

Evaluation präventiver Praxis gegen Rechtsextremismus

Positionen, Konzepte und Erfahrungen

Michaela Glaser & Silke Schuster (Hrsg.)



Arbeits- und Forschungsstelle
Rechtsextremismus
und Fremdenfeindlichkeit



HELMUT KROMREY

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

Vorbemerkungen

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

Gleichartige methodologische Rahmenbedingungen für wissenschaftliche Begleitung und Evaluation: Angewandte Forschung

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

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

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

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

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

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

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

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

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

Begriffsklrung: Was ist Evaluation?

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

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

- Er bezieht sich zum einen auf eine „symbolische“ Ebene: Als aus dem Amerikanischen ubernommene Bezeichnung fr Bewertung / Bewerten sowie fr Denkmodelle bzw. Konzepte systematischer Bewertungsverfahren. Wenn wir uns ber Methoden oder Standards der Evaluation¹ unterhalten, beziehen wir uns auf diese Ebene.
- Der Begriff kann aber auch auf ein spezifisches empirisches Handeln, auf einen Prozess zielorientierten Informationsmanagements gemnzt sein: Auf eine (methodisch kontrollierte) verwertungs- und bewertungsorientierte Form des Sammelns und Auswertens von Informationen. Wenn von Begleitforschung die Rede ist, dann steht dieser Aspekt im Vordergrund.
- Und schlielich kann damit auch das Resultat dieses Prozesses, ein Werturteil ber das Zielobjekt des Evaluierens (das „Evaluationsprodukt“) angesprochen werden. Dies ist der Fall, wenn wir eine Aussage hren, das Programm X oder das Institut Y sei „positiv“ oder „negativ“ evaluiert worden.

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

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

¹ Siehe z. B. die von der Deutschen Gesellschaft fr Evaluation beschlossenen „Standards fr Evaluation“ (DeGE-val 2002), die in einem laufenden Revisionsprozess weiter entwickelt werden (www.degeval.de).

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

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

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

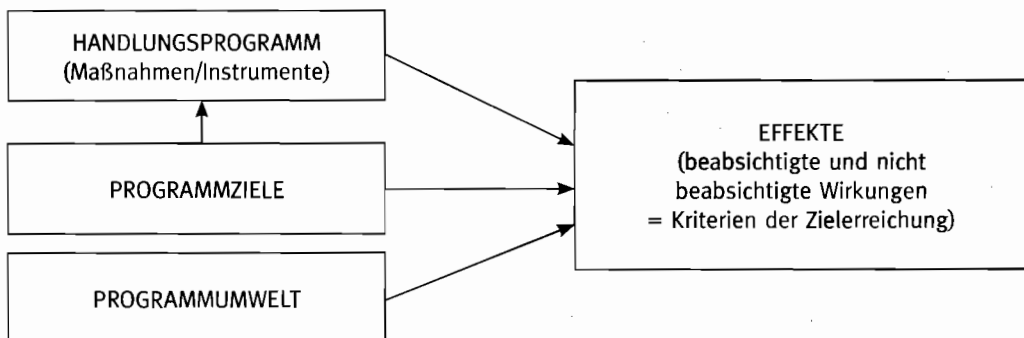
Programm angestrebten Sollwerten zu vergleichen. Sie ist somit kein eigentliches „Werturteil“, sondern eine technologische Vergleichsaussage. Forschung und Evaluierung fallen in diesem Konzept definitionsgemäß zusammen.

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

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

-
- 2 Dieser Typ von Evaluationsforschung wird üblicherweise „Programmevaluation“ genannt. Das ist insofern korrekt, als das Design neben der Erhebung und Analyse evaluationsrelevanter empirischer Daten auch Verfahren der Ableitung datenbasierter Werturteile (also Evaluationen) bereitstellt.
 - 3 Soweit alternative Evaluationstypen in ihrem Vorgehen von diesem Konzept abweichen, verwenden sie für die empirische Informationsgewinnung die üblichen Methoden quantitativer und/oder qualitativer Sozialforschung, auf die hier nicht eingegangen werden kann (vgl. dazu Lehrbücher, etwa Kromrey 2006; Lamnek 1995).

Abbildung 1: Programm-Evaluation



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

An konkreten Forschungsaufgaben folgen daraus:

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

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

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

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

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

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

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

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

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

Ziel- und Maßnahmenanalyse

Nimmt man die gängigen Definitionen von „Interventionsprogramm“ oder „Handlungsprogramm“ beim Wort⁴, dann müssten die Kenntnisse, die man sich durch die Forschung erhofft, bei den Akteuren weitestgehend schon vorhanden sein: Neben einem widerspruchsfreien Zielsystem müsste zuverlässiges Praxiswissen existieren, um auf der Basis von Daten über die gegebene Ausgangssituation die erforderlichen Maßnahmen und Instrumente zur Erreichung der Zielsituation zu bestimmen. Solche Kenntnisse über Ziel-Mittel-Relationen müssten zudem technologisch verwertbar sein; d.h. die als strategisch wichtig erkannten Variablen müssten dem Eingriff der Programmdurchführenden zugänglich sein. Mayntz weist jedoch mit Recht darauf hin, „das nur im Ausnahmefall ein Programm zu Beginn des Implementationsprozesses als konkrete, faßbare Einheit vorliegt“ (1980, S. 4). So sind die Ziele oft nicht eindeutig und nicht konkret, sondern vage und leerformelhaft formuliert. Das kann ganz bewusst im Prozess der Programmaushandlung geschehen sein (um einen Kompromiss zwischen widerstreitenden Interessen zu ermöglichen oder um Konflikte zwischen parlamentarischen Koalitionspartnern zu vermeiden). So wird die Aufgabe der Präzisierung aus der (öffentlich-politischen) Zielfindungs- in die (weniger öffentliche) Implementierungsphase verschoben. Besonders problematisch ist dies, wenn am Programmvollzug mehrere Ebenen beteiligt sind (Bund/Länder/Kommunen/andere Träger). Auf jeder Ebene von Akteuren kann die Ausfüllung der Ziel-Leerformeln in unterschiedlicher Weise geschehen. Gleiches kann bei der Zuordnung von Maßnahmen (Instrumenten) der Fall sein, die der Zielerreichung dienen sollen. So können im Prinzip unterschiedliche Maßnahmen mit dem gleichen Programm vereinbar erscheinen. Eine mögliche Folge ist, dass bestimmte Träger ihre bereits auf Vorrat bestehenden „Schubladenpläne“ unter das neu beschlossene Programm subsumieren. In solchen Fällen entstehen Diskrepanzen zwischen offiziellen (manifesten) und verdeckten (latenten) Programmzielen. Außerdem können Ziele, die in ihrer vagen Formulierung als miteinander vereinbar erscheinen, sich bei der Konkretisierung als im Widerspruch zueinander stehend erweisen. Des Weiteren können zu Beginn gesetzte Ziele (selbst wenn sie präzise formuliert waren) sich im Laufe der Programmrealisierung ändern oder in ihrer Gewichtung verschieben, etwa weil sich wichtige Rahmenbedingungen für das Programm in nicht erwarteter Weise entwickelt haben. In manchen Programmen schließlich finden sich eher Kataloge von Maßnahmen als eindeutige Ziele, so dass unbestimmt bleibt, was mit dem Programm letztlich erreicht werden soll, welches also die angestrebten Effekte sind.

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

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

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

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

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

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

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

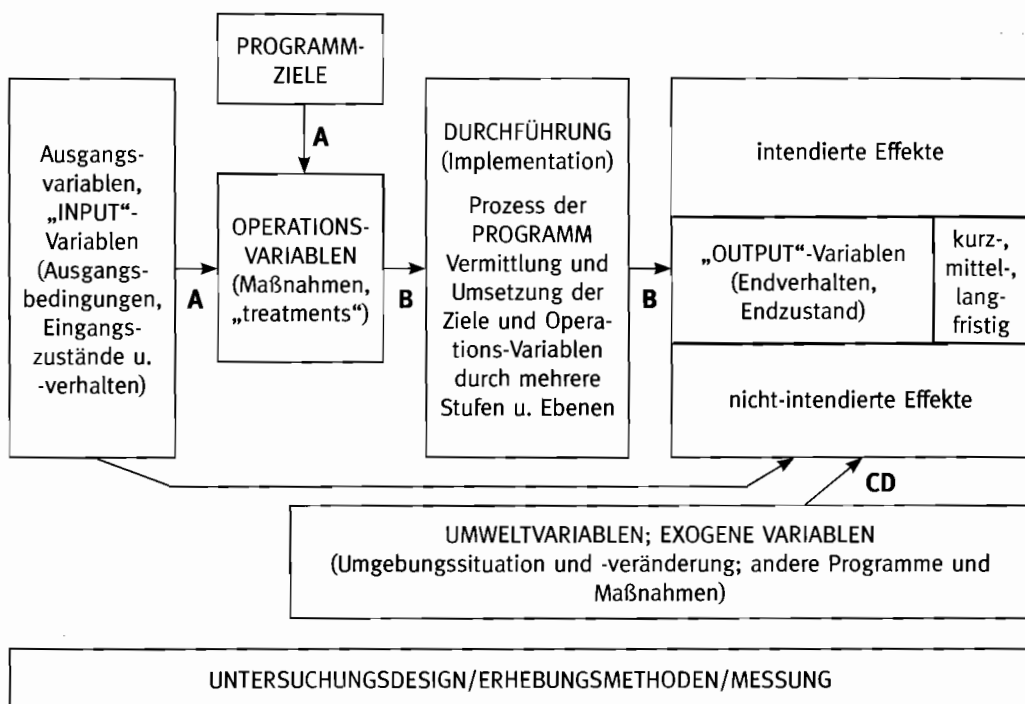
Konzipierung der Wirkungen (Modell des Wirkungsfeldes)

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

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

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

Abbildung 2: Grobes Variablenmodell einer Evaluationsstudie



Erläuterungen:

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

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

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

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

Forschung – im Normalfall nur näherungsweise (in Form diverser quasi-experimenteller Ansätze) realisierbar⁵.

Das Werturteilsdilemma in der Evaluation

Wenn von Evaluation – in welchem Zusammenhang und in welcher spezifischen Begriffsbedeutung auch immer – gesprochen wird, ist damit eine bewertende Aussage über einen Gegenstand oder Sachverhalt gemeint. Kommt – wie hier dargestellt – als Erkenntnisinstrument auch empirische Forschung ins Spiel, kann das Ziel von Evaluation präzisiert werden als: empirisch gestützte Gewinnung von Bewertungen mit intersubjektivem Geltungsanspruch, häufig zugespitzt auf die Forderung: forschungsgestützte Gewinnung von Qualitätsaussagen. Das Ziel ist in jedem Fall eine normative Aussage, ein „Werturteil“. Für das Konzept „Evaluation durch Forschung“⁶ führt dies zu einem gravierenden Problem, sind doch nach herrschender Meinung in Methodologie und Wissenschaftstheorie normative Aussagen – Werturteile generell und damit auch Qualitätsurteile im Besonderen – empirisch nicht begründbar. Überraschenderweise wird dieses Dilemma – abgesehen von ganz seltenen Ausnahmen⁷ – im Diskussionskontext von Evaluationsforschung selten thematisiert. Aus der Perspektive der analytisch-homologischen Wissenschaftstheorie fällt die Antwort nicht schwer; der Anspruch, „empirisch gestützte Bewertungen mit intersubjektivem Geltungsanspruch“ zu gewinnen, ist methodologisch nicht legitimierbar:

- Die unmittelbare empirische Begründung von Bewertungen durch Forschung ist nicht möglich; auch aus korrekten empirischen Beschreibungen und Analysen sind normative Aussagen nicht ableitbar.
- Die Geltungsbegründung empirischer Analysen folgt einer anderen Logik als die Geltungsbegründung normativer Aussagen. Für die ersteren gibt es in der Wissenschaftstheorie klare Regeln, für die letzteren nicht.
- Möglich ist es lediglich, die Forschung möglichst genau auf den Zweck „Bewertung“ auszurichten, indem wenn schon nicht direkt bewertende, so doch bewertungsrelevante Informationen gesammelt und systematisiert werden⁸.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

⁹ siehe Fußnote 4.

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

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

Evaluation als Qualitätsmessung: das methodologische Problemkind

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

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

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

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

Evaluation durch Befragung: das erkenntnistheoretische Problemkind

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

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

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

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

Sofern dies zuträfe¹⁰, wäre die Evaluation per Befragung der Königsweg zur Lösung aller Probleme der Evaluationsforschung – auch der Werturteilsproblematik, denn die Bewertungen nehmen hier die „per Betroffenheit dazu Legitimierten“ vor. Die Forschung selbst bleibt neutral; sie erhebt, systematisiert und analysiert lediglich. Zwar sind „Messungen“ per Befragung nicht so problemlos wie dies dem Laien häufig erscheint (vgl. Kromrey 2006, S. 257ff.). Doch sofern systematische Verzerrungen vermieden werden können, lassen sich in der Tat bei hinreichend großer Befragtenzahl und bei repräsentativer Datenbasis individuelle Unterschiede – wie im obigen Zitat behauptet – herausmitteln“. Einzulösen sind hierfür lediglich durch das Erhebungsinstrument und in der Befragungssituation einige formale methodologische Voraussetzungen die allerdings nicht ohne weiteres als erfüllt gelten können:

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

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

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

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

Alternative 1: Qualität ist ein „objektives“ Merkmal eines Sachverhalts¹¹, dessen Ausprägung durch abbildende subjektive Wahrnehmung ohne systematische Verzerrung „gemessen“ werden kann, so dass bei hinreichend großer Zahl von Messungen der Erwartungswert dem „wahren objektiven Wert“ entspricht. Oder:

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

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

Fazit: Trennung von Forschung und Evaluation

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

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

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

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

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

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

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

Literaturverzeichnis

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

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