinandersetzung mit struktureller, institutioneller und interaktioneller Diskriminierung*. Opladen: VS Verlag für Sozialwissenschaften
- Klingelhöfer, Susanne/Brüggemann, Ulrich (2003): *Das Aktionsprogramm „Maßnahmen gegen Rechtsextremismus und Gewalt“ – Darstellung und kritische Bilanz der wissenschaftlichen Begleitung*. In: Lynen von Berg, Heinz/Roth, Roland (Hrsg.): *Maßnahmen und Programme gegen Rechtsextremismus wissenschaftlich begleitet. Aufgaben, Konzepte und Erfahrungen*. Opladen: VS Verlag für Sozialwissenschaften, S. 77–93
- Lüders, Christian/Haubrich, Karin (2003): *Qualitative Evaluationsforschung*. In: Schewpe, Cornelia (Hrsg.): *Qualitative Sozialforschung in der Sozialpädagogik*. Opladen: Leske + Budrich, S. 305–330.
- Mannheim, Karl (1964): *Beiträge zur theorie der Weltanschauungsinterpretation*. In: Ders.: *Wissenssoziologie*. Neuwied, S. 91-154 (Original: 1921-1922 In: *Jahrbuch für Kunstgeschichte* XV, 4)

- Nentwig-Gesemann, Iris (2006): Dokumentarische Evaluationsforschung. In: Flick, Uwe (Hrsg.): Qualitative Evaluationsforschung. Reinbek bei Hamburg: Rowohlt, S. 159–183
- Pingel, Andrea/Rieker, Peter (2003): Evaluation pädagogischer Praxis gegen Rechtsextremismus: Potenziale und Chancen. In: Lynen von Berg, Heinz/Roth, Roland (Hrsg.): Maßnahmen und Programme gegen Rechtsextremismus wissenschaftlich begleitet. Aufgaben, Konzepte und Erfahrungen. Opladen: VS Verlag für Sozialwissenschaften, S. 103–118
- Rommelspacher, Birgit/Polat, Ülger/Wilpert, Czarina (2003): Die Evaluation des CIVITAS-Programms. In: Lynen von Berg, Heinz/Roth, Roland (Hrsg.): Maßnahmen und Programme gegen Rechtsextremismus wissenschaftlich begleitet. Aufgaben, Konzepte und Erfahrungen. Opladen: VS Verlag für Sozialwissenschaften, S. 51–75
- Roth, Roland/Lynen von Berg, Heinz/Benack, Anke (2003): Programme und Maßnahmen gegen Rechtsextremismus und Fremdenfeindlichkeit – Fragen und Anmerkungen zu ihrer wissenschaftlichen Begleitung. In: Lynen von Berg, Heinz/Roth, Roland (Hrsg.): Maßnahmen und Programme gegen Rechtsextremismus wissenschaftlich begleitet. Aufgaben, Konzepte und Erfahrungen. Opladen: VS Verlag für Sozialwissenschaften, S. 9–24
- Schwandt, Thomas A. (1997): Evaluation as Practical Hermeneutics. In: Evaluation, 3 (1), 69–83
- Schwandt, Thomas A. (2002): Evaluation Practice Reconsidered. New York/Bern/Berlin/Bruxelles/Frankfurt am Main/Oxford/Wien: Peter Lang Publishing Group
- Stake, Robert E. (1978): The Case Study Method of Social Inquiry. In: Educational Researcher 7 (2). S. 5–8
- Stake, Robert E. (2000): Program Evaluation, Particularly Responsive Evaluation. In: Madaus, Georg F./Scriven, Michael/Stufflebeam, Daniel L. (Hrsg.): Evaluation Models. Viewpoints on Educational and Human Services Evaluation. Boston/Dordrecht/London: Kluwer Academic Publishers, S. 343–362

## Begleitforschung und Evaluation – fast das Gleiche, und doch etwas Anderes!

### Vorbemerkungen

„Wissenschaftliche Begleitung“ und „Evaluierung“ sind zu gängigen Bestandteilen von Politik und Praxis geworden. Und nicht selten (insbesondere im Kontext der Durchführung innovativer Politikprogramme oder von Modellprojekten) geht im Sprachgebrauch der Projektträger wie der finanzierenden Institutionen das Eine in das Andere über: Wissenschaftliche Begleitung ist Evaluation und Evaluation ist wissenschaftliche Begleitung. Diese Gleichsetzung ist jedoch problematisch. Natürlich besteht – sofern mit „Evaluieren“ ein datenbasiertes Fällen von Qualitäts- oder Erfolgsurteilen gemeint ist – methodisch eine relativ große Schnittmenge zwischen beiden Aufgaben: Beide müssen bei der Informationssammlung und -auswertung auf das Instrumentarium der empirischen Sozialforschung zurückgreifen und sich an deren Regelsystem ausrichten. Während es sich aber bei Begleitforschung um „nichts weiter“ als um „ganz gewöhnliche“ empirische Forschung handelt (parallel zur Projektdurchführung werden die zur Beschreibung und Bewertung relevanten Informationen gesammelt), kann das „Evaluieren“ unter Rückgriff auf verschiedenartige Konzepte und auf unterschiedliche Weise mit unterschiedlichen Zielen von unterschiedlichen Instanzen vorgenommen werden. Und während für die Geltungsbegründung der Resultate von Begleitforschung eindeutige methodologische Standards existieren (wie Gültigkeit, Zuverlässigkeit, Repräsentativität), existiert für den Nachweis der Geltung normativer Aussagen (und darum handelt es sich bei „Evaluationen“) keine allgemein anerkannte Begründungslogik. Ganz im Gegenteil gilt nach vorherrschender wissenschaftstheoretischer Auffassung für empirische Forschung das Postulat der „Wertfreiheit“ – genauer: Des Ausschlusses jeder nichtwissenschaftlichen Wertbasis aus dem Kanon der zulässigen Argumente für die Begründung empirischer Aussagen. Für das damit angesprochene Dilemma – zwischen einerseits dem Objektivitätsanspruch an wissenschaftlich begründete Aussagen empirischer Forschung und andererseits dem Anspruch an die empirische Evaluation zu bewertenden Aussagen mit intersubjektiver Geltung zu gelangen – existiert bisher kein überzeugender Lösungsvorschlag.

### **Gleichartige methodologische Rahmenbedingungen für wissenschaftliche Begleitung und Evaluation: Angewandte Forschung**

Wenn eingangs in Hinblick auf wissenschaftliche Begleitung von „ganz gewöhnlicher“ empirischer Forschung die Rede war, so ist dies unter methodologischen Gesichtspunkten nicht ganz zutreffend; denn wir haben es hier nicht mit „normaler“, diszipliniert orientierter Forschung oder mit Grundlagenwissenschaft zu tun (für die die empirische Methodologie

entwickelt wurde), sondern mit angewandter (oder anwendungs- bzw. praxisorientierter) Forschung. Zwar sind die Methoden und Instrumente empirischer Informationsgewinnung in beiden Fällen dieselben; unterschiedlich sind jedoch die Rahmenbedingungen ihres Einsatzes. Von angewandter Sozialforschung wird allgemein gesprochen, wenn ein erfahrungswissenschaftliches Untersuchungsvorhaben auf die unmittelbare Lösung praktischer, gesellschaftlich-politischer Probleme hin angelegt ist (vgl. Habermehl 1992). Es geht darum, gezielt und systematisch Grundlagen für außerwissenschaftliche Entscheidungsprozesse bereitzustellen (vgl. Clemens 2000, S. 215). Dazu zählt auch die wissenschaftliche Projektbegleitung.

Als zentrale Funktionen disziplinorientierter Sozialforschung gelten: „Aufklärung“ (enlightenment) durch Veröffentlichung, Wissensvermehrung, Erklärung, Verallgemeinerung. Praxisorientierte Sozialforschung dagegen ist geprägt durch Aufgaben wie Wissensübertragung (Transfer) durch Beratung, durch Handlungsorientierung, Beeinflussung, Situationsbezogenheit. Unterschiede finden wir auch bei der Art der Reaktionen auf die Forschung. Bei disziplinorientierter Sozialwissenschaft existiert kein direktes Feedback (außer bei Rezensionen von Veröffentlichungen), es geht ihr um die wertungsneutrale Erweiterung des Wissensbestandes. Praxisorientierte Sozialforschung muss dagegen mit direktem Feedback durch ihre Nutzer rechnen; schließlich hat sie ja bedarfsorientiert Informationen bereit zu stellen und aufzubereiten. Wer schon Evaluation betrieben hat, weiß, wie viel Zündstoff in jedem dieser für die angewandte Forschung aufgeführten Punkte steckt. Umso höhere Anforderungen an Fach- wie an Sozialkompetenz müssen die Anwendungsforscherinnen und -forscher erfüllen, um in diesem „verminten“ Arbeitsfeld bestehen zu können!

Was ist es, das anwendungsorientierte Forschung – und damit auch Begleitforschung und Evaluation – methodisch so besonders anspruchsvoll macht? Was sind die spezifischen Schwierigkeiten, die Praxisforschung von der Grundlagenwissenschaft unterscheidet? Einen wesentlichen und konsequenzenreichen Unterschied finden wir in einem zentralen Charakteristikum ihres „Gegenstands“: Der Komplexität sozialer Sachverhalte. Grundlagenforschung mit Zielen wie Theorietest oder Theorieentwicklung löst das Komplexitätsproblem durch konsequente Vereinfachung der Untersuchungssituation – bis hin zur Konstruktion bewusst realitätsferner Laborsituationen, um den Einfluss einzelner Wirkfaktoren zu isolieren. Im Rahmen anwendungsorientierter Forschung verbietet sich ein solches Vorgehen. Reale soziale Sachverhalte sind kein der Vereinfachung zugängliches Experimentierfeld; sie sind und bleiben komplexe Realität. Eine angemessene Beschreibung und Diagnose kann in diesem Rahmen immer auch nur komplex ausfallen – oder sie ist zwangsläufig falsch und führt ebenso zwangsläufig zu falschen Schlussfolgerungen.

Eine Privilegierung der Grundlagenwissenschaft ist darüber hinaus deren erkenntnistheoretisch gewollte Autonomie bei der Wahl ihres „Gegenstands“. Sie kann (bis zu einem gewissen Grad) ihre Themen unter wissenschaftlichen Gesichtspunkten wählen und deren Relevanz unter Hinweis auf angestrebte „reine“ Erkenntnis begründen. Sie kann und soll sich bei allen zu treffenden Entscheidungen an wissenschaftsimmanenten Normen – und

nur an diesen – orientieren (insbesondere Erkenntnisfortschritt als Selbstzweck, Gültigkeit, Zuverlässigkeit, Generalisierbarkeit der Resultate); außerwissenschaftliche Interessen werden in den Entdeckungs- und Verwertungskontext ausgegliedert. Das ist bequem! Diese Entlastung von gesellschaftlicher Verwertungs- und Nützlichkeitsargumentation ist wesentlicher Bestandteil der wissenschaftlichen Legitimation.

Anwendungsorientierte Forschung muss dagegen nicht nur die Definition und Präzisierung ihrer Fragestellungen an außerwissenschaftlichen Erkenntnisinteressen und Verwertungskontexten ausrichten: Dies ist ihr zentrales Definitions- und Legitimationskriterium. Vielmehr darf auch bei der methodologischen – also empirisch-wissenschaftlichen – Begründung jeder Detailentscheidung innerhalb des Forschungsablaufs die unmittelbare Verwertbarkeit der erzielbaren Resultate nicht aus dem Blick geraten. Der „Elfenbeinturm der Wissenschaft“ ist für diese Forschungsrichtung kein möglicher Zufluchtsort. Ganz im Gegenteil werden wissenschaftsexterne Zweckmäßighkeitsüberlegungen aus dem Verwertungskontext „internalisiert“ und geraten so auf die gleiche Stufe wie die methodologischen Normen der Gültigkeit und Zuverlässigkeit. Dennoch dürfen auch letztere nicht vernachlässigt werden, denn falsche – also empirisch nicht haltbare – Informationen sind genauso wertlos wie dies pragmatisch unbrauchbare Resultate wären. Dieser Spagat ist schwierig!

Die Grundlagenforschung darf sich noch eines weiteren Privilegs erfreuen: Sie kann – zumindest vom Prinzip her – unter Entlastung von aktuellem Zeitdruck arbeiten. Neue Erkenntnisse – „Entdeckungen“ – lassen sich schließlich nicht mit dem Terminkalender erzwingen. Benötigt ein Forschungsprogramm mehr Zeit, dann muss diese eingeräumt werden. Und falls dennoch das vorgesehene Arbeitsprogramm nicht geschafft wurde und bei zeitlich befristeten Projektressourcen nicht eingelöste Desiderate unter „weiterer Forschungsbedarf“ ausgewiesen werden müssen, dann entwertet dies jedenfalls nicht das schon Erreichte. Anders bei anwendungsorientierter Forschung: Hier haben die Ergebnisse im zuvor definierten Umfang zeitpunktgenau vorzuliegen. Verspätet vorgelegte Befunde sind als vorgesehene Grundlage von Beratungen oder zur Absicherung von Entscheidungen nicht mehr verwertbar, sobald der Entscheidungstermin verstrichen ist.

Last, but not least: Grundlagenforschung darf sich „irren“. Damit ist gemeint: Hypothesen, die als Ausgangspunkt gewählt werden, dürfen sich im Verlauf der Forschung als falsch erweisen. Es gilt sogar als besonders wünschenswert, die Suche nach neuer Erkenntnis von „gewagten“ Annahmen aus aufzunehmen. Deren informationsreiches Scheitern ist nicht selten der Startpunkt für grundlegende Erkenntnisse, die eine neue Forschungslinie begründen. Diesen Luxus darf sich angewandte Sozialwissenschaft nicht leisten. Bei der Konzipierung des Forschungsdesigns ist große Sorgfalt darauf zu verwenden, dass die zugrunde liegenden Annahmen und Hypothesen einen hohen Grad empirischer Bewährung aufweisen und dass der Prozess der Gewinnung, Auswertung und Interpretation aller Informationen methodisch abgesichert und mit begleitender Qualitätskontrolle verläuft. Jede falsche Schlussfolgerung im Verwertungskontext, die wegen fehlerhafter Forschungsdaten gezogen wird, hat Konsequenzen für einen nicht absehbaren Kreis von Betroffenen.

## Begriffsklurung: Was ist Evaluation?

Evaluationsforschung ist ein prototypisches Beispiel fur Anwendungsorientierung; fur sie gelten samtliche methodologischen Anforderungen, die im vorigen Abschnitt aufgelistet wurden. Doch darf „Evaluation“ nicht auf Evaluationsforschung reduziert werden; sie umfasst erheblich mehr als „nur“ Forschung. In den Vorbemerkungen wurde bereits auf die Unbestimmtheit und Vielgestaltigkeit des Konzepts hingewiesen. Es stellt sich daher die Frage: Was ist das eigentlich – Evaluation?

Die Antwort auf diese simpel erscheinende Frage fallt allerdings gar nicht so leicht. Schon beim ersten Anlauf zu einer Antwort fallt auf, dass wir es hier mit einem geradezu – zumindest fur nicht alltagssprachliche Zwecke – „schlampigen“ Begriff zu tun haben, der eigentlich in eine wissenschaftliche Terminologie keinen Zugang hatte finden durfen, denn er steht gleichzeitig fur verschiedene Arten von Referenzobjekten:

- Er bezieht sich zum einen auf eine „symbolische“ Ebene: Als aus dem Amerikanischen ubernommene Bezeichnung fur Bewertung/Bewerten sowie fur Denkmodelle bzw. Konzepte systematischer Bewertungsverfahren. Wenn wir uns uber Methoden oder Standards der Evaluation<sup>1</sup> unterhalten, beziehen wir uns auf diese Ebene.
- Der Begriff kann aber auch auf ein spezifisches empirisches Handeln, auf einen Prozess zielorientierten Informationsmanagements gemunzt sein: Auf eine (methodisch kontrollierte) verwertungs- und bewertungsorientierte Form des Sammelns und Auswertens von Informationen. Wenn von Begleitforschung die Rede ist, dann steht dieser Aspekt im Vordergrund.
- Und schlielich kann damit auch das Resultat dieses Prozesses, ein Werturteil uber das Zielobjekt des Evaluierens (das „Evaluationsprodukt“) angesprochen werden. Dies ist der Fall, wenn wir eine Aussage horen, das Programm X oder das Institut Y sei „positiv“ oder „negativ“ evaluiert worden.

Dieser Mangel an begrifflicher Prazision mag harmlos klingen, ist aber verantwortlich fur die Ergebnislosigkeit mancher Diskussionen, in denen die Beteiligten nicht immer bemerken, dass sie wechselseitig uber vollig Unterschiedliches und damit aneinander „vorbei“ reden. Die mangelnde Prazision reicht aber noch einen Schritt weiter: Nicht nur bezieht sich der Begriff „Evaluation“ auf unterschiedliche Arten von Referenzobjekten, sondern seine Bedeutung variiert zudem in verschiedenen Verwendungskontexten erheblich:

- (1) Alltagssprachlich handelt es sich lediglich um ein „wohlklingendes“ und deshalb gern benutztes Fremdwort fur jegliche Art von Bewertung: Irgendjemand bewertet irgend etwas irgendwie unter irgendwelchen Gesichtspunkten.

---

<sup>1</sup> Siehe z. B. die von der Deutschen Gesellschaft fur Evaluation beschlossenen „Standards fur Evaluation“ (DeGE-val 2002), die in einem laufenden Revisionsprozess weiter entwickelt werden ([www.degeval.de](http://www.degeval.de)).

- (2) Im öffentlichen, im weitesten Sinne „politischen“ Sprachgebrauch ist damit die Überprüfung eines Programms oder einer Einrichtung durch ein zu diesem Zweck berufenes Expertengremium gemeint.
- (3) Im wissenschaftlichen Kontext, insbesondere unter empirisch-methodischer Perspektive, verbirgt sich hinter „Evaluation“ das Design für einen spezifischen Typ von Sozialforschung.
- (4) Und schließlich wird ärgerlicherweise auch noch „ganz normale“ Umfrageforschung mit Evaluation gleichgesetzt, falls sie – analog zur Meinungsforschung – bewertende (also evaluierende) Aussagen der Befragten erhebt und aufbereitet. Zielgruppen solcher Evaluationsbefragungen sind Personen, die in besonderer Weise als geeignet angesehen werden, einen zu bewertenden Sachverhalt oder Gegenstand zu beurteilen: Etwa Teilnehmerinnen und Teilnehmer von Lehrveranstaltungen als Evaluatorinnen und Evaluatoren der Qualität der Lehre, Klientinnen und Klienten einer sozialen Dienstleistung als Beurteiler der Dienstleistungsqualität, Wahlberechtigte als Begutachter von Politikern (vgl. „Politbarometer“). Genau besehen handelt es sich hierbei allerdings nicht um methodisch kontrollierte Bewertungen mit dem Anspruch intersubjektiver Geltung wie bei den vorgenannten Typen (2) und (3), sondern entweder um die Erhebung und Aufbereitung subjektiver „Alltagsevaluationen“ (deren zugrunde liegende Kriterien und Standards nicht rekonstruierbar sind) oder um Zufriedenheitsmessung.

Methodisch kontrollierte Evaluation unterscheidet sich von Alltagsevaluation dadurch, dass nicht „irgendetwas“ beurteilt wird, sondern dass genau benannte und empirisch beschreibbare Sachverhalte (Programme, Maßnahmen, Organisationen) und auf diese gerichtete präzise und operationalisierbare Fragestellungen ihr „Gegenstand“ sind. Beurteilt wird nicht von „irgendjemandem“, sondern durch Personen, die zur Bewertung besonders befähigt erscheinen (das können die beauftragten Forscherinnen und Forscher oder Expertinnen und Experten sein oder Beteiligte und/oder Betroffene oder zum Fällen der Urteile speziell legitimierte Gremien wie der intern/extern besetzte Beirat eines Projekts oder ein Fachausschuss in der Kommunalverwaltung oder die Lehrkommission in einem universitären Fachbereich). Evaluiert wird auch nicht „in irgendeiner Weise“ und „nach irgendwelchen Kriterien“, sondern in einem objektivierten Verfahren auf der Basis empirisch abgesicherter Informationen sowie nach explizit auf den Sachverhalt bezogenen und begründeten Beurteilungskriterien und -maßstäben.

Deshalb ist es umso wichtiger, dass die für die Funktion Evaluation wichtigste Frage beantwortet wird: Wer beurteilt nach welchen Kriterien? Anders ausgedrückt: Wer ist die Evaluierungs-Instanz, und im Hinblick auf welche/wessen Ziele wird die Bewertung vorgenommen? Lediglich im Evaluationsdesign der empirischen Programmwirkungsforschung (siehe den nächsten Abschnitt) fällt die Antwort leicht. Die Beurteilungskriterien ergeben sich aus den operationalisierten Zielen des zu evaluierenden Programms. Die ermittelten Zielerreichungsgrade gelten dann als die empirischen Messwerte des Erfolgs/Misserfolgs und die Evaluierung besteht darin, die gemessenen Grade der Zielerreichung mit den vom

Programm angestrebten Sollwerten zu vergleichen. Sie ist somit kein eigentliches „Werturteil“, sondern eine technologische Vergleichsaussage. Forschung und Evaluierung fallen in diesem Konzept definitionsgemäß zusammen.

### **Der methodologische „Goldstandard“ der Evaluationsforschung: Das Design der Programm(wirkungs-)forschung**

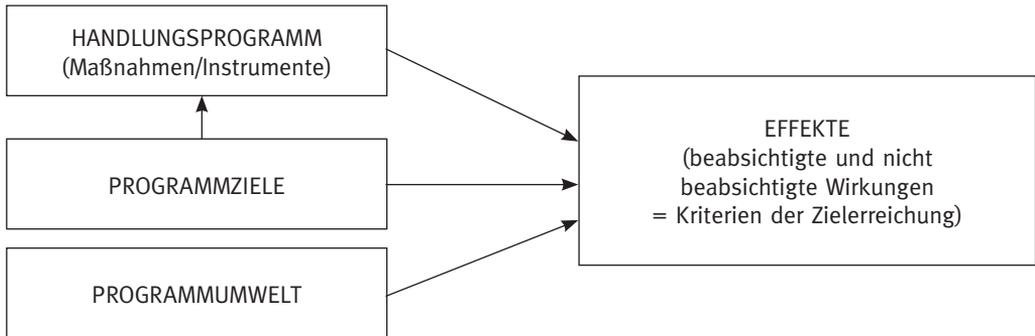
Die Begleitforschung hat – so wurde bisher argumentiert – „lediglich“ empirisch korrekte Informationen zum Projektverlauf und zu den Projektergebnissen zu liefern und dabei den für empirische Forschung geltenden Qualitätsstandards zu genügen. Diese selbstverständliche Anforderung ist jedoch – darauf wurde eingangs hingewiesen – unter den Rahmenbedingungen anwendungsorientierter Forschung nicht so einfach einzulösen wie sie klingt. Dies soll im Folgenden am Konzept der traditionellen Programmwirkungsforschung, dem am komplettesten ausgearbeiteten Design einer „Evaluation durch Forschung“<sup>2</sup> illustriert werden. Die Methodologie der Programmevaluation wurde im Wesentlichen schon in den 1970er und 80er Jahren entwickelt.<sup>3</sup> Das Design ist von der Logik her im Prinzip für unterschiedliche Evaluationsinteressen anwendbar, sei es dass – in Anlehnung an eine von Eleanor Chelimsky (1997, S. 100ff.) vorgeschlagene grobe Einteilung – sie zur Verbreiterung der Basis handlungspraktischen Wissens („Forschungsparadigma“), zu Kontrollzwecken („Kontrollparadigma“) oder zur Entwicklungszwecken („Entwicklungsparadigma“) durchgeführt wird (vgl. Kromrey 2005, S. 36ff.). Je nachdem, welche Interessen und Zwecke im Vordergrund stehen und in welcher Weise die vier Aspekte (Gegenstand – Evaluatorin/Evaluator – Bewertungskriterien – Verfahren) konkretisiert werden, sind zwar die Schwerpunkte unterschiedlich zu setzen, doch bleibt allen Vorhaben gemeinsam, dass sie (mindestens) drei interdependente Dimensionen aufweisen – nämlich Ziele, Maßnahmenprogramm, Effekte – und dass sie (anders als in einem Forschungslabor) von Umgebungseinflüssen nicht abgeschirmt werden können.

---

2 Dieser Typ von Evaluationsforschung wird üblicherweise „Programmevaluation“ genannt. Das ist insofern korrekt, als das Design neben der Erhebung und Analyse evaluationsrelevanter empirischer Daten auch Verfahren der Ableitung datenbasierter Werturteile (also Evaluationen) bereitstellt.

3 Soweit alternative Evaluationstypen in ihrem Vorgehen von diesem Konzept abweichen, verwenden sie für die empirische Informationsgewinnung die üblichen Methoden quantitativer und/oder qualitativer Sozialforschung, auf die hier nicht eingegangen werden kann (vgl. dazu Lehrbücher, etwa Kromrey 2006; Lamnek 1995).

**Abbildung 1: Programm-Evaluation**



Somit kommt auf die Begleitforschung und Evaluation die auf den ersten Blick simpel anmutende Aufgabe, die in obiger Abb. 1 aufgeführten vier Variablenbereiche (Ziele/Maßnahmen/Effekte/Programmumwelt) mit empirischen Daten abzubilden (zu „messen“) und miteinander zu verknüpfen. Das Vorgehen orientiert sich dabei am Modell der Kontrolle der „unabhängigen“ bzw. „explikativen“ Variablen (hier: Maßnahmen des Programms) und der Feststellung ihrer Effekte auf genau definierte „abhängige“ Variablen (Zielerreichungskriterien).

An konkreten Forschungsaufgaben folgen daraus:

- Messung der „unabhängigen Variablen“, d.h.: das Handlungs-Programm mit seinen einzelnen Maßnahmen ist präzise zu erfassen;
- Identifizierung und Erfassung von Umwelt-Ereignissen und -Bedingungen, die ebenfalls auf die vom Programm angestrebte Zielsituation Einfluss nehmen könnten (exogene Einflüsse);
- Messung der „abhängigen Variablen“, d.h.: das Wirkungsfeld (beabsichtigte und nicht-beabsichtigte Effekte) ist zu identifizieren, die Wirkungen sind anhand definierter Zielerreichungs-Kriterien (operationalisierter Ziele) zu messen.

Die Aufgabe der Datenerhebung besteht für die gesamte Dauer des Programmablaufs in einem – so Eekhoff u.a. (1977, S. 11ff.) – „Monitoring“ der Instrumentvariablen (Programm-Input), der exogenen Einflüsse und der Zielerreichungsgrade (Output). Bei diesem dreifachen „Monitoring“ handelt es sich somit um methodologisch problemlose deskriptive Forschungsaktivitäten. Wesentlich schwerer zu lösen ist die darauf folgende analytische Aufgabenstellung: Die festgestellten Veränderungen im Wirkungsfeld des Programms sind aufzubrechen:

- in jene Teile, die den jeweiligen Maßnahmen als deren Wirkung zurechenbar sind,
- und in die verbleibenden Teile, die als Effekte exogener Einflüsse (Programm-Umwelt) zu gelten haben.

Die eigentliche „Erfolgskontrolle“ oder „Evaluation“ beinhaltet nach diesem Modell zwei Aspekte:

- Analyse der Programmziele und ihrer Interdependenzen (Präzisierung eines Zielsystems einschließlich der Festlegung des angestrebten Zielniveaus) sowie Zuordnung der Instrumente zur Zielerreichung (Maßnahmen des Programms);
- Vergleich der den einzelnen Maßnahmen zurechenbaren Effekte mit den angestrebten Zielniveaus.

Das damit skizzierte Modell einer kausalanalytisch angeleiteten Programmevaluations- und Wirkungsforschung wirkt in sich schlüssig und einleuchtend und scheint nur noch einer weiteren Differenzierung hinsichtlich der Methodik zu bedürfen. Bei näherem Hinsehen allerdings wird erkennbar, dass es von anspruchsvollen Voraussetzungen über den Gegenstand der Untersuchung wie auch von Voraussetzungen bei den Programm durchführenden Instanzen und der Evaluation selbst ausgeht. Diese mögen zwar bei Vorhaben der Grundlagenforschung (vereinzelt) gegeben sein, sind jedoch in Programmforschungsprojekten wenig realitätsnah, so dass die Forscherinnen und Forscher fortwährend zu Kompromissen zwischen methodologischen Normen und empirischer Praxis gezwungen sind. Drei dieser meist implizit gelassenen Voraussetzungen sind besonders hervorzuheben, da deren Erfüllung eine wesentliche Bedingung dafür ist, das Forschungsprogramm empirischer Kausalanalysen überhaupt anwenden zu können:

- Vor der Entwicklung des Forschungsdesigns muss Klarheit über die Untersuchungsziele – bezogen auf einen definierbaren und empirisch abgrenzbaren Untersuchungsgegenstand – bestehen. Für die Dauer der Datenerhebung dürfen sich weder die Untersuchungsziele noch die wesentlichen Randbedingungen des Untersuchungsgegenstandes in unvorhersehbarer Weise ändern.
- Vor der Entwicklung des Forschungsdesigns müssen des Weiteren begründete Vermutungen (Hypothesen) über die Struktur des Gegenstandes wie auch über Zusammenhänge und Beziehungen zwischen dessen wesentlichen Elementen existieren, nach Möglichkeit in Form empirisch bewährter Theorien. Erst auf ihrer Basis kann ein Gültigkeit beanspruchendes Indikatorenmodell konstruiert, können geeignete Messinstrumente entwickelt, kann über problemangemessene Auswertungsverfahren entschieden werden.
- Die Forscherinnen und Forscher müssen die Kontrolle über den Forschungsablauf haben, um die (interne und externe) Gültigkeit der Resultate weitestgehend sicherzustellen.

Im Normalfall der Begleitforschung zu Programm-Implementationen oder gar zu Modellversuchen neuer Techniken, neuer Schulformen, zur Erprobung alternativer Curricula oder Lernformen u. ä. ist keine einzige dieser Bedingungen voll erfüllt. Die Untersuchungssituation weist vielmehr in dieser Hinsicht erhebliche „Mängel“ auf (vgl. Kromrey 1988). Die im Fol-

genden skizzierte Methodologie der Programmevaluation ist daher weniger ein Real- als ein Idealtyp, an den anzunähern die Forschung sich je nach gegebener Situation bemühen wird.

## Ziel- und Maßnahmenanalyse

Nimmt man die gängigen Definitionen von „Interventionsprogramm“ oder „Handlungsprogramm“ beim Wort<sup>4</sup>, dann müssten die Kenntnisse, die man sich durch die Forschung erhofft, bei den Akteuren weitestgehend schon vorhanden sein: Neben einem widerspruchsfreien Zielsystem müsste zuverlässiges Praxiswissen existieren, um auf der Basis von Daten über die gegebene Ausgangssituation die erforderlichen Maßnahmen und Instrumente zur Erreichung der Zielsituation zu bestimmen. Solche Kenntnisse über Ziel-Mittel-Relationen müssten zudem technologisch verwertbar sein; d.h. die als strategisch wichtig erkannten Variablen müssten dem Eingriff der Programmdurchführenden zugänglich sein. Mayntz weist jedoch mit Recht darauf hin, „das nur im Ausnahmefall ein Programm zu Beginn des Implementationsprozesses als konkrete, faßbare Einheit vorliegt“ (1980, S. 4). So sind die Ziele oft nicht eindeutig und nicht konkret, sondern vage und leerformelhaft formuliert. Das kann ganz bewusst im Prozess der Programmaushandlung geschehen sein (um einen Kompromiss zwischen widerstreitenden Interessen zu ermöglichen oder um Konflikte zwischen parlamentarischen Koalitionspartnern zu vermeiden). So wird die Aufgabe der Präzisierung aus der (öffentlich-politischen) Zielfindungs- in die (weniger öffentliche) Implementierungsphase verschoben. Besonders problematisch ist dies, wenn am Programmvollzug mehrere Ebenen beteiligt sind (Bund/Länder/Kommunen/andere Träger). Auf jeder Ebene von Akteuren kann die Ausfüllung der Ziel-Leerformeln in unterschiedlicher Weise geschehen. Gleiches kann bei der Zuordnung von Maßnahmen (Instrumenten) der Fall sein, die der Zielerreichung dienen sollen. So können im Prinzip unterschiedliche Maßnahmen mit dem gleichen Programm vereinbar erscheinen. Eine mögliche Folge ist, dass bestimmte Träger ihre bereits auf Vorrat bestehenden „Schubladenpläne“ unter das neu beschlossene Programm subsumieren. In solchen Fällen entstehen Diskrepanzen zwischen offiziellen (manifesten) und verdeckten (latenten) Programmzielen. Außerdem können Ziele, die in ihrer vagen Formulierung als miteinander vereinbar erscheinen, sich bei der Konkretisierung als im Widerspruch zueinander stehend erweisen. Des Weiteren können zu Beginn gesetzte Ziele (selbst wenn sie präzise formuliert waren) sich im Laufe der Programmrealisierung ändern oder in ihrer Gewichtung verschieben, etwa weil sich wichtige Rahmenbedingungen für das Programm in nicht erwarteter Weise entwickelt haben. In manchen Programmen schließlich finden sich eher Kataloge von Maßnahmen als eindeutige Ziele, so dass unbestimmt bleibt, was mit dem Programm letztlich erreicht werden soll, welches also die angestrebten Effekte sind.

---

<sup>4</sup> Programme sind komplexe Handlungsmodelle, die auf die Erreichung bestimmter Ziele gerichtet sind, die auf bestimmten, den Zielen angemessen erscheinenden Handlungsstrategien beruhen und für deren Abwicklung finanzielle, personelle und sonstige Ressourcen bereitgestellt werden (vgl. Hellstern/Wollmann 1983, S. 7; grundlegend: Mayntz 1980).

Mag es bei der Formulierung von Programmzielen im politischen Aushandlungsprozess durchaus funktional sein, diese bewusst vage und mehrdeutig zu lassen, so darf die Interpretation empirischer Befunde für Evaluationszwecke dagegen nicht „aushandelbar“ sein. Das bedeutet, dass ungenaue Zielformulierungen im Zuge der Designentwicklung konkretisiert werden müssen. Präzise Aussagen über die genannten Aspekte – Ziele/Instrumente/Ziel-Mittel-Relationen – sind schon rein forschungstechnisch eine unabdingbare Voraussetzung, um überhaupt eine Evaluation im hier verstandenen Sinne vornehmen zu können. Die Forschenden sind somit im Falle „unvollständiger“ Programme gezwungen, Lücken zu schließen und die notwendigen Präzisierungen vorzunehmen sowie im Falle widersprüchlicher Formulierungen Konflikte und Unverträglichkeiten zwischen den Zielen herauszuarbeiten und zu beseitigen – selbst auf die Gefahr hin, dass damit das Programm in Teilen zu einem „Konstrukt des Forschers“ (Mayntz 1980, S. 4) wird.

Ein erster Ansatz ist eine eher „technische“ Zielanalyse in der Absicht, die Liste der Programmziele zu komplettieren, ein hierarchisches Zielsystem zu konstruieren (ausgehend von Oberzielen über Haupt-, Teil- und Unterziele bis hin zu Indikatoren, die als Näherungskriterien den Grad der Zielerreichung zu messen gestatten) und die jeweils einzusetzenden Instrumente zuzuordnen (vgl. Hellstern/Wollmann 1983, S. 11ff.). Ausgangspunkt sind die vorhandenen Angaben im Programm; daneben sind relevante ergänzende Dokumente (Beratungsprotokolle, Grundsatz-Aussagen der beteiligten Organisationen/Parteien etc.) heranzuziehen sowie ggf. Beteiligte am Entscheidungsprozess zu befragen. Ein weiterer Zugriff bietet sich über die Personen und Institutionen an, die für die Implementation verantwortlich sind (insbesondere wenn sich der Vollzug des Programms über mehrere Ebenen erstreckt). Eine Zielpräzisierung aus der Sicht der Beteiligten bietet bereits zu Beginn wichtige Anhaltspunkte, ob und in welcher Weise es im Verlaufe des Programmvollzugs zu Zielverschiebungen kommen dürfte.

Das erstellte Zielsystem hat drei formalen Kriterien zu genügen: Konsistenz, Operationalisierbarkeit, Praktikabilität. Ein inkonsistentes System von Zielen (das also nicht in sich geschlossen und logisch widerspruchsfrei ist) kann nicht Grundlage für eine rationale Analyse und für die Bewertung eines Programms sein. Nicht-operationalisierbare Ziele sind nicht durch Daten abbildbar, können somit nicht Gegenstand empirischer Forschung sein. In Bezug auf den Zielinhalt sind Ziele operational, wenn sich gegenüber einer bestehenden Ausgangssituation die angestrebte Zielsituation genau herleiten und anhand geeigneter Indikatoren (Zielerreichungskriterien) messen lässt. Praktikabel schließlich sind Ziele dann, wenn sie auf praktisches Handeln gerichtet sind und ihre Verwirklichung kontrolliert werden kann. Zielaussagen zu Sachverhalten, die im Rahmen des Programms nicht Gegenstand von planenden und handelnden Eingriffen sein können, sind für die Praktikerinnen und Praktiker irrelevant; Forschungsbefunde zu Aspekten, die im Zuständigkeitsbereich der Programmdurchführenden nicht veränderbar sind, tragen in diesem Kontext nicht zu praxisrelevantem Wissen bei und sind aus dieser Sicht „wertlos“.

Eine wichtige Unterscheidung, die nicht in allen Programmen vorgenommen wird, ist die zwischen Zielen und Maßnahmen/Instrumenten. Ziele geben an, was erreicht werden

soll. Instrumente sind die Hilfsmittel, die einzusetzen sind, um die Ziele zu erreichen. Diese eindeutig klingende Abgrenzung ist jedoch nicht so simpel wie es den Anschein hat, insbesondere dann nicht, wenn mehrere Ebenen an der Durchführung beteiligt sind. Je nach Betrachtungsperspektive kann ein und derselbe Sachverhalt ein Ziel oder aber eine Maßnahme sein.

Allerdings ist kein Evaluationsvorhaben so umfassend realisierbar, dass alle aus den verschiedenen Beteiligten-Perspektiven erstellbaren Ziel-Mittel-Systeme als alternative Maßstäbe an die Bewertung angelegt werden können. Somit ist die vorherige bewusste und begründete Entscheidung über den Verwendungszusammenhang notwendig: Wer ist Adressat der Evaluationsaussagen und welchem Zweck sollen die Ergebnisse dienen? Offensichtlich kann also die Ziel-/Maßnahmen-Analyse niemals eine (interessenneutrale) Rekonstruktion und Präzisierung „des Programms“ sein, sondern immer nur eine Perspektive, unter der das komplexe Gefüge Programm/Beteiligte/Umwelt betrachtet und untersucht wird.

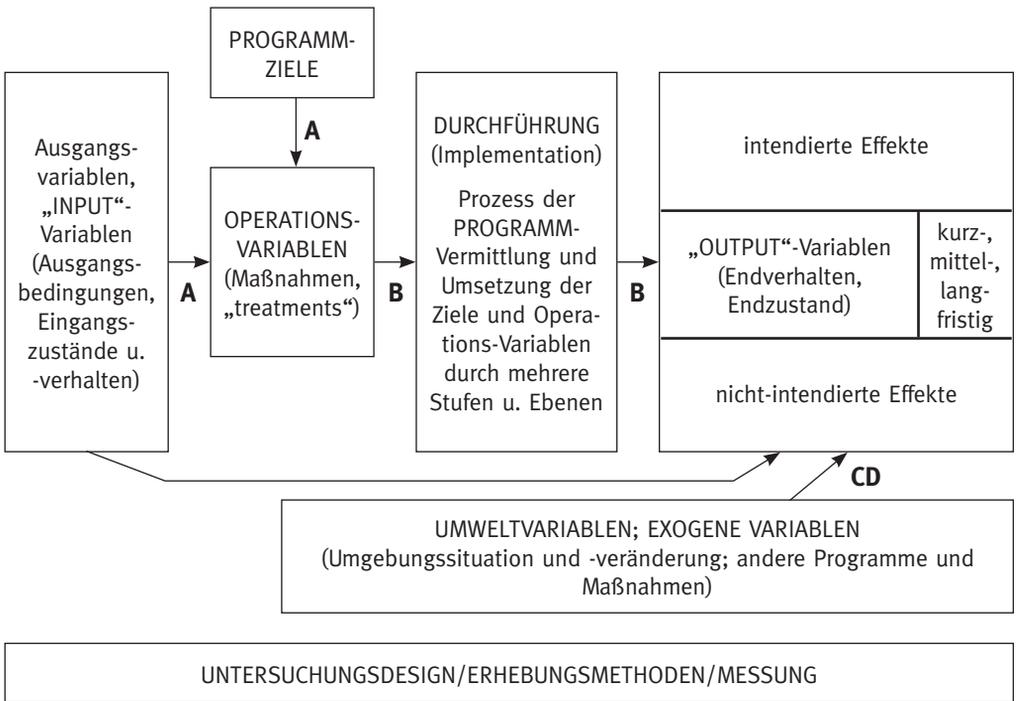
### **Konzipierung der Wirkungen (Modell des Wirkungsfeldes)**

Mit der Ziel- und Maßnahmenanalyse ist zwar die Voraussetzung für die Evaluierung, nicht jedoch für die Wirkungsanalyse geschaffen. Programmforschung erfolgt nicht in einer Laborsituation, in der jede einzelne Maßnahme isoliert auf ihre Effekte hin untersucht und in der die Wirkung der isolierten Maßnahme von allen übrigen Einflüssen („Störgrößen“) abgeschirmt werden könnte. Die Forschung hat es nicht mit einfachen Kausalketten zu tun (Maßnahme  $X_1$  bewirkt über die intervenierenden Zwischenschritte  $Y_1$  und  $Y_2$  die Veränderung der Zielsituation von  $Z_0$  nach  $Z_1$ ), sondern mit einem komplexen Wirkungsfeld. Insbesondere wirkt eine Maßnahme nicht (trennscharf) nur auf ein Ziel und ist ein Ziel nicht (monokausal) nur durch eine Maßnahme erreichbar. Die Wirkungen treten nicht sämtlich zu gleicher Zeit ein; es ist zwischen kurz-, mittel- und langfristigen Effekten zu unterscheiden. Schließlich kann es neben den beabsichtigten auch zu ungeplanten Wirkungen und Wechselwirkungen kommen. Die Programmumwelt (Maßnahmen anderer Programme oder sozialer Akteure) können auf den Verlauf des zu evaluierenden Programms in erwünschter wie auch in unerwünschter Richtung Einfluss nehmen (und umgekehrt, das Programm kann seine Umwelt verändern).

Ziel des zu konzipierenden Wirkungsmodells ist es also, möglichst alle für die Beurteilung des Programmverlaufs relevanten (potenziellen) Wirkungen und Wirkungszusammenhänge auf gedanklicher Ebene vorab zu explizieren. Kriterien für die Entscheidung, was das „relevante Wirkungsfeld“ sein soll, liefern die Programmziele. Benötigt werden aber darüber hinaus begründete Vermutungen und empirisch abgesicherte Theorien, die geeignet sind, die wechselseitigen Beziehungen zwischen Zielen, Maßnahmen und Umfeld zu antizipieren. Dieses Denkmodell eines Wirkungsfeldes, die formale Darstellung einer Theorie des Handlungsprogramms und seiner Einbettung in soziale Realität, ist die Basis für die Entwicklung eines „maßgeschneiderten“ Forschungsdesigns. Dieses Design soll einerseits

unter methodologischen Gesichtspunkten möglichst hohen Standards der empirischen Wissenschaft genügen, zugleich aber unter den durch das Programm gesetzten Rahmenbedingungen realisierbar sein. Jede ernstzunehmende Grundlagenforschung würde unter solchen Bedingungen versuchen, die Fragestellung so weit einzuschränken, dass möglichst alle wesentlichen (potenziellen) Einflussgrößen methodisch kontrolliert werden können. Evaluationsforschung darf genau diese Strategie – sollen die Ergebnisse ernst genommen werden – nicht verfolgen, sondern muss möglichst viel Komplexität in ihrer Untersuchungsanlage mitberücksichtigen. Das zu Beginn skizzierte Modell der Programmforschung stellt sich nach diesen Überlegungen nun wie folgt dar (wobei die methodischen Probleme der Indikatorenbildung und -messung in jedem der Variablenfelder sich nicht in anderer Weise stellen als bei jeder empirischen Untersuchung und daher hier nicht gesondert behandelt werden):

**Abbildung 2:** Grobes Variablenmodell einer Evaluationsstudie



**Erläuterungen:**

Die mit A beschrifteten Pfeile symbolisieren, dass sich die Maßnahmen an den Ausgangszuständen („Input“) sowie an den Sollzuständen („Ziele“) zu orientieren haben. B besagt, dass die Maßnahmen durchgeführt werden und „Wirkungen“ zeigen. C soll in Erinnerung

bringen, dass sich zwischen Beginn und Ende des Programms „autonome“ Veränderungen der Ausgangszustände einstellen können (sog. „Reifungsprozesse“). D schließlich weist auf die Notwendigkeit der Berücksichtigung externer Effekte (Wirkungen der Programmumwelt) hin. Alle diese Merkmale sind im Monitoring über die Laufzeit des Programms in geeigneter Weise zu „messen“; dem muss das Untersuchungsdesign Rechnung tragen. In der sozialen Realität sind selbstverständlich weitere Zusammenhänge wirksam. So kommen die Programmziele nicht unbeeinflusst von der Ausgangslage (die als „problematisch“ definiert worden ist) sowie der sozialen Umwelt zustande. Auch existieren Wechselbeziehungen zwischen dem „Wirkungsfeld“ des Programms und seiner Umwelt. Diese Aspekte bleiben jedoch in Evaluationsprojekten im Normalfall unberücksichtigt.

Diese der Forschungslogik des empirischen Theorie- bzw. Hypothesentests entlehnte Konzeption führt jedoch zu einem gewissen Paradoxon, wenn sie – wie hier – auf angewandte Forschung in einem Aufgabenfeld übertragen wird, das charakteristischerweise „Neuland“ ist. Zu evaluierende Programme, sozialwissenschaftlich zu begleitende Modellvorhaben bewegen sich in einem Praxisfeld, in dem es gerade sowohl an theoretischem als auch an Erfahrungswissen mangelt (andernfalls wäre ihre Evaluierung überflüssig). Bei der Konzipierung der Untersuchung sind dennoch die wesentlichen Wirkungszusammenhänge zu antizipieren und zwar – anders als bei Forschungen zum Zwecke der Theorieentwicklung – nicht „versuchsweise“, um dann empirisch auf ihre Haltbarkeit überprüft zu werden, sondern unmittelbar anwendungsbezogen. Das Paradoxon besteht nun in folgendem: Einerseits müssen unter methodologischen Gesichtspunkten empirisch bestätigte Theorien über die Struktur des Untersuchungsfeldes verlangt werden; sie sind als Basis für die Entwicklung eines auf Wirkungsmessungen angelegten Designs erforderlich. Andererseits existiert logischerweise bei „neuen“ Untersuchungsfeldern das empirisch gesicherte Wissen noch nicht; es kann erst durch die noch durchzuführende Untersuchung gewonnen werden.

Dieses Dilemma ist nur durch einen Verstoß gegen die Methodologie traditioneller empirischer Forschung auflösbar. Den oben festgestellten „Mängeln“ der Untersuchungssituation (fehlende Konstanz des Programms und der Rahmenbedingungen) kann nur mit einem analogen „Mangel“ des Designs begegnet werden: Das zu Beginn formulierte Wirkungsmodell ebenso wie das darauf zugeschnittene Forschungsdesign sind „veränderungsoffen“ anzulegen. Die Forschungslogik strukturtestender Verfahren (einschließlich standardisierter Erhebungsmethoden) ist zu ergänzen um Verfahren strukturentdeckender Forschung (etwa das von Glaser/Strauss entwickelte Konzept zur „Entdeckung einer gegenstandsbezogenen Theorie“, vgl. Kromrey 1994b). Das zu entwickelnde und schrittweise zu optimierende Design hat jedoch nicht nur – wie bei jeder empirischen Datenerhebung – zu gewährleisten, dass die in Abb. 3 dargestellten Variablengruppen möglichst zuverlässig und verzerrungsfrei gemessen werden können, es muss darüber hinaus die Voraussetzungen für eine Kausalanalyse der Daten schaffen. Idealtypisch geeignet ist die (feld)experimentelle Vorgehensweise. Sie darf dementsprechend als der „Königsweg“ der Programmforschung gelten, ist allerdings – wegen des Primats des Programms vor der

Forschung – im Normalfall nur näherungsweise (in Form diverser quasi-experimenteller Ansätze) realisierbar<sup>5</sup>.

## Das Werturteilsdilemma in der Evaluation

Wenn von Evaluation – in welchem Zusammenhang und in welcher spezifischen Begriffsbedeutung auch immer – gesprochen wird, ist damit eine bewertende Aussage über einen Gegenstand oder Sachverhalt gemeint. Kommt – wie hier dargestellt – als Erkenntnisinstrument auch empirische Forschung ins Spiel, kann das Ziel von Evaluation präzisiert werden als: empirisch gestützte Gewinnung von Bewertungen mit intersubjektivem Geltungsanspruch, häufig zugespitzt auf die Forderung: forschungsgestützte Gewinnung von Qualitätsaussagen. Das Ziel ist in jedem Fall eine normative Aussage, ein „Werturteil“. Für das Konzept „Evaluation durch Forschung“<sup>6</sup> führt dies zu einem gravierenden Problem, sind doch nach herrschender Meinung in Methodologie und Wissenschaftstheorie normative Aussagen – Werturteile generell und damit auch Qualitätsurteile im Besonderen – empirisch nicht begründbar. Überraschenderweise wird dieses Dilemma – abgesehen von ganz seltenen Ausnahmen<sup>7</sup> – im Diskussionskontext von Evaluationsforschung selten thematisiert. Aus der Perspektive der analytisch-homologischen Wissenschaftstheorie fällt die Antwort nicht schwer; der Anspruch, „empirisch gestützte Bewertungen mit intersubjektivem Geltungsanspruch“ zu gewinnen, ist methodologisch nicht legitimierbar:

- Die unmittelbare empirische Begründung von Bewertungen durch Forschung ist nicht möglich; auch aus korrekten empirischen Beschreibungen und Analysen sind normative Aussagen nicht ableitbar.
- Die Geltungsbegründung empirischer Analysen folgt einer anderen Logik als die Geltungsbegründung normativer Aussagen. Für die ersteren gibt es in der Wissenschaftstheorie klare Regeln, für die letzteren nicht.
- Möglich ist es lediglich, die Forschung möglichst genau auf den Zweck „Bewertung“ auszurichten, indem wenn schon nicht direkt bewertende, so doch bewertungsrelevante Informationen gesammelt und systematisiert werden<sup>8</sup>.

---

5 Dazu sowie zu Alternativen der Wirkungszurechnung vgl. Kromrey 2001, S. 119ff. sowie ders. 2005, S. 45ff. Für das Experimentaldesign sowie quasi-experimentelle Ansätze kann auf Frey/Frenz 1982 verwiesen werden.

6 Wie zu Beginn angesprochen, existieren nebeneinander zahlreiche unterschiedliche Verständnisse von Evaluation, von relativ wenig formalisierter Evaluierung durch Experten (auch als Basis von professioneller Beratung) über responsive, stakeholder-orientierte, partizipative Evaluationsansätze bis hin zu unterschiedlichen forschungsbasierten Evaluationen. Dieser Text beschränkt sich auf die Wertproblematik im Rahmen von Evaluation durch Forschung (ungenauer: „Evaluationsforschung“).

7 Beispielsweise Christian Lüders wie auch Wolfgang Beywl in Flick 2006.

8 Dies geschieht z. B. in der Hochschulevaluation und/oder bei Akkreditierungen nach dem bekannten mehrstufigen Modell des Peerreviews. Die Forschung hat hier nur die Funktion des Informationszulieferers; das Evaluieren (d.h. das Fällen der Werturteile) geschieht durch Expertinnen und Experten oder durch ein dazu legitimes Gremium oder durch Aushandeln zwischen den beteiligten Parteien.

Soll also unmittelbar „durch Forschung“ evaluiert werden, müssen Strategien gefunden werden, mit deren Hilfe erreichbar wird, dass die empirischen Daten einen quasi normativen Charakter erhalten, so dass sie sozusagen „für sich selbst sprechen“. Eine explizite Geltungsbegründung daraus abgeleiteter Wertaussagen durch die Forschung wäre dann nicht mehr notwendig.

### **Exkurs: Die Argumentationslogik analytisch-nomologischer, „wertfreier“ Forschung**

Für wissenschaftstheoretische Argumentationen kann das von Hempel und Oppenheim konzipierte Schema wissenschaftlicher Erklärung (Hempel/Oppenheim 1948) als verbindliches Gerüst gelten:

- Explanans:*
- (1) Es gilt (mindestens) ein nomologisches Gesetz (z. B.: „Wenn A und B, dann C“)
  - (2) Die in der Wenn-Komponente genannten Randbedingungen sind empirisch erfüllt (z. B.: „A und B liegen vor“)

- 
- Explanandum:*
- (3) Singulärer Satz, der den zu erklärenden Sachverhalt beschreibt (z. B. „C liegt vor“).

Gegeben ist das zu erklärende „singuläre Ereignis“ (3), gesucht ist das „Explanans“ (1 und 2). Bei dieser Art von Erklärung muss (3) deduktiv-logisch aus (1) und (2) folgen, wobei (2) aus der Wenn-Komponente und (3) aus der Dann-Komponente des nomologischen Gesetzes abgeleitet wird.

Das Hempel-Oppenheim-Erklärungsschema ebenso wie die Übernahme der Struktur für andere Erkenntniszwecke impliziert als erkenntnistheoretische Basis den erkenntnistheoretischen Realismus. Das heißt, unterstellt wird auf der Gegenstandsseite eine „real existierende“ Welt, gekennzeichnet durch Merkmale wie Ordnung, Struktur und Tatsachenautonomie, die Geltung von Regelmäßigkeiten bzw. Gesetzmäßigkeiten und Kausalität. Unterstellt wird zudem auf der Seite des erkennenden Subjekts die prinzipielle, wenn auch möglicherweise unvollständige und teilweise fehlerbehaftete Erkennbarkeit dieser Realität durch Wahrnehmungssinne sowie unterstützende Instrumente.

Das Ziel empirischer Wissenschaft ist die Erkenntnis der „wahren“ Strukturen und Gesetzmäßigkeiten der Realität sowie ihre Dokumentation in Theorien. Erreicht werden soll dies durch eine Strategie des „kontrollierten Ratens“ über das Aufstellen erkenntnisleitender ex ante-Hypothesen und deren Konfrontation mit der (objektiven) Realität, abgebildet in (subjektiven) Wahrnehmungsdaten. Die darin implizierte Subjektivität wird kontrolliert durch strenge methodologische Regeln, die ein möglichst weitgehendes Ausschalten wahrnehmungs-beeinflussender Bedingungen bei der empirischen Erkenntnisgewinnung gewährleisten sollten („Objektivierung“ der Verfahren, intersubjektive Nachprüfbarkeit).

Eine besondere Bedeutung kommt der Trennung deskriptiver (und damit „objektivierbarer“) von normativen Aussagen zu, deren subjektiver Charakter methodologisch nicht aufhebbar und deren intersubjektive Geltung daher mit empirischen Mitteln nicht begründbar ist. Die normativen Elemente der Erkenntnisgewinnung werden daher aufgeteilt in solche, die die normative Basis der Forschung bilden (der Forschung vorgelagerte wissenschaftsimmanente Werte), und den nichtwissenschaftlichen Interessen und Werten, die aus dem wissenschaftlichen Begründungskontext ausgelagert und in den Entstehungs- und Verwertungskontext des Forschungsprojekts verwiesen werden. In genau dieser pragmatischen Strategie zur Handhabung des Werturteilsproblems wird in der „herkömmlichen empirischen Forschung“ auch die Lösung des Bewertungsdilemmas in der Evaluationsforschung gesucht. Drei der dabei verfolgten Strategien sollen im Folgenden dargestellt werden.

### **„Wertneutrale Evaluation“ im Design der Programmevaluation: der methodologische Idealtypus**

Die oben skizzierte Forschungslogik einer auf Wertfreiheit verpflichteten Wissenschaft ist bruchlos auf das Konzept „Evaluation durch Forschung“ anwendbar, sofern es sich bei ihrem Gegenstand um ein ausgearbeitetes „Programm“<sup>9</sup> handelt, das in Form explizit formulierter Programmziele die für eine „empirische Bewertung“ notwendige normative Basis bereits in die Evaluation mitbringt. Wie im Kapitel zur Programmwirkungsforschung dargestellt, berücksichtigt das Design dieses Evaluationstyps drei Dimensionen des zu bewertenden Gegenstands – Ziele, Maßnahmen, Effekte – sowie die programmexterne Umwelt als Quelle möglicher „Störvariablen“ (s. Abb. 1).

Auch dieses Konzept ist im Sinne der Hempel-Oppenheim'schen Logik interpretierbar. Zum einen hat schon das „Programm“ als „technologische Aussage“ formal die gleiche Struktur wie eine „Erklärung“. Zum anderen orientiert sich auch das Design der Evaluation an diesen Komponenten:

- Sowohl die existierenden Randbedingungen (2) als auch der Ist-Zustand der Zielvariablen (3) sind vor Programmbeginn –  $t_0$  – empirisch zu beschreiben.
- Während der Programmlaufzeit sind die Veränderungen der Randbedingungen (2) zu erfassen („monitoring“ sowohl der Eingriffe durch die im Programm vorgesehenen Maßnahmen als auch anderer relevanter Veränderungen in der Programmumwelt).
- Schließlich ist sicherzustellen, dass der Zustand der Zielvariablen (3) nach Programmdurchführung –  $t_1$  – wiederum empirisch beschrieben wird, so dass Art und Ausmaß der Veränderungen feststellbar sind.

---

<sup>9</sup> siehe Fußnote 4.

Der Gegenstand, den es zu bewerten gilt, ist natürlich als Gegenstand nicht wertneutral oder zweckfrei. Ganz im Gegenteil: Das Programm soll etwas erreichen. Damit wird auch das Konzept „Evaluation als Programmwirkungsforschung“ mit dem Wertproblem konfrontiert. Für die Forschung ist es aber dadurch gelöst worden, dass es in den „Entstehungskontext“ verlagert wurde (Programmziele als normative Basis), wodurch die eigentliche „Evaluation“ einen deskriptiven (und somit „wertneutralen“) Charakter erhält: Der Bewertungsprozess reduziert sich hier auf einen Vergleich der vom Programm gesetzten Sollwerte (Zielerreichungskriterien) mit den gemessenen (und den Maßnahmen zurechenbaren) Effekten im Wirkungsfeld des Programms. Solche Aussagen lassen sich in vollem Umfang auf die Logik der Geltungsbegründung empirischer Faktenbehauptungen stützen. Sofern von der Evaluation differenzierte „Qualitätsurteile“ gefordert werden, bestehen diese in Aussagen über die Zielerreichung, ggf. konkretisiert durch die Dimensionen Effektivität (Wirkungsgrad der jeweiligen Maßnahmen) und Effizienz (Kosten-Wirksamkeits-Relation). Dass die Realisierung dieser Aufgaben riesige Schwierigkeiten bereitet, tut der überzeugenden Logik des Modells keinen Abbruch.

Anders fällt das Urteil allerdings aus, wenn es um die praktische Bedeutung für ein breites Anwendungsspektrum und für den „alltäglichen“ Evaluationsbedarf geht. Die methodologischen Schwierigkeiten sind nämlich – das sollte in der Darstellung deutlich geworden sein – so umfassend, dass diese überzeugende Logik nur unter sehr einschränkenden, nur ganz selten erfüllbaren Bedingungen praktisch einsetzbar ist. Überwiegend muss die Evaluation sich mit mehr oder weniger guten Annäherungen an das idealtypische Modell zufrieden geben: z. B. in Form quasi-experimenteller Designs mit relativ vielen Zugeständnissen an die interne Validität, projektbegleitendem Monitoring mit erst in der Auswertungsphase an die Experimentallogik angelehnter statistischer Analyse, mit Zeitreihenbetrachtungen der Zielvariablen u.Ä. Es liegt daher nahe, nach komplett andersartigen Ersatzlösungen zu suchen. Eine solche Ersatzlösung folgt aus der Überlegung, „Qualität“ nicht erst anhand von „Effekten“, also als Folge der Eigenschaften des zu bewertenden Gegenstands oder Sachverhalts zu interpretieren, sondern unmittelbar zu messen.

### **Evaluation als Qualitätsmessung: das methodologische Problemkind**

Wenn es gelänge, am zu evaluierenden Gegenstand oder Sachverhalt Qualitätsmerkmale zu bestimmen und präzise zu definieren, hätten wir einen direkten Weg, die Evaluation methodologisch zu „objektivieren“, d.h. am zu bewertenden „Objekt“ festzumachen. Wir könnten uns damit zugleich den schwierigen Umweg über die Messung von Outcome-Variablen und die Zurechnung ihrer Veränderungen als Zielerreichungs-Kriterien ersparen. Die implizite und unmittelbar einleuchtend scheinende Annahme bei dieser Überlegung ist: Wenn die Qualität des zu bewertenden Sachverhalts hoch ist, dann werden auch seine Wirkungen positiv sein; oder anders formuliert: dann werden auch die mit ihm verknüpften Ziele erreicht werden. In diesem Fall hätte die Forschung lediglich die – forschungsmethodisch zur Alltagsroutine zählende – Aufgabe zu erfüllen, „Qualität“ durch einen Satz

qualitätsrelevanter Merkmale ausdifferenzieren und durch geeignete Indikatoren so zu operationalisieren, dass an ihnen situationsunabhängige „Qualitätsmesswerte“ abgelesen werden können. Auch bei dieser Strategie wird die normative Basis der Evaluation in den Entstehungskontext ausgelagert, in dem der normative Begriff „Qualität“ von dazu legitimer Seite festzulegen ist.

Methodologisch ist diese Aufgabe allerdings gar nicht so simpel, wie es auf den ersten Blick erscheinen mag. Die „Sozialindikatorenbewegung“ in den 1970er Jahren hat sich damit intensiv auseinandergesetzt und eine Liste von (ebenfalls schwer erfüllbaren) Anforderungen an gültige und praxisverwertbare Indikatorensysteme formuliert (vgl. Werner 1975). Grundlegender ist jedoch eine andere Problematik, die bei dieser Argumentation häufig übersehen wird: Neben dem Zutreffen der o. g. impliziten Annahme eines direkten Zusammenhangs zwischen Qualitätsmerkmalen und Zielerreichung (was empirisch geprüft werden kann) muss nämlich eine weitere, eine erkenntnistheoretische (und damit axiomatische) Voraussetzung erfüllt sein: Qualität muss als direkte Eigenschaft des Objekts verstanden werden können (wie etwa Größe, Gewicht, Farbe usw.); bzw. methodologisch formuliert: Das Konstrukt „Qualität“ ist so zu definieren, dass seine Dimensionen als Merkmale des Gegenstands erscheinen. Schon eine oberflächliche semantische Analyse lässt erkennen, dass „Qualität“ eben nicht als Merkmal des zu bewertenden Objekts selbst zu verstehen ist, sondern als ein relationales Merkmal: als Eignung, Brauchbarkeit, Güte in Bezug auf bestimmte Ziele und Zwecke sowie auf bestimmte Nutzer- und Klientengruppen.

Damit (und wenn wir die methodologischen Anforderungen der Sozialindikatorenbewegung einbeziehen) haben wir aber wieder das Anforderungsniveau des Modells der Programmevaluation erreicht – ohne allerdings über deren methodologische Problemlösung zu verfügen, nämlich: vom Programm vordefinierte Ziele und Maßnahmen. Anders als zu Beginn postuliert, bleibt – wenn sich das Konstrukt „Qualität“ nicht aus dem Objekt selbst herleiten lässt – die Wert-Entscheidungs-Frage weiterhin unbeantwortet. Die normative Basis für die Bewertung durch Qualitätsindikatoren setzt (und das heißt: die eigentliche Evaluation betreibt) diejenige Instanz, die festlegt, was als „Qualität“ gelten soll und welche Qualitätskriterien und -standards anzulegen sind. Angesichts dieser erneuten Problematik verwundert es nicht, wenn die Forschung sich aus diesen Dilemmata zu befreien versucht, indem sie Evaluation auf das reduziert, was sie unbestritten kann: Befragungen durchführen.

### **Evaluation durch Befragung: das erkenntnistheoretische Problemkind**

Statt aufwändige, methodisch kontrollierte Evaluation durch Programmforschung zu betreiben (deren Anwendungsvoraussetzungen selten erfüllbar sind) oder Qualitätsindikatoren zu messen (deren Gültigkeit fragwürdig ist), wird die Bewertung von Maßnahmen (oder Sachverhalten oder Dienstleistungen) bei dieser Strategie per „Betroffenenbefragung“ ermittelt. Die Adressatinnen, Adressaten und Nutzerinnen und Nutzer, Kundinnen und Kunden sowie Klientinnen und Klienten sind – so wird argumentiert – die von den zu

evaluierenden Leistungen ganz konkret „Betroffenen“ und daher in der Lage, aus eigener Erfahrung auch deren Qualität sachverständig und zuverlässig zu beurteilen. Befragungen erscheinen erheblich weniger anspruchsvoll – sowohl hinsichtlich des Aufwands der Durchführung als auch hinsichtlich der Strategie der Objektivierung: Sind die erbrachten Dienstleistungen „schlecht“, werden auch die Beurteilungen auf einer vorgegebenen Skala negativ ausfallen und umgekehrt. Befragt man eine hinreichend große Zahl von „Betroffenen“ und berechnet pro Skala statistische Kennziffern (etwa Mittelwerte oder Prozentanteile), dann kommen – so die weitere Argumentation – individuelle Abweichungen der einzelnen Urteilenden darin nicht mehr zur Geltung. So hieß es z. B. in einem Dozentenkurs des HDZ Essen schon 1980 optimistisch zum Thema Lehrevaluation:

„Urteile (Schätzungen) von Studenten über die Lehre ... sind – wenn man etwa 20 bis 30 Studenten urteilen lässt – zuverlässig wie professionelle Testverfahren“.

Und: „Sie sind von anderen Merkmalen der Studenten selbst und der Dozenten wenig beeinflusst“ (Schmidt 1980, S. 51f.).

Sofern dies zuträfe<sup>10</sup>, wäre die Evaluation per Befragung der Königsweg zur Lösung aller Probleme der Evaluationsforschung – auch der Werturteilsproblematik, denn die Bewertungen nehmen hier die „per Betroffenheit dazu Legitimierten“ vor. Die Forschung selbst bleibt neutral; sie erhebt, systematisiert und analysiert lediglich. Zwar sind „Messungen“ per Befragung nicht so problemlos wie dies dem Laien häufig erscheint (vgl. Kromrey 2006, S. 257ff.). Doch sofern systematische Verzerrungen vermieden werden können, lassen sich in der Tat bei hinreichend großer Befragtenzahl und bei repräsentativer Datenbasis individuelle Unterschiede – wie im obigen Zitat behauptet – herausmitteln“. Einzulösen sind hierfür lediglich durch das Erhebungsinstrument und in der Befragungssituation einige formale methodologische Voraussetzungen die allerdings nicht ohne weiteres als erfüllt gelten können:

- der „Gegenstand“ (das Objekt) der Beurteilung ist eindeutig definiert,
- das zu messende „Merkmal“ ist eindeutig definiert und operationalisiert,
- eine „Mess-Skala“ existiert und ist eindeutig definiert,
- die Befragten sind in der Lage, den „Gegenstand“ intersubjektiv übereinstimmend zu identifizieren, das zu messende „Merkmal“ intersubjektiv übereinstimmend zu erkennen und die „Mess-Skala“ in intersubjektiv übereinstimmender Weise anzuwenden.

Im Falle der Erhebung von Evaluationen wird die Situation zusätzlich dadurch schwieriger, dass es sich bei den zu messenden Merkmalen um die oben genannten „qualitätsrelevanten Merkmale“ (oder „Qualitätskriterien“) handelt, durch die der Begriff „Qualität“ operationalisiert wird. Und die hierauf anzuwendende Mess-Skala ist die Bezugsgröße, auf der das „Ausmaß“ von Qualität angebbare ist (also der „Qualitätsstandard“). Damit aber

---

<sup>10</sup> Leider ist dieser Optimismus bei Lehrevaluationen nicht gerechtfertigt, wie differenzierte statistische Analysen von Daten aus Veranstaltungsbefragungen belegen (vgl. Kromrey 1994a).

sind wir auch bei der indirekten Qualitätsmessung per Befragung wieder mit dem gleichen Problem konfrontiert wie beim Ansatz der direkten Messung von Objektqualität durch Indikatoren. Das statistische „Ausmitteln“ von Messungenauigkeiten setzt die Existenz eines „wahren Wertes“ voraus, von dem die einzelnen Messwerte lediglich „zufällig“ abweichen. Bezogen auf die Beurteilungsvariation zwischen den einzelnen Befragten heißt dies: Um auf diese Weise zu einem gültigen Qualitätsmaß zu kommen, muss die Annahme gerechtfertigt sein, dass es einen „wahren“ Qualitätswert für den zu beurteilenden Sachverhalt gibt, um den die einzelnen Antworten „zufällig“ streuen. Diese Annahme wäre aber nur dann haltbar, wenn eines der beiden folgenden erkenntnistheoretischen Axiome zuträfe:

Alternative 1: Qualität ist ein „objektives“ Merkmal eines Sachverhalts<sup>11</sup>, dessen Ausprägung durch abbildende subjektive Wahrnehmung ohne systematische Verzerrung „gemessen“ werden kann, so dass bei hinreichend großer Zahl von Messungen der Erwartungswert dem „wahren objektiven Wert“ entspricht. Oder:

Alternative 2: Qualität ist ein „intersubjektiv gültiges“ Konzept, über das alle Menschen in gleicher Weise verfügen. Anders formuliert: Alle Menschen bewerten nach gleichen Kriterien und Standards in gleicher Weise. In konkreten Situationen auftretende Unterschiede zwischen Bewertenden sind als Zufallsvariation anzusehen, so dass bei hinreichend großer Zahl von Messungen der Erwartungswert dem „wahren subjektiven Wert“ entspricht.

Grundlage für die erste Alternative ist die im erkenntnistheoretischen Realismus (vom frühen Empirismus bis zum Gründer des Wiener Kreises, vgl. Schlick 1918, S. 111) vertretene Überzeugung von der Möglichkeit abbildender Wahrnehmung der Realität: Das Wahrgenommene steht in einem genauen Entsprechungsverhältnis zum Wirklichen (vgl. ebd. 1918). In diesem Fall wäre „Qualität“ durch standardisierte Befragung „objektiv“ messbar. Allerdings könnte sie dann auch durch direkte Indikatorenmessung (s.o.) „objektiviert“ werden. Grundlage für die zweite Alternative wäre der erkenntnistheoretische Idealismus, am kompromisslosesten konzipiert in Platons „Ideenlehre“: Hinter der sinnlich wahrnehmbaren Welt stehen (als das „in Wahrheit Seiende“) die „Ideen“, die zwar der direkten Wahrnehmung nicht zugänglich, aber der unsterblichen Seele des Menschen von Anfang an mitgegeben sind. Erkenntnis besteht nach dieser Vorstellung im Wiedererkennen der allgemeingültigen Konzepte (der „Ideen“) in den empirischen (Einzel-)„Erscheinungen“ (vgl. Beckmann 1994). Sofern „Qualität“ der Status einer solchen „Idee“ zukäme (analog zu Gerechtigkeit, Gleichheit, Heldentum, Liebe etc.), wäre sie durch standardisierte Befragung „intersubjektiv“ messbar.

## **Fazit: Trennung von Forschung und Evaluation**

Wenn die Versuche wenig überzeugend erscheinen, das Wertproblem der Evaluationsforschung dadurch zu entschärfen, dass man die für das „wissenschaftliche Evaluieren“ erfor-

---

<sup>11</sup> Dies ist die identische Voraussetzung, die auch im Konzept der direkten Qualitätsmessung durch Indikatoren erfüllt sein muss.

derliche Wertbasis aus dem Begründungskontext empirischer Forschung hinausverlagert (in den Entstehungskontext oder in die extern vorzunehmende Definition eines normativen Begriffs von „Qualität“ oder in nicht intersubjektiv nachvollziehbare Bewertungsprozesse von „Betroffenen“), dann bieten sich eigentlich nur zwei Alternativen für eine „wissenschaftliche Evaluation“ an.

Die eine bestünde darin, die Evaluation als einen Spezialfall aus dem Aufgabengebiet einer wertneutral verfahrenen empirischen Forschung auszusondern und ihr die zusätzliche Aufgabe der (nach wissenschaftlicher Methodologie verfahrenen, intersubjektiv nachprüfbaren) Ableitung von Wertaussagen zuzuschreiben. Überlegungen in dieser Richtung werden von Christian Lüders (2006) angestellt. Eine überzeugende Methodologie ist allerdings derzeit (noch) nicht erkennbar.

Die andere Alternative bestünde darin, Evaluieren und Forschen klar zu trennen: Evaluation und Forschung statt Evaluation durch Forschung. Diese Strategie wird z. B. im Konzept der Hochschulevaluations-„Peerreviews“ sowie in Akkreditierungsverfahren für neue Studiengänge verfolgt. Der Forschung wird dabei die Aufgabe zugeschrieben, alle für die Bewertung von Programmen, Maßnahmen etc. relevanten Informationen unter Einsatz des bewährten empirischen Instrumentariums zu erheben, zu analysieren und für Bewertungs- und Entscheidungsprozesse aufzubereiten. Die Funktion des Evaluierens dagegen sowie der Ableitung möglicher Konsequenzen für das Evaluationsobjekt wird dafür explizit legitimierten Gremien zugewiesen. Dass dieses Modell realisierbar ist und die Akzeptanz von Evaluation erhöht, zeigen die Erfahrungen mit diesem Modell bei der mehrstufigen Hochschulevaluation, bspw. in der vom Verbund Norddeutscher Universitäten praktizierten Variante: Selbstbeschreibung/Selbstevaluation – Peerreview – Auswertende Konferenz (vgl. Nordverbund 2004).

Bezogen auf wissenschaftliche Begleitung im Allgemeinen würde daraus folgen: Bei der Präzisierung – und je nach Projektfortschritt auch Fortschreibung – der Ziele sowie der Entscheidung über die Festlegung geeigneter Messgrößen (Zielerreichungs-Indikatoren) kann und muss die wissenschaftliche Begleitung mitwirken. Sie kann und darf aber diese Entscheidung nicht stellvertretend für das Projekt treffen. Mitwirken sollte sie (und methodologisch gesehen: muss sie), da geänderte Ziele und Zielerreichungskriterien Konsequenzen für das Forschungsdesign haben. Sobald diese Voraussetzungen gegeben sind, kann die Begleitforschung – ähnlich wie die Programmwirkungsforschung – auch die Aufgabe des Evaluierens in Form des Vergleichs von Ist- und Sollwerten übernehmen. Unter pragmatischen Gesichtspunkten und im Interesse der Akzeptanz der Evaluationen kann es aber selbst in diesem Fall geboten sein, das eigentliche Bewerten der erreichten Zwischen- und Abschlussresultate an eine nicht in den laufenden Entwicklungsprozess eingebundene Instanz zu delegieren (Expertinnen/Experten, Nutzerinnen/Nutzer, Betroffene, durch legitimierte Vertreterinnen und Vertreter der Beteiligten und Betroffenen gebildeter Projektbeirat). Dabei können beide „Parteien“ ihre besonderen Stärken und ihre Legitimation in die Waagschale werfen. Für die Geltungsbegründung von Tatsachenbehauptungen ist die unabhängige und unparteiische Forschung zuständig; ihre Aussagen

legitimieren sich durch methodisch kontrollierte, intersubjektiv nachprüfbar Verfahren der Informationsgewinnung sowie durch Einhaltung der methodologischen Regeln und der methodischen Standards empirischer Wissenschaft. Mit der Geltungsbegründung von Bewertungsaussagen sind Instanzen zu betrauen, die sich durch Überparteilichkeit, Unabhängigkeit und Glaubwürdigkeit der Bewertenden auszeichnen; ihre Aussagen legitimieren sich durch Einhaltung der Standards für Evaluationen (vgl. DeGEval 2002).

## Literaturverzeichnis

- Beckmann, Jan P. (1994): Einführung in die Erkenntnistheorie, Hagen: Fernuniversität Hagen (Kurs 3303)
- Beywl, Wolfgang (2006): Evaluationsmodelle und qualitative Methoden. In: Flick, U. 2006, S. 92–116
- Chelimsky, Eleanor (1997): Thoughts for a new evaluation society. „Keynote speech“ at the UK Evaluation Society conference in London 1996. In: Evaluation, 3/1, S. 97–109
- Clemens, Wolfgang (2000): Angewandte Sozialforschung und Politikberatung. In: Clemens, Wolfgang; Jörg Strübing, Jörg (Hrsg.): Empirische Sozialforschung und gesellschaftliche Praxis. Bedingungen und Formen angewandter Forschung in den Sozialwissenschaften, Opladen: S. 211–232
- Deutsche Gesellschaft für Evaluation (DEGEval) (2002): Standards für Evaluationen, Köln: ([www.degeval.de](http://www.degeval.de))
- Eckhoff, Johann/Muthmann, Robert/Sievert, Olaf (1977): Methoden und Möglichkeiten der Erfolgskontrolle städtischer Entwicklungsmaßnahmen, Bonn-Bad Godesberg, Schriftenreihe „Städtebauliche Forschung“, Bd. 03.060
- Flick, Uwe (Hrsg.) (2006): Qualitative Evaluationsforschung. Konzepte, Methoden, Umsetzungen. Reinbek bei Hamburg: Rowohlt
- Frey, Siegfried/Frenz, Hans-Georg (1982): Experiment und Quasi-Experiment im Feld. In: Patry, Jean-Luc (Hrsg.): Feldforschung. Bern/Stuttgart, S. 229–258
- Habermehl, Werner (1992): Angewandte Sozialforschung. München/Wien
- Hellstern, Gerd-Michel/Wollmann, Helmut (1983): Evaluierungsforschung. Ansätze und Methoden, dargestellt am Beispiel des Städtebaus, Basel/Stuttgart
- Hempel, Carl Gustav/Oppenheim, Paul (1948): Studies in the Logic of Explanation. In: ZS Philosophy and Science, Vol. 15, S. 135–175
- Kromrey, Helmut (1988): Akzeptanz- und Begleitforschung. Methodische Ansätze, Möglichkeiten und Grenzen. In: Massacommunicatie, 16/3, S. 221–242
- Kromrey, Helmut (1994a): Wie erkennt man „gute Lehre“? Was studentische Vorlesungsbefragungen (nicht) aussagen. In: Empirische Pädagogik, 1994/2, S. 153–168
- Kromrey, Helmut (1994b): Strategien des Informationsmanagements in der Sozialforschung. Ein Vergleich quantitativer und qualitativer Ansätze. In: Angewandte Sozialforschung 18/3, S. 163–184

- Kromrey, Helmut (2001): Evaluation – ein vielschichtiges Konzept. Begriff und Methodik von Evaluierung und Evaluationsforschung. Empfehlungen für die Praxis. In: Sozialwissenschaften und Berufspraxis, 24. Jg., H. 2, S. 105–131
- Kromrey, Helmut (2005): Evaluation – ein Überblick. In: Schöch, Heidrun (Hrsg.): Was ist Qualität? Die Entzauberung eines Mythos, Berlin 2005: Wissenschaftl. Verlag (Schriftenreihe Wandel und Kontinuität in Organisationen, Band 6), S. 31–85
- Kromrey, Helmut (2006): Empirische Sozialforschung. Modelle und Methoden der standardisierten Datenerhebung und Datenauswertung, 11. Aufl., Stuttgart: Lucius&Lucius, utb 1040
- Lamnek, Siegfried (1995): Qualitative Sozialforschung. 3. Aufl., München/Weinheim (Band 1: Methodologie; Band 2: Methoden und Techniken)
- Lüders, Christian (2006): Qualitative Evaluationsforschung – Was heißt hier Forschung? In: Flick, Uwe: Qualitative Evaluationsforschung. Konzepte, Methoden, Umsetzungen. Reinbek bei Hamburg: Rowohlt, S. 33–62
- Mayntz, Renate (1980): Die Entwicklung des analytischen Paradigmas der Implementationsforschung. In: dies. (Hrsg.): Implementation politischer Programme, Königstein/Ts., S. 1–17
- Nordverbund (Hrsg.) (2004): 10 Jahre Evaluation von Studium und Lehre. Verbund Norddeutscher Universitäten. Verbund Materialien Band 16, Hamburg
- Schlick, Moritz (1918): Allgemeine Erkenntnislehre (2. Aufl. 1925), unveränd. Nachdruck Frankfurt: Suhrkamp, 1978, S. 110–111
- Schmidt, J. (1980): Evaluation. I. Evaluation als Diagnose, Essen
- Sjoberg, Gideon (1983): Politics, Ethics and Evaluation Research. In: Struening, Elmer L./ Brewer, M.B. (eds.): Handbook of Evaluation Research, Beverly Hills, London, S. 65–88
- Werner, Rudolf (1975): Soziale Indikatoren und politische Planung. Einführung in Anwendungen der Makrosoziologie, Reinbek: rororo

## Evaluation in wenig formalisierten pädagogischen Settings

### Erste Annäherungen

Zugegeben: Die Überschrift dieses Beitrages klingt nicht gerade elegant, reichlich unbestimmt und in jedem Fall gewöhnungsbedürftig. Und doch, sie trifft das Problem. Die leitende Annahme dabei ist, dass es eine Reihe von pädagogischen Handlungsfeldern gibt, die durch eine dominierende Form des pädagogischen Settings geprägt sind und dass es daneben andere pädagogische Felder gibt, bei denen dies – aus welchen Gründen auch immer – nicht der Fall ist. Will man sich aber letzteren evaluativ zuwenden, sieht man sich neben den ohnehin schon bekannten Klippen bei Evaluationsstudien unweigerlich mit weiteren Herausforderungen konfrontiert. Um diese These zu begründen, wenden wir uns in einem ersten Schritt der Frage zu, womit es Programmevaluationen – und darauf sollen sich die folgenden Überlegungen konzentrieren – im Kern zu tun haben. Dabei sollte deutlich werden, dass neben dem Evaluationsbegriff, mit dem sich üblicherweise und sehr zu Recht die einschlägige Diskussion beschäftigt, sich bei näherer Hinsicht auch der Programmbegriff als relativ voraussetzungsvoll erweist. In einem zweiten Schritt möchten wir zeigen, dass gerade in pädagogischen Feldern diese Voraussetzungen oft nicht gegeben sind. Wir greifen dazu auf die Idee der pädagogischen Form zurück. In den daran anschließenden Überlegungen diskutieren wir einige daraus resultierende Anforderungen für Programmevaluationen, bevor in dem abschließenden vierten Schritt ein möglicher Ausweg, aber auch damit einhergehende neue methodologische Herausforderungen vorgestellt werden.

### Programme als Gegenstand von Evaluationen

Unter Evaluation verstehen wir – im Anschluss an M. Scrivens Betonung der Bewertungsaufgabe von Evaluation (Scriven 1991, S. 139) – die forschungsbasierte Bestimmung des Wertes eines Gegenstandes (vgl. Haubrich/Lüders 2004, S. 318ff.). Programmevaluation bezeichnet dementsprechend alle empirischen Verfahren zur Bestimmung des Wertes eines Programms. Wenn man – z. B. in Anlehnung an J. M. Owen und P. J. Rogers – unter Programmen ausgewählte Aktivitäten in einer bestimmten Abfolge versteht, die das Ziel verfolgen, erwünschte Effekte auf Seiten der Beteiligten – vorsichtig formuliert – anzuregen<sup>1</sup>, dann wird sichtbar, dass mit Programm je nach Praxis- und Diskursfeld zumindest auf der semantischen Ebene ganz Unterschiedliches gemeint sein kann: Methoden, Arbeitsformen, Didaktiken, Strategien, Ansätze, Verfahren, Instrumente, Techniken, Therapien u. ä.. Ohne hier auf die Details näher einzugehen, kann man festhalten, dass all diesen

---

<sup>1</sup> Im Original: „In other words, selected activities, presented in a given sequence, should produce desired effects on the participants“ (Owen/Rogers 1999, S. 24).

Begriffen gemeinsam ist, dass sie eine zunächst vergleichsweise personenunabhängige Ebene fachlichen Handelns in den Blick nehmen. Thematisiert wird mit diesen Begriffen das „Handwerkszeug“ fachlich kompetenten Handelns, das man in der Ausbildung, z. B. in Veranstaltungen zur Methodenlehre, in Praxiswerkstätten und Praktika einüben und lernen kann und das den Fachkräften zunächst unabhängig von der individuellen Fallkonstellation bzw. der konkreten Situation zur Verfügung steht, das aber immer auch dem jeweiligen Kontext mehr oder weniger angepasst werden muss. Es ist nur eine Frage der in den Praxisfeldern jeweils dominierenden Semantiken und der mit ihnen verbundenen Differenzen sowie der Abstraktionsebene, welche Begrifflichkeit jeweils zu bevorzugen wäre.

Um Programme evaluieren zu können, müssen diese identifizierbar sein. Es muss ein Wissen darüber vorhanden sein, mit Hilfe welcher Aktivitäten welche Ziele vor dem Hintergrund welcher Rahmenbedingungen und unter Nutzung welcher Ressourcen auch immer in einem überschaubaren Zeitraum erreicht werden sollen. Etwas konzeptioneller formuliert: Es bedarf eines fachlichen Konzeptes bzw. in der Sprache der Evaluationsdiskussion formuliert: einer Programmtheorie. Das Konzept bzw. die Programmtheorie formulieren nicht nur das fachlich-konzeptionell, strategisch begründete Zusammenwirken von jeweils zu berücksichtigenden Ausgangs- und Rahmenbedingungen, notwendigen Ressourcen und den jeweiligen Arbeitsschritten, um die Programmziele zu erreichen; darüber hinaus formulieren sie anzustrebende Ziele. Programmtheorien sind also in zweifacher Hinsicht besondere Theorien. Es handelt sich unvermeidlich um normative Theorien, die auf „umzu“-Motiven beruhen (vgl. zuerst Haubrich 2006; sowie Haubrich in Vorbereitung) und es handelt sich insofern um empirisch unterfütterte Theorien, als sie beanspruchen, die modellhafte „Logik“ einer realen Praxis zu formulieren. Dabei ist klar, dass Programmtheorien nicht 1:1 in der Praxis umgesetzt werden. Eher muss man wohl in Dimensionen von zugrundeliegenden Regelstrukturen denken, die sich in vielfältigen Konkretionen manifestieren<sup>2</sup>.

## Pädagogische Formen

Derartige Regelstrukturen haben – wie bereits angedeutet – in der pädagogischen Praxis unterschiedliche Namen. Neben den oben bereits erwähnten, die man alle auch in pädagogischen Feldern antreffen kann, greift die pädagogische Tradition immer wieder gerne zu Metaphern aus dem Bereich der bildenden und darstellenden Kunst, des Handwerks, der

---

2 Man mag angesichts dieser Überlegungen Zweifel bekommen, ob der Begriff Programmtheorie glücklich gewählt ist, da er Assoziationen an den in den Sozialwissenschaften vertrauten Theoriebegriff weckt, aber zugleich nur wenig mit ihm gemein hat. Die Tatsache, dass sich dieser Begriff weithin eingebürgert hat, tröstet dabei nur wenig. Es ist an der Zeit, an dieser Stelle genauer hinzusehen (vgl. Haubrich in Vorbereitung). Es sei angemerkt, dass mit dem hier skizzierten Verständnis von Programmevaluation ein Typus in den Vordergrund gerückt wird, den Beywl u.a. als „Programmtheorie-gesteuerte Evaluation“ bezeichnet haben und den sie gegenüber insgesamt 12 „Modellen von Evaluation“ abgrenzen (Beywl u.a. 2004, S. 65ff.; ebd., S. 83). Zugleich gilt, um die eigene Argumentation nicht zu sehr zu bescheiden, dass es auch für andere Evaluationsmodelle in Abhängigkeit von der Fragestellung nützlich sein könnte, über eine gute Programmtheorie zu verfügen (ebd., S. 34).

Rhetorik sowie aus anderen Lebensbereichen zurück<sup>3</sup>. Zum Teil sind die Begriffe eng mit bestimmten Handlungsfeldern verbunden, wie z. B. Didaktik mit Unterricht und Schule; andere Begriffe werden davon unabhängig verwendet. Zugleich sind die Begriffe nicht völlig beliebig. Sie transportieren Differenzen, z. B. hinsichtlich des für pädagogisches Handeln wichtigen Aspekts der Standardisierung des Vorgehens vs. des individuellen Fallbezugs.

Es ist hier nicht der Ort, diese Begriffe im Einzelnen zu diskutieren und zu systematisieren. Für unsere Zwecke reicht es, wenn man eine abstrakte Ebene erklimmt, um ihre Gemeinsamkeiten in den Blick zu nehmen. Wir greifen dabei auf die Überlegungen von K. Prange und G. Strobel-Eisele (2006) zurück, die an dieser Stelle den Begriff der pädagogischen Form verwenden. „Form ist das, was sich bei aller Verschiedenheit nach Alter, Geschlecht, sozialer Schicht, Zeitlage usw. als dasjenige identifizieren lässt, was immer als Gemeinsames gegeben sein muss, damit man von Erziehung sprechen kann“ (ebd., S. 38). Aus dieser Sicht sind Formen pädagogischen Handelns „ein Ergebnis der Methodisierung der Erfahrung“ (ebd., S. 26). Elementare Formen wären z. B. das Zeigen, die Übung, die Darstellung, die Aufforderung oder die Rückmeldung. Arrangement, Spiel, Arbeit, Erlebnis, Strafe sind Beispiele für komplexe Formen und die Schulerziehung wäre ein Beispiel für institutionalisierte Formen (vgl. Prange/Strobel-Eisele 2006). Pädagogische Formen sind dabei immer auf die Ermöglichung von Erziehung, Lernen und Bildung fokussiert. Das unterscheidet sie von anderen Formen. Zwar spielen Persönlichkeiten, individuelle Haltungen, situative Rahmenbedingungen, kulturelle und soziale Hintergründe und vieles andere immer auch eine nicht zu vernachlässigende Rolle; Unterricht oder z. B. Beratung kommen jedoch nur dann zustande, wenn die Beteiligten der jeweiligen Form entsprechend agieren und entsprechend gegenseitig kalkulierbar handeln. Würden die Beteiligten nicht im Großen und Ganzen formentsprechend handeln, hätte man es mit klassischen Krisenexperimenten im Sinne H. Garfinkels zu tun, aber nicht mehr mit pädagogischen Settings – was nicht bedeutet, dass man nicht auch aus Krisenexperimenten lernen könnte.

Um Missverständnisse zu vermeiden sei noch angemerkt, dass entgegen dem im Deutschen schnell aufkeimenden statischen Beigeschmack des Formbegriffs, dieser in dem hier verwendeten Sinne Strukturen, Verfahren bzw. Prozesse, die ihnen innewohnenden Absichten bzw. Ziele und die darin eingebetteten bzw. handelnden Akteure umfasst. Derartige Formen pädagogischen Handelns müssen von den Beteiligten immer wieder situativ reproduziert, genauer neu konstituiert werden. Ganz in diesem Sinne kann man z. B. bei den Ethnomethodologen und den Konversationsanalytikern lernen, wie derartige Formen durch die Alltagspraxis der Beteiligten immer wieder reproduziert und aufrechterhalten werden (vgl. z. B. Mehan 1979). Und von der Analyse kommunikativer Gattungen könnte man – würden sie sich mehr für pädagogische Formen interessieren – erfahren, wie diese im Alltag kommunikativ realisiert werden (vgl. Bergmann/Luckmann 1999 a und 1999 b; vgl. hierzu ausführlich auch Kade/Seitter 2007 a und 2007 b).

---

3 Gesprochen wird von der vollendeten Kunst der Erziehung (J. F. Herbart), von Bauformen (K. Prange), Skripten, Dramaturgien, Drehbüchern, Regieplänen u.Ä.

Die pädagogische Form ist dabei insofern eine eigentümliche Form, weil ihr auf der einen Seite eine gewisse methodische Komponente innewohnt, sie sich aber andererseits nicht darauf verlassen darf. H.-E. Tenorth hat vor diesem Hintergrund pädagogische Formen erneut kürzlich als „paradoxe Technologien“ bezeichnet, „weil sie [die pädagogische Form, Ch. L./K. H.] angesichts der Struktur von Unterricht und Lernen ganz besondere Probleme zu lösen hat: das Nicht-Planbare zu planen, einen festen Rahmen für offene Ergebnisse zu geben, mit der Alltäglichkeit von Überraschungen zu rechnen und das, Überraschungsfähigkeit, zur Routine werden zu lassen“ (Tenorth 2006, S. 588; vgl. auch Tenorth 1992, 1999).

Immerhin erweist sich Praxis als durchaus erfinderisch bei der Bewältigung derartiger Paradoxien. Und wenn man der Frage nachgeht, wie es Praxis gelingt, trotz allem Nicht-planbaren eine Form zu gewinnen, stößt man auf einige zentrale Konstruktionsprinzipien (vgl. Tenorth 2006, S. 588 bzw. Tenorth 1999, S. 262 ff.). Dazu gehört, dass die Praxis in ihrem Alltag eine eigene Form ausdifferenzieren muss und dass dabei die Lerngelegenheiten in zeitlicher, sachlicher und sozialer Hinsicht geordnet werden müssen. Es bedarf eigener, den Prozessen angemessener Gütekriterien und man „muss den derart strukturierten Prozess begleiten, von Zäsur zu Zäsur, die der Prozess selbst erzeugt, und zwar so, dass man offen bleibt für Neues, ohne von Überraschungen und Enttäuschungen aus der Bahn geworfen zu werden“ (Tenorth 2006, S. 588). Schließlich benötigt man einen Modus der Strukturierung des Prozesses, der weniger auf Personen als vielmehr auf Strukturen bzw. Organisationen abhebt.

Es ist vermutlich kein Zufall, dass H.-E. Tenorth immer wieder auf Unterricht und die praktischen Erfahrungen der Lehrkräfte zurückgreift, um seine Thesen zu plausibilisieren. Diese Praxis ist nicht nur in vielfältiger Weise umstritten, was dann gelegentlich Anlass zum Widerspruch erzeugt (vgl. Tenorth 2004, 2006), sondern wie wohl kaum eine zweite durch die institutionelle Rahmung der Schule, die Organisation und die Rhythmisierung des Unterrichts sowie durch zahlreiche Didaktiken vergleichsweise hochgradig formalisiert. Und umgekehrt mag man es wohl auch kaum als Zufall abtun, dass sowohl bei K. Prange und G. Strobel-Eisele als auch bei H.-E. Tenorth z. B. das weite Feld der außerschulischen Angebote der Jugendhilfe bzw. der Sozialpädagogik nicht einmal als Beispiele genannt werden. Ein Grund dafür wäre nahe liegend: Sie sind deutlich weniger formalisiert. Seinen theoretischen Ausdruck und diskursiven Widerhall findet dies in den zahlreichen Debatten der letzten Jahrzehnte, die die außerschulischen erziehungswissenschaftlichen Teildisziplinen, allen voran die Sozialpädagogik, umgetrieben haben: Sowohl die Professionalisierungs- und Handlungskompetenzdebatte wie auch die Suche nach den zentralen Codes des Systems Soziale Arbeit, die Diskussionen um Fallbezug vs. Methodisierung- bzw. Standardisierbarkeit, die jüngeren Versuche, eine rekonstruktive Sozialarbeit zu etablieren und vieles andere lässt sich vor diesem Hintergrund auch als Suche nach der eigenen Form begreifen.

## Herausforderungen für Programmevaluationen

Evaluation ist schon in sich ein in mehrfacher Hinsicht heikles Geschäft. Dies gilt auch, wenn man das Feld auf den Bereich externe Programmevaluation einschränkt. Die umfangreiche einschlägige und immer noch anschwellende Literatur gibt über die Voraussetzungen, Hürden und möglichen Lösungswege ebenso beredte Auskunft wie die Bemühungen sachgerechte Standards, um wenigstens die größten Klippen zu umschiffen, und Gütekriterien zu formulieren. Nicht zuletzt dokumentieren Forschungsberichte und Erfahrungsberichte aus der Praxis, die mit diesem Typus einhergehenden Herausforderungen. Es ist hier nicht der Ort, diese im Detail aufzulisten und zu diskutieren.

Wichtiger ist hier eine andere Überlegung. Es geht um die Evaluationen pädagogischer Programme oder etwas konkreter und im Anschluss an die vorangegangenen Überlegungen, um die Evaluation unterschiedlich komplexer Formen pädagogischen Handelns. Wenn man akzeptiert, dass es in der pädagogischen Fachpraxis offenbar ein Spektrum gibt von einerseits vergleichsweise formalisierten Settings, die gesättigt von Erfahrung identifizier- und beschreibbare Formen des fachlichen Handelns ausbilden konnten, bis hin zu wenig formalisierten, eher fluiden Settings andererseits, bei denen man nicht so genau weiß, welche Momente für die Erreichung der anvisierten Ziele und der zugrundeliegenden Absichten ausschlaggebend sind, dann haben Evaluationen von Fachpraxen, die nahe am zweiten Pol liegen, mit erheblichen zusätzlichen Problemen zu kämpfen.

Exemplarisch kann man dies an einem in der Diskussion zur Evaluation pädagogischer Praxis immer wieder auftauchenden Thema verdeutlichen, der Frage nach dem „richtigen“ Zeitpunkt der evaluativen Beobachtung. Dabei wird mit Bezug auf Prozesse lebenslangen Lernens und non-formaler bzw. informeller Bildung darauf verwiesen, dass man im Prinzip immer und überall lernen kann und man deshalb nie wisse, ob und wann pädagogische Angebote auf Seiten der Subjekte zu nachhaltigen Veränderungen führen würden und ob nicht ganz andere Anlässe zu Lern- und Bildungsprozessen führen könnten. Garniert werden diese Einwände üblicherweise mit Geschichten aus dem Leben, denen gemeinsam ist, dass in pädagogischen Settings ermöglichte Einsichten lange nach Beendigung der Angebote zu Veränderungen geführt haben und die zeigen, welche vielfältigen Bildungsprozesse außerhalb, manchmal auch trotz der pädagogischen Settings möglich sind. Systematisch gesprochen akzentuieren diese Argumentationen jene Momente pädagogischen Handelns, die sich – im Anschluss an das obige Zitat von H.-E. Tenorth – immer wieder in positiver wie in negativer Hinsicht als nicht planbar und überraschend herausstellen. Abgesehen davon, dass diese in der Sache selbst begründeten Argumente, wenn man sie verabsolutiert, dazu führen, dass absurderweise Evaluationen pädagogischer Praxis entweder nicht mehr oder nur noch post mortem denkbar sind, verweisen sie auch auf die ungeklärte Bedeutung der zeitlichen Dimension in wenig formalisierten pädagogischen Settings (zur Periodisierung pädagogischen Handelns vgl. Luhmann/Schorr 1990). Umgekehrt bedeutet dies, dass Formen pädagogischen Handelns genau genommen nur dann zum richtigen Zeitpunkt evaluiert werden können, wenn sie für sich geklärt haben, wann redlicherweise von beobacht-

baren und ihnen zurechenbaren Ergebnissen ausgegangen werden kann. Wo dies nicht geschieht, wird das Tor der beliebigen ansetzbaren, aber niemals befriedigenden Zeitpunkte für Evaluation geöffnet.

Nun gibt es aber ganz offensichtlich diese wenig formalisierten pädagogischen Settings. Dazu gehören neben den bewusst auf Innovation und Erprobung angelegten Settings, die z. B. im Kontext von Modellprogrammen oder in Phasen der Organisationsentwicklung und der Umstrukturierung darauf abzielen, neue Formen auszubilden, z. B. auch weite Teile der offenen Jugendarbeit. Das hat zunächst wenig – wie immer wieder unterstellt – mit nicht erfolgter Professionalisierung oder anderen Defiziten zu tun, sondern liegt zunächst in der Struktur des Feldes selbst begründet. Die starke Betonung des Lebensweltbezuges und der Beteiligung als zentrale fachliche Standards wirken nicht gerade dauerhaft formbildend und zwingen die Fachkräfte geradezu, vorhandene Formen immer wieder in Frage zu stellen.

Dabei ist es zunächst gleichgültig, ob die Evaluation vorrangig struktur-, prozess- oder ergebnisorientiert angelegt ist bzw. es sich – um eine Unterscheidung von J. M. Owen und P. J. Rogers aufzunehmen – eher um eine proaktive (proactive), eine klärende (clarificative), eine interaktive (interactive), eine dokumentierende (monitoring) oder eine wirkungsfeststellende (impact) Evaluation handelt (Owen/Rogers 1999, S. 39ff.; Beywl u.a. 2004, S. 39ff.). Auch im Fall – um nur ein Beispiel zu nennen – von strukturbezogenen Evaluationen von Angeboten offener Jugendarbeit bedarf es eines konzeptionellen Rahmens, mit dessen Hilfe entschieden werden kann, welche Strukturen in welcher Qualität für das jeweilige pädagogische Setting von zentraler Bedeutung sind und welche Aspekte dabei eine eher randständige Rolle spielen. Wenn aber in der jüngeren Debatte zur Jugendarbeit wiederholt die Strukturmomente Freiwilligkeit, Offenheit und Diskursivität als zentrale Strukturbedingungen von Jugendarbeit vorgestellt werden (Sturzenhecker 2004)<sup>4</sup> und Jugendarbeit zugleich als „organisierte Anarchie“ (ebd., S.446) beschrieben wird, kann man erahnen, vor welchen Problemen Evaluationen in diesem Feld stehen. Dies gilt umso mehr, wenn in dem gleichen Text eine Seite weiter die „Ausbildung eines Sets von Handlungskompetenzen, die man sich aneignen muss, wenn man unter den Bedingungen der Institution handeln will“ als das „geheime Curriculum“ (ebd., S. 447) der Strukturmerkmale dieses Feldes bezeichnet wird. Dies hat zur Folge, dass unsere fiktive Evaluation nicht mehr unterscheiden kann zwischen funktional offenbar notwendigen Strukturmerkmalen und unverzichtbaren Handlungskompetenzen des Fachpersonals.

Im Grundsatz trifft diese Schwierigkeit auch für die pädagogische Rechtsextremismusprävention zu: Zwar haben wir es in diesem Handlungsfeld häufig mit Bildungs- und Beratungsformaten zu tun, die stärker strukturiert sind als die sehr unstrukturierte offene Jugendarbeit. Doch spielen hier Lernziele eine große Rolle, die neben der Vermittlung bestimmter Wissensbestände vor allem die Stärkung emotionaler und handlungsbezogener

<sup>4</sup> Nur zum Vergleich: W. Thole (2000, S. 260) nennt folgende Strukturmerkmale: Freiwilligkeit, Partizipation, Integration, Lebensweltorientierung, Gruppenorientierung, Biografieorientierung, Zeitorientierung, Alltagsorientierung, Förderung von Selbstwertkompetenzen, Regionalisierung und Dezentralisierung.

Kompetenzen anstreben. Deshalb kommen in diesem Lernfeld auch verstärkt Lernarrangements zum Einsatz, die solche emotionalen und handlungsorientierten Lernebenen einbeziehen und im Vergleich etwa zum schulischen Lernen deutlich weniger formalisiert und sehr viel ergebnisoffener angelegt sind (vgl. Glaser 2007). Bei Angeboten, die sich direkt an rechtsorientierte Jugendliche wenden, dominieren wiederum Angebotsformen, die sich an den Grundsätzen der mobilen und offenen Jugendarbeit orientieren und in ihren Arbeitsabläufen entsprechend gering strukturiert sind.

Es braucht nicht lange ausgeführt werden, dass die damit verbundenen Schwierigkeiten in sehr viel stärkerem Maße erst recht für alle Varianten der Wirkungsanalyse gelten. Dies macht beispielsweise die jüngst veröffentlichte Metanalyse ausgewählter Studien aus dem Bereich erzieherischer Hilfen von Th. Gabriel, S. Keller und T. Studer (2007) deutlich. Ausgewertet wurden die 11 Studien unter der Perspektive der Wirkungen erzieherischer Hilfen. Der vergleichende Blick belegt, dass die Fachdiskussion weit davon entfernt ist, ein empirisch belastbares Wissen darüber zu besitzen, welche Struktur- und Prozessmerkmale auch nur mittelfristig erfolgversprechend sein dürften. Die von Th. Gabriel u.a. mühsam herausdestillierten Prozess- und Strukturmerkmale wie die Qualität der Hilfeplanung, die Dauer der Hilfestellung, die Kontinuität der sozialen Bezüge und der Grad der Partizipation der jungen Menschen und der Eltern, die Stabilität der Platzierung und die Qualität der sozialen Bezüge in den Einrichtungen und der Betreuung, die therapeutische und klinische Professionalität, der Stellenwert der Elternarbeit, die Dauer der Legalbewährung, die schulische und Ausbildungssituation der Jugendlichen und die Offenheit der Einrichtung gegenüber ihrem Umfeld sind auf der einen Seite nicht sonderlich überraschend und auf der anderen Seite noch lange nicht so belastbar und differenziert, dass Effekte begründet zugerechnet werden können (vgl. Gabriel u.a. 2007, S. 29ff.).

## Auswege

Will man an dieser Stelle nicht die Segel streichen, bedarf es alternativer Strategien. Denn wenn die These zutrifft, dass auch derart wenig bzw. fluide formalisierte Praxisfelder nur dann angemessen im Sinne von pädagogischen Formen evaluiert werden können, wenn man über eine entsprechende empirisch verlässliche „Theorie“ des Programms bzw. der Form verfügt, bedeutet dies, dass diese, wenn sie nicht vorliegt, vor Beginn der Evaluation oder als ein Schritt im Evaluationsdesign unvermeidlich selbst erzeugt werden muss. Dabei fällt eine zunächst nahe liegende Option gleichsam von selbst aus: Die Flucht in die Theorie. Es liegt nicht in der Sache selbst begründet; vielmehr lassen die zuvor erwähnten Hinweise über den Stand der Diskussion und die vielfältigen Suchbewegungen zusammen mit dem wenig elaborierten Forschungsstand alle Bemühungen auf eine theoretische Antwort schnell scheitern. Es gibt sie in den meisten Fällen nicht.

Bliebe also nur der empirische Zugang. Sein Gewinn könnte sich aus der keineswegs unplausiblen Annahme ergeben, dass die Praxis mitunter mehr über sich selbst weiß als ihr häufig bewusst ist. Weil man schlecht leugnen kann, dass auch in den wenig formalisierten

Praxisfeldern nicht nur beschreibbare, sondern von erfahrenen Fachkräften auch ausreichend kalkulierbare Ergebnisse erzielt werden, darf man davon ausgehen, dass es auf Seiten der Praxis auch in diesen Feldern gleichsam trotz aller Offenheit, Dynamik, Schwierigkeiten und Paradoxien ein tacit knowledge über die eigenen Gelingensbedingungen gibt, das bislang weder von Forschung noch von Seiten der Theorie angemessen rekonstruiert werden konnte<sup>5</sup>. In diesem Sinne könnte es viel versprechend sein, mit explorativen qualitativen Zugängen eben jenes implizite Wissen zu heben und zu systematisieren, um auf diesem Weg zu einer beschreibenden Annäherung an die jeweilige Form pädagogischen Handelns zu kommen.

Genau dies ist auch die Kernidee einer Reihe von Ansätzen, die sich in den letzten 15 Jahren vorrangig in der amerikanischen Diskussion zur Programmevaluation zum Teil unter unterschiedlichen Begrifflichkeiten herausgebildet haben. Unter Begriffen wie Logisches Modell (logic model), Theorie der Veränderung (theory of change), Ereignisketten (chain of events), Programmbaum (vgl. Beywl 2007, S. 33ff.); bzw. in den 70er Jahren dem sog. CIPP-Modell von D. Stufflebeam (1972 und 2003) wird der Versuch der vorgängigen Rekonstruktion bzw. Modellierung der gleichsam inneren Form des jeweils zu evaluierenden Settings unternommen<sup>6</sup>. Die leitende Frage, die diesen Rekonstruktionen zu Grunde liegt, lautet: Wie funktioniert das Programm? Oder in den Worten von J. M. Owen und P. J. Rogers: „What are the intended outcomes of this program and how is the program designed to achieve them? What are the underlying rationale for this program? What program elements or structures need to be modified to maximise program potential to achieve the intended outcomes? Is the program plausible“ (Owen/Rogers 1999, S. 43). Konzepte, wie die genannten, stellen hilfreiche aufmerksamkeitsleitende Heuristiken dar, an Hand derer die jeweilige Projektpraxis rekonstruiert und in einen „logischen“ Zusammenhang gebracht werden kann<sup>7</sup>.

Wie angedeutet, soll in diesem Beitrag der Blick auf die empirische, wissensgenerierende Beantwortung der Fragen gelenkt werden, was wiederum bedeutet, dass die Antwort,

---

5 Dies provoziert unweigerlich die Nachfrage, wie diese Diskrepanz zu erklären sei. Neben dem Verweis auf die vergleichsweise kurze Geschichte dieser Praxisfelder in ihrer heutigen Form, dem generellen, wiederholt notierten Problem, dass es sich bei „Erziehung um ein noch wenig verstandenes Phänomen“ handelt, das „begrifflich nicht konsolidiert ist“ (Herzog 2002, S. 65), spielt sicherlich die „Last der großen Hoffnungen“ (Müller 1985) sowohl von Seiten der Fachpraxis selbst wie auch der Beteiligten und der Öffentlichkeit eine nicht zu unterschätzende Rolle. Angemerkt sei, dass wir die Erinnerung an die Studie von B. Müller in diesem Kontext einer Diskussionsbemerkung von Th. Gabriel auf einem DJI-Workshop im Mai 2007 in München verdanken.

6 Vergleiche auch [www.uwex.edu/ces/pdande/evaluation/evallogicmodel.html](http://www.uwex.edu/ces/pdande/evaluation/evallogicmodel.html) und [www.uwex.edu/ces/pdande/evaluation/evallogicbiblio.html](http://www.uwex.edu/ces/pdande/evaluation/evallogicbiblio.html)

7 In Bezug auf die Art und Weise, wie dabei Antworten gefunden werden und welche Art von Antworten gesucht werden, gibt es allerdings erhebliche Differenzen (vgl. Rogers u.a. 2000). Zentral sind diesbezüglich einerseits der Stellenwert wissenschaftlicher Theorien (vgl. Chen 1990; Haubrich in Vorbereitung) und andererseits die Erwartungen und Zwecke, wobei im amerikanischen Sprachraum der Aspekt der Programmmodellierung auch zur späteren Planung und Steuerung dominiert (z. B. die Beiträge in der Zeitschrift *Evaluation and Program Planning*, vgl. [www.elsevier.com/locate/evalprogplan](http://www.elsevier.com/locate/evalprogplan)).

also die „Programmtheorie“ keine wissenschaftliche Theorie im systematischen Sinne sein kann, sondern eher empirisch gesättigte, systematisierte und zu einer Form verdichtete Beschreibungen bzw. genauer: Rekonstruktionen umfasst (vgl. Haubrich in Vorbereitung). Gleichsam stellvertretend für die und mit der Praxis werden im Vorfeld oder zu Beginn der Evaluation auf der Basis empirischer Analysen die mehr oder weniger implizit enthaltenen Annahmen über die innere Form des Programms expliziert, um auf diese Weise eine Grundlage für die Evaluation zur Verfügung zu haben<sup>8</sup>.

Die bislang am DJI tastenden Versuche und dabei gemachten Erfahrungen mit diesem Vorgehen sind vielfältig (vgl. Haubrich/Frank 2000; Haubrich/Vossler 2001 sowie Klingelhöfer i. d. Bd.). Zunächst gilt es festzuhalten, dass das Verfahren für alle Beteiligten relativ ressourcenaufwändig ist und es nur dann erfolgversprechend ist, wenn die Fachkräfte aus der Praxis nicht nur irgendwie beteiligt, sondern zentral eingebunden sind. Dies setzt nicht nur die Bereitschaft voraus, die eigene Praxis reflexiv Externen zu beschreiben, sondern auch die mindestens zeitweilige Möglichkeit, auf Distanz zu dieser Praxis zu gehen. Zugleich belegen die Erfahrungen, dass die Praxis die Ergebnisse üblicherweise als sehr hilfreich und entlastend wahrnimmt. Der Grund hierfür ist das immer wieder auftretende Phänomen, dass im Prozess der Beschreibung und Rekonstruktion der eigenen Praxis es fast notwendigerweise zu einer Reduktion der „großen Hoffnungen“ auf realistischere erreichbare Ziele kommt. Aus der z. B. zunächst versprochenen Förderung demokratischer Gegenwelten im ländlichen ostdeutschen Raum wird dann das sehr viel kleinteiligere, dafür aber realistische und überprüfbare Erfolgskriterium der zeitweiligen, in diesem Rahmen aber möglichst kontinuierlichen Bindung der Jugendlichen an die Einrichtung und ihre Beteiligung an internen, demokratischen Programmentscheidungen, um auf diese Weise positive Erfahrungen mit dem zivilen Austragen und Bewältigen von Interessenskonflikten zu erwerben. Damit einher geht die nicht zu vernachlässigende Wahrnehmung der Fachkräfte, dass auf diese Weise die Evaluation etwas mit ihrer alltäglichen Realität zu tun hat und sich nicht über ihre Köpfe hinweg vollzieht.

## Methodische Herausforderungen

Den damit angedeuteten praktischen Gewinnen steht allerdings eine Reihe von offenen methodischen wie auch methodologischen, inhaltlichen Fragen gegenüber (vgl. zum Folgenden Haubrich in Vorbereitung). Man übertreibt nicht, wenn man zugibt, dass die

---

8 Spätestens an dieser Stelle dürfte die Frage auftauchen, ob diese Rekonstruktionen Teil der Evaluation sind, als Voraussetzungen derselben anzusehen sind oder selbst bereits eine Form der Evaluation sind. Angesichts des mittlerweile weitgehend entgrenzten Verständnisses von Evaluation mag man geneigt sein, die empirische Rekonstruktion der inneren Logik von Programmen selbst als Evaluation, z. B. als „clarificative evaluation“ (Rogers/Owen 1999, S. 42) zu bezeichnen. Demgegenüber tendieren wir dazu, um gleichsam nicht das „Kerngeschäft“ der Evaluation – das systematische und transparente Bewerten – zu vernebeln, die Differenz zwischen Evaluation und der empirischen Klärung ihrer Voraussetzungen im Sinne der Rekonstruktion von z. B. Logischen Modellen deutlich zu betonen, wenngleich diese Klärungsprozesse Teil eines Evaluationsdesigns sein können.

Empirie noch nicht ausgereift ist. Dies beginnt schon bei der Frage, was die empirische Grundlage für derartige Rekonstruktionen darstellt. Soweit zu sehen, geraten in den meisten Fällen an erster Stelle die unmittelbar agierenden Fachkräfte oder die für die Programmsteuerung relevanten Akteure in den Blick. Mit ihnen werden ausführliche Interviews oder noch weitergehend Workshops durchgeführt, mit dem Ziel einer gemeinsamen Explikation der mehr oder weniger impliziten Programmannahmen. Mitunter werden Dokumente zum Programm analysiert; schon teilnehmende Beobachtungen oder Ethnografien der Programmrealisierung sind in diesem Kontext kaum anzutreffen. Bei den Interviews handelt es sich meistens um eine Art von explorativen, gegenstandsbezogenen Expertinnen- und Experteninterviews. Auffallend ist, dass bislang die Adressatinnen und Adressaten keine systematische Rolle spielen.

Eine zweite Frage betrifft den Analysegegenstand. Auch wenn oben von *tacit knowledge* bzw. implizitem Wissen gesprochen wurde, so wäre es doch eine eigene Diskussion wert, zu klären, um welche Art von Wissen es sich handelt und was genau methodologisch in diesen Prozessen rekonstruiert wird. Der derzeit wohl weitestgehende Vorschlag in der deutschsprachigen Evaluationsdebatte stammt von R. Bohnsack und seinen Mitarbeiterinnen und Mitarbeitern. Vor dem Hintergrund seines Konzeptes der dokumentarischen Interpretation wurde dort ein Zugang der dokumentarischen Evaluationsforschung entwickelt, der auf die Rekonstruktion der impliziten, kollektiv geteilten, atheoretischen, handlungspraktischen, konjunktiven Wissensbestände abzielt (vgl. z. B. Nentwig-Gesemann u.a. 2007; Fritzsche/Bohnsack i. d. Bd.). Mit der zweiten Frage eng verbunden ist drittens der Aspekt der Auswertung. Während für das Konzept der dokumentarischen Evaluationsforschung ausführlich beschriebene Verfahren vorliegen, sucht man andererseits vergeblich nach entsprechenden Beschreibungen. Oftmals wird von Inhaltsanalysen gesprochen, doch ist dieser Begriff mittlerweile so weit entleert, dass sich darunter alles Mögliche verbergen kann. Viertens schließlich wäre die Frage der Gütekriterien zu beantworten. Virulent ist dabei nicht nur der Aspekt der empirischen Qualitätskriterien wie z. B. der Validität der Rekonstruktion; angesichts der zuvor schon erwähnten spezifischen paradoxen Struktur des Feldes stellt sich darüber hinaus auch die Frage nach der Form der Form, also den Konstruktionsprinzipien der Rekonstruktion und den damit einhergehenden Gütekriterien. Als Frage formuliert: Wie stringent und intern konsistent kann die Form einer Praxis rekonstruiert werden, von der man weiß, dass sie offen für Nichtplanbares und Überraschungen sein muss, die also unvermeidlich mit Kontingenzen zu tun hat? Denn letztlich unterliegt die Anwendung logischer Modelle, die als schematische Modellierungen der Wirklichkeit (Form der Form) immer zugleich die in aller Regel komplexen Programmstrukturen deutlich vereinfachen, einer gewissen Beschränkung: Modelle bilden immer nur Ausschnitte der Wirklichkeit unter einer bestimmten Perspektive ab (vgl. Haubrich 2006b, S. 119).

Neben diesen methodischen Aspekten muss schließlich die Frage nach dem Status der in den logischen Modellen rekonstruierten Wissensbestände auf der einen Seite für sich genommen und als Grundlage von Programmevaluationen auf der anderen Seite aufgeworfen

werden. Dahinter verbirgt sich die unvermeidliche Frage, ob und in welchem Umfang man den professionellen Selbstbeschreibungen trauen darf, und zwar in doppelter Hinsicht: Zum einen verfügen selbstverständlich auch Fachkräfte über „blinde Flecken“ und es gibt auch in professionellen Kontexten hartnäckige Routinen und Deutungsmuster, die sich bei kritischer Hinsicht eigentlich nicht halten lassen. Zum anderen stellt sich die Frage, was eine Evaluation, die sich auf die wie auch immer rekonstruierten professionellen Wissensbestände bezieht, mehr sein kann, als die Bestätigung eben dieser.

Beide kritischen Einwände sind ernst zu nehmen. Allein den explizierten Selbstbeschreibungen zu trauen, bedeutet in der Tat eine Engführung der Untersuchungsperspektive. Wie man allerdings implizite Wissensbestände erschließen kann, lässt sich letztlich nur anhand des konkreten empirischen Vorgehens in einer Evaluationsstudie, also der jeweiligen Erhebungs- und Auswertungsverfahren begründen. Hier fehlt es bislang – auch in der internationalen Debatte – an einschlägigen methodologisch reflektierten Zugängen und Vorbildern. Am DJI wurden hierzu in drei Studien, die Logische Modelle als Grundlage des Evaluationsdesigns gewählt haben (Haubrich/Frank 2000, Vossler/Obermaier 2003 und Klingelhöfer u. a. i. E. sowie Klingelhöfer i. d. Bd.), jeweils unterschiedliche Zugänge erprobt<sup>9</sup>, aus deren Vergleich man für die Weiterentwicklung der Fachdebatte im Sinne eines praxeologischen Vorgehens der Entwicklung qualitativer Methoden lernen kann.

Beide Fragen – ebenso wie die zuvor angesprochene Qualität der Modelle (Form der Form) – wären jedoch vor allem dann folgenreich, wenn die sich daran anschließende Programmevaluation in naiver, verdinglichender Weise auf diese in den Logischen Modellen abgebildeten Annahmen über das Programm stützen würde, also die Beschreibungen der Fachkräfte für gleichsam bare Münze nehmen würde. Demgegenüber ist es wichtig darauf hinzuweisen, dass die Rekonstruktionen der inneren Logik der Programme zunächst den Charakter einer vorläufigen Heuristik haben, die die Aufmerksamkeit der Evaluation lenkt, aber nicht fixiert. Methodisch bedeutet dies, dass man Logische Modelle, vor allem dann, wenn man sich über ihren Status noch nicht im Klaren ist, nur mit offenen im Sinne von nicht-standardisierten Verfahren evaluieren kann. Auf diese Weise kann der Blick offen gehalten werden für nicht-intendierte Folgen des Programms wie auch für Aspekte, die das logische Modell selbst erweitern, verändern bzw. Anlass geben, es ggf. gänzlich zu reformulieren. So hat die Evaluation durchaus die Möglichkeit, das vorliegende Logische Modell auf der Basis des verfügbaren Wissens vor dem Hintergrund des jeweiligen Programms zu erweitern bzw. zu verändern, um eine realistischere Modellierung der Programmrealität zu erreichen. Je nach Untersuchungsgegenstand, Fragestellung und methodischem Vorgehen sind in jedem Falle Schritte in die Erhebung einzubeziehen, die diesem modellhaften und in manchen Fällen (bspw. bei Modellvorhaben) vorläufigen Charakter Logischer Modelle

---

<sup>9</sup> Die zuvor aufgeworfenen Fragen sowie insbesondere auch die Frage nach dem Stellenwert von Selbstbeschreibungen und den theoretischen und methodologischen Grundlagen des Vorgehens in den Evaluationsstudien „Integration junger Ausländerinnen und Ausländer“ (Haubrich/Frank 2000) sowie „Interkulturelles Netzwerk der Jugendsozialarbeit im Sozialraum“ (Vossler/Obermaier 2003) werden ausführlich erörtert in Haubrich (in Vorbereitung).

Rechnung tragen und der Gefahr vorbeugen, dass Modelle überprüft werden, die die Realität der Programme verfehlen. Bei den am DJI durchgeführten Studien handelte es sich in allen Fällen um mehrjährige Begleitprojekte, in denen Logische Modelle über mehrere aufeinander folgenden Erhebungsphasen hinweg weiterentwickelt und zudem ergänzend weitere Verfahren zur Beschreibung der untersuchten Programme eingesetzt wurden (teilnehmende Beobachtung, Interviews mit Kooperationspartnern etc.). Zugegeben werden muss allerdings, dass mit der Entscheidung, Logische Modelle als Grundlage des Evaluationsdesigns einzusetzen, ein bestimmter Zugang der Beschreibung der Realität gewählt wird und damit manches in den Blick gerät und anderes ausgeblendet bleibt. Damit wird eine Beobachtungsperspektive gewählt, die aus der Fragestellung der Erhebung und den Besonderheiten des Untersuchungsgegenstandes abzuleiten ist und sicher nicht in allen Fällen für sich beanspruchen kann, den Königsweg darzustellen.

Wem – vor allem aus der Ecke der Evaluationsforschung kommend – dies alles sehr spezifisch für den pädagogischen Diskurs erscheinen mag, der trifft den zentralen Punkt unserer Überlegungen. Evaluation wie andere Forschung auch kann nicht umhin, den Spezifika ihres jeweiligen Gegenstandes gerecht zu werden und dies bedeutet im konkreten, sich auf die jeweiligen Strukturen und Prozesse – und seien sie noch so paradox und kontingent – konzeptionell und empirisch einzulassen<sup>10</sup>.

## Literaturverzeichnis

- Bergmann, Jörg/Luckmann, Thomas (Hrsg.) (1999a): Kommunikative Konstruktion von Moral. Bd. 1: Struktur und Dynamik der Formen moralischer Kommunikation. Opladen: Westdeutscher Verlag
- Bergmann, Jörg/Luckmann, Thomas (Hrsg.) (1999b): Kommunikative Konstruktion von Moral. Bd. 2: Von der Moral zu den Moralen. Opladen: Westdeutscher Verlag
- Beywl, Wolfgang/Speer, Sandra/Kehr, Jochen (2004): Wirkungsorientierte Evaluation im Rahmen der Armut- und Reichtumsberichterstattung. Eine Perspektivstudie. Im Auftrag des BMGS. Bonn (Zugriff 19.05.2007)
- Beywl, Wolfgang (2006): Demokratie braucht wirkungsorientierte Evaluation – Entwicklungspfade im Kontext der Kinder- und Jugendhilfe. In: Projekt eXe (Hrsg.): Wirkungsevaluation in der Kinder- und Jugendhilfe. Einblicke in die Evaluationspraxis. München: DJI, S. 25–46
- Chen, Huey-Tsyh (1990): Theory-driven Evaluations. Newbury Park, London, New Dehli, Sage

---

<sup>10</sup> Als Gegenbeispiel vgl. Kuper 2005.

- Gabriel, Thomas/Keller, Samuel/Studer, Tobias (2007): Wirkungen erzieherischer Hilfen – Metaanalyse ausgewählter Studien (Wirkungsorientierte Jugendhilfe, Bd. 3). Münster, ISA: [www.wirkungenzeigen.de/seiten/material/wojh\\_schriften\\_heft\\_3.pdf](http://www.wirkungenzeigen.de/seiten/material/wojh_schriften_heft_3.pdf) (03.06.2007)
- Glaser, Michaela (2007): Schulische und außerschulische Bildung gegen Rechtsextremismus. Stellungnahme im Rahmen der Anhörung im Innenausschuss des Hessischen Landtags zum Thema „Frühzeitige Bekämpfung von rechtsextremistischen Tendenzen bei Jugendlichen“, [www.juergen\\_froemrich.de/cms/files/dokbin/181/181821.stellungnahmen\\_anhhoerung\\_rechtsextremism.pdf](http://www.juergen_froemrich.de/cms/files/dokbin/181/181821.stellungnahmen_anhhoerung_rechtsextremism.pdf) (6.11.07)
- Haubrich, Karin (2006a): Evaluation von Reformen in der Kinder- und Jugendhilfe – praktische Probleme und methodologische Herausforderungen (Teil II). Vortrag im Rahmen des 23. wissenschaftlichen Kongresses der Deutschen Vereinigung für Politische Wissenschaften. Münster, 29.09.2006
- Haubrich, Karin (2006b): Wirkungsannahmen sichtbar machen: Cluster-Evaluation innovativer multizentrischer Programme. In: Projekt eXe (Hrsg.): Wirkungsevaluation in der Kinder- und Jugendhilfe – Einblicke in die Evaluationspraxis. München: DJI, S. 101–122
- Haubrich, Karin (in Vorbereitung): Evaluation innovativer multizentrischer Programme: Möglichkeiten und Grenzen programmtheoriebasierter Cluster-Evaluation am Beispiel von Bundesmodellprogrammen in der Kinder- und Jugendhilfe
- Haubrich, Karin/Frank, Kerstin (2000): Vom Aufsuchen zur beruflichen Integration. Evaluationsstudie zum Bundesmodellprogramm „Mobile Jugendsozialarbeit für junge Menschen ausländischer Herkunft“. DJI-Arbeitspapier Nr. 1–156. München
- Haubrich, Karin/Lüders, Christian (2004): Evaluation – mehr als ein Modewort. In: Recht der Jugend und des Bildungswesens H. 52, S. 316–337
- Herzog, Werner (2002): Zeitgemäße Erziehung. Die Konstruktion pädagogischer Wirklichkeit. Weilerswist, Velbrück
- Kade, Jochen/Seitter, Wolfgang (Hrsg.) (2007a): Umgang mit Wissen. Recherchen zur Empirie des Pädagogischen. Bd 1. Pädagogische Kommunikation. Opladen Farmington Hills: Barbara Budrich Verlag
- Kade, Jochen/Seitter, Wolfgang (Hrsg.) (2007b): Umgang mit Wissen. Recherchen zur Empirie des Pädagogischen. Bd. 2. Pädagogisches Wissen. Opladen Farmington Hills: Barbara Budrich Verlag
- Kuper, Harm (2005): Evaluation im Bildungssystem. Stuttgart, Kohlhammer
- Luhmann, Niklas/Schorr, Eberhard (Hrsg.) (1990): Zwischen Anfang und Ende. Fragen an die Pädagogik. Frankfurt (Main): Suhrkamp
- McLaughlin, John A./Jordan, Gretchen B (1999): Logic Models: A Tool for Telling Your Program's Performance Story. In: Evaluation and Program Planning. 22, S. 65–72.
- Mehan, Hugh (1979): Learning Lessons. Social Organization in the Classroom. Cambridge, Massachusetts, London, Harvard University Press
- Müller, Burkhard (1985): Die Last der großen Hoffnungen. Methodisches Handeln und Selbstkontrolle in sozialen Berufen. Weinheim, München: Juventa

- Nentwig-Gesemann, Iris/Fritzsche, Bettina/Bohnsack, Ralf (2007): Konflikte friedlich lösen. Dokumentarische Evaluationsforschung am Beispiel von zwei Peer-Mediationsprojekten. In: Schröder, Ute B./Streblow, Claudia (Hrsg.): Evaluation konkret. Fremd- und Selbstevaluationsansätze anhand von Beispielen aus Jugendarbeit und Schule. Opladen: Barbara Budrich Verlag, S. 65–105
- Owen, John M./Rogers, Patricia J.(1999): Program Evaluation. Forms and Approaches. London/Thousand Oaks/New Dehli: Sage
- Prange, Klaus/Strobel-Eisele, Gabriele (2006): Die Formen des pädagogischen Handelns. Eine Einführung. Stuttgart: Kohlhammer
- Rogers, Patricia J./Hacsi, Timothy A./Petrosino, Anthony/Huebner, Tracy A. (Hrsg.) (2000): Program Theory in Evaluation: Challenges and Opportunities. New Directions for Evaluation, No. 87. San Francisco, Josey-Bass
- Scriven, Michael (1991): Evaluation Thesaurus. Newbury Park/London/New Dehli: Sage
- Stufflebeam, Daniel L. (1972): Evaluation als Entscheidungshilfe. In: Wulf, Christoph (Hrsg.) Evaluation. Beschreibung und Bewertung von Unterricht, Curricula und Schulversuchen. München, Piper, S. 113–145
- Stufflebeam, Daniel L (2003).: The CIPP Model for Evaluation – an update. Annual Conference of the Oregon Program Evaluators Network. Portland [www.wmich.edu/evalctr/pubs/CIPP-ModelOregon10-03.pdf](http://www.wmich.edu/evalctr/pubs/CIPP-ModelOregon10-03.pdf) (Zugriff 05.05.2007)
- Sturzenhecker, Benedikt (2004): Strukturbedingungen von Jugendarbeit und ihre Funktionalität für Bildung. In: neue praxis H. 34, S. 444–454
- Tenorth, Heinz-Elmar (1992): Paradoxa, Widersprüche und die Aufklärungspädagogik. Versuch, die pädagogische Denkform vor ihren Kritikern zu bewahren. In: Oelkers, Jürgen (Hrsg.): Aufklärung, Öffentlichkeit und Bildung: pädagogische Beiträge zur Moderne (28. Beiheft der Zeitschrift für Pädagogik) Weinheim, Basel: Beltz, S. 117–134
- Tenorth, Heinz-Elmar (1999): Technologiedefizit in der Pädagogik? Zur Kritik eines Missverständnisses. In: Fuhr, Thomas/Schultheis, Klaudia (Hrsg.): Zur Sache der Pädagogik. Untersuchungen zum Gegenstand der allgemeinen Erziehungswissenschaft. Bad Heilbrunn: Klinkhardt, S. 252–266
- Tenorth, Heinz-Elmar (2004): Bildungsminimum und Lehrfunktion. Eine Apologie der Schulpflicht und eine Kritik der „therapie“-orientierten pädagogischen Professionstheorie. In: Gruehn, Sabine/Kluchert, Gerhard/Koinzer, Thomas (Hrsg.): Was Schule macht. Schule, Unterricht und Werteerziehung: theoretisch, historisch, empirisch. Achim Leschinsky zum 60. Geburtstag. Weinheim: Beltz/Deutscher Studienverlag, S. 15–29
- Tenorth, Heinz-Elmar (2006): Professionalität im Lehrerberuf. Ratlosigkeit der Theorie, gelingende Praxis. In: Zeitschrift für Erziehungswissenschaft H. 9, S. 580–597.
- Thole, Werner (2000): Kinder- und Jugendarbeit. Eine Einführung. Weinheim, München: Juventa 2000
- Vossler, Andreas/Obermaier, Andrea (2003): Netze knüpfen – Integration fördern. Evaluationsstudie zum Bundesmodellprogramm „Interkulturelles Netzwerk der Jugendsozialarbeit im Sozialraum“. München: DJI.

## Der soziale Faktor<sup>1</sup> – Ergebnisse einer explorativen Untersuchung zu sozialen Beziehungen in der Evaluation von Maßnahmen gegen Rechtsextremismus<sup>2</sup>

Die an einer Evaluation Beteiligten sind in ein Netz sozialer Beziehungen eingebunden, die die Voraussetzungen, Bedingungen und Ergebnisse der Evaluation maßgeblich prägen. Je nach Ziel sowie konzeptioneller und methodischer Anlage einer Evaluation wird diesen Beziehungen in der Fachliteratur unterschiedlicher Stellenwert zuerkannt.

Bei Evaluationen, die eher summativ angelegt sind (vgl. Scriven 1991), die methodisch eher quantitativ ausgerichtet sind und sich an wertneutralen Modellen orientieren, geht man tendenziell davon aus, dass sie außerhalb sozialer Interessenkonflikte durchgeführt werden (vgl. Beywl 2006, S. 97). Bei Evaluationskonzepten dieses Typs dominiert eine technokratische Betrachtungsweise, die Evaluation vor allem als messtechnischen Prozess betrachtet und soziale Prozesse ausblendet bzw. ihr Funktionieren voraussetzt. Bezugspunkte für Qualität sind hier vor allem die Kriterien wissenschaftlicher Forschung.

Wenn Evaluationen vor allem formativ angelegt sind (vgl. Scriven 1991), wenn sie in erster Linie qualitative Methoden verwenden und die Klärung von Bewertungsfragen als Teil des Evaluationsprozesses verstehen, wird den sozialen Beziehungen zwischen den Beteiligten erheblicher Stellenwert zuerkannt. Dies ist vor allem im Rahmen von Evaluationen der vierten Generation („fourth level evaluation“) relevant (vgl. Guba/Lincoln 1989, S. 21ff.), die die unterschiedlichen Sichtweisen verschiedener Beteiligter zum Ausgangspunkt ihrer Analysen machen. Einerseits bezieht sich dieses Verständnis auf die Perspektiven der Beteiligten, die gleichberechtigt neben denen der Evaluierenden und der Auftraggeber stehen sollen (vgl. Klawe 2006, S. 136). Andererseits betont man die Notwendigkeit, die Beteiligten bei der Gestaltung einer Evaluation einzubeziehen. So sollen z. B. bei einer nutzenfokussierten Evaluation die Beteiligten, die die Ergebnisse nutzen und umsetzen sollen, frühzeitig in die Entscheidungsprozesse zur Gestaltung der Evaluation einbezogen werden, wofür die Kommunikation zwischen Evaluierenden und Nutzerinnen und Nutzern entsprechend gestaltet werden muss (vgl. Patton 1997, S. 20ff.). Auch im Rahmen solcher Konzepte, mit denen partizipative Evaluationen angestrebt werden, betont man die Bedeutung der Mitwirkung der Beteiligten an einer Evaluation (vgl. Ulrich/Wenzel 2004, S. 35).

Die Analyse sozialer Beziehungen gewinnt auch dadurch an Bedeutung, dass Evaluationen als soziale Prozesse bzw. Prozesse sozialer Interaktion angesehen werden (vgl. Widmer 2004, S. 88), die vielfältigen Aushandlungsprozesse zwischen den Beteiligten und Betroffenen bein-

---

1 Der Titel dieses Beitrags spielt auf das Konzept des „persönlichen Faktors“ an, mit dem Michael Patton die Qualität der Beziehung zwischen Evaluierenden und zentralen Beteiligten bezeichnet, die er als ausschlaggebend für die Nutzung einer Evaluation erachtet (Patton 1997, S. 39ff.).

2 Konstruktive Kritik und hilfreiche Anmerkungen zu einer früheren Fassung dieses Beitrags verdanke ich Michaela Glaser, Karin Haubrich und Silke Schuster.

halten. Diese Aushandlungen können sich auf die Initiierung einer Evaluation, auf die Festlegung des Designs, die Steuerung und Begleitung der Evaluation sowie die Rückkopplung und Nutzung der Ergebnisse beziehen (vgl. Haubrich u.a. 2006, S. 12f.). Bei der Gestaltung und Steuerung einer Evaluation wird neben dem Auftraggeber auch dem Beirat oder der Projektgruppe eine zentrale Steuerungsfunktion zugesprochen, d.h. aus diesem Gremium kommen Vorschläge zur Gestaltung der Erhebungsinstrumente und hier werden Entscheidungen über die Schwerpunktsetzung einer Evaluation getroffen. Daneben gibt es eine Vielzahl Beteiligter und Betroffener, die eine periphere Funktion für die Steuerung eines Evaluationsvorhabens haben können (vgl. Erzberger 2006, S. 26f.). Eine ausgewogene und vertrauensvolle Kommunikation zwischen den Beteiligten gilt als Voraussetzung für eine erfolgreiche Evaluation. Die Evaluierenden haben im Rahmen dieser Kommunikations- und Aushandlungsprozesse vielfältige, zum Teil konträre Rollen (vgl. Wottawa/Thierau 2003, S. 46). Sie gelten entweder als Moderatorinnen und Moderatoren, die zwischen den Interessen und Perspektiven der Beteiligten vermitteln oder die Vorstellungen des Auftraggebers kanalisieren müssen. Oder ihnen werden – je nach Sichtweise – die Rollen von Forschungshandwerkern, die praktikable Umsetzungsvorschläge machen und umsetzen oder von „Resonanzkörpern“ bzw. „Erhebungsinstrumenten“ zuerkannt (vgl. Erzberger 2006, S. 22). In jedem Fall werden von ihnen hohe kommunikative und soziale Kompetenzen gefordert und sie sollen angemessen und flexibel auf die Einflusskräfte im Feld reagieren, z. B. um Stakeholder sowie potenzielle Nutzer einzubinden und Widerstände gegen eine Evaluation zu überwinden (vgl. Beywl 2006, S. 103ff.).

Außerdem werden Vorstellungen zu den sozialen Beziehungen zwischen Evaluation und den gesellschaftlichen Bereichen Politik/Verwaltung, Wissenschaft und Öffentlichkeit formuliert (Widmer 2004, S. 88ff.).

- Politik/Verwaltung seien häufig Auftraggeber, aber auch Nutzer von Evaluationen. Da Evaluatoren von Auftraggebern abhängig seien, wird dieses Verhältnis als Abhängigkeitsverhältnis beschrieben. Die Beziehung zwischen Evaluation und Politik/Verwaltung sei vor allem ökonomisch und juristisch geprägt.
- Die Wissenschaft könne von Evaluation profitieren, da diese die Wissenschaft legitimiere und empirische Evidenz ihrer Erkenntnisse produziere, während die Evaluation auf die Wissenschaft als Lieferantin von Wissensbeständen angewiesen sei. Die Beziehung zwischen Wissenschaft und Evaluation sei damit durch den gegenseitigen Austausch von Wissen und Reputation geprägt.
- Auf die Interaktion mit der Öffentlichkeit sei Evaluation zunächst nicht angewiesen, eine verantwortungsvolle Evaluation solle die Öffentlichkeit aber nicht nur als Nutzer, sondern auch als Diskurspartner ernst nehmen, der ein Gegengewicht zu Politik und Wissenschaft darstellt. Die soziale Beziehung zwischen Öffentlichkeit und Evaluation sei damit eine ethisch-moralische.

In Hinblick auf das Verhältnis zwischen Evaluation und Wissenschaft werden allerdings auch andere Vorstellungen vertreten. So findet sich die Position, Evaluierende unterschied-

den sich aufgrund ihrer Abhängigkeit von Auftraggebern grundlegend von Wissenschaftlerinnen und Wissenschaftlern, die unabhängig seien und nur nach Wahrheit suchten (vgl. Wottawa/Thierau 2003, S. 37). Selbst wenn man diesem idealisierten Bild von Wissenschaft kritisch gegenüber steht, kommt man nicht umhin, festzustellen, dass Evaluierende innerhalb des Wissenschaftssystems teilweise auf erhebliche Akzeptanzprobleme stoßen.

Konkretisierende Hinweise zu den sozialen Beziehungen zwischen den an einer Evaluation Beteiligten und zwischen Evaluation und anderen gesellschaftlichen Bereichen sowie zu den damit verbundenen Schwierigkeiten finden sich in Berichten zur Umsetzung von Evaluationen (vgl. z. B. die Fallbeispiele in Joint Committee/Sanders 1999). Dabei werden u.a. erhebliche Interessenkonflikte zwischen den Beteiligten deutlich, die von Evaluierenden moderiert werden müssen, was mit erheblichen Anforderungen verbunden sein kann. Notwendig kann es sein, mit Ängsten und Misstrauen der Beteiligten umzugehen, da eine Evaluation sich auf die Arbeitsbedingungen und die Machtverhältnisse in den jeweiligen Einrichtungen auswirken kann (vgl. Klawe 2006, S. 135; Mensching 2006, S. 341). In diesem Zusammenhang werden Hinweise für die konstruktive Gestaltung der Beziehungen zwischen Evaluierenden und Auftraggebern sowie zwischen Evaluierenden und Beteiligten gegeben: Ein Vertrauensverhältnis aufbauen, kontinuierliche Informationen über den Stand der Forschung und das Vorgehen der Evaluation geben, Stakeholder an der Planung beteiligen (Mensching 2006, S. 350ff.). In diesem Rahmen gelten auch Rückmeldungen an die Adressaten und die kommunikative Validierung von Ergebnissen durch diese als wichtig (vgl. Eichler/Merkens 2006, S. 317).

Auf besondere Weise stellen sich Fragen nach den sozialen Beziehungen im Rahmen von Evaluationen, die Programme und Projekte der Prävention gegen Rechtsextremismus und Fremdenfeindlichkeit untersuchen – was in ersten Publikationen zum Themenfeld bisher jedoch kaum Berücksichtigung findet (vgl. z. B. die Beiträge in Lynen von Berg/Roth 2003). Wir sind in diesem Arbeitsfeld mit affektiv und normativ hoch aufgeladenen Themen konfrontiert sowie zum Teil auch mit verhärteten Frontstellungen und Bewertungen. Wer in diesem Zusammenhang Informationen gibt oder vorenthält, positioniert sich in einem brisanten Feld und ist sich bewusst, dass dies mit erheblichen Konsequenzen für sein Sozialleben verbunden sein kann. Außerdem ist davon auszugehen, dass rechtsextreme und fremdenfeindliche Einstellungen im Rahmen einer Evaluation nicht immer offengelegt werden, da bekannt ist, dass entsprechende Haltungen oft nicht salonfähig sind (vgl. Widmer/Hirschi 2007, S. 262). Es zeigt sich sogar, dass moralische Bewertungen in diesem Themenfeld teilweise so dominant sind, dass es als unschicklich gilt, gut gemeintes zu kritisieren, wenn es schlecht gemacht wird (vgl. Lynen von Berg u.a. 2007). Schließlich ist zu beachten, dass Evaluationen hier zum Teil Folge hohen politischen Handlungsdrucks sind und unter permanenter Beobachtung durch Politik und Öffentlichkeit stehen, was mit unrealistischen Erfolgserwartungen verbunden sein kann (vgl. Frindte/Preiser 2007, S. 34 sowie Palloks i. d. Bd.). Dieser Druck und die daraus resultierenden Erwartungen können zu Aufträgen führen, die aus wissenschaftlicher Perspektive nicht sinnvoll und wünschenswert sind. Für die sozialen Beziehungen derjenigen, die mit Evaluationen auf diesem Feld befasst sind, ist dies mit besonderen Herausforderungen verbunden.

Um die Relevanz dieser sozialen Beziehungen einschätzen zu können und Hinweise auf die Bedingungen und Aspekte zu erhalten, die für ihre Gestaltung von Bedeutung sind, werden in diesem Beitrag die Erfahrungen von Evaluierenden ausgewertet. Auf dieser Grundlage sollen erste Erkundungen zu diesen Fragen auf dem Feld der Evaluation im Bereich der Prävention von Rechtsextremismus und Fremdenfeindlichkeit unternommen und dargestellt werden. Im Zentrum dieses Beitrags stehen nicht Evaluationskonzepte sondern die Evaluationspraxis, was insofern von Bedeutung ist, als Konzepte in der Praxis häufig nur eingeschränkt umgesetzt werden (können) und sich in der Praxis auch Herausforderungen stellen, denen auf konzeptioneller Ebene nicht immer die notwendige Aufmerksamkeit gewidmet wird. Im Folgenden wird zunächst eine erste Systematisierung sozialer Beziehungen im Rahmen von Evaluationen vorgestellt. Anschließend wird die Materialgrundlage dieses Beitrags dargestellt. Im Hauptteil des Beitrags wird auf verschiedene Beziehungen spezieller eingegangen: Soziale Beziehungen zwischen Evaluierenden und Evaluierten, zwischen Evaluierenden und Auftraggebern sowie zwischen Evaluierenden und sonstigen Beteiligten und Betroffenen. Abschließend werden die erzielten Erkenntnisse bilanziert sowie zentrale Herausforderungen benannt.

### **Soziale Beziehungen im Rahmen von Evaluation: Erste Systematisierungen**

An Evaluationen sind verschiedene Akteure beteiligt, die in der Fachliteratur als Beteiligte und Betroffene (oder „Stakeholder“) bezeichnet werden. Um die sozialen Beziehungen zwischen diesen Akteuren zu analysieren, wird in diesem Beitrag zwischen verschiedenen Beteiligten auf begrifflicher Ebene folgendermaßen differenziert.

- *Evaluierende* sind die Wissenschaftlerinnen und Wissenschaftler, die Evaluationsforschung durchführen und in dieser Eigenschaft mit den anderen Akteuren in Kontakt treten.
- Als *Evaluierte* werden diejenigen bezeichnet, deren Einschätzungen im Rahmen einer Evaluation erfragt werden oder deren Verhalten beobachtet wird, um auf dieser Grundlage Aussagen über ein Programm oder Projekt zu treffen.
- *Auftraggeber* erteilen den Evaluierenden den Auftrag zur Durchführung einer Evaluation und sind für die Finanzierung der Evaluation verantwortlich.
- Als *sonstige Beteiligte und Betroffene* werden diejenigen bezeichnet, die mit der Konzeption, Steuerung oder Nutzung einer Evaluation zu tun haben oder von einer Evaluation betroffen sind, ohne Auftraggeber oder Evaluierende zu sein.

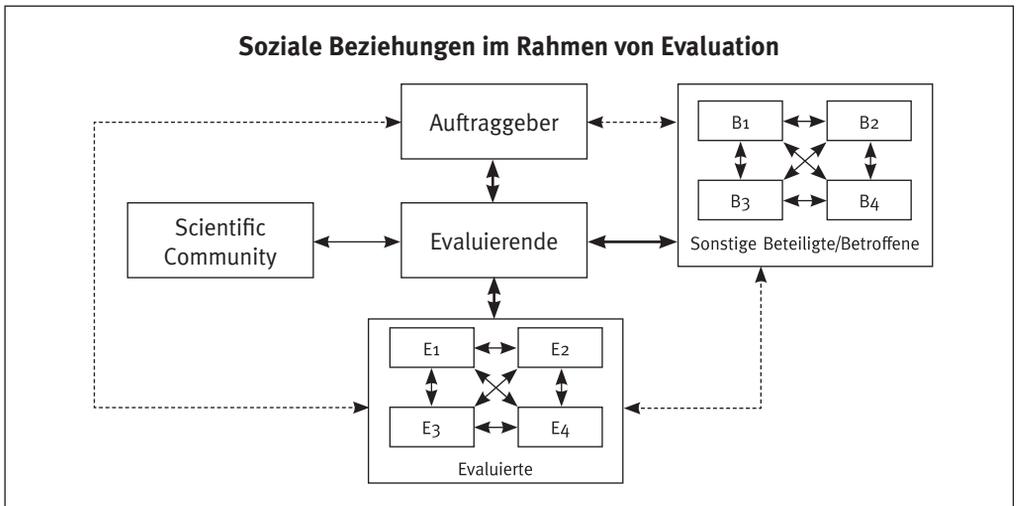
Diese begriffliche Unterscheidung stellt einen nur teilweise befriedigenden Versuch zur Differenzierung zwischen den an einer Evaluation beteiligten sozialen Akteuren dar. Weder sind die Begriffe immer trennscharf: So können z. B. sonstige Beteiligte und Betroffene unter bestimmten Umständen (s. u.) zu Evaluierten werden. Noch sind die gewählten Begriffe durchweg konsistent: So werden strenggenommen nicht diejenigen, deren Einschätzungen im Rahmen einer Evaluation erfragt werden oder deren Verhalten beobachtet wird, evaluiert, sondern die Projekte oder Programme, an denen sie beteiligt sind. Diese, im

Rahmen einer Evaluation Befragten oder Beobachteten, werden in diesem Beitrag dennoch als Evaluierete bezeichnet, da sie es sind, die sich häufig auf dem Prüfstand wähen und die von den Ergebnissen u. U. in hohem Maße abhängig sind. Wenn man an den sozialen Beziehungen im Rahmen von Evaluationen interessiert ist, muss man diesen Deutungen und Abhängigkeiten Rechnung tragen. Den relevanten sozialen Prozessen wird man nicht gerecht, wenn man bei diesen Personen von „Datenträgern“ (Wottawa/Thierau 2003, S. 59) spricht oder Distanz zwischen Personen und Programm anmahnt und darauf beharrt, lediglich das Programm, nicht jedoch die damit assoziierten Personen würden evaluiert.

Wie oben bereits deutlich wurde, zeichnen sich soziale Beziehungen im Rahmen einer Evaluation im Vergleich zu den Beziehungen im Kontext sonstiger Sozialforschung durch Besonderheiten aus.

- Evaluierende suchen sich ihre Untersuchung in der Regel nicht selbst, sondern werden beauftragt – der Zugang zum Untersuchungsfeld kann dabei auch von den Auftraggebern abhängig sein.
- Evaluierete sind teilweise abhängig von Bewertungen, die im Rahmen von Evaluationen abgegeben werden (oder fühlen sich zumindest von diesen Bewertungen abhängig), da Evaluationen erhebliche Konsequenzen für sie bzw. die Programme oder Projekte haben, an denen sie beteiligt sind.
- Evaluationen finden in einem Feld statt, das durch diverse Beteiligte und Betroffene geprägt ist, die in die Untersuchungen einzubeziehen sind.

Alle Beteiligten stehen damit in einem Geflecht sozialer Beziehungen zueinander, das erhebliche Konsequenzen für eine Evaluation haben dürfte.



**Abbildung 1:** Geflecht sozialer Beziehungen

Dieser Beitrag konzentriert sich auf die sozialen Beziehungen, die die Evaluierenden zu anderen Akteuren unterhalten – zu den Evaluierten, zu Auftraggebern sowie zu sonstigen Beteiligten und Betroffenen (vgl. dicke Pfeile). Vernachlässigt werden dabei sowohl die Beziehungen zwischen diesen anderen Akteursgruppen als auch die innerhalb von Akteursgruppen, z. B. zwischen verschiedenen Evaluierten (vgl. gestrichelte Pfeile). Auch das Verhältnis zwischen Evaluierenden und dem wissenschaftlichen Umfeld kann in diesem Beitrag nur kurz gestreift werden.

### **Die Materialgrundlage: Sechs Evaluationsuntersuchungen**

Die folgenden Ausführungen basieren auf fünf ausführlichen Interviews mit Kolleginnen und Kollegen, die Programme, Projekte oder Maßnahmen gegen Rechtsextremismus oder für Toleranz evaluieren oder evaluiert haben<sup>3</sup>. Ergänzend wurden die Erfahrungen herangezogen, die im Rahmen von Projektbegleitungen der Arbeitsstelle Rechtsextremismus und Fremdenfeindlichkeit gemacht wurden<sup>4</sup>. Diese Daten geben Einblick in die Komplexität der Evaluationspraxis auf diesem Feld und sensibilisieren für die damit verbundenen Fragen, Schwierigkeiten und Potenziale.

Das in diesem Band in den Blick genommene Arbeitsfeld zeichnet sich durch eine bunte Vielfalt von Ansätzen und Herangehensweisen aus, die nicht immer ohne weiteres bestimmten Typen oder „Schulen“ zugeordnet werden können. Da es uns in diesem Zusammenhang vor allem um die Exploration von Erkenntnissen zu Bedingungen, Erfahrungen und Ergebnissen geht, haben wir uns bemüht, ein möglichst breites Spektrum ganz unterschiedlicher Evaluationsansätze, -designs und -konstellationen einzubeziehen. Aufgrund der Heterogenität unseres Samples müssen Abstriche hinsichtlich der Möglichkeiten des systematischen Vergleichs zwischen den einbezogenen Evaluationsstudien in Kauf genommen werden.

An dieser Stelle zunächst eine kurze Darstellung der Evaluationen und Gesprächspartnerinnen/-partner, zu denen bzw. aus deren Perspektive Informationen aus den Interviews vorliegen:

Drei Evaluationen beziehen sich auf einzelne Projekte, die entweder in Form von Projektwochen an einer Einrichtung oder als Curricula an verschiedenen Einrichtungen umgesetzt werden. Eine dieser Evaluationen ist zum Interviewzeitpunkt bereits seit einigen Jahren abgeschlossen, die beiden anderen noch nicht. Die Daten werden bei einer dieser Evaluationen vor allem per Gruppendiskussion mit Jugendlichen und flankierend durch Beobachtungen erhoben sowie bei der zweiten Evaluation durch Telefoninterviews und im Rahmen moderierter Gruppenarbeit mit Schülerinnen und Schülern. Die dritte Evaluation

---

3 Diese Interviews wurden in den Jahren 2005 und 2006 von Silke Schuster geführt. Den Interviews liegt ein Leitfaden zugrunde, sie wurden per Tonband aufgezeichnet und anschließend verschriftet.

4 Bezogen wird sich dabei auf Erkenntnisse, die von Michaela Glaser und Peter Rieker in den Jahren 2004 bis 2006 im Rahmen wissenschaftlicher Begleitungen zu Angeboten Interkulturellen Lernens gesammelt wurden (vgl. Glaser/Rieker 2006).

verwendet einen sehr breiten Methodenmix: Vorher-Nachher-Befragungen per Fragebogen sowie qualitative Interviews mit solchen Adressaten, die extreme Positionen vertreten, Interviews, Fragebogenerhebungen, Protokolle und Auswertungsgespräche sowie Beobachtungen.

Eine vierte Evaluation bezieht sich auf ein großes, heterogenes, auf mehrere Jahre angelegtes Programm. Die Programmevaluation bezieht dabei die geförderten Projekte ein, nicht jedoch die Programmsteuerung. Daten zu allen geförderten Projekten werden mittels eines standardisierten Fragebogens jährlich erhoben; die Projektverantwortlichen ausgewählter Projekte werden zusätzlich in Interviews befragt – in einigen Fällen werden auch jugendliche Teilnehmerinnen und Teilnehmer der geförderten Projekte interviewt und gelegentlich werden auch Beobachtungen durchgeführt.

Die fünfte Evaluation bezieht sich auf eine begrenzte Anzahl sehr verschiedenen Interventionen, die innerhalb eines bestimmten Zeitraums auf ganz unterschiedlicher Grundlage durchgeführt werden, aber nicht Bestandteile eines gemeinsamen Programms oder Projektes sind. Das Evaluationsteam führt Interviews mit den für diese Interventionen Verantwortlichen sowie mit Fachleuten, die diese Vorhaben und die durch sie angeregten Prozesse beurteilen sollen. Während die zuvor genannten Evaluationen sich auf präventive Maßnahmen beziehen, die auf die Förderung von Demokratie, Toleranz und Vielfalt gerichtet sind, bezieht diese fünfte Evaluation sich auf Interventionen, die sich zumeist gegen manifeste Ausprägungen von Rechtsextremismus richten.

Die wissenschaftlichen Begleitungen der Arbeitsstelle Rechtsextremismus und Fremdenfeindlichkeit beziehen sich auf 22 Projekte und Einrichtungen, die ein breites Spektrum an Angebotstypen Interkulturellen Lernens repräsentieren und unter ganz unterschiedlichen Bedingungen initiiert, gefördert und durchgeführt werden. Diese Projekte wurden nach inhaltlichen Gesichtspunkten ausgewählt, wobei es darum ging, unterschiedliche Projekttypen und Lernformate einzubeziehen. Im Rahmen dieser Begleitungen, an denen die Projekte sich freiwillig beteiligten, wurden Interviews mit verschiedenen Akteuren geführt, Projekteinheiten beobachtet und Workshops mit den Projektverantwortlichen durchgeführt und ausgewertet (vgl. Glaser/Rieker 2006, S. 15ff. sowie Glaser i. d. Bd.).

In Hinblick auf die sozialen Beziehungen im Rahmen von Evaluationen sind verschiedene Aspekte der Evaluationsanlage relevant (vgl. Haubrich u.a. 2006). Im Kontext der Evaluationen, auf die sich dieser Beitrag bezieht, scheinen vor allem Differenzen zwischen dem Auftraggeber und den Evaluierten von Bedeutung. Diese Evaluationen können hinsichtlich des Auftraggebers und in Bezug auf die Frage, wer die Evaluierten sind, folgendermaßen zugeordnet werden.

## Übersicht: Auftraggeber und Evaluierete

| Evaluationen | Auftraggeber „desinteressiert“ | Auftraggeber verpflichtend | Evaluierete Durchführende | Evaluierete Adressaten | Evaluierete sonst. Beteil. |
|--------------|--------------------------------|----------------------------|---------------------------|------------------------|----------------------------|
| A            | x                              |                            | x                         |                        | x                          |
| B            |                                | x                          |                           | x                      | x                          |
| C            |                                | x                          | x                         | x                      |                            |
| D            |                                | x                          | x                         |                        |                            |
| E            | x                              |                            | x                         | x                      |                            |
| F            | x                              |                            | x                         | x                      | x                          |

### Auftraggeber

In Hinblick auf die konkreten Evaluationen haben die Evaluierenden in drei Fällen einen „desinteressierten“<sup>5</sup> Auftraggeber, der in keiner direkten Beziehung zu den Evaluierten und den sonstigen Beteiligten steht. Selbstverständlich haben auch diese Auftraggeber Interessen – z. B. daran, dass die Möglichkeiten und Grenzen einer Evaluation erprobt werden oder dass grundlegende Informationen zu einem Arbeitsfeld generiert werden – doch diese beziehen sich nicht auf die Bewertung eines bestimmten Programms. Die Evaluierenden und die sonstigen Beteiligten sind in diesen Fällen nicht Teil eines zu evaluierenden Programms, sondern müssen durch die Evaluierenden eigenständig recherchiert, zur Beteiligung motiviert und eingebunden werden. Die Konstellation, mit der wir in den drei Fällen mit „desinteressierten“ Auftraggebern konfrontiert sind, ist für Evaluationen eher untypisch. Üblicherweise sind die Auftraggeber die für ein Programm oder Projekt Verantwortlichen, die durch die Evaluation Erkenntnisse und Ergebnisse zum Programm bzw. Projekt erhalten möchten – so wie in den drei anderen Fällen unseres Untersuchungssamples. Evaluierete und sonstige Beteiligte sind mit dem Auftraggeber dabei also durch ein gemeinsames Vorhaben verbunden, in dessen Rahmen die Beteiligung an der Evaluation zum Teil verbindlich festgelegt ist, d.h. Teil der Förderbedingungen ist.

In Hinblick auf die Gewinnung von Evaluierten zeigten sich Differenzen zwischen diesen unterschiedlichen Auftragskonstellationen. Bei Evaluationen mit „desinteressierten“ Auftraggebern, in denen keine Pflicht zur Mitwirkung bestand, waren die angefragten Akteure in einzelnen Fällen nicht bereit, sich an der Evaluation zu beteiligen. Durch eine Mischung aus Druck und Anreizen konnte in diesen Fällen zum Teil dennoch die Beteiligung dieser

5 Der Begriff „desinteressiert“ geht in der hier intendierten Bedeutung auf Alfred Schütz zurück, der damit die Haltung eines Wissenschaftlers oder einer Wissenschaftlerin bezeichnet, der in Bezug auf das untersuchte Feld lediglich Erkenntnisinteressen hat, nicht jedoch in das Geflecht alltagspraktischer Interessen verstrickt ist (vgl. Schütz 1971, S. 43).

Akteure erreicht werden. Umgekehrt wurde es bei den Evaluationen, in denen eine Mitwirkungspflicht bestand, vermieden, unwillige Akteure zur Beteiligung zu verpflichten. Hintergrund dieses Vorgehens ist die Einschätzung, dass eine Evaluation nur auf der Grundlage freiwilliger Beteiligung sinnvoll ist. Unabhängig von der Art des Auftrags und vom Grad der Verpflichtung zur Beteiligung, sehen sich Evaluationen also generell vor der Aufgabe, Akteure für eine freiwillige Teilnahme zu gewinnen.

## Evaluierte

Im Kontext der einbezogenen Evaluationen sind die Evaluierten zum Teil diejenigen, die eine Maßnahme durchführen, teilweise die Adressatinnen/Adressaten dieser Maßnahme sowie in einigen Fällen auch sonstige Beteiligte. Gemeint sind also diejenigen, die in Erhebungen (Befragungen, Beobachtungen) einbezogen sind und deren Einschätzungen bzw. deren Verhalten Grundlage der Bewertung durch die Evaluierenden sind. In einem Fall werden Durchführende und sonstige Beteiligte befragt, in zwei Fällen Durchführende und Adressaten, in einem Fall nur Durchführende und einmal sowohl Adressaten als auch sonstige Beteiligte. Während es bei Adressatenbefragungen um Informationen über jeweils eigene Empfindungen oder Einschätzungen geht, beziehen sich die Befragungen von Durchführenden und sonstigen Beteiligten in der Regel nicht auf diese selbst, sondern sie sollen vor allem Auskunft über Entwicklungen geben, die sie bei den Adressatinnen/Adressaten beobachtet haben. Generell lässt sich feststellen, dass überwiegend Durchführende befragt werden, obwohl sich die Indikatoren für den Erfolg eines Programms in erster Linie auf die Adressatinnen/Adressaten beziehen.

Dieses Ungleichgewicht spiegelt sich auch in der Art und den Bedingungen der Beteiligung. Während die Durchführenden sich in der Regel auf freiwilliger Grundlage an einer Evaluation beteiligen, entsteht bei der Befragung von Adressaten und sonstigen Beteiligten zum Teil der Eindruck einer unfreiwilligen Teilnahme. Dies dürfte auch damit zusammenhängen, dass die Durchführenden zumeist an den Entscheidungen, ob eine Einrichtung oder ein Projekt an der Evaluation mitwirkt, beteiligt sind, während die sonstigen Beteiligten und vor allem die Adressaten (in der Regel Jugendliche) dabei nicht einbezogen werden.

## Beziehungen zwischen Evaluierenden und Evaluierten

Unabhängig davon, welcher Gruppe die Evaluierten angehören, betonen die Evaluierenden die große Bedeutung einer guten Beziehung zu ihnen. Zum Teil wird dabei die Notwendigkeit einer „persönlichen Beziehung“ hervorgehoben, wobei nicht immer klar ist, ob damit ein direkter Kontakt zu den Evaluierten gemeint ist – d.h. dass z. B. Erhebungen nicht delegiert werden können – oder ob eine über den professionellen Rahmen hinausgehende Beziehung erstrebenswert erscheint. Als Ausdruck und Ergebnis einer guten Evaluationsbeziehung sehen es die Evaluierenden an, wenn sie durch die Evaluierten nicht nur positive

Erfolgsmeldungen erhalten, sondern auch offen und ehrlich über schwierige und unerwünschte Aspekte der zu evaluierenden Arbeit informiert werden. Auch das Engagement der Evaluierten gilt als Gradmesser für die Evaluationsbeziehung: Mehrere Evaluationsteams berichten von einer ausgeprägten Kooperationsbereitschaft von Evaluierten – in der Regel waren dies die Durchführenden einer Maßnahme. Zum Teil engagierten sie sich auch in ihrer Freizeit, unterstützten die Evaluierenden bei der Motivierung von Untersuchungsteilnehmern und beteiligten sich auch auf andere Weise weit über das erwartbare Maß hinaus. Solch großes Engagement wird verschiedentlich damit erklärt, die Evaluierten hätten den Eindruck geäußert, von der Evaluation in hohem Maße zu profitieren, z. B. dadurch, dass sie zur Selbstreflexion oder zur inhaltlichen Weiterentwicklung angeregt wurden. Auf der anderen Seite äußern Evaluierende verschiedentlich den Eindruck, mehr als die Evaluierten von der Evaluation zu profitieren. Vor diesem Hintergrund sehen sie die Verpflichtung, den Evaluierten eine Gegenleistung für ihre Beteiligung bieten zu müssen. Gegenleistungen können Rückmeldungen zu den im Rahmen der Evaluation gewonnenen Eindrücken oder Ergebnissen beinhalten, sie können aber auch in Form von finanziellen Vergütungen, Fortbildungen, Konfliktmoderationen, Vorträgen oder Gesprächsangeboten erfolgen. Wie verschieden solche Gegenleistungen auch aussehen mögen, so haben die Evaluierenden jedenfalls den Eindruck, sich nicht auf die Position neutraler Informationssammlerinnen/-sammler und -verarbeiterinnen/-verarbeiter beschränken zu können, sondern selbst etwas geben zu müssen. Als weitere Bedingung gelingender Kooperation wird die Freiwilligkeit der Beteiligung an einer Evaluation hervorgehoben. Es entsteht somit der Eindruck, dass eine Evaluationsbeziehung von den Evaluierenden dann als gut eingeschätzt wird, wenn Evaluierte und Evaluierende auf freiwilliger und reziproker Basis zusammenarbeiten.

Plausibel wird die große Bedeutung, die die Evaluierenden einer guten Beziehung zu den Evaluierten beimessen dadurch, dass nicht nur der Kooperation großer Stellenwert beigemessen wird, sondern in Hinblick auf die Ergebnisse regelrechte Ko-Produktionsbeziehungen beschrieben werden. So wird beispielsweise in der folgenden Interviewpassage, die der Illustration der engagierten Kooperation dienen soll, beschrieben, dass Lehrerinnen und Lehrer sich auf Fragen der Evaluierenden nicht mehr an die Inhalte der evaluierten Lerneinheiten erinnern konnten:

„Oh, was sollen wir jetzt dazu sagen, ihr müsst uns helfen, was haben wir überhaupt noch gemacht?“ Also, sogar die Lehrer haben's vergessen, aber sobald der Anstoß da war, da ging es wieder ratter..ratter..ratter.

Die Evaluierenden waren – in diesem Fall als Interviewende – offenbar nicht nur durch ihre Fragen, sondern auch dadurch, dass sie ihren Gesprächspartnern Anregungen für die Antworten gaben, direkt an der Produktion von Ergebnissen beteiligt. Es geht also nicht nur darum, dass Evaluierende und Evaluierte so miteinander in Kontakt gebracht werden müssen, dass die eine Seite Fragen stellen kann, die durch die andere Seite beantwortet

wird; darüber hinaus muss von beiden Seiten ein Arbeitsbündnis gestaltet werden, das als tragfähige Basis für die gemeinsame Ergebnisproduktion taugt.

Die Interviews verdeutlichen, dass nicht immer solch gute Beziehungen zwischen Evaluierenden und Evaluierten realisiert werden können. Vor allem dann, wenn Evaluierete befürchten, dass ihre Arbeit kontrolliert und gemäß eines Rankings bewertet werden soll, sind sie kaum zu einer konstruktiven Mitarbeit zu bewegen. Auch in der Untersuchung der Arbeitsstelle Rechtsextremismus und Fremdenfeindlichkeit waren Pädagoginnen und Pädagogen teilweise stark verunsichert, wenn wir uns nach Möglichkeiten der Begleitung ihres Angebots erkundigen. Einige lehnen dies unter Hinweis auf eine mögliche Störung des pädagogischen Prozesses oder der Zielgruppen ab. Dann wenn wir als wissenschaftliche Begleiterinnen/Begleiter Zugang zu einem Angebot erhielten, erkundigten sich Projektverantwortliche oder Adressaten mitunter so defensiv danach, ob sie alles richtig gemacht hätten, dass es schwer fiel, sich nicht als Kontrolleurin/Kontrolleur zu fühlen. Die Befürchtung, dass eine Evaluation mit negativen Konsequenzen verbunden sein könnte, trug in einem unserer Fälle offensichtlich dazu bei, dass sich eine Klasse im Rahmen einer „Generalprobe“ auf einen Unterrichtsbesuch vorbereitete: Die Auswertung einer vorangegangenen Projekteinheit war zwei Tage zuvor einstudiert worden, ebenso wie die Äußerungen einzelner Schülerinnen und Schüler, die die Ergebnisse ihrer Arbeitsgruppen präsentierten.

Hindernisse werden immer dann beschrieben, wenn negative Konsequenzen einer Evaluation befürchtet werden, d.h. wenn z. B. die Gefahr gesehen wird, mit kritischen Einschätzungen die Öffentlichkeit oder (potenzielle) Geldgeber gegen ein Projekt einzustimmen. Entsprechende Befürchtungen kommen auch in einem unserer Beobachtungsprotokolle zum Ausdruck.

„Bitte bitte schließen Sie nicht unseren Club – sonst sind auch viel mehr Jugendliche da“, fleht eine junge Migrantin, als der Beobachter einen interkulturellen Schülerclub besucht. Von einem anderen Mädchen bekommt er bei der gleichen Gelegenheit eine Cola spendiert – wobei ihn der Verdacht beschleicht, sie habe durch diesen spendablen Akt möglicherweise die befürchtete Schließung des Clubs abwenden wollen.“

Gefördert werden können solche Befürchtungen dadurch, dass nicht alle Evaluierten ausreichend über die Aufgaben und Kompetenzen einer Evaluation sowie über die Verwendung der Ergebnisse informiert wurden – während die Durchführenden einer Maßnahme in der Regel gut informiert werden, ist dies bei den Adressatinnen und Adressaten häufig nicht der Fall. Mitunter bestehen bei den Evaluierten auch gar keine klaren Vorstellungen zu einer Evaluation, dann z. B., wenn diese mit Presseberichterstattung gleichgesetzt wird. Schwierigkeiten werden auch in solchen Fällen beschrieben, in denen Evaluierende nicht allen Beteiligten Rückmeldungen zu ihren Ergebnissen geben konnten oder Erwartungen der Evaluierten – z. B. in Hinblick auf Praxisberatung – enttäuschten. Belastungen werden

auch mit den Kontextbedingungen in Zusammenhang gebracht, unter denen Befragungen oder Gespräche stattfinden: So können unzureichende Räumlichkeiten und Zeitknappheit zu Anspannungen, Verstimmungen oder wenig aussagekräftigen Ergebnissen beitragen. Besonders problematisch gestalten sich Evaluationsbeziehungen dann, wenn mit einzelnen Evaluierten gar keine Vorgespräche geführt wurden und diese auch nicht um ihr Einverständnis gebeten wurden. Dies ist häufiger dann berichtet worden, wenn Jugendliche Zielgruppe einer Evaluation sind: Vorabsprachen werden dann zum Teil nur mit Lehrkräften oder anderen erwachsenen Betreuerinnen und Betreuern geführt. Derart übergangene Jugendliche können eine Evaluation unter Umständen wirkungsvoll blockieren.

Vor dem Hintergrund der Erfahrungen mit entsprechenden Schwierigkeiten werden verschiedene Möglichkeiten beschrieben wie die Entwicklung einer guten Beziehung zwischen Evaluierenden und Evaluierten unterstützt werden kann. Als wichtig gelten vor allem klärende Vorgespräche, in denen die Beteiligten sich wechselseitig über ihre Vorhaben informieren. Die Evaluierten formulieren ihre Erwartungen – und möglicherweise auch ihre Befürchtungen – in Hinblick auf die Evaluation und geben Einblick in die (geplanten) Arbeiten der Einrichtung bzw. des Projektes. Die Evaluierenden klären dabei über ihren Arbeitsauftrag auf, legen ihr Vorgehen offen und berichten wie die Ergebnisse verwendet werden sollen. Falls Diskrepanzen zwischen den Erwartungen und Plänen der Beteiligten deutlich wurden, konnten diese zum Teil offengelegt und manchmal auch aufgelöst werden. Zum Teil wurden auch formale Vereinbarungen zur Regelung der wechselseitigen Ansprüche und Verpflichtungen getroffen. Gemeinsame Gespräche über die Evaluation wurden in einigen Fällen auch institutionalisiert, z. B. in Steuerungsgruppen, denen auch Evaluierte angehören können. In solchen Steuerungsgruppen kann das Design einer geplanten Evaluation diskutiert und ggf. auch gemeinsam geplant werden und es wird auch berichtet, dass die Auswahl der zu evaluierenden Einheiten gemeinsam mit Vertreterinnen und Vertretern der Evaluierten erfolgt. Schließlich wird es als wichtig erachtet, den Evaluierten Rückmeldungen zu den erzielten Ergebnissen zu geben und auch die Möglichkeit einzuräumen, zu diesen Ergebnissen Stellung zu nehmen. Vorgespräche, Steuerungsgruppen und Rückmeldungen werden deshalb als wichtige Bestandteile einer Evaluation angesehen, weil sie den Evaluierten Möglichkeiten der Beteiligung einräumen und verhindern, dass diese sich lediglich als Objekte einer Evaluation fühlen. Zu wenig berücksichtigt werden dabei allerdings häufig die Adressaten, die ebenfalls systematisch einbezogen und beteiligt werden sollten.

## **Beziehungen zwischen Evaluierenden und Auftraggebern**

Im Rahmen der vorliegenden Untersuchung werden ganz unterschiedliche Beziehungen zwischen Evaluierenden und Auftraggebern deutlich. Vor allem kann zwischen Evaluationen differenziert werden, in denen es gemäß der oben getroffenen Unterscheidung „desinteressierte“ und solchen, in denen es „verpflichtende“ Auftraggeber gibt. Im Falle der „desinteressierten“ Auftraggeber finden diese in den Interviews kaum Erwähnung. Wenn über sie berichtet wird, dann im Zusammenhang kleinerer Unterstützungsleistungen für

die Evaluation oder im Kontext von Rückmeldungen zu den Ergebnissen. In Hinblick auf die Ergebnisse der jeweiligen Evaluationen sind diese Auftraggeber durchaus interessiert, sie greifen aber kaum in die Gestaltung der Evaluationen ein, so dass ihre Beziehungen zu den Evaluierenden vergleichsweise wenig Berührungs- und Reibungspunkte aufweisen.

Die „verpflichtenden“ Auftraggeber haben in Hinblick auf den Evaluationsgegenstand eigene Interessen, weil sie für dieses Projekt oder Programm verantwortlich sind. Die Interessiertheit der Auftraggeber beschränkt sich in diesen Fällen nicht auf die inhaltlichen Ergebnisse, sondern erstreckt sich auf die Forschungsfragen, auf das Design und die Methoden der jeweiligen Evaluationen. In den drei hier einbezogenen Fällen, die entsprechende Auftraggeber haben, lassen sich allerdings ganz unterschiedliche Formen der Zusammenarbeit erkennen.

- **Evaluierende als Verantwortliche:** In einem Fall werden Forschungsfragen und Design von Evaluierenden und Auftraggeber gemeinsam entwickelt, wobei der Eindruck entsteht, dass der Auftraggeber sich mit den Vorschlägen der Evaluierenden weitgehend einverstanden erklärt hat.
- **Evaluierende als Vermittelnde:** In einem anderen Fall hat der Auftraggeber in Hinblick auf die Forschungsfragen klare Vorgaben formuliert, sich im Rahmen von Diskussionen der Steuerungsgruppe, in denen von sonstigen Beteiligten andere Interessen formuliert worden waren, aber mit einem Kompromiss einverstanden erklärt, in den weitere Fragestellungen eingeflossen sind. Die Evaluierenden beschreiben sich dabei in der Rolle der Vermittelnden.
- **Evaluierende als Dienstleistende:** In einem weiteren Fall hat der Auftraggeber gemäß den Angaben der Evaluierenden klare Vorgaben gemacht, die innerhalb des Evaluationszeitraums noch erweitert wurden, ohne dass die Evaluierenden die Möglichkeit gesehen haben, sachlich begründete Bedenken zur Geltung bringen zu können.

Ungeachtet dieser ganz unterschiedlichen Positionen, in denen Evaluierende gemäß ihrem eigenen Verständnis gegenüber Auftraggebern agieren, können sowohl Bedingungen der Beziehungen zwischen Evaluierenden und Auftraggebern benannt werden, die für Evaluationen als günstig eingeschätzt werden, als auch solche, die für Schwierigkeiten verantwortlich gemacht werden.

Als förderlich und hilfreich für eine Evaluation beschreiben die befragten Evaluatorinnen und Evaluatoren bestimmte Bedingungen: Wenn Auftraggeber kontinuierlich durch gleichbleibende Ansprechpartner vertreten werden und wenn diese Ansprechpartner als verlässlich, kooperationswillig und auf persönlicher Ebene auch als freundlich erlebt werden. Die Kooperation mit Auftraggebern wird dann als gelungen beschrieben, wenn diese sich auf Fragestellungen und Designs einlassen, die aus Sicht der Evaluierenden realistisch und sinnvoll sind. Zum Teil wird es auch als bereichernd erlebt, wenn Auftraggeber in Hinblick auf den konzeptionellen Zuschnitt und Fragestellungen einer Evaluation eigene Vorstellungen entwickeln. Wichtige Unterstützung einer Evaluation durch Auftraggeber kann z. B.

auch darin bestehen, dass Zugänge zu den Evaluierten eröffnet werden. Auch dann, wenn im Anschluss an die Rückmeldung von (Zwischen-)Ergebnissen einer Evaluation gemeinsam Schlussfolgerungen für die praktische Arbeit oder den Fortgang der Evaluation getroffen werden, wird dies verschiedentlich als positiv beschrieben.

Schwierigkeiten mit Auftraggebern werden in solchen Fällen beschrieben, in denen Vorgaben für eine Evaluation vom Auftraggeber als nicht verhandelbar festgelegt werden, was besonders dann als belastend erlebt wird, wenn diese Vorgaben aus fachlich evaluatorischer Sicht als nicht sinnvoll oder unrealistisch eingeschätzt werden. Auch eher globale Vorgaben von Auftraggebern, die Raum für unterschiedliche und wechselnde Auslegungen von Vorgaben an eine Evaluation bieten, können zu Belastungen der Beziehung zwischen Auftraggeber und Evaluierenden führen. Zusätzliche Probleme können sich dann ergeben, wenn die Ansprüche an eine Evaluation oder ihre Ziele im Untersuchungszeitraum durch den Auftraggeber verändert werden.

### **Beziehungen zwischen Evaluierenden und sonstigen Beteiligten/Betroffenen**

Während recht klar umrissen ist, wer Auftraggeber und Evaluierte sind, sind wir bei den sonstigen Beteiligten und Betroffenen mit Unklarheiten konfrontiert. Bei den hier einbezogenen Evaluationen treten nicht nur ganz unterschiedliche Akteure als (potenziell) Beteiligte und Betroffene in Erscheinung, sondern auch die Art und Weise der Beteiligung variiert erheblich.

- **Indirekt Beteiligte/Betroffene:** Teilweise werden im Rahmen von Evaluationen solche Personen als relevant beschrieben, die als indirekt Beteiligte und Betroffene bezeichnet werden können, so z. B. Eltern oder Schulbehörden, die ihre Zustimmung zu einem Evaluationsvorhaben geben müssen. Ihre Bedeutung ist aus Sicht der Evaluierenden auf diese Funktion beschränkt, die Beziehung zu ihnen wird gemäß ihrer Bereitschaft zur Zustimmung beurteilt und sie haben für die Evaluation in der Regel ansonsten keine erkennbare Bedeutung. Vereinzelt gibt es jedoch Hinweise darauf, dass diese indirekt Beteiligten und Betroffenen einer Evaluation potenziell noch mehr zu bieten haben: So können z. B. Eltern Rückmeldungen zu Entwicklungen ihrer Kinder auch über längere Zeiträume geben, die im Rahmen von Evaluationen üblicherweise nicht berücksichtigt werden können. Wenn diese Einschätzungen in einer Evaluation Berücksichtigung finden, dann würden die indirekt Beteiligten allerdings selbst zu Evaluierten.
- **Direkt Beteiligte/Betroffene:** Andererseits werden Fälle beschrieben, in denen sonstige Beteiligte und Betroffene in den Fokus einer Evaluation rücken und sich zumindest ihrem Selbstverständnis gemäß als Evaluierte verstehen. Diese Variante wird aus dem Kontext außerschulischer Bildungsmaßnahmen mit Schülerinnen und Schülern berichtet, in denen die für diese Lerneinheiten nicht verantwortlichen Lehrerinnen und Lehrer sich auf dem Prüfstand sehen. Aber auch in Gemeinden, in denen Interventionsmaßnahmen gegen Rechtsextremismus durchgeführt werden, sehen sich beteiligte

und betroffene Akteure oft als Evaluierter. Festzustellen ist dies dann, wenn Beteiligte emotional stark engagiert sind und wenn der thematische Fokus einer Maßnahme zu Polarisierungen einlädt. Je nach Interessenlage können dann entweder Unterstützung und Mitwirkungsbereitschaft oder Tendenzen zur Verweigerung seitens Beteiligter und Betroffener festgestellt werden. Die Beteiligten werden dann zu Evaluierter, wenn sie in Befragungen oder Beobachtungen einbezogen werden.

- **Beteiligte/Betroffene in einem Beirat:** Ein besonderer Fall liegt dann vor, wenn Beteiligte und Betroffene zusammen mit Vertreterinnen und Vertretern von Auftraggebern und Evaluierter in einem Beirat in die Koordination eines Evaluationsvorhabens eingebunden sind. In solch einem Beirat können Fragestellung, Design und Methoden einer Evaluation festgelegt werden und aus solch einer Gruppe kann Hilfe beim Zugang zu den Evaluierter geleistet werden. Aus diesen Zusammenhängen wird von engagierten und konstruktiven Diskussionen berichtet sowie von wichtigen Unterstützungen für eine Evaluation. In einem Fall kommt den Evaluierender dabei die Funktion von Vermittelnden zwischen den unterschiedlichen Interessen der Beteiligten zu.

So unterschiedlich diese Modelle auch sind, so zeigen sich doch in allen Fällen produktive Potenziale von Einbindung und Partizipation sonstiger Beteiligter. Am besten zu erkennen sind diese im Fall eines Beirates, auch weil sie hier am ehesten formal geregelt sind. Im Fall der indirekt Beteiligten deutet sich an, dass es gewinnbringend sein könnte, ihre Perspektive systematischer für die Evaluation verfügbar zu machen, d.h. sie zu direkt Beteiligten und Evaluierter zu machen. Die Beziehungen zu direkt Beteiligten bieten Chancen für interessante Perspektivenerweiterungen, sie sollten jedoch besser geklärt und auch in die Steuerung eines Evaluationsvorhabens eingebunden werden. Im Gegenzug werden dann negative Konsequenzen beschrieben, wenn Beteiligte und Betroffene im Rahmen einer Evaluation nicht eingebunden sind, d.h. vorab nicht informiert, in Fragen der Organisation nicht konsultiert und in Hinblick auf die Präsentation von Ergebnissen nicht berücksichtigt werden. Dies kann dazu beitragen, dass die Ergebnisse einer Evaluation durch diese unberücksichtigten Beteiligten nicht zur Kenntnis genommen und in ihren Verantwortungsbereichen auch nicht umgesetzt werden.

### **Abschließende Bemerkungen**

Die vorliegende Analyse hat den aus der Fachliteratur bekannten Befund bestätigt, dass Evaluierender in ein Geflecht spezifischer sozialer Beziehungen eingebunden sind (vgl. Widmer 2004). Dabei sind sie mit ganz unterschiedlichen Erwartungen verschiedener Beteiligter konfrontiert. So haben z. B. Evaluierter und Auftraggeber häufig verschiedene Interessen. Zum Teil kommen die Evaluierender dabei in die Rolle von Vermittelnden, die gefordert sind, die richtige Balance zwischen diesen konkurrierenden Interessen und Ansprüchen zu finden (vgl. Erzberger 2006; Wottawa/Thierau 2003). Manchmal geraten sie dabei zwischen die „Mühlsteine“ dieser Interessen und Ansprüche, ohne dass es gelingt,

diese miteinander zu vermitteln. In jedem Fall werden an Evaluierende hohe Anforderungen in Hinblick auf ihre sozialen und kommunikativen Kompetenzen gerichtet (vgl. Beywl 2006).

Unsere Ergebnisse zeigen, dass diese Anforderungen mit besonderen Belastungen verbunden sind. Evaluierende beklagen verschiedentlich den Mangel an kollegialem Austausch und kompetenter Beratung – dies spricht dafür, dass auf dem Arbeitsfeld der Evaluation weniger Diskussionen und kollegialer Austausch als in anderen Forschungsbereichen existieren. Möglicherweise hängt dies mit einer stark ausgeprägten Konkurrenz auf diesem Feld zusammen, auf dem die Abhängigkeit von Aufträgen vergleichsweise hoch ist und es kaum auftragsfreie Forschung gibt. Der Mangel an kollegialem Austausch wird von den Evaluierenden wahrscheinlich auch deswegen als besonders gravierend erlebt, da es in Hinblick auf Evaluationen keine klaren Qualifikationsprofile gibt. Außerdem entsteht der Eindruck, dass fachwissenschaftliche Standards (vgl. DeGEval 2001; Joint Committee/Sanders 1999) in der Praxis der Evaluation noch nicht in der Weise etabliert sind, dass sie in schwierigen Situationen Orientierung bieten könnten.

Auf dem Feld der Evaluation von Maßnahmen und Programmen gegen Rechtsextremismus sind die geschilderten Anforderungen und Belastungen mit besonderer Brisanz verbunden. Evaluierende stehen hier gegenwärtig häufig im Licht der Öffentlichkeit und sehen sich stark divergierenden Interessen ausgesetzt, die nicht nur fachlich, sondern auch politisch motiviert sein können (vgl. auch Palloks i. d. Bd.). Wie in diesem Beitrag deutlich wurde, können die polarisierenden und affektiv aufgeladenen Einschätzungen, Ansprüche und Befürchtungen auch die Mitwirkung (bzw. Nicht-Mitwirkung) der an einer Evaluation Beteiligten bestimmen. Einerseits macht dies auf besondere Anforderungen an Evaluation in diesem thematischen Feld aufmerksam, angesichts derer es z. B. hilfreich sein könnte, die politischen Haltungen und Orientierungen aller Beteiligten zu Beginn einer Evaluation zu klären. Andererseits stellen sich hier Fragen, die auch für Evaluationen auf anderen Feldern relevant sind, besonders deutlich, z. B. in Hinblick auf die für die Beteiligung an Evaluation häufig postulierte Freiwilligkeit, in Bezug auf die Folgen unfreiwilliger Teilnahme und hinsichtlich der Kooperation im Rahmen von Evaluationsbeziehungen.

- **Freiwilligkeit:** Es wurde deutlich, dass die freiwillige Beteiligung der Evaluierten nicht nur als erstrebenswert, sondern als unverzichtbare Voraussetzung einer gelingenden Evaluation anzusehen ist. Allerdings muss man sich die Frage stellen, welche Qualität Evaluationsergebnisse haben, die lediglich auf Einschätzungen solcher Beteiligter beruhen, die der Beteiligung nicht nur explizit zugestimmt haben, sondern damit möglicherweise eigene Interessen verfolgen sowie über das Evaluationsdesign und die Auswahl der zu evaluierenden Einheiten mitbestimmt haben. Generell ist von ganz unterschiedlichen Graden und Ausprägungen von Freiwilligkeit auszugehen: Freiwilligkeit auf der Grundlage eigenen Erkenntnisinteresses; Freiwilligkeit aufgrund öffentlichen Drucks zur Legitimation; Freiwilligkeit auf der Basis einer eingegangenen Verpflichtung; Freiwilligkeit, die durch Konformität oder Abwesenheit von explizitem Widerstand zustande kommt. Es ist davon auszugehen, dass diese Formen von Freiwilligkeit zu unterschied-

lichen Kooperationsbeziehungen beitragen und es wäre wichtig, diese im Rahmen zukünftiger Analysen zu reflektieren und zwischen ihnen zu differenzieren.

- **Unfreiwillige Beteiligung:** Massive Probleme bei der Durchführung einer Evaluation werden immer dann berichtet, wenn die Freiwilligkeit nicht gewährleistet ist bzw. wenn einzelne Beteiligte gar nicht erst in die Lage versetzt wurden, sich für die Beteiligung an einer Evaluation zu entscheiden. Auch spricht einiges dafür, dass die Qualität der Ergebnisse einer Evaluation mit unfreiwilligen Teilnehmenden ungenügend ist. Obwohl über die Forderung nach freiwilliger Beteiligung also weitgehend Konsens besteht, wurde in verschiedenen der hier analysierten Evaluationen deutlich, dass zumindest einzelne Evaluierete nicht auf freiwilliger Grundlage beteiligt waren, in der Regel handelt es sich dabei um die Adressaten eines Programms. Zu wenig ist bislang darüber bekannt, welche Auswirkungen die unfreiwillige Beteiligung an einer Evaluation hat und inwieweit die auf unfreiwilliger Grundlage erzielten Ergebnisse aussagekräftig sind. Sinnvoll könnte es auch sein, Möglichkeiten zu erschließen, mehr über die Motive zur Nicht-Beteiligung an einer Evaluation zu erfahren.
- **Kooperation:** Deutlich und plausibel beschreiben die von uns befragten Evaluierenden die Notwendigkeit verlässlicher und vertrauensvoller Kooperationsbeziehungen zwischen Evaluierenden und Evaluierten, die eine gewisse Reziprozität erfordern. Für die gewährten Einblicke und Informationen erscheint es aus Sicht der Evaluierenden notwendig, den Evaluierten Gegenleistungen zu gewähren, sei es in Form von Informationen, Beratung oder dadurch, dass man selbst als verlässlicher Partner agiert. In diesem Zusammenhang werden Loyalitätskonflikte deutlich – und zwar dann, wenn durch die Evaluierenden kritische und distanzierte Einschätzungen vorgenommen werden sollen. Diese Kooperationsbeziehungen erfordern einerseits Loyalität und andererseits Kritik- und Distanzfähigkeit. Die vielfältigen Dimensionen dieser Beziehungen zeigen sich z. B. darin, dass diese Loyalitätskonflikte in einem Versprecher, der in einem von uns geführten Interview geäußert wird, zum „Koalitionskonflikt“ werden. Das heißt, die Beteiligten sind aufeinander angewiesen, aber sie müssen sich auch voneinander abgrenzen, um ihre unterschiedlichen Aufgaben zu erfüllen.

Diese hier angedeuteten Diskrepanzen lassen sich nicht durch einfache Entscheidungen für oder gegen einzelne Positionen auflösen, sondern sie erfordern die kontinuierliche Reflexion dieser Aspekte sowie einen ständigen Balanceakt bei der jeweils angemessenen Bestimmung der eigenen Position und des Vorgehens bei einer Evaluation. Hilfreich könnte es dabei sein, wenn Beteiligung und Kooperation durch eine klare Arbeitsteilung geregelt wären, bei der die Evaluierenden Expertinnen und Experten für Fragen der Evaluation sind, d.h. Design, Sample und Methoden maßgeblich bestimmen und auf der Grundlage empirischer Untersuchungen Ergebnisse formulieren. Dabei sind die Evaluierenden auch gefordert, Informationen über die Perspektiven derjenigen einzubeziehen, die sich nicht freiwillig beteiligen wollen – sowie über die Gründe für diese Nicht-Beteiligung. Eine Möglichkeit wäre es bspw., im Rahmen einer Evaluation die Prozesse zu berücksichtigen,

die zu Beteiligung und Nicht-Beteiligung führen. Eine andere Möglichkeit liegt in multimedialen bzw. multiperspektivischen Evaluationsdesigns, die nicht nur die Perspektiven einzelner Beteiligter berücksichtigen, sondern z. B. durch Beobachtungen eigene Zugänge zum Evaluationsgegenstand erschließen.

Diese und andere Fragen, die im Rahmen des vorliegenden Beitrags nur angedeutet werden konnten, sollten im Rahmen zukünftiger Untersuchungen geklärt werden. Dabei wäre es auch notwendig, eine systematische Evaluationsforschung – im Sinne von Forschung zu den Voraussetzungen, Bedingungen und Verläufen von Evaluation – stärker zu etablieren.

## Literaturverzeichnis

- Beywl, Wolfgang (2006): Evaluationsmodelle und qualitative Methoden. In: Flick, Uwe (Hrsg.): Qualitative Evaluationsforschung. Konzepte, Methoden, Umsetzung. Reinbek: Rowohlt, S. 92–116
- Deutsche Gesellschaft für Evaluation (DeGEval) (2001): Standards für Evaluation. [www.degeval.de/calimero/tools/proxy.php?id=70](http://www.degeval.de/calimero/tools/proxy.php?id=70) (28.6.2007)
- Eichler, Dirk/Merkens, Hans (2006): Organisationsforschung mit qualitativen Methoden – Erfahrungen aus der Evaluation eines freien Jugendhilfeträgers. In: Flick, Uwe (Hrsg.): Qualitative Evaluationsforschung. Konzepte, Methoden, Umsetzung. Reinbek: Rowohlt, S. 301–318
- Erzberger, Christian (2006): Steuerung externer Evaluation durch Kommunikation. In: Projekt eXe (Hrsg.): Aushandlungsprozesse zur Steuerung externer Evaluation. München: DJI, S. 19–27. [www.dji.de/evaluation/eXe\\_Aushandlungsprozesse.pdf](http://www.dji.de/evaluation/eXe_Aushandlungsprozesse.pdf) (24.8.2007)
- Frindte, Wolfgang/Preiser, Siegfried (2007): Prävention gegen Rechtsextremismus. In: Aus Politik und Zeitgeschichte, H. 11, S. 32–38
- Glaser, Michaela/Rieker, Peter (2006): Interkulturelles Lernen als Prävention von Fremdenfeindlichkeit. Ansätze und Erfahrungen in Jugendbildung und Jugendarbeit. Halle: DJI
- Guba, Egon G./Lincoln, Yvonna S. (1989): Fourth Generation Evaluation. Newbury Park: Sage
- Haubrich, Karin/Struhkamp, Gerlinde/Lüders, Christian (2006): Einleitung: Aushandlungsprozesse in externen Evaluationen – Aufarbeitung eines „stillen“ Themas. In: Projekt eXe (Hrsg.): Aushandlungsprozesse zur Steuerung externer Evaluation. München: DJI, S. 5–18. [www.dji.de/evaluation/eXe\\_Aushandlungsprozesse.pdf](http://www.dji.de/evaluation/eXe_Aushandlungsprozesse.pdf) (24.8.2007)
- Joint Committee on Standards for Educational Evaluation/Sanders, James R. (Hrsg.) (1999): Handbuch der Evaluationsstandards. Opladen: Leske + Budrich
- Klawe, Willy (2006): Multiperspektivische Evaluationsforschung als Prozess-Wirkungsrekonstruktion aus der Sicht der Beteiligten. In: Projekt eXe (Hrsg.): Wirkungsevaluation in der Kinder- und Jugendhilfe. Einblicke in die Evaluationspraxis. München: DJI, S. 125–142
- Lynen von Berg, Heinz/Palloks, Kerstin/Steil, Armin (2007): Interventionsfeld Gemeinwesen. Evaluation zivilgesellschaftlicher Strategien gegen Rechtsextremismus. Weinheim/München: Juventa

- Lynen von Berg, Heinz/Roth, Roland (Hrsg.) (2003): Maßnahmen und Programme gegen Rechtsextremismus wissenschaftlich begleitet. Aufgaben, Konzepte und Erfahrungen. Opladen: Leske + Budrich
- Mensching, Anja (2006): Zwischen Überforderung und Banalisierung – zu den Schwierigkeiten der Vermittlungsarbeit im Rahmen qualitativer Evaluationsforschung. In: Flick, Uwe (Hrsg.): Qualitative Evaluationsforschung. Konzepte, Methoden, Umsetzung. Reinbek: Rowohlt, S. 339–362
- Patton, Michael Quinn (1997): Utilization-Focused Evaluation. The New Century Text. Thousand Oaks/London/New Delhi: Sage
- Scriven, Michael (1991): Evaluation Thesaurus. Newbury Park: Sage
- Schütz, Alfred (1971): Wissenschaftliche Interpretation und Alltagsverständnis menschlichen Handelns. In: Schütz, A.: Gesammelte Aufsätze I. Den Haag: Nijhoff, S. 3–54
- Ulrich, Susanne/Wenzel, Florian M. (2004): Partizipative Evaluation. In: Uhl, Katrin/Ulrich, Susanne/Wenzel, Florian M. (Hrsg.): Evaluation politischer Bildung. Ist Wirkung messbar? Gütersloh: Verlag Bertelsmann Stiftung, S. 27–48
- Widmer, Thomas (2004): Qualität der Evaluation – Wenn Wissenschaft zur praktischen Kunst wird. In: Stockmann, Reinhard (Hrsg.): Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder. Opladen: Leske + Budrich, S. 83–109
- Widmer, Thomas/Hirschi, Christian (2007): Herausforderungen der Evaluation von Maßnahmen gegen Rechtsextremismus. In: LEGES, H. 2, S. 255–274
- Wottawa, Heinrich/Thierau, Heike (2003): Lehrbuch Evaluation. Bern u.a.: Verlag Hans Huber

## Improvisation, Partizipation und die Frage der Wirkungen – Eine Untersuchung zu Evaluationen der pädagogischen Arbeit gegen Rechtsextremismus<sup>1</sup>

Evaluationen erfahren seit längerem eine erhöhte Aufmerksamkeit in vielen Gesellschaftsbereichen, Politik- und Praxisfeldern. Gleichzeitig bleibt unscharf was unter Evaluation in den jeweiligen gesellschaftlichen Teilbereichen zu verstehen ist (vgl. Haubrich/Lüders 2004, S. 316). Fast könnte man behaupten, dass es so viele verschiedene Auffassungen von Evaluationen gibt wie Evaluationen existieren. Auch im Zusammenhang der pädagogischen Auseinandersetzung mit Rechtsextremismus und Fremdenfeindlichkeit stellt sich die bislang nicht befriedigend beantwortete Frage nach Qualität und Wirksamkeit unterschiedlicher Arbeitsansätze und Herangehensweisen.

Mittlerweile (u.a. durch die Förderung des Bundes von Modellprogrammen zur Rechtsextremismusprävention angeregt) liegen in diesem spezifischen Handlungsfeld eine Vielzahl an externen Projekt- und Programmevaluationen vor (Eser-Davolio 2000; Schrödter 2004; Wagner/Homburg 2004; Klingelhöfer u.a. 2006; Lynen von Berg u.a. 2007, vgl. auch die Beiträge i. d. Bd.). Insgesamt fehlt es aber an einer Systematisierung und Aufbereitung dessen, wie und in Bezug auf was Projekte und Programme evaluiert, welche Kriterien zur Bewertung und Überprüfung von Zielen herangezogen werden und wie letztlich versucht wird Wirkungen der Aktivitäten nachzuweisen.

Vor dem Hintergrund dieser Fragestellung hat die Arbeitsstelle Rechtsextremismus und Fremdenfeindlichkeit am Deutschen Jugendinstitut in einer Recherche 15 externe Evaluationen identifiziert und zunächst dokumentenanalytisch<sup>2</sup> untersucht. Wir wählten Evaluationen aus, auf die wir anhand veröffentlichter Dokumente und Materialien aufmerksam wurden<sup>3</sup>. Bei dieser Auswahl bemühten wir uns, ein breites Spektrum an Ansätzen, Designs und Konstellationen einzubeziehen. Wir berücksichtigten nur solche Studien, die jüngeren Datums waren und eine ausreichende Materialbasis bereitstellten. Mehrheitlich handelte es sich um Evaluationen von Modellprojekten bzw. -programmen. In drei Fällen hatten wir es mit Evaluationen von etablierten Bildungsprogrammen zu tun.

Im Zentrum unseres Interesses standen zunächst Fragen nach dem angewandten Evaluationsdesign, dem Evaluationszweck, den Evaluationszielen sowie dem Qualitäts- bzw.

- 1 Ich bedanke mich herzlich bei Michaela Glaser und Dr. Peter Rieker für ihre konstruktive Kritik und ihre wertvollen Anmerkungen zu einer früheren Fassung dieses Artikels.
- 2 Eine sorgfältige Sichtung des Materials und dessen strukturierte Aufbereitung verdanken wir Sabine Diederich.
- 3 Die Evaluationsberichte waren i. d. R. schwer zugänglich. Wir haben die Berichte meistens auf den Internetseiten der zu evaluierenden Projekte gefunden, wobei diese häufig nach Beendigung des Projekts wieder aus dem Netz genommen wurden. Daneben sind wir über Publikationslisten von Forschenden auf entsprechende Dokumente gestoßen, wobei diese Berichte häufig nicht einsehbar waren bzw. vom Auftraggeber erst „freigegeben“ werden mussten.

Erfolgsbewertungsprozess. Die Analyse der Dokumente lieferte erste Einschätzungen zum Feld und half weitere relevante Fragestellungen zu identifizieren. Auf dieser Basis konnten fünf Evaluationsprojekte für vertiefende Interviews ausgewählt werden. Um also mehr über die Erfahrungen von Evaluationen, ihre Voraussetzungen und Bedingungen zu erfahren, führten wir fünf ausführliche leitfadengestützte Interviews mit Kolleginnen und Kollegen, die Programme, Projekte oder Maßnahmen gegen Rechtsextremismus oder für Toleranz evaluieren bzw. evaluiert haben. Hauptaugenmerk der Interviews war es neben den oben bereits genannten Fragen auch, die Potenziale und Schwierigkeiten von Evaluationen zu klären. Die Interviews lieferten aussagekräftiges Material und geben insbesondere Einblick in die Komplexität der Evaluationspraxis auf diesem Feld. Deshalb stellen diese mündlichen Informationen die zentrale Materialgrundlage des folgenden Berichts dar.

Im Rahmen der ersten Auswertungen zeigten sich außerdem Fragestellungen, denen wir in einem vertiefenden Schritt mit erfahrenen Evaluationsexpertinnen und -experten gemeinsam nachgehen wollten. Zu diesem Zweck luden wir im November 2006 zu einem Workshop ein<sup>4</sup>. Bei diesem Treffen wurden zu Beginn zentrale Aspekte der DJI-Untersuchung Kolleginnen und Kollegen aus Wissenschaft und Praxis vorgestellt. Anschließend wurden in einem Werkstattgespräch die Frage nach förderlichen und hinderlichen Bedingungen von Evaluationen sowie eigene Anregungen und erprobte Lösungswege diskutiert. Im Anschluss an ein Referat von Christian Hirschi, der Ergebnisse einer Metaevaluation zu Maßnahmen gegen Rechtsextremismus<sup>5</sup> vorstellte (vgl. Hirschi 2006), wurde die Frage nach Spezifika im Feld der Evaluation präventiver Arbeit gegen Rechtsextremismus erörtert. Ausgewählte Ergebnisse des Workshops fließen kursorisch in die folgenden Darstellungen mit ein.

Bevor nun zentrale Ergebnisse der untersuchten Evaluationen zu den Kategorien angewandte Evaluationsansätze und -designs, partizipative Evaluation und Wirkungsfeststellung vorgestellt und Differenzen in Positionen und diskrepante Erfahrungen geschildert werden, sollen die Evaluationen skizziert werden, zu denen Informationen aus den Interviews vorliegen, u.a. um die Verschiedenheit und Komplexität der Fälle zu illustrieren<sup>6</sup>.

In drei Fällen werden einzelne Projekte evaluiert, die entweder in Form von Projektwochen an einer Einrichtung oder als Curricula an verschiedenen Einrichtungen umgesetzt werden. Bei einer dieser Evaluationen erheben die Evaluierenden die Daten in erster Linie per Gruppendiskussion mit Jugendlichen und zusätzlich durch Beobachtungen bei Projekttagen. Die Daten werden bei der zweiten Evaluation durch Telefoninterviews mit

- 
- 4 Wir bedanken uns an dieser Stelle recht herzlich bei Dr. Bettina Fritzsche, Christian Hirschi, Jochen Kehr, Susanne Klingelhöfer, Ralf Kuhnke, Gunnar Lemmer, Dr. Christian Lüders, Veronika Lukas, Mareike Schmidt und Karsten Speck für ihre konstruktiven Diskussionsbeiträge und Schilderungen eigener Praxiserfahrungen.
  - 5 Diese Untersuchung wurde gemeinsam mit Dr. Thomas Widmer am Institut für Politikwissenschaft der Universität Zürich im Rahmen des Nationalen Forschungsprogramms 40+ „Rechtsextremismus Ursachen und Gegenmaßnahmen“ realisiert und durch den Schweizer Nationalfonds gefördert.
  - 6 Die folgende Darstellung orientiert sich an einer von Peter Rieker verfassten Charakterisierung der untersuchten Evaluationen.

Lehrerinnen und Lehrern und anhand moderierter Gruppensitzungen mit Schülerinnen und Schülern generiert. Eine dritte Evaluation setzt einen sehr breiten Methodenmix ein: Vorher-Nachher-Befragungen per Fragebogen sowie qualitative Interviews mit Schülerinnen und Schülern, die extreme Positionen vertreten, Interviews, Fragebogenerhebungen, Protokolle und Auswertungsgespräche mit Lehrerinnen und Lehrern sowie Unterrichtsbeobachtungen durch das Evaluationsteam. Die befragten Evaluatorinnen und Evaluatoren sind in einem Fall auch diejenigen, die das Programm entwickelt haben, dessen Umsetzung sie untersuchen; sie haben die Lehrenden entsprechend fortgebildet und evaluieren nun die Umsetzung des Programms. In einem anderen Fall interviewten wir ein Evaluationsteam, dessen Tätigkeit sich auf die Konzeption und Umsetzung der Evaluation beschränkt. Bei der dritten Evaluation sprachen wir mit einem Kollegen, der in einem stark arbeitsteiligen Evaluationsprozess für die Konzeption und datentechnische Fragen verantwortlich war, mit der Umsetzung der Evaluation allerdings nichts zu tun hatte.

In einem vierten Fall wird ein großes, heterogenes, auf mehrere Jahre angelegtes Programm evaluiert. Die Programmevaluation zielt auf die geförderten Projekte. Die Programmsteuerung wird dabei nicht berücksichtigt. Anhand eines standardisierten Fragebogens werden in regelmäßigen Zeitabständen Daten zu allen geförderten Projekten erhoben. Darüber hinaus werden Projektverantwortliche ausgewählter Projekte in Interviews befragt; in einigen Fällen interviewen die Evaluierenden auch jugendliche Teilnehmerinnen und Teilnehmer und führen teilnehmende Beobachtungen durch.

In einem fünften Fall werden eine begrenzte Anzahl sehr verschiedener Interventionen, die innerhalb eines bestimmten Zeitraums auf ganz unterschiedlicher Grundlage durchgeführt werden, aber nicht Bestandteile eines gemeinsamen Programms oder Projektes sind, evaluiert. Das Evaluationsteam führt Interviews mit den für diese Interventionen Verantwortlichen sowie mit lokalen und überregionalen Fachleuten, die diese Vorhaben und die durch sie angeregten Prozesse beurteilen sollen. Während die zuvor genannten Evaluationen sich um präventive Maßnahmen, also vor allem um die Förderung von Demokratie, Toleranz und Vielfalt bemühen, bezieht diese fünfte Evaluation sich auf Interventionen, die sich zumeist gegen manifeste Ausprägungen von Rechtsextremismus richten.

### **Improvisierte Ansätze und aufwendige Designs**

In unserem Material wird eine Vielfalt an Evaluationsansätzen und Designs sichtbar, die durch ein großes Spektrum an eingesetzten Methoden, theoretischen Bezugspunkten und involvierten Zielgruppen gekennzeichnet ist. Verschiedentlich betonen unsere Gesprächspartnerinnen und -partner, dass sie sich bei ihren Evaluationen nicht an bestehenden Evaluationsmodellen oder -verfahren orientieren, sondern eher improvisieren, sich zum Teil an anderen Evaluationen orientieren und auf diese Weise eigene Ansätze zusammenstellen. Beispielsweise wird einmal das Evaluationsdesign auf der Basis eines erprobten Verfahrens der empirischen Sozialforschung entworfen, ein anderes Mal werden gängige Methoden der Sozialforschung mit pädagogischen Methoden kombiniert, so

dass ein neues, innovatives Design entsteht. In einer anderen Evaluation kommen bis zu sechs verschiedene methodische Zugänge zum Einsatz. In der Regel werden im Rahmen einer Evaluation drei, mindestens zwei unterschiedliche Methoden zur Datengewinnung kombiniert – oder wenn nur eine Methode (z. B. Interview) eingesetzt wird, wird diese aber mindestens mit zwei, in der Regel mit drei unterschiedlichen Zielgruppen durchgeführt. Hier wird ein starkes Bemühen sichtbar, unterschiedliche Perspektiven von Akteuren und differente Materialien so weit wie möglich in die Evaluation zu integrieren.

Im Zusammenhang mit der Frage nach dem gewählten Design berichten uns unsere Gesprächspartnerinnen und -partner von großen Mengen an Datenmaterial. In drei Fällen wird beschrieben, dass zu viele Daten erhoben wurden, die nicht angemessen ausgewertet werden können. Zusätzlich benennen zwei befragte Teams das Zusammenspielen der mittels verschiedener Methoden generierten Daten als äußerst wichtigen, aber auch sehr komplexen Prozess. In nur einem Fall wird die Datengrundlage als zu schmal empfunden. Es entsteht der Eindruck, dass gelegentlich versucht wird, mittels „überfrachteter“ Methodensettings das hochkomplexe evaluative Gefüge, bestehend aus Erwartungen der Auftraggeber, der Betroffenen und Beteiligten, der Evaluierenden sowie einer Unsicherheit in Bezug auf die „richtige“ methodische und theoretische Anlage der Evaluation, in den Griff zu bekommen. Die praktische Konsequenz dieser Überfrachtung ist, dass am Ende zu wenig Zeit und Ressourcen bleiben, um das ganze Material sorgfältig auswerten zu können.

Neben den vielfältigen Herangehensweisen fällt auf, dass in der Regel innerhalb einer Evaluation verschiedene Ziele verfolgt werden. Es soll sowohl etwas über Wirkungen und Ergebnisse in Erfahrung gebracht werden, als auch etwas über den Prozess der Umsetzung. Dabei fallen die Zielformulierungen sehr global aus und sind nur schwach operationalisiert. In der Mehrheit der Fälle sind die Ziele der Evaluation an den Zielen der Projekte orientiert. In einem Fall geben die Interviewten an, dass die Zielvorstellungen zwischen den Hauptbeteiligten ungeklärt und für die Interviewten widersprüchlich seien und man unterschiedliche Erwartungen mit der Evaluation verbinde. In einem anderen Fall wird die Evaluationszielsetzung unabhängig von den Zielen der Interventionen aus einer professionsorientierten Perspektive, nämlich der Sozialen Arbeit formuliert. In der Regel aber wird nicht ersichtlich, von wem die Ziele stammen und wer an ihrer Formulierung beteiligt ist.

Insgesamt zeichnet sich ein Bild von Evaluationen, die mehrheitlich formativ angelegt sind und in der Regel qualitative Methoden verwenden. Vermutlich weil es sich bei den Gegenständen um Modellprojekte und -programme handelt, die einen Entwicklungs- und Reifeprozess durchlaufen und deswegen eher auf Prozessdaten angewiesen sind als auf bilanzierende Ergebnisse (vgl. Lüders/Haubrich 2006, S. 16). Gleichzeitig sehen sich die Evaluierenden mit der häufig von Auftraggebern kommenden Anforderung konfrontiert, trotz dieser „Unreife“ der Programme, Aussagen zu deren Qualität und Wirksamkeit zu treffen. Die Schwierigkeiten, die sich daraus ergeben, zeigen sich zuallererst in der Praxis: Die Evaluierenden stehen vor erheblichen Anforderungen: Einerseits sollen sie Ergebnisse für die am Projekt Beteiligten liefern, um Programme zu verbessern, gleichzeitig Aussagen zur

Wirksamkeit der Maßnahmen treffen, um den politischen Erwartungen zu entsprechen und darüber hinaus sollen die dabei angewandten Methoden und Verfahren wissenschaftlichen Ansprüchen genügen. Der Vermittlung und Positionierung zwischen und in diesen verschiedenen Welten mit ihren je eigenen Ansprüchen und Logiken: Der Praxis, der Wissenschaft und der Auftraggeber- bzw. Fördererseite kommt im Evaluationsprozess und gerade für die Evaluierenden selbst eine entscheidende Bedeutung zu. (Zu den Herausforderungen mit denen Evaluierende in diesem komplexen Gefüge sozialer Beziehungen konfrontiert sind vgl. Rieker i. d. Bd.)

### **Partizipative Evaluation in der präventiven Arbeit gegen Rechtsextremismus**

Mitsprache und Mitbestimmung in Evaluationsprozessen ist keine junge Forderung an die Gestaltung von Evaluationen. Erste Ansätze „partizipativer“ Evaluation zeigten sich in den Vereinigten Staaten seit Beginn der 1970er Jahre (vgl. Dimbath/Schneider 2006, S. 111). Ausgehend von der Kritik an experimentellen Designs entwickelten sich Evaluationsansätze, die zunehmend den sozialen Kontext und die Wertvorstellungen der Beteiligten und Betroffenen (Stakeholder) mit in ihren Untersuchungsplan einbeziehen<sup>7</sup>. Ein Konzept partizipativer Evaluation für den Bereich der politischen Bildung haben Ulrich und Wenzel für den deutschsprachigen Raum entworfen (vgl. Ulrich/Wenzel 2003, S. 16; Ulrich/Wenzel 2004). In Anlehnung an die von Guba und Lincoln entwickelte „Fourth Generation Evaluation“ fordern die Verfasser, Beteiligte und Betroffene an einer Evaluation so weitreichend wie möglich am Evaluationsgeschehen zu beteiligen. So sollen nicht nur im Sinne herkömmlicher Evaluation unterschiedliche Beteiligte und Betroffene als Informationslieferanten einbezogen sondern – dem konstruktivistischen Forschungsparadigma folgend – grundlegend an der Gestaltung der Evaluation beteiligt werden (vgl. Klawe 2006). Fragestellung, Design, Datenerhebung und Auswertung sind in diesem Konzept Produkte eines gemeinsamen kommunikativen Aushandlungsprozesses für dessen Gelingen die Evaluierenden verantwortlich sind

In unseren Interviews werden Partizipation und Nutzenorientierung als bedeutsame Charakteristika mehrfach betont. Die Wahl des partizipativen Zugangs wird zum einen damit begründet, dass die Verwendung der Evaluationsergebnisse befördert würde; zum anderen wird angegeben, dass der Gegenstand der Evaluation für diese Ausrichtung eine entscheidende Rolle spiele. So erklärt ein Interviewpartner bspw., dass gerade der für Toleranzerziehung und Bekämpfung von Fremdenfeindlichkeit konstituierende Aspekt der Akzeptanz unterschiedlicher Interessen und Sichtweisen auch für die zugehörige Evaluation bestimmend sein müsse.

In denjenigen Fällen, in denen uns ausführlich von partizipativen Prozessen berichtet wird, handelt es sich einmal um regelmäßige Treffen einer sog. „Steuerungsgruppe“, der

---

<sup>7</sup> Zu nennen sind hier bspw. das CIPP-Modell nach Daniel Stufflebeam, der Countenance-Ansatz, der auf Robert E. Stake zurückgeht, der Stakeholder-Ansatz sowie das Konzept der Empowerment-Evaluation nach David M. Fetterman (vgl. ausführlicher Dimbath/Schneider 2006, S. 112ff.)

ausgewählte Vertreterinnen und Vertreter aller Beteiligengruppen (Auswahl durch die Evaluierenden und den Auftraggeber) angehören und in der die Fragestellung und das Design der Evaluation diskutiert und gemeinsam festgelegt werden. Die Datenerhebung und Auswertung liegt hier in den Händen der beauftragten Evaluatorinnen und Evaluatoren. In diesem Fall kommt es nach Angaben des Befragten zu einem Interessenkonflikt zwischen Steuerungsgruppe und Auftraggeber, der zentrale Fragestellungen der Evaluation berührt. Durch eine Ausweitung der Evaluationsfragestellung konnte in diesem Fall der Konflikt gelöst werden. Ein anderes Mal wird zu Beginn der Evaluation in Vorgesprächen mit den Projektumsetzenden das Evaluationsdesign festgelegt, wobei das methodische Verfahren durch die Evaluierenden vorgegeben wird. Anschließend werden in bestimmten zeitlichen Abständen Ergebnisse an die Projektumsetzenden rückgemeldet, um Fragen zu klären und Konsequenzen für die weitere pädagogische und evaluative Arbeit gemeinsam zu besprechen. In zwei Fällen, in denen uns mündliche Informationen vorliegen, ist das Design vorab durch die Evaluierenden festgelegt und anschließend werden die Projektdurchführenden über Zweck und Verlauf der Evaluation informiert. Bei einer dieser Evaluationen werden sie zusätzlich für ihre Mitarbeit bezahlt und umgekehrt stellen sich die Evaluierenden für Vorträge oder andere Aktivitäten zur Verfügung. In diesem Fall wird uns davon berichtet, dass dieses Vorgehen bewusst gewählt worden sei, um ein offenes Arbeitsklima herzustellen und um potenzielle Ängste gegenüber der Evaluation auf Seiten der Beteiligten und Betroffenen zu minimieren. Im anderen Fall, der im Sample vertretenen Programmevaluation, werden die Projektdurchführenden an der Erstellung und Aktualisierung der Evaluationsinstrumente regelmäßig beteiligt. Nur in einem Interview erhalten wir keine Angaben darüber, ob und wie Beteiligte und Betroffene an der Evaluation partizipierten.

Unsere Ergebnisse zeigen, dass gerade im Bereich der pädagogischen Auseinandersetzung mit Fremdenfeindlichkeit und Rechtsextremismus und der Förderung von Demokratie und Toleranz, Beteiligungsprozessen in der Evaluation eine besondere Bedeutung zugewiesen wird. Das lässt sich mit dem Gegenstand, auf den sich die Evaluationen beziehen, begründen. Zentrale Ziele und inhärente Gestaltungsprinzipien vieler Angebote und Projekte in diesem Bereich sind nun mal die Förderung von Perspektivenübernahme, Toleranz und Verständnis, unterschiedliche Sichtweisen und Kulturen sowie die Einbindung von Adressatinnen und Adressaten in die pädagogische Arbeit. Die hohe Relevanz partizipativer Evaluation kann daher als Gegenstück zu dieser Orientierung gedeutet werden. Gleichzeitig haben wir aber den Eindruck, dass neben der hohen Bedeutung, der Partizipation in diesem Feld zukommt, praktische Umsetzungsformen noch wenig entwickelt sind. Was im Einzelfall genau unter der Bezeichnung Partizipation verstanden wird und wie dies in der Praxis umgesetzt wird differiert – wie oben beschrieben – stark. Partizipative Evaluation ist auch hier wohl am ehesten – wie Haubrich u.a. schreiben, ein „Ober- und Sammelbegriff für alle Formen und Verfahren der gezielten Einbindung von Beteiligten und Betroffenen einer Evaluation“ (Haubrich u.a. 2005, S. 4). In unseren Fällen reichen diese Einbindungsstrategien von einem schlichten in Kenntnis setzen über den Evaluationsplan, über die Beteiligung am Erstellen von Evaluationsinstrumenten bis hin zu Möglichkeiten

einer gezielten Einflussnahme auf Fragestellung, Zielsetzung und Design der Evaluation. Zukünftig wäre es interessant zu untersuchen, in welcher Hinsicht Partizipationsformen in diesem Feld weiterentwickelt werden können.

Der Einbezug von Beteiligten und Betroffenen bietet nicht nur Chancen, sondern stellt Evaluierende auch vor diverse Schwierigkeiten. Einerseits vermag eine rege Beteiligung von unterschiedlichen Stakeholdern die zukünftige Ergebnisverwendung zu fördern. Die Evaluierenden erhalten Zugang zu relevanten Informationen, Evaluationsinstrumente können verbessert, Kontakte hergestellt (z. B. zu Interviewpartnerinnen und -partnern) und Ergebnisse aus unterschiedlichen Perspektiven diskutiert werden. Andererseits sehen sich die Evaluierenden mit der Schwierigkeit konfrontiert, auftretende Konflikte moderieren zu müssen und viel Zeit in die Kontaktpflege zu investieren. Eng mit der Frage nach den unterschiedlichen Partizipationsformen ist die Frage verbunden, welche neuen Anforderungen in diesem Zusammenhang auf Evaluierende zukommen. Sie müssen zumindest geschult sein in Konfliktmoderation, hervorragende Kommunikationsfähigkeiten mitbringen und im besten Fall über ein breites Repertoire an Moderationstechniken verfügen.

Unsere Ergebnisse deuten insgesamt darauf hin, dass theoretische und konzeptionelle Forderungen nach gezielter Einbindung von Beteiligten und Betroffenen häufig an den Rahmenbedingungen der Praxis scheitern. Zum einen brauchen Partizipationsprozesse Zeit und Personal. Gerade diese beiden Aspekte vermissen unsere Interviewpartnerinnen und -partner häufig. Kommen dann für die Evaluierenden zu ihren meist umfassenden forschungspraktischen Aufgaben zusätzliche Aufgaben wie die Identifikation von Interessenvertreterinnen und -vertretern aus den Reihen der Beteiligten und Betroffenen, die Moderation von Begleit- oder Beiratsgruppen, der kontinuierliche Informationsaustausch und die Rückmeldung von Ergebnissen sowie die gemeinsame Diskussion und Interpretation der erhobenen Daten hinzu, müssten die meisten Evaluationsteams mindestens personell aufgestockt, gezielt nach Aufgabenfeldern differenziert (vgl. Klawe 2006, S. 140) oder Untersuchungszeiten verlängert werden.

### **„Wir haben zwar keine Wirkungen finden können, aber ...“ Zur Wirkungsfrage in der Evaluation präventiver Arbeit gegen Rechtsextremismus**

Wirkungsevaluation ist eine spezielle Form der Evaluation. Bei der Feststellung von Wirkungen geht es in Abgrenzung zur reinen Ergebnisbeobachtung darum, Ergebnisse ursächlich auf ein Projekt, eine Maßnahme oder ein Programm zurückzuführen (vgl. Haubrich u.a. 2005, S. 3) oder mit den Worten der Autoren Beywl, Speer und Kehr: „Resultate qualifizieren sich dann als Wirkungen, wenn sie auf das Programm ursächlich rückführbar sind, also nachvollziehbare theoretische Annahmen sowie empirische Daten über die Verbindung von Maßnahmen/Aktivitäten und Resultaten beigebracht werden“ (Beywl u.a. 2004, S. 18). Diese Begriffsdefinitionen sagen jedoch noch nichts darüber aus wie Wirkungen, insbesondere im Feld der pädagogischen Arbeit gegen Rechtsextremismus, identifiziert und empirisch nachgewiesen werden können. Die Feststellung von Wirkungen im o. g. Sinne,

ist – wie die Fachliteratur zum Thema bestätigt – ein höchst voraussetzungsvoller Prozess (vgl. ebd., S. 113; Projekt eXe 2006).

Die Aussagen, die uns zum Thema Wirkungsfeststellung aus dem Material vorliegen, sind sehr widersprüchlich. Einerseits werden Angaben zu Wirkungen der begleiteten Programme und Projekte gemacht, andererseits stoßen wir häufig auf Fragen, Unsicherheiten und Skepsis, wenn wir nach Einschätzungen zu Möglichkeiten der Wirkungsfeststellung und Beurteilung von Qualität pädagogischer Arbeit fragen. Wirkungsfeststellung gilt zunächst unter den befragten Expertinnen und Experten in den Bereichen, in denen es um weiche Faktoren wie Einstellungsveränderungen, Erhöhung der selbstreflexiven Fähigkeiten sowie Steigerung interkultureller Sensibilität – eben um Bildungs- und Lernprozesse geht, als besonders schwierig. Dabei wird uns einerseits das Identifikationsproblem (Passung zwischen beobachteter Wirkung und Indikator) und andererseits das Rückführbarkeitsproblem (Wirkung auf Programm rückführbar) als Schwierigkeit geschildert. Zwei Aussagen auf die Frage wie versucht wurde, Wirkungen der pädagogischen Maßnahmen nachzuweisen, verdeutlichen die Suchbewegungen der Evaluierenden:

„Also zum einen, die Ziele [der Evaluation, d. Verf.] sind auf dieser Wirkungsebene, was auch immer – wie man die messen kann und ob das immer so sinnvoll ist.“

„Es ist eben schwierig in diesem Bereich, wenn es um Einstellungen geht, wenn es um Reflexion geht, festzulegen wie man die sehen kann (...) wir haben nur Annahmen über Wirkungen erhalten, keine Wirkungen an sich.“

Diese Unsicherheiten und Schwierigkeiten führen dazu, dass die Evaluierenden teilweise ihren Anspruch, Wirkungen zu ermitteln, abschwächen und eher Evaluationen betreiben, die sich der Wirkungsfrage zwar annähern, aber zugleich auch die damit verbundenen Herausforderungen problematisieren. Zweifel gegenüber der Möglichkeit Wirkungen – im Sinne eines empirischen Wirkungsnachweises – von Lern- und Reflexionsprozessen, Orientierungs- und Einstellungsveränderungen festzustellen, durchzieht alle unsere Interviews<sup>8</sup>.

Gleichzeitig wird aber die Untersuchung zu Wirkungen von Maßnahmen, Projekten und Programmen von den befragten Expertinnen und Experten als zentrales Anliegen ihrer Studien genannt. Um Wirkungen nachzuweisen, setzen die Evaluierenden in unserem Sample verschiedene Verfahren ein. In der Regel wird erst über die Befragungen deutlich, in welcher Art und Weise die Evaluierenden zu Bewertungen oder Aussagen über Wirkungen der untersuchten Projekte gelangen. In den Berichten wird wenig Raum dafür verwendet, den

---

<sup>8</sup> Zu ähnlichen Ergebnissen kam eine Studie aus Nordrhein-Westfalen bei der Politische Bildnerinnen und Bildner, also in der Regel Programmanbieter, nach ihren Einschätzungen zu Chancen und Grenzen der Wirkungsfeststellung befragt wurden. Das Ergebnis erbrachte eine tiefe Skepsis gegenüber möglicher Wirkungsforschung bei gleichzeitigem Optimismus und Zuversicht, dass die Angebote trotzdem wirksam wären (vgl. Alheim 2005, S. 22).

Bewertungsprozess ausführlich zu dokumentieren. Die genaueren Konzeptionen dessen, was als Wirkung gefasst wird und deren Operationalisierungen unterscheiden sich in unseren Fällen erheblich voneinander.

In unserem Sample zeigen sich drei grundsätzliche Möglichkeiten, um Aussagen über Wirkungen zu treffen bzw. Bewertungen vorzunehmen:

Die Evaluierenden befragen die Akteure im Feld zu ihrer Wirkungseinschätzung und stützen sich anschließend bei der Interpretation dieser Daten zentral auf diese Aussagen. Dabei handelt es sich in unseren Fällen hauptsächlich um Lehrerinnen und Lehrer, welche die Wirkung auf Schülerinnen und Schüler einschätzen sollen. Einmal werden zusätzlich zu den Einschätzungen der Programmbeteiligten, Auskünfte von externen Expertinnen und Experten über Distanzierungen rechtsextremer Jugendlicher von der Szene und über Abwanderungen von Anführern einbezogen. In diesem Fall wird angenommen, dass Expertinnen und Experten „objektiver“ als Projektakteure bzw. frei von Eigeninteresse antworten. Ein anderes Mal wird auf Grundlage der Wirkungsauskünfte der Projektbeteiligten ein Fragebogen für Schülerinnen und Schüler entwickelt und eingesetzt.

Eine zweite Möglichkeit stellt die ebenfalls datenbasierte, aber unabhängig von den im Feld existierenden Bewertungen, Ermittlung von Wirkungen dar. In unserem Sample werden in diesem Zusammenhang Wirkungen von den Evaluierenden als Irritationen bei Jugendlichen beschrieben, wobei unklar bleibt was genau darunter zu verstehen ist. Über einen längeren Zeitraum werden Gruppendiskussionen mit Jugendlichen geführt mit dem Ziel, Hinweise auf Einstellungsveränderungen zu erhalten. Wie die Interviewten im Nachhinein erwähnen, haben verschiedene Bedingungen und Konstellationen verhindert, dass sie eindeutige Aussagen über Veränderungen in den Orientierungen der Jugendlichen treffen konnten. Gleichzeitig haben die Forschenden mehrere vom Projekt nichtintendierte Effekte beobachten können, was somit doch für hervorgerufene Veränderungen spricht. Zum Thema Irritationen und Nachdenklichkeiten anstelle des Nachweises von Orientierungsveränderungen wird uns Folgendes berichtet:

„Es gibt sehr dichte Passagen zum Besuch in X [einer sozialen Einrichtung für gesellschaftliche Randgruppen, d. Verf.], die sie [die Jugendlichen] sehr beschäftigt haben und glaube ich schon, vieles in Frage gestellt haben von dem, was sie vorher gedacht haben. Allerdings ist es schwierig das sozusagen, als Veränderung im Vergleich zu vorher so klar festzuhalten, weil das sind so Erfahrungen, die dann teilweise auch schon in der ersten Runde erzählt wurden. Und wo dann deutlich wurde, sie haben das irritierend erlebt. Aber ob sich ihre Orientierungen verändert haben, das weiß ich da natürlich nicht. Also ich kann über diese Irritation viel sagen, aber wenn man jetzt Orientierungen als so tief verwurzelte Einstellungen zur Welt, also wo es um ganz grundlegende Sachen geht, wie gehe ich mit Leuten um, die ich anders wahrnehme, als ich selber es bin. Das ist so was Fundamentales, wo es halt wirklich schwierig ist ranzukommen.“

Als eine dritten Möglichkeit, Wirkungen zu ermitteln bzw. Bewertungen vorzunehmen, treffen die Evaluierenden auf der Basis von mit den Projektverantwortlichen gemeinsam entwickelten Indikatoren Einschätzungen zur Qualität und Wirksamkeit einer Maßnahme. Wobei in diesen Fällen der Prozess der Indikatorenbildung als äußerst voraussetzungsvoll und zeitaufwendig geschildert wird.

Auffällig bei allen Evaluationen unseres Samples ist, dass sie ihren eingangs formulierten Anspruch, Wirkungen festzustellen, im Verlauf des Interviews stark relativieren. Außerdem bezieht keine Evaluation theoretische Erkenntnisse in ihre Untersuchung mit ein, um unabhängig von den subjektiven Bewertungen der Akteure im Feld, zu Wirkungsaussagen zu gelangen. Diese Ergebnisse aus der Evaluationspraxis zeigen einmal mehr, dass Wirkungsforschung zum Thema Rechtsextremismusprävention ein Forschungsfeld ist, das praktisch noch ganz am Anfang steht. Sowohl die Frage wie man Wirkungen angemessen beobachten kann als auch die Frage nach den spezifischen Feldbedingungen sind nicht ausreichend beantwortet. Erstere ist zugleich in der gesamten Kinder- und Jugendhilfe eine Frage, um die sich die aktuelle Fachdebatte bemüht (vgl. Lüders/Haubrich 2006; Beywl 2006).

Als Erklärung für die Schwierigkeiten wurden folgende Überlegungen genannt: Die Ermittlung von Wirkungen sei deshalb so kompliziert, weil es sich bei den untersuchten Projekten zumeist um Entwicklungsprojekte handele und in diesen Fällen seien abschließende Aussagen zur Wirksamkeit schlichtweg nicht möglich. Andererseits wird das Ermitteln von Bewertungen als Kunst beschrieben, weil mit jedem Evaluationsprojekt neu ausgehandelt werden müsse, welche Kriterien und Werte man als Grundlage der Evaluation verwende. Um das herauszufinden seien die Interessen und Werte der Programmbeteiligten entscheidend. Den Evaluierenden käme dabei die Aufgabe zu, diesen Abstimmungsprozess zu fördern. Als eine weitere Schwierigkeit wurde der Umgang mit diskrepanten Wirkungsaussagen von Akteuren aufgeführt.

Wir haben es bei der Frage, wie Wirkungen beobachtet werden können, mit einer hoch komplexen und voraussetzungsvollen Aufgabe zu tun. Es zeigt sich, dass sehr verschiedene Auffassungen über angemessene Verfahren und Vorgehensweisen in der Fachwelt herrschen. Und auch in der übergeordneten Fachdebatte zur Wirkungsevaluation finden wir anstelle von Einigkeit vielfältige Verfahren und Ansätze (vgl. Beywl u. a. 2004; Projekt eXe 2006). An einigen Stellen finden sich in unserer Untersuchung Hinweise darauf, dass die Wirkungen eines Projektes, Programms oder einer Maßnahme nicht nur vom Konzept, seiner angewandten Methodik und Vorgehensweisen abhängen, sondern auch durch konkrete Umsetzungsbedingungen wie bspw. die Haltung der Pädagoginnen und Pädagogen oder der Umfang und die Intensität der Umsetzung beeinflusst werden. So haben wir bei zwei Evaluationen Qualitätskriterien gefunden, die den Stellenwert externer Faktoren für den Erfolg oder die Qualität pädagogischer Arbeit gegen Rechtsextremismus betonen. Dabei handelt es sich in beiden Fällen um die Faktoren Qualifikation und/oder pädagogisches Know-how bzw. Fingerspitzengefühl der Pädagoginnen und Pädagogen sowie um die Möglichkeit einer kontinuierlichen und langfristigen Arbeit mit derselben Gruppe. In der Untersuchung von Glaser und Rieker zu Angeboten Interkulturellen Lernens werden

ebenfalls als Qualitätsbedingungen die Faktoren „Dauer und Kontinuität“ sowie die „hohen fachlichen und persönlichen Anforderungen“, die an die Umsetzenden solcher Angebote gestellt werden, aufgeführt (vgl. Glaser/Rieker 2006, S. 97 u.S. 108). Wenn dem so ist und persönliche Kompetenzen von Pädagoginnen und Pädagogen neben deren fachlichen Fähigkeiten eine entscheidende Rolle bei der wirksamen Umsetzung von Maßnahmen und Lernprozesse gegen Fremdenfeindlichkeit und Rechtsextremismus zukommen, dann stellt sich hier für die Evaluation dieser Angebote, die in der Regel auf die Übertragbarkeit von erfolgreichen Strategien, Konzepten und Modellen zielt die Frage, wie mit dieser Herausforderung zukünftig umzugehen ist. So müsste man bspw. zunächst genauer untersuchen, welche persönlichen und fachlichen Kompetenzen förderlich für die Umsetzung pädagogischer Angebote gegen Rechtsextremismus sind, um dann zu klären, welche professionellen Anforderungen sich daraus in der Rechtsextremismusprävention ergeben<sup>9</sup>. Erst in einem zweiten Schritt wäre dann die Überlegung anzustellen, was das für die Evaluation in diesem Feld bedeutet.

## Fazit

Evaluatorinnen und Evaluatoren haben die anspruchsvolle Aufgabe, die hohe Relevanz und die großen Erwartungen, die mit Evaluationen einhergehen, in realistische und umsetzbare Verfahren zu bringen. Dazu bedarf es entsprechender Kompetenzen und Rahmenbedingungen. In unseren analysierten Fällen ist das vor allem dann gelungen, wenn die Evaluierenden über ausreichende Vorerfahrungen in der Evaluation verfügten, der Evaluationsgegenstand gut abgegrenzt und entsprechend definiert war, das Evaluationsdesign nicht zu aufwendig konzipiert, ausreichende personelle wie auch zeitliche Ressourcen zur Verfügung standen sowie Beteiligte und Betroffene so weit wie möglich am Evaluationsprozess beteiligt wurden. In denjenigen Fällen, in denen einzelne oder mehrere dieser Bedingungen nicht erfüllt waren bzw. nicht erfüllt werden konnten, zeigten sich Schwierigkeiten und Umsetzungsprobleme.

Wie unsere Ergebnisse zeigen, bergen vor allem die widersprüchlichen und teilweise divergierenden Anforderungen, denen Evaluierende ausgesetzt sind, erhebliche Schwierigkeiten für eine gelingende Umsetzung von Evaluationen. Deshalb wäre es wünschenswert, dass die Ansprüche an Evaluationen im Vorfeld besser und realistischer eingeschätzt werden. Gleichzeitig sollte aber auch über Wege nachgedacht werden, welchen Beitrag Evaluierende selbst zu dieser Klärung leisten können. Gerade sehr aufwendige und überladene Evaluationsdesigns sind eine Folge hoher und widersprüchlicher Anforderungen an Evaluationen. Hier wäre es bspw. zu begrüßen, wenn zukünftig über wissenschaftliche Standards zur angemessenen Umsetzung von Evaluationen nachgedacht würde.

---

<sup>9</sup> In Bezug auf Angebote Interkulturellen Lernens arbeiten z. B. Michaela Glaser und Peter Rieker in ihrer Studie einige solcher Kompetenzen heraus (vgl. dies. 2006, S. 102–106).

Partizipativer Evaluation kommt gerade im Bereich der Prävention von Rechtsextremismus eine hohe Bedeutung zu, diese ist aber nicht immer einfach einzulösen. Aus der Forderung nach umfassender Partizipation ergeben sich für Evaluierende hohe Anforderungen an deren Qualifikation. Im Fall der Beteiligung unterschiedlicher Stakeholder am Evaluationsprozess müssen von Seiten der Evaluierenden bspw. Interessenskonflikte moderiert und viel Zeit in die Kontaktpflege investiert werden. Diese sollten dann über das notwendige Handwerkzeug verfügen und Kenntnisse in Konfliktmoderation sowie gute Kommunikations- und Moderationsfähigkeiten mitbringen.

Die Heterogenität und Vielfältigkeit der verwendeten Evaluationsansätze und -modelle bergen ein großes Potenzial in sich. Demgegenüber tragen die häufige Intransparenz von Bewertungsprozessen und die zum Teil ungünstigen Rahmenbedingungen bei Evaluationen nicht dazu bei, in dem Feld vergleichbare oder gar übertragbare Ergebnisse zu produzieren. Deshalb sollten zukünftig im Rahmen von Evaluationen die Grundlagen und Vorgänge von Bewertungsprozessen systematisch dokumentiert und stärker offengelegt werden, so dass die Qualität und Reichweite von Evaluationsergebnissen einzuschätzen ist.

Letztlich sieht es so aus, dass Evaluation pädagogischer Arbeit gegen Rechtsextremismus ein Bereich des Ausprobierens und Suchens ist, in dem im stärkeren Maße Debatten über Begrifflichkeiten und Verfahren zu führen sind. Erste Schritte, um eine Kultur der Evaluation zu befördern, wären bspw. die inhaltliche Qualifikation der Evaluierenden zu verbessern und den Fachaustausch zu stärken und Schwächen zu fördern. Gerade eine verstärkte fachliche Diskussion über Chancen und Grenzen bspw. zum Thema Wirkungsevaluation oder zu Erfahrungen mit partizipativen Ansätzen in diesem Feld, könnte Impulse zu einer Selbstklärung liefern und helfen, zukünftig die Position von Evaluierenden zu stärken und dazu beitragen, eine Kultur der Evaluation in diesem Feld zu stärken.

## Literaturverzeichnis

- Alheim, Klaus (2005): Wirkungsoptimismus und Messbarkeitskepsis. Evaluation und Wirkungsforschung in der politischen Erwachsenenbildung. In: Praxis Politische Bildung, 9. Jg., H. 1, S. 17–25
- Beywl, Wolfgang/Speer Sandra/Kehr, Jochen (2004): Wirkungsorientierte Evaluation im Rahmen der Armuts- und Reichtumsberichterstattung. Perspektivstudie im Auftrag des Bundesministeriums für Gesundheit und Soziale Sicherung und durchgeführt von Univation – Institut für Evaluation in Köln. [www.univation.org/download/Evaluation\\_der\\_Armuts-\\_und\\_Reichtumsberichterstattung.pdf](http://www.univation.org/download/Evaluation_der_Armuts-_und_Reichtumsberichterstattung.pdf), (4.07.2007)
- Beywl, Wolfgang (2006): Demokratie braucht wirkungsorientierte Evaluation – Entwicklungspfade im Kontext der Kinder- und Jugendhilfe. In: Projekt eXe (Hrsg.): Wirkungsevaluation in der Kinder- und Jugendhilfe. Einblicke in die Evaluationspraxis. München: DJI, S. 25–48
- Dimbath, Oliver/Schneider, Werner (2006): Partizipative Evaluation in der politischen Bildung. In: Zeitschrift für Evaluation. H. 1, S.109–134

- Eser Davolio, Miryam (2000): Fremdenfeindlichkeit, Rassismus und Gewalt. Festgefahrenes durch Projektunterricht verändern, Bern/Stuttgart/Wien: Haupt Verlag
- Glaser, Michaela/Rieker, Peter (2006): Interkulturelles Lernen als Prävention von Fremdenfeindlichkeit. Ansätze und Erfahrungen in Jugendbildung und Jugendarbeit. Halle: DJI
- Haubrich, Karin/Holthusen, Bernd/Struhkamp, Gerlinde (2005): Evaluation – einige Sortierungen zu einem schillernden Begriff. In: DJI-Bulletin 72plus
- Haubrich, Karin/Lüders, Christian (2004): Evaluation – mehr als ein Modewort? In: Recht der Jugend und des Bildungswesens. Jg. 52, H. 3, S. 316–337
- Hirschi, Christian (2006) unveröffentlichtes Vortragsmanuskript „Maßnahmen gegen Rechtsextremismus in der Schweiz: Evaluation und ihre Reichweite“ vorgelegt zum Expertenworkshop der Arbeitsstelle Rechtsextremismus und Fremdenfeindlichkeit am 9. November 2006 in Halle.
- Klawe, Willy (2006): Multiperspektivische Evaluationsforschung als Prozess – Wirkungsrekonstruktion aus der Sicht der Beteiligten. In: Projekt eXe (Hrsg.): Wirkungsevaluation in der Kinder- und Jugendhilfe. Einblicke in die Evaluationspraxis. München: DJI, S. 125–142
- Klingelhöfer, Susanne/Schmidt, Mareike/Schuster, Silke/Brüggemann, Ulrich (2006): Ergebnisbericht zu den qualitativen Analysen des Programms „Entimon – gemeinsam gegen Gewalt und Rechtsextremismus“, 2002 bis 2006. Unveröff. Sachstandbericht. Halle: DJI
- Lynen von Berg, Heinz/Palloks, Kerstin/Steil, Armin (2007): Interventionsfeld Gemeinwesen. Evaluation zivilgesellschaftlicher Strategien gegen Rechtsextremismus. Weinheim/München: Juventa
- Lüders, Christian/Haubrich, Karin (2006): Wirkungsevaluation in der Kinder- und Jugendhilfe: Über hohe Erwartungen, fachliche Erfordernisse und konzeptionelle Antworten. In: Projekt eXe (Hrsg.) Wirkungsevaluation in der Kinder- und Jugendhilfe. Einblicke in die Evaluationspraxis. München: DJI, S. 5–24
- Projekt eXe (Hrsg.) (2006): Wirkungsevaluation in der Kinder- und Jugendhilfe. Einblicke in die Evaluationspraxis. München: DJI
- Schrödter, Mark (2004): Die Einheit interkultureller Jugendhilfe. Strukturbestimmung und Rekonstruktion. Dissertation Universität Bielefeld
- Ulrich, Susanne/Wenzel, Florian M. (Hrsg.) (2003): Partizipative Evaluation. Ein Konzept für die politische Bildung. Gütersloh: Verlag Bertelsmann Stiftung
- Ulrich, Susanne/Wenzel, Florian M. (2004): Partizipative Evaluation. In: Uhl, Katrin/Ulrich, S./Wenzel, F. M. (Hrsg.) Evaluation politischer Bildung. Ist Wirkung messbar? Gütersloh: Verlag Bertelsmann Stiftung, S. 27–48

## Autorinnen- und Autorenverzeichnis

**Ralf Bohnsack**, Prof. Dr. rer. soc., Dr. phil. habil., Dipl.-Soziologe, Arbeitsfelder: Rekonstruktive Sozialforschung, praxeologische Wissenssoziologie, dokumentarische Methode, Gesprächsanalyse, Bild- und Filminterpretation, Evaluationsforschung, Milieu-, Generations-, Jugend- und Devianzforschung. Freie Universität Berlin, Leiter des Arbeitsbereichs Qualitative Bildungsforschung, E-Mail: bohnsack@zedat.fu-berlin.de

**Matthias Drilling**, Prof. Dr., Sozialgeograph, Forschungen zu den Themen Gewalt und Armut im städtischen Kontext sowie Kooperationen von Schule und Sozialer Arbeit. Hochschule für Soziale Arbeit der Fachhochschule Nordwestschweiz, Institut für Sozialplanung und Stadtentwicklung, E-Mail: matthias.drilling@fhnw.ch

**Miryam Eser Davolio**, Dr., Erziehungswissenschaftlerin, Forschung zu Rechtsextremismus, Fremdenfeindlichkeit, politische Bildung, Integration und Gewaltprävention. Hochschule für Soziale Arbeit der Fachhochschule Nordwestschweiz, Institut für Sozialplanung und Stadtentwicklung, E-Mail: miryam.eser@fhnw.ch

**Bettina Fritzsche**, Dr., Erziehungswissenschaftlerin, Forschung zu qualitativen Methoden, Evaluationsforschung, Jugendforschung, Geschlechterforschung. Zurzeit Gastprofessur „Allgemeine und historische Erziehungswissenschaft“ am Institut für Erziehungswissenschaft, Technische Universität Berlin, E-Mail: bettina.fritzsche@tu-berlin.de

**Michaela Glaser**, Soziologin M.A., Arbeits- und Forschungsfelder: Wissenschaftliche Begleitforschung, Prävention von Fremdenfeindlichkeit, Rechtsextremismus und Antisemitismus, Interkulturelles Lernen. Wissenschaftliche Referentin der Arbeits- und Forschungsstelle Rechtsextremismus und Fremdenfeindlichkeit am Deutschen Jugendinstitut, Außenstelle Halle, E-Mail: glaser@dji.de

**Haubrich, Karin**, Dipl.-Soziologin, Arbeitsschwerpunkte: Evaluation in der Kinder- und Jugendhilfe, Cluster-Evaluation, Programmtheorie-Evaluation, theoretische Grundlagen der Evaluation. Mitglied des Vorstandes der DeGEval – Gesellschaft für Evaluation; Wissenschaftliche Referentin am Deutschen Jugendinstitut in München im Projekt „Strategien und Konzepte externer Evaluation in der Kinder- und Jugendhilfe (eXe)“, E-Mail: haubrich@dji.de

**Susanne Klingelhöfer**, Diplom-Pädagogin und Romanistin. Arbeits- und Forschungsfelder: Programmevaluation, interkulturelles und interreligiöses Lernen. Wissenschaftliche Mitarbeit in der Programmevaluation „Vielfalt tut gut“ und „kompetent. für Demokratie“ am Deutschen Jugendinstitut, Außenstelle Halle, E-Mail: klingelhoefer@dji.de

**Helmut Kromrey**, Prof. (i. R.) Dr., Sozialwissenschaftler, Forschung zu berufliche Bildung/ Weiterbildung, Neue Medien, Städtebau; Arbeitsschwerpunkte: Methodologie empirischer Forschung, Evaluation. Freie Universität Berlin, Adjunct Professor of Sociology an der Graduate School of Management der Universität Educatis (Altdorf/Schweiz), E-Mail: mail@kromrey.de

**Christian Lüders**, Dr., Erziehungswissenschaftler, Forschungsschwerpunkte: Kinder- und Jugendhilfeforschung, Evaluationsforschung, qualitative Sozialforschung, Theorie der Sozialpädagogik. Leiter der Abteilung Jugend und Jugendhilfe am Deutschen Jugendinstitut, München, E-Mail: lueders@dji.de

**Kerstin Palloks**, Dipl. Sozialwissenschaftlerin; Arbeitsfelder: Evaluation und Evaluationsforschung, Forschungen zu Armut, Arbeitsmarktintegration, Gewaltprävention und Rechtsextremismus. Freiberufliche Evaluatorin und Beraterin für Praxisprojekte; bis Frühjahr 2007 wissenschaftliche Mitarbeiterin am IKG der Universität Bielefeld im Projekt „Wissenschaftliche Begleitforschung des CIVITAS-Programms“, E-Mail: kpalloks@web.de

**Peter Rieker**, PD Dr., Soziologe, Forschung zu Rechtsextremismus, Fremdenfeindlichkeit, abweichendem Verhalten, Migration und interethnischen Beziehungen. Projektleiter der Arbeits- und Forschungsstelle Rechtsextremismus und Fremdenfeindlichkeit sowie der Programmevaluation „Vielfalt tut gut“ und „kompetent. für Demokratie“ am Deutschen Jugendinstitut, Außenstelle Halle, E-Mail: rieber@dji.de

**Stephan Schack**, Dipl. Sozialpädagoge (FH), Trainer und Berater für Demokratiepädagogik, Interkulturelle Verständigung und Partizipation, Projektmanager im Projekt „mitWirkung! – eine Initiative zur Stärkung der Kinder- und Jugendpartizipation der Bertelsmann Stiftung, Gütersloh, E-Mail: stephan.schack@gmx.de

**Erich Schäfer**, Prof. Dr., Soziologe, Lehre und Forschung in den Feldern außerschulische Jugend- und Erwachsenenbildung, Medienpädagogik sowie wissenschaftliche Weiterbildung. Fachhochschule Jena, Fachbereich Sozialwesen, E-Mail: erich.schaefer@fh-jena.de

**Silke Schuster**, Dipl. Pädagogin, Arbeitsfelder: Evaluation und wissenschaftliche Begleitforschung, Jugend- und Erwachsenenbildung, Interkulturelles und antirassistisches Lernen. Freie Wissenschaftlerin und Autorin, ehem. wissenschaftliche Referentin am Deutschen Jugendinstitut, E-Mail: schuster.silke@gmx.de

