

Karin Haubrich

Rekonstruktive Programmtheorie – Evaluation multizentrischer Entwicklungsprogramme

Reconstructive Program Theory – Evaluation of Multi-site Developmental Programs

Zusammenfassung: In dem Beitrag wird der Frage nachgegangen, wie innovative Entwicklungsprogramme mit mehreren Projekten, die jeweils weitgehend autonom innovative sozialpädagogische Konzepte entwickeln und erproben, trotz ihrer Heterogenität und Veränderungsoffenheit als Programm evaluiert werden können. Am Beispiel von Bundesmodellprogrammen in der Kinder- und Jugendhilfe werden spezifische Merkmale solcher Entwicklungsprogrammen als Gegenstand einer Programmevaluation herausgearbeitet. Darauf aufbauend wird eine Konzeption rekonstruktiver Programmtheorie-Evaluation vorgestellt, die eine empirisch fundierte theoretische Beschreibung des Evaluationsgegenstandes als „Programmcluster heterogener Projekte“ ermöglicht.

Schlagerworte: Programmevaluation, multizentrische Programme, programmtheoriebasierte Evaluation, Cluster-Evaluation, Expert(innen)interviews, pädagogisches Wissen

Abstract: This article is concerned with the question of how to evaluate multi-site developmental programs given the presence, and heterogeneity, of several relatively autonomous projects developing and testing innovative concepts in social work, and the emergent character of the program. First, the main characteristics of developmental programs as „evaluands“ are described, with a number of model programs in the field of child and youth welfare serving as an example. Second, a reconstructive approach to program theory evaluation is proposed. It provides a concept of reconstructive program evaluation that allows for an empirically well-grounded theoretical description of the evaluand, i.e., a program that is conceived of as a cluster of heterogeneous projects.

Keywords: program evaluation, multi-site programs, program theory-based evaluation, cluster evaluation, expert interviews, pedagogic knowledge

1. Einleitung

Innovation und Verbesserung, die Erprobung neuer Wege und die Anpassung bestehender sozialstaatlicher Interventionen an veränderte gesellschaftliche Rahmenbedingungen sind ein Anliegen vieler zeitlich befristeter sozialpoliti-

scher Programme. Evaluationen werden ihnen zur Seite gestellt, um die gewonnenen Einsichten und Erfahrungen wissenschaftlich abzusichern. Doch solche Programme erweisen sich als anspruchsvolle oder – metaphorisch gesprochen – als „dornige“ Evaluationsgegenstände. Durch die Evaluation sollen Einsichten gewonnen werden, die geeignet sind, zur fachlichen Weiterentwicklung eines Arbeitsfeldes beizutragen oder sozialpolitische Entscheidungen vorzubereiten. Daher umfassen solche Programme meist mehrere Projekte oder lokale Standorte – mit anderen Worten: Es handelt sich um *innovative multizentrische Programme*. Sind solche Programme als innovative *Entwicklungsprogramme* konzipiert, gestehen sie den Projekten eine gewisse Autonomie in der Umsetzung der inhaltlichen Programmkonzeption zu und erlauben Entwicklungen und Veränderungen der Projektkonzeptionen im zeitlichen Verlauf.

Um ein innovatives, multizentrisches Programm in systematischer Weise und intersubjektiv nachvollziehbar evaluieren zu können, stellt sich die Forschungsaufgabe, einen solch komplexen Gegenstand theoretisch zu beschreiben. Erst auf dieser Grundlage kann dann in einem zweiten Schritt ausgehend von den jeweiligen Fragestellungen und den theoretischen Vorannahmen über den Untersuchungsgegenstand ein Evaluationsdesign entwickelt werden, das neben den Erhebungs- und Auswertungsverfahren vor allem auch Auskunft darüber geben sollte, wie die Evaluation Bewertungen ermöglichen will und wie sie sich zu Fragen der Verwendung ihrer Ergebnisse verhält.

Die Evaluation komplexer Entwicklungsprogramme steht vor Anforderungen, die in der derzeitigen Evaluationsdebatte nur ungenügend berücksichtigt werden. Rekonstruktive und interpretative Forschungszugänge (bspw. grounded theory) können hier hilfreiche Anknüpfungspunkte bieten. Sie ermöglichen es, zu einer empirisch fundierten theoretischen Beschreibung des Evaluationsgegenstandes zu gelangen, die sowohl der sozialen Wirklichkeit von Entwicklungsprogrammen wie auch den Fragestellungen der Evaluation angemessen ist. Im Folgenden wird eine Konzeption rekonstruktiver Programmtheorie-Evaluation vorgestellt, die in zwei Evaluationsstudien von Bundesmodellprogrammen in der Kinder- und Jugendhilfe¹ entwickelt und erprobt wurde.

Zunächst werden am Beispiel von Bundesmodellprogrammen die spezifischen Herausforderungen an die Evaluation von Entwicklungsprogrammen skizziert. Dann wird ein Evaluationskonzept vorgestellt, das den Ansatz der Cluster-Evaluation (u.a. Sanders 1997; Haubrich 2006) mit einem programmtheoriebasierten Vorgehen verknüpft. Dieses ermöglicht die empirisch fundierte theoretische Beschreibung eines komplexen Programms als Ausgangspunkt für weitere Erhebungs- und Auswertungsschritte zur Bewertung modellhaft erprobter „fachlicher Strategien“, d.h. sozialpädagogischer Handlungskonzepte, die in mehreren Projekten in ähnlicher Weise umgesetzt wurden.

2. Entwicklungsprogramme und ihre Herausforderungen für die Evaluation: Das Beispiel Bundesmodellprogramme

2.1. Bundesmodellprogramme als politisches Instrument des Bundes

Modellprogramme wurden bereits ab Mitte der 1960er Jahre als ein Instrument innovativer Politik auf Bundesebene eingesetzt (vgl. Häußler et al. 1988, S. 32). Mit ihrer Hilfe sollen Reformvorhaben in verschiedenen gesellschaftlichen Bereichen versuchsweise erprobt werden, bevor sie evtl. zu einer flächendeckenden Umsetzung gelangen.

In der Kinder- und Jugendhilfe sind die Einsatzmöglichkeiten von Bundesmodellprogrammen auf den Zuständigkeitsbereich des Bundes beschränkt, der im Kinder- und Jugendhilfegesetz (SGB VIII) geregelt wird. Nach § 83 SGB VIII ist es die Aufgabe der „fachlich zuständigen obersten Bundesbehörde“, die Fachpraxis in der Kinder- und Jugendhilfe anzuregen und zu fördern, soweit dies „(...) von überregionaler Bedeutung ist und ihrer Art nach nicht durch ein Land allein wirksam gefördert werden kann“ (Wiesner 2006, S. 1508).

In den Modellförderungsgrundsätzen des Bundesministeriums für Jugend, Familie, Frauen und Gesundheit (abgedruckt in Dietzel 1988, S. 27-30) werden Modellvorhaben beschrieben als „Maßnahmen oder Einrichtungen, die Erkenntnisse bringen sollen im Hinblick auf die Entwicklung, Verwirklichung sowie Erprobung *neuer* Methoden oder Konzeptionen oder die Überprüfung sowie Weiterentwicklung *bestehender* Methoden oder Konzeptionen“ (Hervorh. i. Orig, ebd.: 27). Modellprogramme in der Kinder- und Jugendhilfe haben somit eine *Innovations- und Erprobungsfunktion*. Reformkonzepte, fachliche Strategien und Arbeitsformen, die so noch nicht Bestandteil der Regelpraxis sind, können entwickelt und/oder auf ihre praktische Bewährung hin überprüft werden. Erfahrungen können auf diese Weise „stellvertretend“ für bestimmte Angebots- und Leistungsbereiche in der Jugendhilfe gemacht werden (vgl. ebd., S. 14). Das folgende Beispiel soll dies veranschaulichen.

Beispiel: Das Bundesmodellprogramm „Interkulturelles Netzwerk“

Das Bundesmodellprogramm „Interkulturelles Netzwerk der Jugendsozialarbeit im Sozialraum“ sollte im Vorfeld politischer Reformen in der Zuwanderungspolitik Ansätze zur Zusammenführung der zuvor getrennten Aufgabenfelder der Integration von Aussiedler(inne)n und Ausländer(inne)n entwickeln. Da sich die Lebensbedingungen beider Herkunftsgruppen einander angleichen, wurde die Separierung von Angeboten für unterschiedliche Herkunftsgruppen zunehmend in Frage gestellt.

In der Konsequenz sollten Ende der 1990er Jahre Jugendgemeinschaftswerke zu Jugendmigrationsdiensten weiterentwickelt werden. Im Kontext der Novellierung der Förderungsgrundsätze für diese Einrichtungen zielte ein Programmschwerpunkt auf die Öffnung der Jugendgemeinschaftswerke und ihrer Unterstützungsleistungen für Aussiedlerjugendliche für neue Zielgruppen, nämlich alle Jugendlichen mit Migrationshintergrund unter 27 Jahren. Der zweite Pro-

grammschwerpunkt war der Aufbau eines interkulturellen Netzwerks der Jugendsozialarbeit im Sozialraum. Die zwölf Modellprojekte sollten über Vernetzungsarbeit, Öffnung bestehender Angebote und Einrichtungen für junge Migrant(inn)en sowie Einflussnahme auf kommunale Planungsprozesse auf eine Verbesserung der regionalen Angebotsstruktur für diese Zielgruppe hinwirken (vgl. Haubrich/Vossler 2001, S. 22).

Die Projekte erprobten unterschiedliche Wege der sozialräumlichen Vernetzung: Einige führten Bestands- und Bedarfsanalysen als Basis für die Netzwerk- und Lobbyarbeit durch, manche bauten neue Netzwerke auf und koordinierten diese, andere brachten sich als gleichberechtigte Partner in bestehende Strukturen ein und wieder andere stärkten ihr Profil als interkulturelle Fachstelle und boten Serviceleistungen an (vgl. Vossler/Obermaier 2003, S. 57-67).

Die Programmevaluation verfolgte zwei Fragestellungen:

- Welche Konzepte und Strategien zur Netzwerkarbeit und Verbesserung der regionalen Angebotsstruktur werden mit welchem Erfolg bzw. unter welchen Schwierigkeiten erprobt?
- Welche Schritte zur interkulturellen Öffnung und Weiterentwicklung der Jugendgemeinschaftswerke werden unternommen, welche Erfahrungen werden dabei gemacht und welche Veränderungen werden erreicht?

Die Anregungsfunktion des Bundes impliziert, dass Ergebnisse der Modellförderung auf die Regelpraxis übertragbar sein sollen. Um die lokal in Einzelprojekten eines Modellprogramms gewonnenen Erfahrungen zu bündeln, wird daher die Modellförderung mit einer wissenschaftlichen Begleitung oder Evaluation verknüpft. Sie hat die Aufgabe, fundierte Erkenntnisse über die Grenzen und Möglichkeiten sowie die Erfolge und Misserfolge der erprobten Konzeptionen zu gewinnen und Aussagen zur Übertragbarkeit bewährter neuer Arbeitsformen in *andere lokale Kontexte* unter Bedingungen der *Regelpraxis* zu ermöglichen (vgl. ebd., S. 15).

2.2. Implikationen für die theoretische Beschreibung des Evaluationsgegenstandes

Betrachtet man Bundesmodellprogramme unter der Perspektive einer Programmevaluation, d.h. als ein *komplexes Handlungsmodell*, dem ein *planvolles Handeln* zugrunde liegt, das *Aktivitäten und Ziele* konzeptionell auf der Grundlage benennbarer *Ressourcen* und einer zu verändernden (evtl. auch zu erhaltenden) *Ausgangssituation* verknüpft (vgl. Hellstern/Wollmann 1983, S. 7; Kromrey 1995, S. 315; Haubrich 2009), so werden schnell einige Schwierigkeiten für die Anwendung dieser evaluationstheoretischen Heuristik erkennbar. Schließlich ist anfänglich das „komplexe Handlungsmodell“ eines Modellprogramms lediglich strukturell – im Sinne eines Erprobungsraumes für zu entwickelnde innovative Konzepte – beschreibbar, die konkrete inhaltliche Ausgestaltung des Erprobungsauftrages obliegt jedoch den einzelnen Projekten innerhalb des Modellprogramms, die neue Konzepte entwickeln, erproben und diese im zeitlichen Verlauf auch verändern können. Es handelt sich also um „*emergente*“² Programme. Die angestrebten Ziele – des „ziel“-gerichteten Handlungsmodells – werden nicht vorab festgelegt, sondern die konkreten Strategien, Ak-

tivitäten, Zwischenschritte und Teilziele zur Erreichung der Programmziele (bspw. sozialräumliche Vernetzung) werden im Programmverlauf an den einzelnen Projektstandorten *entwickelt*. Hinzu kommt, dass die Projekte im Programmverlauf „experimentierend“ vorgehen, d.h. dass sie auf sich wandelnde Rahmenbedingungen reagieren, aus Fehlern lernen und entsprechende Konzeptanpassungen vornehmen können.

Aus dem emergenten Charakter überregional angelegter Modellprogramme resultieren spezifische Anforderungen an die theoretische Beschreibung des Untersuchungsgegenstandes:

Wenn das zugrunde liegende Handlungsmodell nicht vorab theoretisch beschrieben werden kann und nicht von einer linearen Verknüpfung zwischen Programmplanung auf der politischen Ebene und „programmgetreuer“ Implementation auf der fachpraktischen Ebene auszugehen ist, bedarf es einer empirischen Untersuchung der tatsächlichen Programmumsetzung im Zusammenwirken der unterschiedlichen beteiligten Akteure (Ministerium, Freie Träger, Einrichtungen und Projekte). Erst auf einer solchen empirischen Basis lässt sich der Untersuchungsgegenstand theoretisch beschreiben.

Wenn das Modellprogramm veränderungsoffen ist, muss auch die Beschreibung des Untersuchungsgegenstandes veränderungsoffen angelegt sein, um Entwicklungsprozesse abbilden zu können.

Wenn unter dem Dach einer Programmkonzeption heterogene lokale Projekte versammelt sind, damit aus kontextbezogenen Lösungen gelernt werden kann und eine möglichst breite Basis für Innovation gegeben ist, dann steht die Programmevaluation vor der Herausforderung, den Evaluationsgegenstand so zu beschreiben, dass das Gesamtprogramm in seinen Kerndimensionen fassbar wird und die Umsetzungsformen und Projektaktivitäten diesen Dimensionen zugeordnet werden können.

Bisher jedoch werden in der Evaluationsfachdebatte die besonderen Merkmale multizentrischer Entwicklungsprogramme und daraus resultierende Anforderungen an die Evaluation kaum systematisch diskutiert. Verbreitete Ansätze der Evaluation multizentrischer Programme wie die „multisite evaluation“ (vgl. Fitzpatrick/Sanders/Worthen 2004, S. 466-471; Straw/Herrell 2002) gehen von einer möglichst gleichförmigen Implementation eines vorab beschreibbaren zentralen Programmmodells durch die einzelnen Projekte aus. Robin S. Turpin und James M. Sinacore beschreiben die Zwecksetzung solcher Evaluationen folgendermaßen:

„(...) the use of multiple sites, in increasing the size of the samples, elevates the statistical power of analyses and hence the validity and reliability of findings. (...) Evaluation of programs at multiple locations is also advantageous when it is important to generalize program effects across a diverse range of individuals (...)“ (1991, S. 5).

Die dazu erforderlichen Voraussetzungen des Untersuchungsgegenstandes sind bei Entwicklungsprogrammen üblicherweise nicht gegeben. Um ein solches Programm evaluieren zu können und – worauf das Augenmerk in diesem Beitrag liegt – auf einer empirischen Grundlage theoretisch beschreiben zu können, wurde in zwei Evaluationsstudien (siehe FN 1) eine Verknüpfung von Cluster-Evaluation mit einem programmtheoriebasierten Vorgehen entwickelt.

3. Programmtheoriebasierte Cluster-Evaluation

3.1. Cluster-Evaluation – ein Evaluationsansatz für Entwicklungsprogramme

In den 1990er Jahren wurde im Rahmen von Programmen der W.K. Kellogg Foundation (WKKF) ein damals neuer Ansatz multizentrischer Evaluation entwickelt: Die so genannte Cluster-Evaluation (vgl. Barley/Jenness 1993; WKKF 1995; Worthen/Schmitz 1997; Fitzpatrick/Worthen/Sanders 2004, S. 475-481). Sie grenzt sich explizit gegenüber „multisite evaluations“³ oder „cross-site evaluations“ (Straw/Herrell 2002, S. 7) ab, welche sich typischerweise auf prospektiv angelegte Programme beziehen, deren Implementation *kontrolliert bzw. standardisiert* erfolgt. Vielmehr soll sie anwendbar sein auf Programme, deren einzelne Projekte relativ *autonom* sind und jeweils *eigene kontextbezogene Strategien* entwickeln, um die übergeordneten Programmziele zu erreichen (vgl. Sanders 1997, S. 396). Da die WKKF soziale Verbesserungen anstrebt und dabei den Menschen helfen will, sich selbst zu helfen, fördert sie „empowerment“ der Fachkräfte vor Ort und räumt den geförderten Organisationen ein hohes Maß an Autonomie, Selbstständigkeit und Gestaltungsfreiheit ein (vgl. Fitzpatrick/Sanders/Worthen 2004, S. 479).

Der Begriff des „Clusters“ bezieht sich auf den *Gegenstand* der Evaluation: „A *cluster* is defined as a strategic collection of grants exploring or developing new approaches to an issue related to program area issues or strategy, with moderate focus on sustainability“ (Hervorh. i. Orig., Russon 2005: 66). Vier Schlüsselmerkmale (WKKF 1991, zit. nach O’Sullivan/O’Sullivan 1998, S. 22-23) beschreiben die programmatischen Eckwerte der Cluster-Evaluation:

- „1. It looks across a group of projects to identify *common threads and themes* that, having cross-confirmation, take on greater significance;
2. It seeks not only to learn *what happened* with respect to a group of projects, but *why* those things happened;
3. It happens in a *collaborative* way that allows all players – projects, foundation, and external evaluators – to contribute to and participate in the process so that what is learned is of value to everyone; and
4. The relationship between the projects and the external evaluators conducting the cluster evaluation is confidential. Evaluation findings of the cluster evaluation are reported back to the Foundation only in aggregate for the entire cluster and never for the individual projects. (...)“ (Hervorh. i. Orig., Fitzpatrick/Sanders/Worthen 2004, S. 477).

Innerhalb dieses Rahmens sind unterschiedliche konzeptuelle und methodische Vorgehensweisen möglich (vgl. WKKF 1995, S. 9; O’Sullivan/O’Sullivan 1998, S. 23; Fitzpatrick/Sanders/Worthen 2004, S. 479). So resümieren Rita G. O’Sullivan und John M. O’Sullivan: „Kellogg has defined the cluster evaluation components but has not prescribed the evaluation approach to accompany the process“ (1998, S. 22).

3.2. Verknüpfung von Cluster-Evaluation und Programmtheorie-Evaluation

Der Ansatz der Cluster-Evaluation legt nicht ein spezifisches methodisches Vorgehen fest. Für die Entwicklung eines Evaluationsdesigns wurde in zwei Evaluationsstudien eine Verknüpfung mit einem programmtheoriebasierten Vorgehen erprobt. Dieses beinhaltet, dass die meist implizite Programmtheorie – d.h. die zugrunde liegenden Annahmen, wie über das Programm Veränderungen hervorgebracht werden – durch die Evaluierenden expliziert wird und auf der Grundlage dieser *theoretischen Beschreibung des Evaluationsgegenstandes* die weiteren Erhebungs- und Auswertungsschritte konzipiert werden (vgl. auch Rogers u.a. 2000, S. 5-6; Weiss 1997a, S. 510). Programmtheorie ist dabei nicht immer – sogar eher selten – sozialwissenschaftliche Theorie. Programme im sozialen Bereich basieren häufig auf nicht explizierten, mehr oder weniger plausiblen Annahmen, wie über ein Programm Veränderungen hervorgebracht werden.

Im Kern waren es zwei Überlegungen, die ein programmtheoriebasiertes Vorgehen nahelegten:

- *Erstens* sollten Erkenntnisse über modellhaft erprobte und übertragbare Angebote und Strategien auf der Programmebene gewonnen werden. Im Mittelpunkt des Erkenntnisinteresses standen nicht die Projekte als „Fälle“ des Programms mit ihrer spezifischen Komposition unterschiedlicher und aufeinander abgestimmter Aktivitäten, sondern die jeweils von ihnen eingesetzten *Strategien* als kleinteilige Umsetzungsformen des Programms. Will man diese Strategien entdecken, verstehen und beschreiben, so sind nicht nur ihre konkreten Handlungsschritte zu benennen, sondern auch deren begründete konzeptionelle Verknüpfung mit den angestrebten Zielen im Hinblick auf bestimmte Zielgruppen und Kontexte.
- *Zweitens* sollte sich die Evaluation nicht allein auf die Beschreibung erprobter Strategien beschränken, sondern diese auch bewerten, d.h. Aussagen über erfolgreiche oder weniger erfolgreiche Zugänge treffen. Dies bedeutet, dass die Strategien daraufhin untersucht werden sollten, *ob* und *wie* sie zur Erreichung der Programmziele *beigetragen* haben, aber auch, welche nicht intendierten Wirkungen dabei aufgetreten sind. Eine solche ursächliche Zurechnung setzt jedoch neben der Bestimmung der Strategien und Handlungsschritte (als „Ursachen“) und der angestrebten Ergebnisse (als „Wirkungen“) auch eine Explikation der angenommenen Wirkungszusammenhänge voraus.

Sowohl die Intention, innovative und übertragbare Konzepte zu beschreiben, als auch die Frage nach den Wirkungen erprobter Strategien machen es erforderlich, die „black box“⁴ zwischen Programmimplementation und beobachtbaren Ergebnissen auszuleuchten. Dafür müssen die einzelnen Handlungsschritte und die erforderlichen Ressourcen zur Umsetzung einer fachlichen Strategie bestimmt werden. Auch sind Zielgruppen sowie angestrebte Veränderungen und Kontexte, unter denen fachliche Strategien erprobt werden, zu benennen. Und schließlich müssen Annahmen darüber, wie Aktivitäten zur Erreichung von Zielen bzw. Ergebnissen beitragen, expliziert werden. Eben darauf zielen Ansätze *programmtheoriebasierter Evaluation*, die ihren Ursprung und eine weite Verbreitung im *nordamerikanischen Raum* haben.

Um Komponenten einer Programmtheorie zu veranschaulichen werden gerne graphisch aufbereitete Programm-Modelle herangezogen. Bspw. bildet das folgende einfache „logische Modell“ die Grundstruktur eines Programms ab, indem Komponenten einer Programmtheorie (wie Aktivitäten und Ergebnisse) in einer linearen Abfolge sequentiell verknüpft werden. In der Evaluationspraxis sind solche Programm-Modelle jedoch meist komplexer.

Abb. 1: Grundstruktur eines logischen Modells



Quelle: Adaptiert nach Wholey 1987, S. 77, McLaughlin/Jordan 2004, S. 7-10. sowie United Way of America 1996, S.: 34

Aus der Veränderungsoffenheit von Entwicklungsprogrammen resultiert eine Besonderheit, die für ein programmtheoriebasiertes Vorgehen in der Cluster-Evaluation folgenreich ist: *Fachliche Strategien der Projekte emergieren im Programmverlauf und stehen auch nicht zu Beginn der Evaluation schon fest*. Daher sind für eine Verknüpfung von Cluster-Evaluation mit einem programmtheoriebasierten Vorgehen insbesondere Verfahren einer empirisch begründeten Theorieentwicklung im Verständnis interpretativer oder rekonstruktiver Sozialforschung von Bedeutung. Allerdings zeigte eine Aufarbeitung der Debatte zur Programmtheorie-Evaluation⁵ (vgl. Haubrich 2009), dass trotz der beachtlichen Verbreitung, die programmtheoriebasierte Ansätze in Nordamerika mittlerweile erfahren haben, insgesamt der Frage, *wie* Programmtheorien entwickelt werden können und *wie* die jeweilige Vorgehensweise *theoretisch* und *methodologisch* zu begründen ist, bis heute vergleichsweise wenig Aufmerksamkeit geschenkt wird.

Als eine weitere Herausforderung für eine Anwendung eines programmtheoriebasierten Vorgehens zeigte sich, dass die Besonderheiten multizentrischer Entwicklungsprogramme, die über mehrere Ebenen der Programmumsetzung auf Bundes- und lokaler Ebene mit vielen Akteuren realisiert werden, bislang nicht diskutiert wurden.

3.3. Die impliziten Programmannahmen der Cluster-Evaluation

In der Cluster-Evaluation wird das Verhältnis von Programm und Projekten – hier als „Programmarchitektur“ bezeichnet – in einer ganz spezifischen Weise theoretisch konzipiert. Über das Verhältnis von Programm und Projekten wird angenommen, dass einerseits die thematischen Dimensionen der Programmkonzeption die Projektkonzeptionen beeinflussen bzw. rahmen und andererseits die Projekte Gestaltungsspielräume haben, das Programm zu interpretieren und zu konkretisieren. Sie entwickeln lokal angepasste *Handlungskonzepte* zur Umsetzung der *thematischen Programmdimensionen*. Im zeitlichen Verlauf gewinnt das Programm seine Konturen.

Gegenstand der Cluster-Evaluation ist das Programm als Ganzes (siehe „*WKKF-Schlüsselmerkmale*“). Die Projekte werden trotz ihrer Autonomie und daraus resultierender Heterogenität als Teil eines Programmclusters verstan-

den. Das bedeutet, dass der Untersuchungsgegenstand das *Programmcluster* ist. *Fachliche Strategien und Ergebnisse* werden auf der Ebene des Programmclusters und nicht auf der Ebene der Handlungskonzepte des einzelnen Projektes evaluiert.

In der Konsequenz heißt dies, dass die Projekte als *Programm-Implementationen* verstanden und in eben dieser Eigenschaft zum Untersuchungsgegenstand werden. Projektaktivitäten werden unter dieser Perspektive als *fachliche Strategien* zur Erreichung der *Programmziele*, die in gleicher oder ähnlicher Weise auch von anderen Projekten – als Konkretisierung der *thematischen Programmdimensionen* – umgesetzt werden, analysiert.

Die impliziten Programmannahmen der Cluster-Evaluation lassen sich daher wie folgt zusammenfassen: Das Programm wird im rekursiven *Wechselwirkungsprozess* zwischen den thematischen Programmdimensionen und den heterogenen lokalen Umsetzungsformen – maßgeblich durch *das Handeln der Akteure* in den Projekten und außerdem *im Zusammenwirken* (Information, Diskussion und Aushandlung) der unterschiedlichen Beteiligungsgruppen des Programms – im *zeitlichen Verlauf* der Programmumsetzung konstituiert.

3.4. Programmarchitekturtheorie als Bindeglied zwischen Cluster-Evaluation und Programmtheorie-Evaluation

Die impliziten Programmannahmen der Cluster-Evaluation unterscheiden verschiedene „Wirklichkeitsausschnitte“ des Programms, die jeweils für sich mittels Programmtheorie als fachlich-konzeptionelle Verknüpfung von Handlungsschritten und Zielen beschrieben werden können: (1) die *Programmkonzeption* zur Beschreibung des Erprobungsraumes, (2) die *Handlungskonzepte der Projekte* zur Umsetzung des Erprobungsauftrages und (3) in einer projektübergreifenden und vergleichenden Perspektive die *fachlichen Strategien*.

Um neben den Komponenten einer Programmtheorie und den Annahmen, wie über Aktivitäten unter bestimmten Rahmenbedingungen Ziele erreicht werden sollen, auch jene Annahmen theoretisch zu beschreiben, die das Verhältnis der genannten unterschiedlichen Wirklichkeitsausschnitte eines multizentrischen Programms zueinander explizieren, bedarf es einer Ergänzung des Konzepts der Programmtheorie um eine – wie ich es zur klaren begrifflichen Unterscheidung nenne – *Programmarchitekturtheorie*. Komponenten einer Programmarchitekturtheorie sind die Wirklichkeitsausschnitte oder „Ebenen“ eines multizentrischen Programms – mindestens die „Ebene“ der einzelnen Projekte und die der übergreifenden Programmkonzeption, möglicherweise aber auch weiterer Beobachtungs- oder Analyseebenen. Sie umfasst außerdem Annahmen über deren Zusammenwirken.

Abbildung 2 visualisiert die Theorie der Programmarchitektur eines rekursiven und emergenten multizentrischen Programms und die oben beschriebenen impliziten Annahmen der Cluster-Evaluation über die Programmarchitektur.

Abb. 2: Programmarchitekturtheorie der Cluster-Evaluation

	Cluster-Evaluation	Nicht-standardisierte „Onsite Evaluation at multiple sites“ (Fallstudien mit Synthesebericht)	„Multisite Evaluation“, „Multicenter Clinical Trial“, standardisierte „Onsite Evaluation“
Programmarchitekturannahme	Rekursives multi-zentrisches Programm	Horizontales multi-zentrisches Programm	Vertikales multi-zentrisches Programm
Untersuchungsgegenstand	Programmcluster	Programm als Summe der Projekte	Zentrales Programmmodell
Beobachtungsperspektive	Projekte als Programm-Implementationen (Stichworte: Programmstrategien / Programm-dimensionen)	Projekte als kasuistische (mehr oder weniger heterogene) Programmumsetzungen	Projekte als standardisierte, programmgetreue Implementationen des Programms

Quelle: Eigene Darstellung

Unter dem gemeinsamen Dach eines Programms – der anfangs vorläufigen *Programmkonzeption* – erproben weitgehend autonome Projekte kontextbezogene Aktivitäten und Konzepte. Die *thematischen Dimensionen* des Erprobungsraumes lassen sich zu Beginn nur rudimentär als *vorläufige Programmtheorie* der Programmkonzeption beschreiben. Genau genommen handelt es sich um fachliche und politische Ideen, wie eine Problemlösung erreicht werden könnte. Dann wird ein Cluster aus Projekten zusammengestellt, die diese „Programmvision“ in konkrete Projektkonzeptionen übersetzen und dabei relativ selbstständig agieren – unter Berücksichtigung von Trägerspezifika und lokalen Bedingungen. Bezug nehmend auf die Programmkonzeption werden die Aktivitäten unter der Perspektive untersucht, inwiefern sie Handlungskonzepte zur Erreichung der Programmziele sind (H1, H2, H3 ...).

Die Projekte kooperieren unter einer programmbezogenen Perspektive in einem Netzwerk (Tagungen, Internetplattformen etc.) und tauschen Erfahrungen aus. Mittels Kommunikation und Aushandlung im Netzwerk kann die vorläufige Programmkonzeption nicht nur konkretisiert, sondern auch vor dem Hintergrund ihrer Interpretation durch die Projekte verändert werden. Gegenüber den Programmdimensionen abweichende Deutungen (bspw. Berücksichtigung der Zielgruppe Kinder bei Angeboten eines jugendpolitischen Programms) werden zwischen den Programmakteuren verhandelt und in der Folge entweder zu einem Teil des Clusters oder sie bleiben in der Evaluation unberücksichtigt. Angeregt durch den Informationsaustausch verändern sich zum Teil Projektaktivitäten, womit der „Wechselwirkungsprozess“ zwischen Programmdimensionen und lokalen Umsetzungsformen in eine neue „Schleife“ übergeht.

Die Handlungskonzepte auf Projektebene werden in der Cluster-Evaluation in einer vergleichenden Perspektive zu fachliche Strategien der Programmumsetzung verdichtet. Die fachlichen Strategien beschreiben das „Programmcluster“ als Untersuchungsgegenstand der Cluster-Evaluation.

An diese Klärung konzeptioneller Grundlagen einer programmtheorie-basierenden Cluster-Evaluation schließt sich die Frage an, wie die Evaluation auf einer empirischen Grundlage zu einer theoretischen Beschreibung ihres Untersuchungsgegenstandes gelangt.

4. Rekonstruktion fachlicher Strategien zur theoretischen Beschreibung des Evaluationsgegenstandes

4.1. Programmtheorien auf den „Ebenen“ der Cluster-Evaluation: Programmkonzept, Handlungskonzepte und fachliche Strategien

Mithilfe der Cluster-Evaluation wurden drei Beobachtungs- und Analyseebenen unterschieden, die programmtheoretisch beschrieben werden können. Zwei davon sind als Ebenen der empirischen Beobachtung Gegenstand der Erhebungsschritte der Evaluation: Die Untersuchungsebene der Programmkonzeption und die Untersuchungsebene der Handlungskonzepte der Projekte.

Programmtheorie der Programmkonzeption: Hier kann von einer *Programmtheorie* lediglich in einem sehr basalen Sinne gesprochen werden. Gemeint sind die Annahmen darüber, worin die Problemursachen liegen und welche allgemeinen Konzepte oder Ideen zur Lösung eines gesellschaftlichen Problems erprobt werden sollen. Auf der Beobachtungsebene der Programmkonzeption geht es zunächst um die Explikation der thematischen Programmdimensionen und Programmziele. Dafür wurden bspw. in der Evaluationsstudie „Interkulturelles Netzwerk“ zu Beginn der Evaluation die Erprobungsabsichten aus der Perspektive derjenigen Akteure nachgezeichnet, die an der Gestaltung der Programmkonzeption mitgewirkt haben – und die auch für den weiteren Verlauf des Programms über eine gewisse Definitionsmacht bezüglich der Programmgestaltung verfügten (Ministerium und Trägerverbände auf Bundesebene). Aus vielfältigen und zum Teil diffusen Zielen und Themen wurden die inhaltlichen Kerndimensionen des Programms (Öffnung der Einrichtungen und Vernetzung im Sozialraum) herauskristallisiert, um die Evaluation auf die gemeinsamen Themen des Clusters fokussieren zu können. Die zu Beginn entworfene Beschreibung der Programmkonzeption wird über weitere Erhebungs- und Auswertungsschritte im Programmverlauf konkretisiert und ausdifferenziert.

Programmtheorien der Handlungskonzepte: In der Cluster-Evaluation werden die Projektaktivitäten unter der Perspektive analysiert, inwiefern sie *Implementationsformen des Programms* sind. Hierfür wird zunächst ihre Bedeutung im Projektkontext rekonstruiert. Sind Projektaktivitäten (im Verlauf der Aushandlungsprozesse der Programmteiligen) den Programmdimensionen zurechenbar, dann werden sie in der Cluster-Evaluation als *fachliche Handlungskonzepte der Programmumsetzung* im Sinne von kleinteiligen „Theorien der

Veränderung“ des Programms (i.e. Programmtheorien) interpretiert.⁶ Die Handlungskonzepte der Projekte sind die kleinste „Untersuchungseinheit“, die mittels Programmtheorie durch die Evaluation auf einer empirischen Grundlage rekonstruiert wird. In der Rekonstruktion der Aktivitäten als fachliche Handlungskonzepte liegt der Fokus auf den *potenziell vergleichbaren bzw. übertragbaren Komponenten und verknüpfenden Annahmen*, die den Aktivitäten der Projektmitarbeitenden zugrunde liegen – nicht in den situativen oder persönlichkeitsbezogenen Besonderheiten.

Programmtheorien modellhaft erprobter fachlicher Strategien: Für die Programmevaluation werden die rekonstruierten programmbezogenen fachlichen Handlungskonzepte der Projekte in einer vergleichenden Perspektive projektübergreifend zusammengeführt. Die typisierende Zusammenführung erfolgt mithilfe des *Konstrukts der fachlichen Strategien*. Um wesentliche Merkmale einer fachlichen Strategie herausarbeiten zu können, wird in der Cluster-Evaluation ein Vergleich ähnlicher Handlungskonzepte, die von verschiedenen Projekten erprobt werden, angestrebt. Über die vergleichende Betrachtung gelangt die Cluster-Evaluation zur typisierenden Beschreibung fachlicher Strategien, die an mehreren Standorten oder in unterschiedlichen Kontexten in ähnlicher Weise umgesetzt werden. Auf die fachlichen Strategien richtet sich der Bewertungsauftrag der Cluster-Evaluation: Die Programmevaluation evaluiert das Programm im Hinblick auf die erprobten fachlichen Strategien und geht der Frage nach, ob diese sich als erfolgreich oder weniger erfolgreich erwiesen haben und was man daraus für eine Weiterentwicklung der Fachpraxis und -politik lernen kann.

4.2. Programmtheorien als Repräsentationen professionellen berufsfeldspezifischen Wissens

Fachliches Handeln ist als ein komplexes Phänomen zu begreifen, in dem professionelles Können, fachliches Wissen, berufliche Ethik, Werthaltungen und persönliche Orientierungen miteinander verwoben sind und das zudem in Prozesse der Ko-Produktion eingebettet ist. (Sozial-)Pädagogisches Wissen und Können ist somit anders und vielschichtiger zu verstehen als etwa aus sozialwissenschaftlichen Theorien abgeleitete Annahmen über Phänomene der sozialen Wirklichkeit – also jener Theorien, die Huey-Tsyh Chen und Peter H. Rossi in ihren frühen Veröffentlichungen zur „theory-driven evaluation“ in den Mittelpunkt ihrer Überlegungen rücken (1980, S. 111; 1983, S. 300). Angesichts dieser Vielschichtigkeit professionellen Wissens, seiner Verankerung in wissenschaftlichem Wissen wie auch in handlungspraktischem Erfahrungswissen, seiner Bezogenheit auf den Fall, die Organisation und die Persönlichkeit der Fachkraft und seiner Mischung aus expliziten und impliziten Wissensbeständen, kann nicht davon ausgegangen werden, dass fachliche Handlungskonzepte unmittelbar im Interview abgefragt werden können – wie es bspw. Ray Pawson und Nick Tilley vorsehen (1997, S. 173).

Für einen rekonstruktiven Forschungszugang wird vorgeschlagen, Programmtheorien auf der Beobachtungsebene der Projektaktivitäten als Repräsentationen professionellen berufsfeldspezifischen Wissens der Projektmitarbeitenden empirisch fundiert zu entwickeln. *Programmtheorien fachlicher Hand-*

lungskonzepte repräsentieren die *Annahmen der Sozialpädagog(inn)en* über das Zusammenwirken von Ressourcen, Aktivitäten und Ergebnissen zum Erreichen der Programmziele in benennbaren Kontexten. Diese Annahmen werden als ein konstitutiver Aspekt fachlichen Handelns von Sozialpädagog(inn)en in den Modellprojekten vorausgesetzt. Dies erscheint insofern gerechtfertigt, als gerade im Konzipieren, Entwickeln und Erproben innovativer Strategien und Arbeitsformen die Aufgabe der Sozialpädagog(inn)en bei der Umsetzung des Modellvorhabens liegt.

Ein programmtheoriebasiertes Vorgehen in der Cluster-Evaluation gründet weiter auf der Vorannahme, dass zielgerichtete Handlungsentwürfe als unterstellter inhärenter Bestandteil fachlichen Handelns in einer strukturierten Form als Programmtheorien *zu weiten Teilen* sprachlich repräsentiert, d.h. nachvollziehbar beschrieben und kommuniziert werden können. Das soll keineswegs heißen, dass Handlungsentwürfe *vollständig* zur Sprache gebracht und unmittelbar abgefragt werden können. Es wird allerdings auch nicht angenommen, dass Sozialpädagog(inn)en *grundsätzlich nicht* die Annahmen, die ihrem fachlichen Handeln zugrunde liegen, explizieren oder in Beschreibungen ihrer Praxis zum Ausdruck bringen können.

Ausgehend von der Annahme, dass Wissen – konkreter formuliert: berufsfeldspezifisches, professionell in Handlungssituationen relationiertes Wissen – für das fachliche Handeln der Mitarbeiter(innen) in den Modellprojekten konstitutiv ist (vgl. auch Luckmann 1986: 191f.), wurde in zwei Evaluationsstudien eine spezifische Variante des Expert(innen)interviews („programmtheorie-rekonstruierende ExpertInneninterviews“, vgl. Haubrich 2009) eingesetzt, dessen Grundzüge – wissenssoziologische Grundlagen, Gesprächsführung und Auswertungsschritte – an anderer Stelle ausgearbeitet wurden. Anliegen dieser mehrstündigen (beim ersten Projektbesuch bis zu sechs oder acht Stunden) und auf zwei Tage verteilten Interviews war es, die Fachkräfte zu selbstläufigen Beschreibungen ihrer fachlichen Praxis anzuregen, möglichst ohne durch direktives Nachfragen argumentatives und legitimierendes Antwortverhalten zu provozieren. Im Programmverlauf wurde jedes Projekt mehrfach (ein bis zweimal pro Jahr) befragt.

Über mehrere Auswertungsschritte wurden die rekonstruierten Projektaktivitäten vor dem Hintergrund der Programmdimensionen als in bestimmten Kontexten erprobte programmbezogene Handlungskonzepte interpretiert. Diese wurden in einem weiteren Schritt in einem fallübergreifenden Vergleich zu fachlichen Strategien verdichtet. Die so rekonstruierten Programmtheorien fachlicher Strategien lassen sich verstehen als *typisierte Repräsentationen* von Annahmen der Projektmitarbeiter(innen), die ihrem fachlichen Handeln im Modellprogramm zugrunde liegen. Sie explizieren das konzeptionelle Zusammenwirken von Ressourcen, Aktivitäten, Interaktions- und Kommunikationsformen, Zielgruppen, Kontexten und Ergebnissen vor dem Hintergrund bestimmter Problemdeutungen, Angemessenheitsentscheidungen und wertbezogener Begründungen. Mit anderen Worten: Sie repräsentieren projektübergreifend typisierte fachlich-konzeptionelle Handlungsschemata, die auf die Planung von Handlungsabläufen zur Lösung als problematisch gedeuteter Situationen in bestimmten Kontexten und mit bestimmten Zielgruppen angelegt sind.

5. Wie gelangt eine rekonstruktive Programmtheorie-Evaluation zu Bewertungen?

Das Hauptaugenmerk dieses Beitrages richtete sich auf die Frage, wie ein komplexes multizentrisches Programm – hier ein Bundesmodellprogramm – auf einer empirischen Grundlage theoretisch beschrieben werden kann. Eine Programmtheorie-Evaluation kann sich nicht auf die Beschreibung des Programms beschränken, sondern nimmt die Rekonstruktion der fachlichen Strategien zum Ausgangspunkt, um zu einer Bewertung zu gelangen. Als Grundlage dafür wurden in den beiden Evaluationsstudien die – über Dokumentenanalysen und Interviews zur Programmkonzeption und zu den Projektaktivitäten – erhobenen Ziele aus der Perspektive unterschiedlicher Programmakteure herangezogen. Diese wurden den weiteren Beobachtungen zu den im Programmverlauf erreichten Ergebnissen und Nebenfolgen gegenübergestellt. Da die Programmziele aus verschiedenen Akteursperspektiven je anders formuliert wurden, führte dieses Vorgehen dazu, dass die gleiche Beobachtung bezogen auf unterschiedliche Ziele je anders zu bewerten war.

Um die Prozesse der Umsetzung des Programms, begünstigende und hinderliche Kontextbedingungen sowie Ergebnisse auf Seiten der Zielgruppen beschreiben zu können, wurden unterschiedliche Erhebungsinstrumente eingesetzt. In der Evaluationsstudie „Interkulturelles Netzwerk“ wurden bspw. neben den Expert(inn)eninterviews mit Projektmitarbeitenden, Interviews mit Kooperationspartnern im Sozialraum durchgeführt sowie teilstandardisierte Fragebögen eingesetzt. Die Evaluation entwickelte nach den ersten Projektbesuchen unterschiedliche Erhebungsinstrumente zu bestimmten thematischen Schwerpunkten der Programmumsetzung. Aus diesem „Erhebungsportfolio“ wurden für jedes Projekt die geeigneten ausgewählt.

Das skizzierte Vorgehen veranschaulicht, wie die beschriebene Konzeption einer rekonstruktiven Programmtheorie-Evaluation für multizentrische Entwicklungsprogramme in ein Evaluationsdesign eingebettet werden kann, um zu einer Bewertung fachlicher Strategien zu gelangen. Andere Optionen – wie sie in Wolfgang Beywls (2006) Systematisierung von Evaluationsmodellen nach ihrer Berücksichtigung von Werten aufgezeigt werden – sind aber ebenso denkbar, wie bspw. die moderierte Rückspiegelung der Ergebnisse an einzelne Akteursgruppen zur Vermittlung der Evaluationsergebnisse für unterschiedliche Verwendungskontexte oder die Einrichtung einer unabhängigen Expert(inn)enkommission zur Entwicklung von Empfehlungen für Fachpolitik und -praxis.

Die vorgestellte wissenssoziologisch begründete Konzeption einer rekonstruktiven Programmtheorie-Evaluation erweitert das Spektrum an Zugängen programmtheoriebasierter Evaluation. Sie bietet eine theoretisch begründete und methodologisch reflektierte Konzeption für eine empirisch fundierte Entwicklung von Programmtheorien. Interpretative und rekonstruktive Verfahren zur Entwicklung von Programmtheorie werden in jüngerer Zeit zunehmend im europäischen Evaluationsdiskurs verhandelt. In der nordamerikanischen Debatte bleiben diese jüngeren Entwicklungen noch weitgehend unbeachtet – wenn gleich Huey-Tsyh Chen und Peter H. Rossi (1992, S. 8) als Väter der „theory-driven evaluation“ schon in den 1990er Jahren die Notwendigkeit der Entwicklung qualitativer Zugänge der Entwicklung von Programmtheorie betont haben.

Anmerkungen

- 1 „Mobile Jugendsozialarbeit für junge Menschen ausländischer Herkunft“ (Haubrich/Frank 2000) sowie „Interkulturelles Netzwerk der Jugendsozialarbeit im Sozialraum“ (Vossler/Obermaier 2003).
- 2 Der Begriff der Emergenz weist darauf hin, dass im Zusammenwirken mehrerer Akteure in der Programmumsetzung neue und nicht vorhersagbare Qualitäten entstehen. Zu betonen ist jedoch, dass hier nicht die Argumentation geteilt wird, Entwicklungsprogramme erforderten auch eine formative Entwicklungsevaluation („developmental evaluation“, Patton 1994; 2008: 278;), bei der die Evaluator(inn)en Teil eines Entwicklungsteams sind: „The evaluator’s primary function in the team is to elucidate team discussions with evaluative data and logic, and to facilitate data-based decision-making in the developmental process“ (ders. 1994: 117).
- 3 Sinacore und Turpin (1991) bezogen den Begriff „multisite evaluations“ zunächst auf alle Formen multizentrischer Evaluation. In späteren Veröffentlichungen (Fitzpatrick/Sanders/Worthen 2004; Straw/Herrell 2002) wird der Begriff jedoch meist verwendet, um eine bestimmte Form multizentrischer Evaluation, welche auf einem standardisierten, in aller Regel quasi-experimentellen Design basiert, zu bezeichnen. Dieser Begriffsverwendung schließe ich mich an. Fitzpatrick, Sanders und Worthen schlagen vor „multiple-site evaluations“(2004: 464) als Dachbegriff zu verwenden.
- 4 Der Begriff „black box“ wird meist im Kontext der Kritik an jenen (quasi-)experimentellen Studien verwendet, die Wirkungen als den Zusammenhang zwischen einer Maßnahme und erzielten Ergebnissen konzipieren, ohne die Schritte und Zusammenhänge zu klären, die zwischen beiden Ereignissen liegen (vgl. zum Begriff bspw. Rossi/Lipsey/Freemann 2004: 165, 424; Posavac/Carey 2007: 25; Bezzi 2006: 59).
- 5 Dabei wurde ebenso Literatur zur „theory-driven evaluation“ (insb. Chen 1990; Chen/Rossi 1980), „theory-based evaluation“ (Weiss 1998, S. 46-71; 1997a, b) und „programmtheory-based evaluation science“ (Donaldson 2007) berücksichtigt, wie auch zu nutzungsorientierten Zugängen der Programmtheorie-Evaluation, bspw. der „user-focused theory of change approach“ (Patton 2008, S. 345) oder das „evaluability assessment“ (Wholey 1987), und zur Anwendung von logischen Modellen und anderen strukturierten Schemata zur Abbildung der Programmtheorie (bspw. Frechtling 2007; Knowlton/Phillips 2009)
- 6 Sind sie nicht eindeutig den Programmdimensionen zuordenbar, werden sie gewissermaßen als „Interpretationsvorschlag zum Gesamtprogramm“ behandelt. Über Aushandlungsprozesse in den Netzwerkstrukturen des Programms können sie zu einer Erweiterung der Programmkonzeption führen. Bspw. wenn Sprachkurse im Kindergarten (Zielgruppe: Kinder) in einem Programm der Jugendberufshilfe als Prävention von Integrationsproblemen im Jugendalter eingestuft werden.

Literatur

- Barley, Zoe A./Jenness, Mark: Cluster Evaluation. A Method to Strengthen Evaluation in Smaller Programs with Similar Purposes. *Evaluation Practice* 14 (1993), H. 2, S. 141-147.
- Beywl, Wolfgang (2006): Evaluationsmodelle und qualitative Methoden. In: Flick, Uwe (Hrsg.): *Qualitative Evaluationsforschung. Konzepte, Methoden, Umsetzungen*. Reinbek bei Hamburg, S. 92-116.
- Bezzi, Claudio (2006): Evaluation Pragmatics. *Evaluation* 12(1): 56-76.
- Chen, Huey-Tsyh (1990): *Theory-Driven Evaluations*. Newbury Park/London/New Delhi.
- Chen, Huey-Tsyh/Rossi, Peter H. (1992): Introduction: Integrating Theory into Evaluation practice. In: dies. (Hrsg.): *Using Theory to Improve Program and Policy Evaluations*. New York/Westport/London., S. 1-11.
- Chen, Huey-Tsyh/Rossi, Peter H. (1983): Evaluating With Sense. The Theory-Driven Approach. *Evaluation Review* 7(3): 283-302.

- Chen, Huey-Tsyh/Rossi, Peter H. (1980): The Multi-Goal, Theory-Driven Approach to Validity. A Model Linking Basic and Applied Social Science. *Social Forces* 59(1): 106-122.
- Dietzel, Gottfried, T.W. (1988): Modellvorhaben und ihre Evaluierung – Überlegungen und Erfahrungen im Bundesministerium für Jugend, Familie, Frauen und Gesundheit. In: Dietzel, Gottfried T.W./von Troschke, Jürgen (Hrsg.): Begleitforschung bei staatlich geförderten Modellprojekten – strukturelle und methodische Probleme. Stuttgart/Berlin/Köln/Mainz, S. 13-30.
- Donaldson, Steward I. (2007): Program Theory-Driven Evaluation Science. Strategies and Applications. New York/London.
- Fitzpatrick, Jody L./Worthen, Blaine R./Sanders, James R. (Hrsg.) (2004 [1993]): Program Evaluation. Alternative Approaches and Practical Guidelines. 3. Auflage. Boston [u.a.]
- Frechting, Joy A. (2007). Logic Modeling Methods in Program Evaluation. San Francisco: Jossey-Bass.
- Häußler, Monika/Stöbel, Ulrich/von Troschke, Jürgen/Walterspiel, Gabriela/Wetterer, Angelika (1988): Konzepte und Erkenntnisinteressen der wissenschaftlichen Begleitforschung von Modelleinrichtungen. In: Dietzel, Gottfried T.W./von Troschke, Jürgen (Hrsg.): Begleitforschung bei staatlich geförderten Modellprojekten – strukturelle und methodische Probleme. Stuttgart/Berlin/Köln/Mainz, S. 31-105.
- 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, S. 101-122.
- Karin Haubrich (2009): Sozialpolitische Innovation ermöglichen. Die Entwicklung der rekonstruktiven Programmtheorie-Evaluation am Beispiel der Modellförderung in der Kinder- und Jugendhilfe. Internationale Hochschulschriften, Band 536. Münster/New York/München/Berlin:
- Haubrich, Karin/Frank, Kerstin (2000): Vom Aufsuchen zur beruflichen Integration. Evaluationsstudie zum Bundesmodellprogramm „Mobile Jugendsozialarbeit für junge Menschen ausländischer Herkunft“. München.
- Haubrich, Karin/Vossler, Andreas (2001): Das Modellprogramm „Interkulturelles Netzwerk der Jugendsozialarbeit im Sozialraum“ – Erste Ergebnisse und Erfahrungen. In : BAG JAW (Bundesarbeitsgemeinschaft Jugendsozialarbeit) (Hrsg.): Jugend-Beruf-Gesellschaft. Arbeitsdruck 39. Sozialanalyse. Bonn, S. 22-33.
- Hellstern, Gerd-Michael/Wollmann, Hellmut (1983a): Evaluierungsforschung. Ansätze und Methoden – dargestellt am Beispiel des Städtebaus. Basel/Boston/Stuttgart.
- Knowlton, Lisa Wyatt/Phillips, Cynthia C. (2009): The Logic Model Guidebook. Better Strategies for Great Results. Los Angeles/London/New Delhi/Singapore.
- Kromrey, Helmut (1995): Evaluation. Empirische Konzepte zur Bewertung von Handlungsprogrammen und die Schwierigkeiten ihrer Realisierung. *Zeitschrift für Sozialisationsforschung und Erziehungssoziologie* 15(4): 313-336.
- Luckmann, Thomas (1986): Grundformen der gesellschaftlichen Vermittlung des Wissens. Kommunikative Gattungen. Sonderheft „Kultur und Gesellschaft“, hrsg. v. Neidhardt/Lepsius/Weiß. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 38 (Sonderheft 27): 191-211.
- McLaughlin, John A./Jordan, Gretchen B. (2004): Using Logic Models. In: Wholey, Joseph S./Hatry, Harry P./Newcomer, Kathryn E. (Hrsg.): Handbook of Practical Program Evaluation. 2. Auflage. San Francisco, S. 7-32.
- O’Sullivan, Rita G./O’Sullivan, John M. (1998): Evaluation Voices. Promoting Evaluation from within Programs through Collaboration. *Evaluation and Program Planning* 21(1): 21-29.
- Patton, Michael Q. (2008 [1978]): Utilization-Focused Evaluation. 4. Auflage. Los Angeles/London New Delhi/Singapore.
- Patton, Michael Q. (1994): Developmental Evaluation. *Evaluation Practice* 15(3): 311-319.
- Pawson, Ray/Tilley, Nick (1997): Realistic Evaluation. London/Thousand Oaks/New Delhi.

- Posavac, Emil J./Carey, Raymond G. (2007 [1980]): Program Evaluation. Methods and Case Studies. 7. Auflage. Upper Saddle River.
- Rogers, Patricia J./Petrosino, Anthony/Huebner, Tracy A./Hacsi, Timothy A. (2000): Program Theory Evaluation. Practice, Promise, and Problems. In: Rogers, Patricia J./Hacsi, Timothy A./Petrosino, Anthony/Huebner, Tracy A. (Hrsg.): Program Theory in Evaluation. Challenges and Opportunities. New Directions for Evaluation, Nr. 87. San Francisco, S. 5-13.
- Rossi, Peter H./Lipsey Mark W./Freeman, Howard E. (2004 [1979]): Evaluation. A Systematic Approach. 7. Auflage. Thousand Oaks/London/New Delhi.
- Russon, Craig (2005): Cluster Evaluation. In: Mathison, Sandra (Hrsg.): Encyclopedia of Evaluation. Thousand Oaks/London/New Delhi, S. 66-67.
- Sanders, James R. (1997): Cluster Evaluation. In: Chelimsky, Eleanore/Shadish, William R. (Hrsg.): Evaluation for the 21st Century. A Handbook. Thousand Oaks/London/New Delhi, S. 396-404.
- Sinacore, James M./Turpin, Robin S. (1991): Multiple Sites in Evaluation Research. A Survey of Organizational and Methodological Issues. In: Turpin, Robin S./Sinacore, James M. (Hrsg.): Multisite Evaluations. New Directions for Evaluation, Nr. 50. San Francisco, S. 5-18.
- Straw, Roger B./Herrell, James M. (2002): A Framework for Understanding and Improving Multisite Evaluations. In: Herrell, James M./Straw, Roger B. (Hrsg.): Conducting Multiple Site Evaluations in Real-World Settings. New Directions for Evaluation, Nr. 94. San Francisco, S. 5-15.
- United Way of America (1996): Measuring Program Outcomes. A Practical Approach. Alexandria.
- Vossler, Andreas/Obermaier, Andrea (2003): Netze knüpfen – Integration fördern. Evaluationsstudie zum Bundesmodellprogramm „Interkulturelles Netzwerk der Jugendsozialarbeit im Sozialraum“. München.
- Weiss, Carol H. (1998 [1972]): Evaluation. Methods for Studying Programs and Policies. 2. Auflage. Upper Saddle River.
- Weiss, Carol H. (1997a): How Can Theory-Based Evaluation Make Greater Headway? *Evaluation Review* 21(4): 501-524.
- Weiss, Carol H. (1997b): Theory-Based Evaluation. Past, Present and Future. In: Rog, Debra J./Fournier, Deborah (Hrsg.): Progress and Future Directions in Evaluation. Perspectives on Theory, Practice, and Methods. New Directions for Evaluation, Nr. 76. San Francisco, S. 41-55.
- Wholey, Joseph S. (1987): Evaluability Assessment. Developing Program Theory. In: Bickman, Leonard (Hrsg.): Using Program Theory in Evaluation. New Directions for Program Evaluation, Nr. 33. San Francisco/London, S. 77-92.
- Wiesner, Reinhard (Hrsg.) (2006 [1991]): SGB VIII Kinder und Jugendhilfe. Kommentar. 3. Auflage. München.
- WKKF (1995): W.K. Kellogg Foundation Cluster Evaluation Model of Evolving Practices. Battle Creek.
- WKKF (1991): Rural American Cluster Evaluation Networking Conference. Cluster Studies – Background Information to Be Used in Small Group Discussions. Battle Creek.
- Worthen, Blaine R./Schmitz, Constance C. (1997): Conceptual Challenges Confronting Cluster Evaluation. *Evaluation* 3(3): 300-319.

