

Rainer Strobl · Olaf Lobermeier
Wilhelm Heitmeyer (Hrsg.)

Evaluation von Programmen und Projekten für eine demokratische Kultur

ARBEIT GRENZEN POLITIK HANDLUNG METHODEN GEWALT SPRACHE WISSEN
SCHAFT DISKURS SCHICHT MOBILITÄT SYSTEM INDIVIDUUM KONTROLLE
ZEIT ELITE KOMMUNIKATION WIRTSCHAFT GERECHTIGKEIT STADT WERTE
RISIKO ERZIEHUNG GESELLSCHAFT RELIGION UMWELT SOZIALISATION
RATIONALITÄT VERANTWORTUNG MACHT PROZESS LEBENSSTIL MELIN
QUENZ KUNST UNGLEICHHEIT ORGANISATION NORMEN REGULIERUNG
IDENTITÄT HERRSCHAFT VERGLEICH SOZIALSTRUKTUR BIOGRAFIE KRITIK
WISSEN MASSEN MEDIEN EXKLUSION GENERATION THEORIE HIERARCHIE
GESUNDHEIT NETZWERK LEBENS LAUF KONSUM FREIHEIT BETEILIGUNG
GEMEINSCHAFT INFORMATION WANDEL DIFFERENZ WOHLFAHRTSSTAAT
ETHNIE BERUF RITUAL KÖRPER MODERNISIERUNG GESCHLECHT DEMOKRA
TIE EVOLUTION INTEGRATION KAPITAL REALITÄT KRIEG BILDUNG ALLTAG
KULTUR VERTRAUEN LIEBE WERBUNG GLOBALISIERUNG BEOBACHTUNG
RECHT EXTREMISMUS STATISTIK INTERAKTION KRIMINALITÄT ZUKUNFT
ALTER ERKENNTNIS MORAL RAUM KLASSE STEUERUNG GELD ZIVILISATION
EMPIRIE AUFKLÄRUNG ARMUT ENTSCHEIDUNG TECHNIK MIGRATION ÖFFENT



Springer VS

Evaluation von Programmen und Projekten für eine demokratische Kultur

Rainer Strobl • Olaf Lobermeier
Wilhelm Heitmeyer (Hrsg.)

Evaluation von Pro- grammen und Projekten für eine demokratische Kultur

Herausgeber

Rainer Strobl,
Olaf Lobermeier,
Hannover, Deutschland

Wilhelm Heitmeyer
Bielefeld, Deutschland

ISBN 978-3-531-18306-0

ISBN 978-3-531-19009-9 (eBook)

DOI 10.1007/978-3-531-19009-9

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

Springer VS

© Springer Fachmedien Wiesbaden 2012

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung, die nicht ausdrücklich vom Urheberrechtsgesetz zugelassen ist, bedarf der vorherigen Zustimmung des Verlags. Das gilt insbesondere für Vervielfältigungen, Bearbeitungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in diesem Werk berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, dass solche Namen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürften.

Gedruckt auf säurefreiem und chlorfrei gebleichtem Papier

Springer VS ist eine Marke von Springer DE. Springer DE ist Teil der Fachverlagsgruppe Springer Science+Business Media
www.springer-vs.de

Inhalt

<i>Rainer Strobl, Olaf Lobermeier und Wilhelm Heitmeyer</i> Einleitung.....	7
--	---

I. Evaluationsansätze und Methoden

<i>Huey T. Chen</i> Theory-driven evaluation: Conceptual framework, application and advancement	17
---	----

<i>Thomas Widmer</i> Wirkungsevaluation zu Maßnahmen der Demokratieförderung	41
---	----

<i>Manuel Eisner, Tina Malti, Denis Ribeaud, Barbara Müller</i> Groß angelegte Feldversuche in der kriminologischen Präventionsforschung. Das Zürcher Projekt zur sozialen Entwicklung von Kindern	69
---	----

II. Nutzungsorientierte Evaluationsstrategien im Bereich der Förderung demokratischer Kultur

<i>Wolfgang Beywl, Susanne Giel</i> Nutzungsfokussierte Evaluation am Beispiel eines multizentrischen Programms.....	101
--	-----

<i>Kerstin Sischka</i> Zum Nutzen eines konfliktensiblen Ansatzes in der wissenschaftlichen Projektbegleitung.....	127
--	-----

<i>Rainer Strobl und Olaf Lobermeier</i> Aufbau und Evaluation von Netzwerken für eine demokratische Kultur	155
---	-----

<i>Gesa Schirrmacher</i> Was sich Auftraggebende von Evaluationen und Evaluationspraxis wünschen	179
--	-----

III. Die gesellschaftliche Bedeutung von Evaluation

<i>Reinhard Stockmann</i> Evaluation und Gesellschaft	195
<i>Wilhelm Heitmeyer</i> Evaluationsforschung. Parameter und Auswirkungen für die Gesellschaft	221
Autorinnen und Autoren	233

Einleitung

Rainer Strobl, Olaf Lobermeier und Wilhelm Heitmeyer

Eine attraktive demokratische Kultur ist keine Selbstverständlichkeit. Vielmehr müssen die Bedeutung und die Qualität demokratischer Werte und Normen immer wieder begreifbar und erlebbar gemacht werden, um zu verhindern, dass sich Teile der Bevölkerung vom demokratischen System abwenden. Besonders drastisch zeigt sich die Abwendung von demokratischen Werten und Normen bei denjenigen, die aus rechtsextremistischen Motiven Straf- und Gewalttaten begehen. Der Anziehungskraft rechtsextremer Gruppen mit ihrer Musik und ihren Symbolen auf Jugendliche stehen die lokalen Institutionen und Akteure allerdings oft hilflos gegenüber, weil ihnen wirksame Gegenmaßnahmen fehlen. Ebenso wichtig wären jedoch effektive Strategien, um einer schleichenden Distanzierung von demokratischen Werten und Normen zu begegnen (vgl. Strobl/Lobermeier 2009).

Ob und wie Evaluationen zur Identifizierung, Bewertung und Weiterentwicklung solcher Strategien beitragen können, war Thema der Konferenz „Evaluation von Programmen und Projekten zur Förderung einer pluralistischen und demokratischen Kultur“, die von der proVal Gesellschaft für sozialwissenschaftliche Analyse, Beratung und Evaluation und dem Institut für interdisziplinäre Konflikt- und Gewaltforschung der Universität Bielefeld gemeinsam veranstaltet wurde. Die Konferenz fand vom 16.-18.2.2011 im Zentrum für interdisziplinäre Forschung der Universität Bielefeld statt und wurde von der Deutschen Forschungsgemeinschaft (DFG) gefördert. Die Beiträge in dem vorliegenden Band geben einen Überblick über die für Evaluationen in diesem Themenfeld wichtigen Diskussionsstränge.

Die im Fokus des Bandes stehenden Programme und Projekte weisen einige Besonderheiten auf, die bei der Evaluation berücksichtigt werden müssen. Bezugspunkt dieser Programme und Projekte sind oft rechtsextremistische, antisemitische und fremdenfeindlichen Straf- und Gewalttaten, die sich in vielen deutschen Städten und Kommunen immer wieder ereignen. Daneben geht es aber oft auch um die heimliche Zustimmung breiterer Bevölkerungsschichten zu Elementen einer rechtsextremen Ideologie. Beispielsweise teilt einer aktuellen repräsentativen Umfrage zufolge mehr als ein Drittel der deutschen Bevölkerung fremdenfeindliche Einstellungen, fast 11 Prozent haben antisemitische Einstellungen, und fast 23 Prozent vertreten die Auffassung, dass Muslimen die Zuwanderung

nach Deutschland untersagt werden sollte (vgl. Heitmeyer 2012). Projekte zur Förderung einer pluralistischen und demokratischen Kultur wollen und sollen dem entgegenarbeiten. Hierfür stellt die Bundesregierung seit 1992 Mittel im Rahmen von Bundesprogrammen bereit. Das größte Bundesprogramm, TOLERANZ FÖRDERN – KOMPETENZ STÄRKEN hat ein Fördervolumen von 24 Millionen Euro pro Jahr. Hinzu kommen weitere Programme auf Bundes-, Landes- und auf kommunaler Ebene. Aber werden mit diesem Geld die richtigen Dinge getan? Gelingt es den Projekten, Maßnahmen zu entwickeln, die den Wert der Demokratie bewusst und erlebbar machen und die antidemokratische Tendenzen eindämmen helfen? Wie können solche Programme und Projekte optimiert werden? Evaluationen müssen versuchen, Antworten zu finden.

In der Vergangenheit war es allerdings eher die Regel als die Ausnahme, dass Programme zur Bekämpfung von Rechtsextremismus und zur Stärkung einer demokratischen Kultur insbesondere nach spektakulären Übergriffen und Verbrechen kurzfristig aufgelegt wurden. Evaluationen kommen oft erst dann ins Spiel, wenn das Programm bereits angelaufen ist und die Projekte ihre Arbeit aufgenommen haben. Wenn grundlegende Entscheidungen bereits getroffen sind, können die Kompetenzen von Evaluatorinnen und Evaluatoren für die Konzeptentwicklung oder für die Einbindung relevanter Stakeholder¹ nur noch eingeschränkt zum Tragen kommen. Häufig dominiert dann die Sichtweise des Initiators, was dazu führen kann, dass wichtige Erfahrungen nicht berücksichtigt und vorhandene Potentiale nicht ausgeschöpft werden können. Darüber hinaus ist es auch für den Nachweis von Wirkungen im Rahmen einer summativen Evaluation oft ein erhebliches Problem, wenn keine verlässlichen Daten zur Ausgangssituation vorliegen.

Politische Erwägungen spielen bei Programmen und Projekten zur demokratischen Kultur natürlich immer eine große Rolle. Bei politisch brisanten oder umstrittenen Programmen und Projekten besteht zum Teil die Hoffnung, dass eine Evaluation zur Versachlichung von Debatten und politischen Entscheidungen beitragen kann. Ob diese Hoffnung berechtigt ist, wird im vorliegenden Band durchaus kontrovers diskutiert. Die Einbettung in einen politischen Kontext hat aber auch Auswirkungen auf die Zeithorizonte, die für die Durchführung wissenschaftlicher Analysen zur Verfügung stehen und die zum Teil sehr kurz sind. Daraus können sich erhebliche methodische Probleme und Herausforderungen ergeben, denen sich eine anwendungsbezogene Wissenschaft aber stellen

1 Mit dem englischen Begriff „Stakeholder“ werden Personen bezeichnet, die ein berechtigtes Interesse am Verlauf eines Programms oder Projekts haben. Im deutschen Sprachraum wird dieser Personenkreis zum Teil mit den Begriffen „Beteiligte und Betroffene“ umschrieben. Da aber keine wirklich befriedigende deutsche Übersetzung existiert, hat sich der Begriff „Stakeholder“ in der Fachdiskussion mittlerweile eingebürgert.

muss, wenn sie den relevanten Stakeholdern nützliche und nutzbare Informationen zur Verfügung stellen will. Typische methodische, praktische und grundsätzliche Probleme der Evaluation zeigen sich im Bereich der Bekämpfung von Rechtsextremismus und der Förderung einer demokratischen Kultur deshalb besonders deutlich. Lösungen, die in diesem schwierigen Feld gefunden werden, haben daher für die gesamte Evaluationspraxis eine große Bedeutung.

Die Beiträge in diesem Band behandeln drei zentrale Themenbereiche: 1. Evaluationsansätze und Methoden, 2. nutzungsorientierte Evaluationsstrategien im Bereich der Förderung demokratischer Kultur und 3. die gesellschaftliche Bedeutung der Evaluation. Im ersten Teil geht es unter der Überschrift „Evaluationsansätze und Methoden“ insbesondere um die Frage, wie die Wirksamkeit von Programmen und Projekten sichergestellt und nachgewiesen werden kann. So genügt es im Non-Profit-Bereich nicht, Akzeptanz für bestimmte Maßnahmen herzustellen und eine Nachfrage nach ihnen zu wecken. Vielmehr ist es für die Qualität der Leistungen von Non-Profit-Organisationen ausschlaggebend, dass die eingesetzten Maßnahmen zur Verbesserung einer problematischen Situation beitragen und Veränderungen in die erwünschte Richtung anstoßen (vgl. Stockmann 2006). Diese Ausrichtung der Programme und Projekte und ihrer Evaluation auf Fragen der Wirksamkeit ist unserer Auffassung nach so entscheidend, dass wir in diesem Zusammenhang von einer Wirkungszentrierung sprechen.

Huey Chen beschreibt in seinem Ansatz der theoriegeleiteten Evaluation (*Theory-driven evaluation: Conceptual framework, application and advancement*) die zentralen Schritte eines solchen wirkungszentrierten Vorgehens. Dabei folgt die theoriegeleitete Evaluation dem Grundsatz, dass das Design und die Durchführung einer Evaluation von einer Programmtheorie angeleitet werden müssen. So eine Programmtheorie besteht aus expliziten und impliziten Annahmen der Stakeholder darüber, mit welchen Maßnahmen und unter welchen organisatorischen und sonstigen Voraussetzungen und Rahmenbedingungen wichtige Ziele erreicht und die problematische Ausgangssituation verbessert werden können. Im Unterschied zu einer so genannten Black-Box-Evaluation geht es also nicht nur darum, ob eine Intervention funktioniert, sondern auch darum, wie und unter welchen Rahmenbedingungen sie funktioniert. Auf die Frage, was „funktionieren“ in diesem Zusammenhang genau bedeutet, gibt Chen eine differenzierte Antwort.

Mit den Möglichkeiten und Problemen der *Wirkungsevaluation zu Maßnahmen der Demokratieförderung* setzt sich der Beitrag von Thomas Widmer auseinander. Wirkung beschreibt Widmer als einen „kausal auf eine Intervention beliebiger Form (wie eine Maßnahme, ein Projekt oder ein Programm) rückführbaren Effekt“. Oft seien Evaluationsgegenstände in diesem Themenfeld allerdings in Mehrebenensysteme eingebettet, bei denen Wirkungsketten über zahl-

reiche Wirkungsstufen auf verschiedenen Ebenen abliefen. Widmer versteht Evaluation unter dieser Voraussetzung als eine Dienstleistung, die Transparenz schafft, Nutzen stiftet und eine Bewertung vornimmt. Diese Dienstleistung wird seiner Ansicht nach auf staatlicher Ebene durch den Übergang von hierarchischen zu kooperativen Steuerungsformen verstärkt benötigt. Am Beispiel des Phänomens Rechtsextremismus diskutiert Widmer im Detail, wie ein derart vielschichtiges Phänomen begrifflich zu fassen ist und welche Evaluationsstrategien bei den verschiedenen Teilaspekten geeignet sind. Widmer stellt vier Aspekte heraus, die im Rahmen einer qualitativ hochwertigen Evaluation in diesem Themenfeld zu berücksichtigen sind: Klarheit über den Evaluationsgegenstand, Fokussierung bei der Wirkungsevaluation, Bescheidenheit angesichts hoher Schwierigkeitsgrade sowie Redlichkeit bei der Darstellung der Aussagefähigkeit.

Ein wesentlicher Aspekt bei der Wirkungsevaluation von Programmen und Projekten ist das Phänomen, dass auch experimentell getestete Strategien unter realen Umsetzungsbedingungen nicht zwangsläufig erfolgreich sind. Manuel Eisner, Tina Malti, Denis Ribeaud und Barbara Müller beschreiben in ihrem Beitrag über das *Zürcher Projekt zur sozialen Entwicklung von Kindern* einen groß angelegten Feldversuch in der kriminologischen Präventionsforschung. Bei dem Feldversuch geht es um die Evaluation von zwei Präventionsprogrammen zur Reduktion von aggressivem Verhalten bei Kindern im Grundschulalter. Solche frühkindlichen Programme sind auch im Bereich der Rechtsextremismusprävention durchaus üblich. Die Voraussetzungen für groß angelegte Feldversuche zur Untersuchung der Wirksamkeit von Interventionen unter realistischen Bedingungen sind allerdings sehr anspruchsvoll. Eisner et al. nennen hier unter anderem eine qualitativ hochwertige Dokumentation des Programms, Belege für die Wirksamkeit des Programms unter Laborbedingungen und verschiedene Aspekte der Durchführbarkeit und Umsetzbarkeit des Programms. Wichtig seien ferner Messungen der Umsetzungstreue, um zu erfassen, ob die Intervention in der von den Programmentwicklern vorgesehenen Form implementiert wurde. Neben der Antwort auf die Frage, ob eine Intervention gewirkt hat, sollen groß angelegte Feldversuche Eisner et al. zufolge klären, warum sie gewirkt hat, bei wem sie gewirkt hat und unter welchen Bedingungen sie gewirkt hat.

Im zweiten Teil des Bandes steht der ebenfalls zentrale Aspekt der Nützlichkeit und der Nutzung von Evaluationsergebnissen stärker im Fokus. In ihrem Beitrag zur *nutzungsfokussierten Evaluation* erläutern Wolfgang Beywl und Susanne Giel, wie Nutzenerwartungen und Nutzungsabsichten der Beteiligten und Betroffenen identifiziert und in den Evaluationsstrategien angemessen berücksichtigt werden können. Mit Bezug auf den Ansatz von Patton zeigen Beywl und Giel am Beispiel der Evaluation des Themenclusters „Früh ansetzende Prä-

vention“ im Bundesprogramm VIELFALT TUT GUT, wie eine Evaluationsstrategie mit diesem Schwerpunkt umgesetzt werden kann. Beschrieben werden unter anderem Instrumente zur Identifizierung der wichtigen Beteiligten, Methoden zur Identifizierung von Nutzungsabsichten und Verfahren zur Einbeziehung der Nutzungsabsichten in den Evaluationsplan. Obwohl ein Mindestmaß an Partizipation für dieses Vorgehen unabdingbar ist, können Nutzungsfokussierung und Partizipation nach den Erfahrungen von Beywl und Giel durchaus in einem Spannungsverhältnis stehen.

Kerstin Sischka spricht sich in ihrem Aufsatz „*Zum Nutzen eines konflikt-sensiblen Ansatzes in der wissenschaftlichen Projektbegleitung*“ für eine stärkere Verknüpfung von methodischer Evaluationsexpertise mit der wissenschaftlichen Forschung über das Themenfeld Rechtsextremismus, Jugend und Prävention aus. Anhand der wissenschaftlichen Begleitung von 18 Modellprojekten zur Arbeit mit rechtsextrem gefährdeten Jugendlichen beschreibt sie einen konflikt-sensiblen und psychosozial ausgerichteten Evaluationsansatz. Grundgedanke dieses Ansatzes ist ebenfalls eine nutzen- und nutzungsorientierte Vorgehensweise, die den evaluierten Projekten ein Angebot zur systematischen Praxisreflexion unterbreitet. Ausgehend vom Prinzip der „Konfliktsensibilität“ sollen Projekte mit Hilfe der wissenschaftlichen Begleitung nicht nur ihren Handlungskontext, sondern auch ihre Interaktion mit diesem Kontext reflektieren. Sischka legt dar, wie die wissenschaftliche Begleitung ein Empowerment der Projektmitarbeiter/innen gefördert hat, indem sie selbstreflexive Prozesse unterstützt und einen Rahmen für selbstorganisierten fachlichen Austausch und für Vernetzung bereitgestellt hat. Dass dieser reflexive Ansatz in der Lage ist, unterschiedliche Wirkungsdimensionen zu thematisieren, zeigt Sischka am Beispiel eines sozialkognitiven Bildungscurriculums mit straffälligen Jugendlichen in einer JVA.

Rainer Strobl und Olaf Lobermeier thematisieren in ihrem Aufsatz zur *Netzwerkevaluation* einen für die Stärkung der demokratischen Kultur zentralen Aspekt. Sie unterscheiden zunächst zwischen zufällig entstandenen, „natürlichen“ Vernetzungen und strategisch geplanten Netzwerken. Letztere werden systematisch aufgebaut, um die Durchführung wesentlicher Aufgaben und Prozesse zu erleichtern oder überhaupt erst zu ermöglichen. Eine Evaluation kann den Autoren zufolge bereits bei der Netzwerkplanung wertvolle Unterstützung leisten und z. B. eine Stakeholderanalyse anleiten. Mit Hilfe der Netzwerkanalyse lassen sich Strukturen schon in dieser Phase visualisieren und später mit der tatsächlichen Vernetzung vergleichen. Hierdurch kann anschaulich gezeigt werden, welche wichtigen Beziehungen im aktuellen Netzwerk noch fehlen. Die Autoren betonen jedoch, dass eine Vernetzung kein Selbstzweck ist, sondern immer dazu dienen sollte, das Erreichen relevanter Ziele zu ermöglichen oder hierzu einen Beitrag zu leisten. Die Frage, ob diese inhaltlichen Ziele tatsächlich

erreicht wurden, kann von der eigentlichen Netzwerkevaluation allerdings nicht beantwortet werden. Dies muss daher gesondert geprüft werden, indem geeignete Evaluationsstrategien zum Nachweis von Wirkungen eingesetzt werden.

Während in den zuletzt vorgestellten Beiträgen geschildert wird, wie Evaluationen nützliche und nutzbare Informationen bereitstellen können, widmet sich Gesa Schirrmacher der Frage, *was sich Auftraggebende von Evaluationen und Evaluationspraxis wünschen*. Ein zentrales Problem bei der Nutzung von Evaluationsergebnissen kann Schirrmacher zufolge darin bestehen, dass die Lösung drängender sozialer Probleme nicht so lange hinausgezögert werden kann, bis die Wissenschaft wirklich valide Ergebnisse bereitstellt. Dennoch benötige die Verwaltung solide Ergebnisse zum Nutzen eines Programms oder Projekts. Mit diesen Ergebnissen würden unterschiedliche Ziele verfolgt. Neben einer Verbesserung von Programmen, Projekten oder Policies gehe es auch um Kostenreduktion oder die Legitimation von politischen Entscheidungen. Wichtig, so Schirrmacher, sei vor allem eine klare Kommunikation zwischen Auftraggebenden und Evaluierenden über die Ziele der Evaluation. Des Weiteren gehe es darum, dass Evaluatorinnen und Evaluatoren ihre Vorgehensweise und die Ergebnisse für die Auftraggebenden verständlich und nachvollziehbar darstellen und bei der Formulierung von Empfehlungen auch die Besonderheiten des jeweiligen politischen Kontexts berücksichtigen.

Der dritte Teil des Bandes widmet sich der gesellschaftlichen Bedeutung der Evaluation. Reinhard Stockmann betont in seinem Beitrag *Evaluation und Gesellschaft*, dass Evaluation nicht nur Hochkonjunktur hat, sondern auch dabei ist, sich als eigenständige Disziplin der angewandten Sozialforschung zu etablieren. Der Evaluationsboom führe zu vielfältigen Evaluationen, die die Gesellschaft auf unterschiedlichen Ebenen betreffen. Stockmann ist überzeugt, dass professionell durchgeführte Evaluationen nicht nur die Effizienz und die Effektivität von Projekten und Programmen steigern oder der Legitimierung von Politikstrategien dienen können, sondern darüber hinaus in der Lage sind, einen Beitrag zur gesellschaftlichen Aufklärung zu leisten, indem sie Transparenz über die Ziele und Wirkungen politischer Strategien und Maßnahmen herstellen. Stockmann diskutiert in seinem Beitrag auch die notwendigen Voraussetzungen zur Erfüllung dieser Evaluationsfunktionen. Er plädiert in diesem Zusammenhang unter anderem für eine weitergehende Professionalisierung und Etablierung der Evaluation als wissenschaftliche Forschungsdisziplin. Insgesamt ist Stockmann optimistisch, dass Evaluationen auf vielfältige Art einen gesellschaftlichen Nutzen stiften können.

Wilhelm Heitmeyer nimmt in seinem Beitrag *Evaluationsforschung. Parameter und Auswirkungen für die Gesellschaft* gewissermaßen die Gegenposition ein, wenn er analysiert, welche Auswirkungen die Zunahme an Evaluationsfor-

schung auf öffentliche Diskurse, Akteursgruppen und Institutionen hat. Er ist insgesamt skeptisch, ob Evaluationsforschung wirklich einen Beitrag zur gesellschaftlichen Aufklärung leisten kann. Jede plurale, moderne Gesellschaft produziert Heitmeyer zufolge Zielkonflikte, durch die Bewertungskriterien für Evaluationen unklar werden. Aufgrund der unterschiedlichen Rationalitäten der verschiedenen gesellschaftlichen Teilsysteme stelle sich immer die Frage, welche Interessen welcher Stakeholder berücksichtigt werden sollen. So könne z. B. die Frage, was einer freiheitlichen demokratischen Kultur gut tut, keineswegs eindeutig beantwortet werden. Als Problembereiche identifiziert Heitmeyer eine unzureichende Bereitschaft zur Finanzierung professioneller Evaluationen, die Möglichkeit einer Hemmung von Risikobereitschaft durch den Zwang zur Evaluation, die Gefahr der ideologischen Instrumentalisierung von Evaluationsergebnissen, den Missbrauch von Evaluationen für ein Kontrollregime sowie den Zwang zu einer stärkeren Standardisierung gesellschaftlicher Praxen als Voraussetzung für ihre Evaluierbarkeit. Hinsichtlich der Auswirkungen der Evaluationsforschung auf die Gesellschaft zeichnet Heitmeyer ein ambivalentes Bild. Unter anderem sieht Heitmeyer auch die Gefahr, das Expertenwissen für die gesellschaftliche Diskussion auf Dauer immer unwichtiger wird, wenn die Abhängigkeit von Interessen und Auftraggebern nicht mehr zu verbergen ist. Er plädiert in diesem Zusammenhang für eine verständliche Darstellung und für eine rigorose Veröffentlichung von Evaluationsergebnissen.

Literatur

- Heitmeyer, Wilhelm (2012): Gruppenbezogene Menschenfeindlichkeit (GMF) in einem unsicheren Jahrzehnt. In: ders. (Hg.): Deutsche Zustände. Folge 10. Berlin: Suhrkamp, S. 15-41.
- Stockmann, Reinhard (2006): Evaluation und Qualitätsentwicklung. Eine Grundlage für wirkungsorientiertes Qualitätsmanagement, Münster; New York; München; Berlin: Waxmann.
- Strobl, Rainer; Lobermeier, Olaf (2009): Die Problemstellung: Rechtsextremismus in der Kommune. In: Molthagen, Dietmar; Korgel, Lorenz (Hg.): Handbuch für die kommunale Auseinandersetzung mit dem Rechtsextremismus. Berlin: Friedrich-Ebert-Stiftung, Forum Berlin, S. 15-27.

I. Evaluationsansätze und Methoden

Theory-driven evaluation: Conceptual framework, application and advancement

Huey T. Chen

1 Introduction

There is an impressive amount of literature on theory-driven evaluation published in the past few decades. The literature devoted to this topic includes four volumes of *New Directions for Evaluation* (Bickman 1987, 1990; Rogers, Hasci, Petrosino, & Huebner 2000; Wholey 1987), several books (Chen 1990, 2005; Chen/Rossi 1992; (Connell, Kubisch, Schorr, & Weiss 1995; Fulbright-Anderson, Kubisch, & Connell 1998; Pawson & Tilly 1997) and numerous articles published in various journals (see recent review by (Coryn, Noakes, Westine, & Schoter 2011; Hansen & Vedung 2010). Furthermore, major evaluation textbooks (Patton 1997; Posavac & Carey 2007; Rossi, Lipsey, & Freeman 2004; Weiss 1998) have a chapter(s) introducing the concepts, methodology, and usefulness of theory-driven evaluation. The purpose of this chapter is to discuss the conceptual framework, applications, and new developments of theory-driven evaluation for facilitating further advancement.

2 Conceptual Framework of Program Theory

The tenet of theory-driven evaluation is that the design and application of evaluation needs to be guided by a conceptual framework called program theory (Chen 1990, 2005). Program theory is defined as a set of explicit or implicit assumptions by stakeholders about what action is required to solve a social, educational or health problem and why the problem will respond to this action. The purpose of theory-driven evaluation is not only to assess whether an intervention works or does not work, but also how and why it does so. The information is essential for stakeholders to improve their existing or future programs.

Theory-driven evaluation is sharply different from another type of evaluation, called black-box evaluation. Black-box evaluation mainly assesses whether an intervention has an impact on outcomes. It does not interest in the transformation process between the intervention and outcomes. Similarly, theory-driven

evaluation is also different from method-driven evaluation. Method-driven evaluation uses a research method as a basis for conducting an evaluation. According to method-driven evaluation proposes the design of an evaluation is mainly guided by the predetermined research steps required by a particular method, quantitative, qualitative, or mixed. Unlike method-driven evaluation views evaluation mainly atheoretical, methodological activities, Method-driven evaluation tends to ignore stakeholders' view and concern in evaluation.

As a basis for designing theory-driven evaluation, program theory is a systematic configuration of stakeholders' prescriptive assumptions and descriptive assumptions underlying programs, whether explicit or implicit (Chen 1990, 2005). Descriptive assumptions, called change model, deal with what causal processes are expected to happen to attain program goals. Prescriptive assumptions, called action model, deal with what actions must be taken in a program in order to produce desirable changes. Theory-driven evaluation uses the action model and change model to address contextual factors and planning and implementation issues that are greatly interested to stakeholders.

Change Model: A change model describes the causal process generated by the program. The elements of a change model consist of the following three elements:

Goals and Outcomes: Goals reflect the desire to fulfill unmet needs, as with poor health, inadequate education, or poverty. Outcomes are the concrete, measurable aspects of these goals.

Determinants: To reach goals, programs require a focus, which will clarify the lines their design should follow. More specifically, each program must identify a leverage or mechanism upon which it can develop a treatment or intervention to meet a need. That leverage or mechanism is variously called the determinant or the intervening variable.

Intervention or Treatment: Intervention or treatment comprises any activity (ies) in a program that aims directly at changing a determinant. It is, in other words, the agent(s) of change within the program.

Action Model: An action model is a systematic plan for arranging staff, resources, settings, and support organizations to reach a target group and deliver intervention services. The action model consists of the following elements.

Implementing Organization: Assess, Enhance, and Ensure Its Capabilities: A program relies on an organization to allocate resources, to coordinate activities, and to recruit, train, and supervise implementers and other staff. How well a program is implemented may be related to how well this organization is structured. Initially, it is important to ensure that the implementing organization has the capacity to implement the program.

Program Implementers: Recruit, Train, Maintain Both Competency and Commitment: Program implementers are the people responsible for delivering services to clients: counselors, case managers, outreach workers, school teachers, health experts, and social workers. The implementers' qualifications and competency, commitment, enthusiasm, and other attributes can directly affect the quality of service delivery.

Peer Organizations/Community Partners: Establish Collaborations: Programs often may benefit from, or even require, cooperation or collaboration between their implementing organizations and other organizations. If linkage or partnership with these useful groups is not properly established, implementation of such programs may be hindered.

Intervention and Service Delivery Protocols: Intervention protocol is a curriculum or prospectus stating the exact nature, content, and activities of an intervention – in other words, the details of its orienting perspective and its operating procedures. Service delivery protocol, in contrast, refers to the particular steps to be taken to deliver the intervention in the field.

Ecological Context: Seek Its Support: Some programs have a special need for contextual support, meaning the involvement of a supportive environment in the program's work. Both microlevel contextual support and macrolevel contextual support can be crucial to a program's success. Microlevel contextual support comprises social, psychological, and material supports clients need to allow their continued participation in intervention programs. In addition to microlevel contextual support, program designers should consider the macrolevel context of a program, that is, community norms, cultures, and political and economic processes. These, too, have the ability to facilitate a program's success.

Target Population: Identify, Recruit, Screen, Serve: In the target group element, crucial assumptions at work include the presence of validly established eligibility criteria; the feasibility of reaching and effectively serving a target group; and the willingness of potential clients to become committed to, or cooperative with, or at least agreeable to joining the program. Relationships among the components are illustrated in Figure 1.

Figure 1: The Conceptual Framework of Program Theory

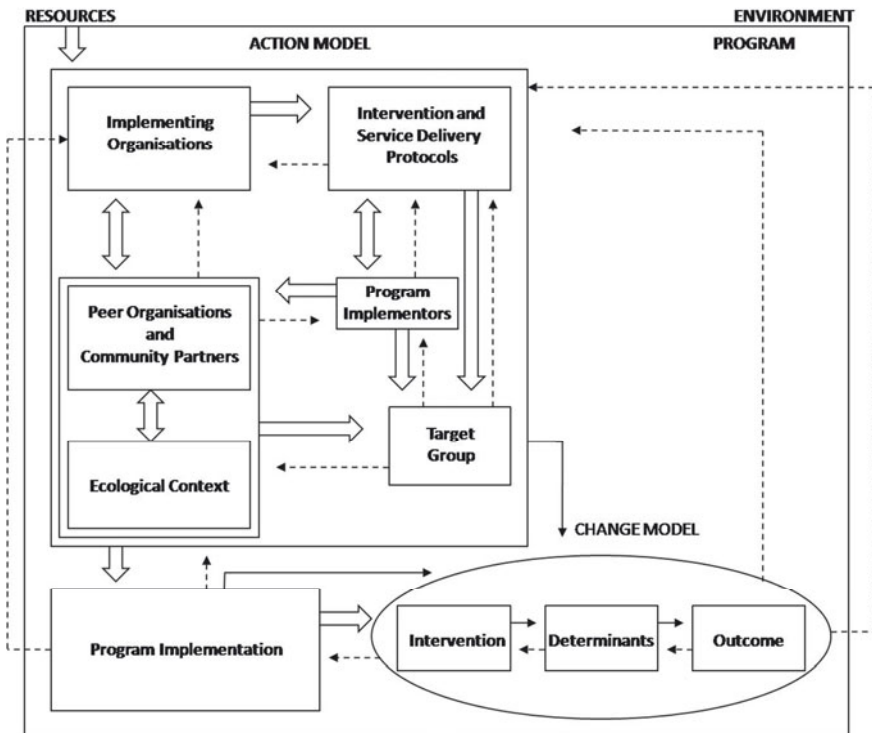


Figure 1 indicates that the action model must be implemented appropriately to activate the „transformation“ process in the change model. For a program to be effective its action model must be sound and its change model plausible; its implementation is then also likely to be doing well. Figure 1 also illustrates evaluation feedback as represented in dotted arrows. Information from implementation can be used to improve the planning or the development of the action model. Similarly, information from the change model can be used to improve the implementation process and the action model. This conceptual framework of program theory should be useful to evaluators charged with designing an evaluation that produces accurate information about the dynamics leading to program success or program failure.

3 Examples of Theory-Driven Evaluation

3.1 Example of Theory-Driven Process Evaluation

Comprehensive theory-driven process evaluation is associated with certain strategies and approaches from the taxonomy. Two evaluations are discussed here to show some of the possible functions of this kind of evaluation.

3.1.1 Evaluating an Anti-Drug Abuse Program.

One comprehensive, theory-driven process evaluation that closely mirrors this handbook's conceptual framework of program theory is an evaluation of a large anti-drug abuse program for middle school students in Taiwan (Chen 1997). The program asked teachers to identify drug-abusing students and provide them with counseling services. A small group of top officials within Taiwan's Ministry of Education had designed the program; under the nation's centralized education system, the Ministry of Education approved appointments and salaries of teachers and administrators. When the program began in January 1991, 3,850 students had been identified as active drug abusers. That number declined sharply, plunging 96 %, to 154 students by June 1991.

The program's huge success led to a theory-driven process evaluation being conducted to examine how the program had been implemented. Hopes were that this program's example could foster the smooth implementation of other programs. The anti-drug abuse program featured a documentary program plan, but it was incomplete in comparison to the action model or program plan illustrated in Figure 1. Acting as facilitators, evaluators convened separate focus group meetings with top officials of the education ministry and with teacher representatives to obtain the information needed to complete the program plan. (The separate meetings acknowledged teachers' tendency to be silent in the presence of top officials, who have much more power than teachers do.) Evaluators played the role of facilitators and consultants, helping these key stakeholders develop their program theory. The final version of the program plan ultimately used for evaluation had been agreed to by both groups; the plan is presented on the left side of Table 1.

Table 1: The Spring Sun Program: normative versus actual

<i>Program domains/ dimensions</i>	<i>Normative</i>	<i>Actual</i>
<i>Goal/outcome</i>	Reduction of student drug use to be verified through urinalysis	Reduction of drug use, but urinalysis collection environment not controlled
<i>Treatment</i>	Primary: provide quality counseling to abusers Secondary: basic drug education	Primary: counseling mainly involved use of threats, admonishment, and/or encouragement not to use Secondary: basic drug education
<i>Implementation Environment</i>		
Target group	All drug abusing students	Only those drug abusing students who were easy to reach
Implementors	Teachers provided with adequate drug treatment training and information	Teachers lacked adequate drug treatment skills and information
Mode of delivery	Compulsory individual counseling	Compulsory individual counseling; but with problems such as lack of plan, format and objective
Implementing organisation	All schools that can adequately implement the program	Smaller schools had difficulties implementing the program
Inter-organisational procedures	Effective centralized school system	Communication gap, mistrust between Ministry of Education and the schools
Micro-context	Eliminate video game arcades	Video game arcades still exist
Macro-context	Strong public support	Strong public support, but problematic education system (elitism)

The program plan entailed mixing research methods – both quantitative and qualitative – to collect data. For example, quantitative methods were applied to rate teachers' satisfaction with a workshop on drug counseling skills sponsored

by the education ministry, whereas qualitative methods were used to probe contextual issues of the teachers' opinions of the workshop. The right side of Table 1 displays empirical findings for the program's real-world implementation; comparison of the program theory to the implementation reveals large discrepancies. The program had been carried out, but the quality of services and the system of implementation were far from being impressive. The discrepancies between plan and implementation resulted from a lack of appropriate counseling training, the overburdening of teachers with counseling work with no change to their usual teaching responsibilities, and lack of communication as well as mistrust between an authoritarian ministry and the teachers. The evaluation results created doubt about how a program without strong implementation achieved a 96 % decrease in drug abuse in schools.

3.2 *Examples of Theory-Driven Outcome Evaluation*

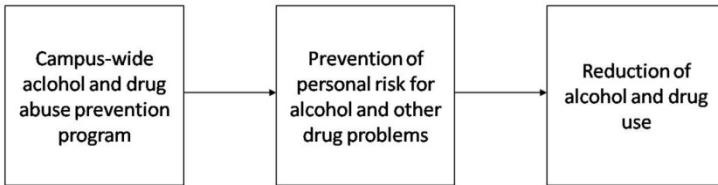
Two basic models of intervening mechanism evaluation predominate in the discipline: linear and dynamic.

3.2.1 The Linear Model

The linear model is currently a very popular application of intervening mechanism evaluation. Linear models assume that the causal relationships among interventions, determinants, and outcomes are unidirectional: intervention affects determinant, and determinant then affects outcome. No reciprocal relationships operate among the variables. In linear models, the number and sequence of the determinants under study determine the model's form. The following causal diagrams illustrate the common linear model forms.

One-Determinant Model. This model, represented by Figure 2, contains a single determinant and is the fundamental model for intervening mechanism evaluation.

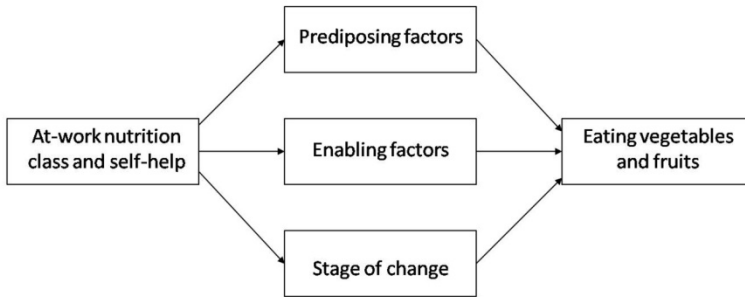
Figure 2: An example of a one-determinant model



The one-determinant model is illustrated here by an evaluation of an alcohol and drug abuse prevention program at a college (Miller, Toscova, Miller, & Sanchez 2000). The intervention consisted of multiple components: print media, videotapes, speakers, referral services, and development of self-control. The determinant was perception of risk, and the outcome was a reduction in alcohol and drug use among the students on the campus where the program was established. As predicted, the data showed that after the interventions, there was heightened awareness on campus of the risks of substance abuse, which in turn reduced alcohol and drug use there. The one-determinant model is relatively easy to construct.

Multiple-Determinant Model, No Sequential Order. Another common linear model is the model with two or more determinants, each affected by the intervention or affecting the outcome, but in no particular sequence. A workplace nutrition program provides an example of the multiple-determinant model (Kristal, Glanz, Tilley, & Li 2000). The intervention featured at-work nutrition classes and self-help. The stakeholders and evaluators selected three determinants: predisposing factors (skills, knowledge, belief in diet-disease relationship), enabling factors (social support, perceived norms, availability of healthful foods) and stage of change (action and maintenance stages being subsequent to the intervention). The outcome variable was dietary change (eating vegetables and fruits). The model of this program is illustrated in Figure 3.

Figure 3: Workplace nutrition program as a multiple determinant. No sequential order.



Kristal and colleagues found that the intervention did enhance predisposing factors as well as the likelihood of entering and remaining in the subsequent stages of change. They also found that the intervention did not affect enabling factors. The program was failing because the intervention was failing to activate one of the three determinants.

Multiple-Determinant Model With Sequential Order. The model containing two or more determinants aligned in a causal order is a multiple-determinant model with a sequential order. That is, certain determinants affect others in a particular sequential order. An example of this kind of linear model is found in an evaluation of a school-based antismoking campaign (Chen, Quane, & Garland 1988). The intervention contained components such as an antismoking comic book, discussions of the health messages the comic book delivered, and parental notification about the intervention program. The determinants of the model, in sequence, were the number of times the comic book was read, and knowledge of the comic book's story and characters. The sequential order indicates that repeated reading of the comic book changed the extent of knowledge about the plot and characters. The sequence is illustrated in Figure 4.