

Institut für Politikwissenschaft
Universität Bern

Interface Institut für Politikstudien
Luzern

A. Balthasar

**Ein Leitfaden für die politikwissenschaftliche
Evaluationsforschung**

Luzern 1995

Inhaltsverzeichnis

1.	Wirkungsmodell	4
2.	Problembenennung	6
	2.1 Eingrenzung der Fragestellung	7
	2.2 Machbarkeitsprüfung	8
	2.3 Bestimmung des Evaluationskontextes	9
	2.4 Festlegung der Evaluationsdimension	10
3.	Untersuchungsdesign	11
	3.1 Ueberblick	12
	3.2 Wichtige Designs	14
	3.2.1 Fallstudie	14
	3.2.2 Breitenuntersuchung	17
	3.2.3 Experimentelle und quasi-experimentelle Designs	18
	3.2.4 Zeitreihenuntersuchung	21
	3.2.5 Konzeptionelle Triangulation	23
4.	Forschungstechniken	25
5.	Umsetzung	25
6.	Literatur	28

Eine erste Fassung dieses Textes wurde im Herbst 1993 für den Unterricht am Institut für Politikwissenschaft der Universität Bern verfasst. Die zahlreichen wertvollen Hinweise von Werner Bussmann und von Frédéric Varone sowie von Studierenden haben mich zu einer Ueberarbeitung veranlasst. Herzlichen Dank allen, die zu seiner Entstehung beigetragen haben. Der Text wird in gekürzter und veränderter Form auch im 'Handbuch Methoden der Evaluationsforschung' erscheinen.

Am Anfang jeder Evaluation stehen Fragen. Politikerinnen fragen zum Beispiel nach den Wirkungen von Subventionen in der Energiepolitik. Politiker interessieren sich für die Erfolge der Gleichstellungspolitik. Beamte zweifeln an den Wirkungen von Privatisierungsmaßnahmen. Oft genügt es, wenn diese Fragen auf der Basis der vorhandenen Kenntnisse beantwortet werden. Oft sind jedoch ausgedehnte, wissenschaftlich fundierte Evaluationen notwendig. Wie werden solche Untersuchungen durchgeführt? Grundlage bildet der normale sozialwissenschaftliche Forschungsprozess, wie er unter anderem von Alemann und Formdran (1990) dargestellt wird. Wissenschaftliche Forschung dient jedoch in erster Linie der Theoriegewinnung. Politikwissenschaftliche Evaluationen beabsichtigen dagegen die Beurteilung von Wirkungen staatlichen Handelns. Zudem muss eine anwendungsorientierte Evaluationsforschung nicht allein wissenschaftlichen Kriterien genügen, sondern auch zu Resultaten führen, die für die Praxis nützlich sind. Dadurch ergeben sich Eigenheiten, welche eine spezifische Einführung in diesen Forschungszweig nahelegen.

Vereinfacht lässt sich der Weg der Evaluation von der Frage zur Antwort in die fünf Etappen "Wirkungsmodell", "Problembenennung", "Untersuchungsdesign", "Forschungstechnik" und "Umsetzung" unterteilen. Allerdings ist das Bild einer schrittweisen Abfolge einzelner Phasen trügerisch. Die Abgrenzung erfolgt weniger anhand einer chronologischen Vorgehensweise, denn aufgrund funktionell unterschiedlicher Arbeitsschritte. In der Regel muss man sich Evaluationsforschung als Prozess vorstellen, bei welchem nebeneinanderliegende Phasen überlappend bearbeitet werden.

Zwei Einschränkungen müssen wir an den Beginn stellen. Zum einen wird der Evaluationsablauf sehr vereinfacht dargestellt und bewusst auf zahlreiche in der praktischen Arbeit wichtige Differenzierungen verzichtet. Auch werden die Schilderung und die Auseinandersetzung mit Konzepten der Literatur beiseite gelassen. Zum anderen beschränken sich die Darlegungen auf die Durchführung von *ex-post* Evaluationen. In der Literatur werden daneben begleitende und *ex-ante* Evaluationen erwähnt (vgl. Arbeitsgruppe Gesetzesevaluation 1991).

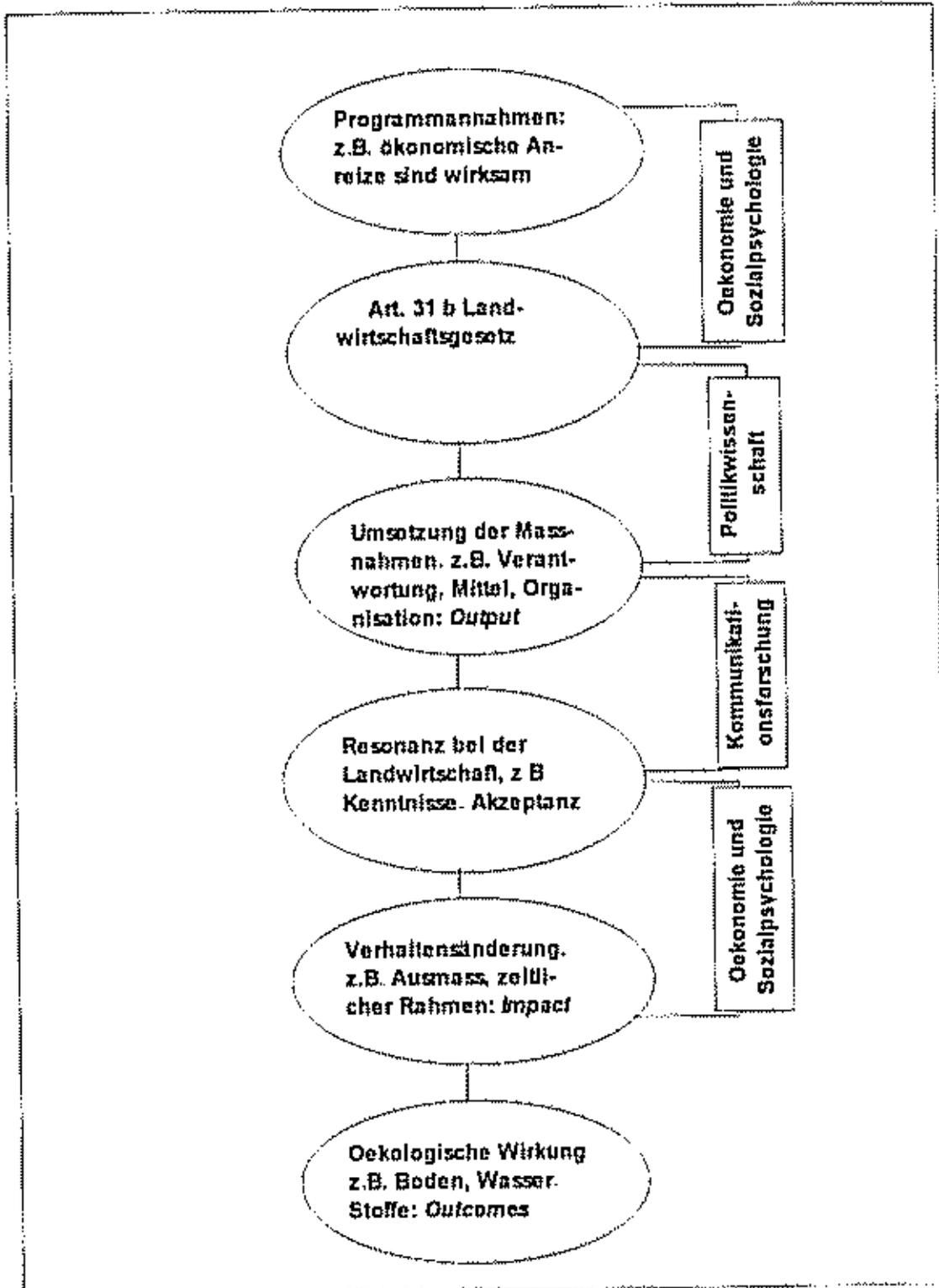
1 Wirkungsmodell

Von den politischen Massnahmen, wie sie in den Gesetzen formuliert sind, ist es im allgemeinen ein langer Weg zur Erreichung der beabsichtigten Wirkungen. Die Umsetzung eines Gesetzes muss organisiert werden: Verantwortungen sind zu klären, Mittel zu sprechen und Abläufe festzulegen. Die Ergebnisse solcher ersten Schritte werden gelegentlich als *Output* bezeichnet. Auch wenn solche Leistungen nachweisbar sind, können die Wirkungen noch auf sich warten lassen. Zuerst muss das Zielpublikum die Massnahme kennen und akzeptieren. Erst wenn es auch im erwarteten Ausmass darauf reagiert, sind zielrelevante Wirkungen oder sogenannte *Impacts* zu erwarten. Politische Massnahmen haben jedoch meist nicht allein Verhaltensänderungen zum Ziel. Die Installation von Solaranlagen ist beispielsweise kein Selbstzweck, sondern dient dazu, die Abhängigkeit der Schweiz von Energieimporten und von der Nuklearenergie zu reduzieren. Ob die Wirkungen also auch die gewünschten gesellschaftlichen Veränderungen, d.h. die gewünschten *Outcomes* hervorgerufen, hängt von den Programmannahmen ab.

Jede politische Massnahme beruht auf - häufig nur impliziten - Vorstellungen der Verantwortlichen über Wirkungszusammenhänge. Mit Hilfe der Analyse politischer Verlautbarungen, amtlicher Dokumente und ähnlicher Unterlagen lassen sich die Wirkungsüberlegungen der Verantwortlichen rekonstruieren. Die Evaluationsteams können diese Vorstellungen als Grundlage für ihre Arbeit übernehmen, sie können sie erweitern oder ersetzen. Daraus entsteht das untersuchungsleitende Wirkungsmodell. Dieses stellt dar, wie der Untersuchungsgegenstand in den Augen der Evaluatorinnen mit relevanten Einflussfaktoren zusammenhängt (Darstellung 1: Beispiel eines Wirkungsmodells)

Die Auseinandersetzung mit den Wirkungsüberlegungen der Verantwortlichen erfolgt meist vor dem Hintergrund von Forschungshypothesen. In unserem Beispiel könnte beispielsweise die Feststellung "Die Massnahme ist bei einem kleinen Teil der Zielgruppe bekannt" eine untersuchungsleitende Hypothese sein. Hypothesen dienen als Leitlinien zum Abfragen des Gegenstandsbereichs. Allerdings ist es so, dass unterschiedliche wissenschaftstheoretischen Konzepte zu unterschiedlichen Definitionen von Hypothesen geführt haben. Vor dem Hintergrund seines empirisch-analytischen Zugangs listet Atlas-Tander (1984) acht Kriterien auf, die seiner Ansicht nach bei der Formulierung von Hypothesen berücksichtigt werden müssen. Dazu

Darstellung 1: Beispiel eines Wirkungsmodells: Die Wirkungen der ökologischen Ausgleichszahlungen nach Artikel 31 b des Landwirtschaftsgesetzes



gehören die Widerspruchsfreiheit, die Falsifizierbarkeit und die Operationalisierbarkeit. Andere Autoren sind in dieser Hinsicht weniger restriktiv (so von Alemann und Formdran 1990)

Forscherinnen und Forscher sind sich jedoch nicht nur uneins über die Definition von Hypothesen. Auseinandersetzungen gibt es auch darüber, wie man zu geeigneten Hypothesen gelangen kann. Für Helmut Seiffert ist jeder Erklärungsversuch, worauf immer er auch gründen mag, eine Hypothese (Seiffert 1969). Andere Autoren leiten Hypothesen deduktiv von theoretischen Vorstellungen über den Untersuchungsgegenstand ab (Alteslander 1984). Generell ist es sicher zweckmässig bei der Erarbeitung eines Wirkungsmodells für eine Evaluation abzuklären, ob aus der bisherigen Forschung Theorien vorliegen, welche für die Beantwortung der spezifischen Fragestellung nutzbringend verwendet werden können. Sind solche theoretische Bezüge möglich, wird die Hypothesenbildung angeregt und die Einordnung des Problems in einen grösseren Zusammenhang ermöglicht.

2. Problembenennung

Kehren wir zu den eingangs gestellten Fragen zurück: Wie ist die Wirksamkeit von Subventionen in der Energiepolitik zu beurteilen? Welche Leistungen hat die Gleichstellungspolitik erbracht? Worin liegen Vor- und Nachteile von Privatisierungsmassnahmen? Auf den ersten Blick erscheinen diese Fragen klar und eindeutig formuliert. Bei näherem Hinsehen ergeben sich jedoch zahlreiche Unklarheiten. Was verstehen die Politikerinnen unter 'Wirkung von Subventionen'? Sind damit die Effekte ganz allgemein gemeint? Oder sind wohl die Leistungen im Bezug zu den Zielen angesprochen? Und nach welchen Gesichtspunkten sollen die Erfolge der Gleichstellungspolitik beurteilt werden? Sollen vor allem die Direktbeteiligten zu Wort kommen? Oder die betroffenen Frauen und Männer? Und wie ist die Wirkung der Privatisierung festzumachen? Sind damit alle Wirkungen und Nebenwirkungen gemeint oder geht es in erster Linie um die wirtschaftliche Effizienz?

Die Fragen, die am Anfang von Evaluationen stehen, können meist nicht ohne Präzisierungen angegangen werden. Wissenschaftlich fundierte Evaluationen können nur auf gezielt formulierte Fragen Antworten geben. Die Festlegung der Forschungsfrage ist der Angelpunkt jeder Studie. Im ersten Schritt einer Evaluation gilt es, ein

politisches oder administratives Problem in eine wissenschaftlich bearbeitbare Forschungsfrage zu gliedern. Diesen Prozess fassen wir unter dem Titel "Problembenennung" zusammen und zählen dazu die Eingrenzung der Fragestellung, die Machbarkeitsprüfung, die Bestimmung des Evaluationskontextes und die Festlegung der Evaluationsdimension. Diese Aspekte hängen eng zusammen und müssen weitgehend gleichzeitig geklärt werden.

2.1 Eingrenzung der Fragestellung

Je mehr wir uns mit einer Evaluationsfragestellung auseinandersetzen, desto vielfältiger und undurchsichtiger erscheint das Wirkungsgeschehen. Welche Subventionen in der Energiepolitik wollen wir untersuchen? Interessiert uns die Anzahl der Auszahlungen? Die Zahl derjenigen, die ohne Subvention nicht gehandelt hätten? Oder ist es die eingesparte Energie, die wir ermitteln wollen? Schliesslich endet jede Problemanalyse in einem Stadium, in welchem alles als faktisch interdependent und relevant wahrgenommen wird. Alle Einflussfaktoren und Zusammenhänge lassen sich jedoch nicht gleichzeitig untersuchen. Eine Auswahl ist notwendig. Dazu ist es zweckmässig, zwischen *deskriptiv*, *normativ* und *kausal-orientierten* Fragestellungen zu unterscheiden.

Bei deskriptiven Fragen steht die Beschreibung von Tätigkeiten, Objekten oder Zusammenhängen im Zentrum (z.B. Wie hat sich die Förderung alternativer Energien in den letzten zehn Jahren verändert?). Unserer Ansicht nach handelt es sich allerdings bei dieser Art von Fragen nicht um eigentliche Evaluationen, weil deskriptive Fragestellungen nicht auf Wertungen abzielen. Evaluationen beinhalten dagegen dem Wortsinn gemäss Beurteilungen. Insbesondere bei umstrittenen Themen werden von auftraggebenden Instanzen jedoch oftmals keine explizit wertenden Aussagen zugelassen, so dass mit deskriptiven "Evaluationen" vorlieb genommen werden muss. Normative Fragestellungen vergleichen Ist- mit Soll-Zuständen (z.B. Werden die Ziele der Energiepolitik erreicht?). Bei kausal-orientierten Fragen werden Wirkungsketten verfolgt (z.B. Welchen Beitrag leisten Subventionen zur Verbreitung von Solaranlagen?).

Die Eingrenzung der Fragestellung wird vielfach durch die Tatsache erschwert, dass der Inangriffnahme eines Evaluationsprojektes oft ein langer, verwaltungsinterner Verhandlungsprozess vorangeht. In der Addition von relevanten Fragen liegt dann vielfach die einzige

Chance. Widerstände gegenüber Evaluationen zu umschiffen. Es entsteht eine Summierung von Fragen, die sich kaum mehr in einem Projekt bearbeiten lassen. Weitere Probleme können sich dadurch ergeben, dass die in der Fragestellung benutzten Begriffe derart schwammig oder mit einer Fülle unterschiedlicher Bedeutungsinhalte versehen sind, dass sich kaum darauf aufbauen lässt. Ausschlaggebend für das Gelingen einer Evaluation ist es, dass die Ausgangsfragen klar formuliert sind und dass realistische Ziele gesteckt werden. Da Evaluationen häufig als Auftragsforschung konzipiert sind, ist in dieser Arbeitsphase die Kooperation und die Verhandlung mit dem Auftraggeber oder der Auftraggeberin besonders wichtig. Nur dadurch können falsche Erwartungen und damit Enttäuschungen vermieden werden.

2.2 Machbarkeitsprüfung

Durch den Praxisbezug und die beurteilende Funktion von Evaluationen stehen WissenschaftlerInnen, welche damit beschäftigt sind, in konfliktären politischen Spannungsfeldern. Das Forscherteam ist plötzlich mit Fragen konfrontiert wie 'Welches sind eigentlich die Motive für die Evaluation?', 'Welche Rolle soll das Evaluatorenteam übernehmen?' oder 'Ist eine unabhängige Untersuchung überhaupt erwünscht oder sollen vorgefasste Meinungen bestätigt werden?' Zur Klärung dieser und weiterer Fragen ist eine Machbarkeitsprüfung notwendig. Es handelt sich dabei um Überlegungen zur Planung und zur Durchführung der Untersuchung.

Ausgehend von einer Beschreibung der interessierenden Maßnahmen sowie von deren Umfeld wird in ersten Gesprächen und in einer ersten Sichtung von Dokumenten der Rahmen der geplanten Arbeit abgeklärt. Dabei müssen sowohl inhaltliche wie auch operationelle Aspekte Beachtung finden. *Inhaltlich* geht es unter anderem um folgende Fragen: Ist die zu evaluierende Maßnahme genügend abgrenzbar? Gibt es Informationen zum Untersuchungsgegenstand oder lassen sich diese in der zur Verfügung stehenden Zeit erheben? Welches sind die Ziele, welche ein allfälliger Auftraggeber mit der Untersuchung verbindet? Geht es um die Lösung einer neuen Problemlage? Geht es um die Beschaffung von Argumenten für eine bekannte Position oder soll bisheriges Verhalten legitimiert werden? Welches ist die Rolle des Evaluators? Sind sie Ratgeber, Informationsbeschaffer oder Schlichter?

Die *operationell* notwendigen Abklärungen betreffen dagegen Themen wie: Sind Widerstände gegenüber der Evaluation zu erwarten, welche eine systematische Bearbeitung des Themas verunmöglichen? Sind die Beteiligten kooperationsbereit? Entsprechen die zur Verfügung stehenden finanziellen Mittel und der festgelegte Zeitrahmen einer sachgerechten Bearbeitung der Fragestellung?

Wesentlich ist in jedem Fall, dass im Vorfeld einer Evaluation abgeklärt wird, ob ein für eine Untersuchung geeignetes Klima besteht (vgl. Rossi et al 1993). Dazu gehört auch die grundsätzliche Bereitschaft der Verantwortlichen, sich auf Lernprozesse einzulassen

2.3 Bestimmung des Evaluationskontextes

Öffentliche Stellen, welche Evaluationen veranlassen, möchten damit in der Regel ihre Kenntnisse über die Wirkungen von politischen Massnahmen verbessern. Dahinter stehen der Wunsch nach einer Kontrolle und - im allgemeinen jedenfalls - nach der Aufdeckung von Verbesserungsmöglichkeiten. Der implizite Widerspruch zwischen 'Kontrollieren' und 'Lernen' bildet den Hintergrund jeder Evaluationsstudie. Hinter der Evaluation von energiepolitischen Massnahmen steht beispielsweise der im Energienutzungsbeschluss von 1990 (Art 24) explizit formulierte Auftrag der Rechenschaftsablage gegenüber dem Parlament. Die Überprüfung der Tätigkeit der Verwaltung durch die Parlamentarische Verwaltungskontrolstelle hat in erster Linie eine kontrollierende Funktion. Evaluationen, in welchen die Motivation der Überprüfung im Vordergrund steht, haben oft gegen die Widerstände von Betroffenen zu kämpfen. Viele Studien sind daran schon gescheitert, denn ohne die Auskunftsbereitschaft der beteiligten Beamten lassen sich zahlreiche Evaluationsfragen nicht beantworten. Die Kontrollstellen haben sich daher die Auskunftspflicht des Personals gesichert. Doch auch dies war der Kooperationsbereitschaft nicht in jedem Fall förderlich.

Obwohl die Kontrolle von Verwaltungshandeln ein aus staatspolitischer Sicht absolut legitimes Ziel von Evaluationen ist, profitiert die Qualität der Ergebnisse davon, wenn die Untersuchung grosses Gewicht auf Lerneffekte bei allen Beteiligten legt. Zu diesem Zweck hat die Evaluationsforschung verschiedene spezielle Verfahren entwickelt. Guba & Lincoln (1987) propagieren beispielsweise die *betroffenen-orientierte Evaluation*. Dabei werden von der Massnahme Betroffene in den Evaluationsprozess einbezogen und auf

diese Weise Akzeptanz und Umsetzbarkeit der Ergebnisse verbessert. Ein gelungenes Beispiel für diese Art von Evaluation bildet die Untersuchung der Wirtschaftsförderung des Kantons Wallis, welche von Michel Ray und Luc Vodoz (1993) durchgeführt wurde. Allerdings ergeben sich auch Probleme bei derartigen Zugängen. So sind die verschiedensten, oft divergierenden Ansprüchen und Interessen zu berücksichtigen. Die Evaluatoren agieren nicht mehr als aussenstehende Beobachter sondern als Mediatoren, Methodologen und Experten. Sie sind ganz in den politischen Aushandlungsprozess integriert. Dadurch besteht die Gefahr, dass sie ihre eigenen Ziele als Massstab zur Beurteilung des Programms heranziehen.

Eine ähnliche Intention wie die betroffenen-orientierte Evaluation verfolgt die *zielfreie Evaluation*. Sie wurde beispielsweise von Scriven propagiert (Scriven 1980). Während die zielorientierte Evaluation mit Hilfe von Indikatoren den Zielerreichungsgrad misst und daher auf eindeutige und stabile Ziele angewiesen ist, beurteilt die zielfreie Evaluation eine Massnahme anhand des gesamten Sets von ausgelösten Wirkungen. Scriven's Motivation für zielfreie Evaluationen stammt daher, dass die Vorgabe von Zielen notwendigerweise zu einer unvollständigen Erfassung der Programmkonsequenzen führen muss. Ausserdem werden nur die Interessen einer kleinen Gruppe von Beteiligten und Betroffenen berücksichtigt. Die Identifizierung der Ursprünge solcher Verzerrungen in der Beziehung zwischen Evaluatoren, Auftraggebern und Zielen ist für Scriven genauso von Bedeutung wie die Entwicklung von praktikablen Methoden zur Kontrolle der identifizierten Fehlerquellen.

2.4 Festlegung der Evaluationsdimension

Im Zuge der Problembenennung gilt es schliesslich, die inhaltliche Ausrichtung der Evaluation festzulegen. Grundsätzlich lassen sich Konsistenz-, Vollzugs- und Wirkungsevaluationen unterscheiden. *Konsistenzevaluationen* analysieren Massnahmen daraufhin, ob sich die intendierten Wirkungen damit überhaupt erzielen lassen. Zum Beispiel wird untersucht, ob die wichtigsten Zielgruppen überhaupt angesprochen werden. Eine andere Frage für eine Konsistenzevaluation wäre, ob die Wirkungen in der vorgesehenen Zeit überhaupt eintreten können. *Vollzugsevaluationen* oder *Prozessevaluationen* untersuchen, ob die Vollzugsträger tatsächlich das tun, was man von ihnen erwartet und ob sie die zur Verfügung stehenden Mittel

(Aktivitäten, Personal, Organisationsstruktur, Kompetenzen, Budget) ordnungsgemäss einsetzen

Bei Wirkungsevaluationen kann zwischen Wirkungs-, Wirksamkeits- und Wirtschaftlichkeitsstudien unterschieden werden. Die Wirkungsstudien beschäftigen sich mit der Untersuchung sämtlicher Programmwirkungen unabhängig davon, wie sie sich zu den gesetzten Zielen verhalten. Besondere Beachtung wird oft Mitnahmeeffekten, Verdrängungseffekten und Nebenwirkungen geschenkt. Wirksamkeits- oder Effektivitätsuntersuchungen vergleichen die Programmwirkungen mit den Programmzielen und beurteilen, ob und in welchem Masse die Ziele erreicht worden sind. Wirtschaftlichkeits- oder Effizienzanalysen beurteilen schliesslich das Kosten-Nutzen-Verhältnis von Massnahmen (Wollawa & Thierau 1990, S. 94)

Gehen wir beispielhaft davon aus, dass die Subventionen in der Energiepolitik der Verbreitung erneuerbarer Energien dienen. Die Untersuchung der Auswirkung, welche Solaranlagen auf das Ortsbild im Tessin haben, wäre ein Beispiel für eine Wirkungsstudie. Eine Wirksamkeitsuntersuchung würde die effektive Zielerreichung untersuchen; sie würde also der Frage nachgehen, wieviele neue Solaranlagen aufgrund der Massnahme erstellt worden sind. Eine Wirtschaftlichkeitsuntersuchung schliesslich würde abklären, ob mit Subventionen pro Quadratmeter installierter Kollektorfläche bei geringem administrativem Aufwand oder mit leistungsabhängigen Beiträgen bei hohem administrativem Aufwand das bessere Kosten-Nutzen-Verhältnis erzielt werden könnte.

3. Untersuchungsdesign

Das Untersuchungsdesign ist das konzeptionelle Herz einer Evaluation. Darüber sind sich die meisten Evaluatoreninnen und Evaluatoren einig. Weniger gross ist die Übereinstimmung hingegen betreffend der Frage, was in einem Evaluationsprojekt unter einem Untersuchungsdesign zu verstehen ist. Die Literatur ist darum reich an Vorschlägen (vgl. z. B. Roth 1993, S. 711 ff.; Fitz-Gibbon & Morris 1987).

Wir schlagen vor, den Begriff "Untersuchungsdesign" eng zu fassen und darunter nur die Bestimmung der Vergleichsebene, d.h. das Analysekonzept und des Forschungsansatzes, d.h. der Approaches, zu subsumieren. Entscheidungen über die Wahl von Forschungstechniken ordnen wir dagegen den nächsten Arbeitsschritt zu. Das

Untersuchungsdesign gibt also noch keine Auskunft darüber mit welchen Techniken die Daten erhoben und ausgewertet werden (vgl. Lamnek 1989, S. 5). Wie bei den anderen Arbeitsschritten gilt auch hier dass die Abgrenzung zwischen Design und Techniken funktionell zu verstehen ist. In diesem Abschnitt geben wir zuerst einen Ueberblick darüber, was wir unter einem Untersuchungsdesign verstehen, danach stellen wir einige wichtige Designs vor.

3.1 Ueberblick

Wenden wir uns zuerst der Vergleichsebene zu. Was ist darunter zu verstehen? Die Evaluationsforschung setzt sich mit der systematischen Beurteilung der Konsistenz, des Vollzugs oder der Wirkungen staatlicher Massnahmen auseinander. Eine wissenschaftlich fundierte Beurteilung ist auf eine Vergleichsgrösse angewiesen. Im Analysekonzept wird die Art des Vergleiches festgelegt. Es lassen sich drei Vergleichsebenen unterscheiden: Die erste Vergleichsmöglichkeit bildet der *Quervergleich* von zwei oder mehreren Untersuchungseinheiten. Eine davon ist einer politischen Massnahme unterstellt, bei der anderen ist dies nicht der Fall. Ein solcher synchron angelegter Vergleich findet sich beispielsweise in den regelmässig veröffentlichten Arbeitslosenstatistiken der OECD. In vertiefenden Betrachtungen werden - aufbauend auf diesen Vergleichszahlen - Beurteilungen vorgenommen und Vorschläge zuhanden der Politik erarbeitet. Die zweite Vergleichsmöglichkeit bildet der *Längsschnitt*. Dort wird die Situation vor der Einführung einer Massnahme mit derjenigen danach verglichen. Solche Untersuchungsansätze finden sich zum Beispiel, wenn die Unfallkosten vor der Einführung des Gurtenobligatoriums mit denjenigen danach verglichen werden. Als dritten Vergleichstyp betrachten wir - anders als andere Autoren - den *Soll-Ist-Vergleich*, obwohl er auf einer anderen Ebene angelegt ist. Dort bilden die Zielvorstellungen des Gesetzgebers, der Beteiligten oder der Betroffenen den Massstab für die Bewertung der Massnahme. Als Beispiel dafür können die jährlichen Paritätslohnberechnungen des schweizerischen Bauernverbandes dienen. Dabei werden die Leistungen der Landwirtschaftspolitik auf der Basis eines Vergleichs der Einkommen einer durchschnittlichen Bauernfamilie und einer durchschnittlichen Arbeiterfamilie verglichen.

Jede dieser drei Vergleichsebenen hat Stärken und Schwächen. Der Soll-Ist-Vergleich lässt sich beispielsweise als sehr flexibel und breit anwendbar charakterisieren. Eine seiner Schwächen liegt je-

doch darin, dass die Ziele in der politischen Praxis meist nur sehr vage formuliert sind. Zudem weisen fast alle Massnahmen ganze Zielkataloge auf. Folglich müssen operationalisierbare Sollgrössen oftmals im Nachhinein ermittelt werden. Bei Quervergleichen stellt sich das Problem, dass die Auswahl von Vergleichseinheiten die Resultate vorwegnehmen kann. Ob wir die Arbeitslosenzahlen der Schweiz mit denjenigen der Nachbarländer vergleichen oder ob wir den ganzen OECD-Raum als Bezugsgrösse nehmen, führt zu wesentlich anderen Ergebnissen. Längsschnittvergleiche haben den Vorteil, dass sie nicht auf andere Untersuchungseinheiten Bezug nehmen müssen. Ihre Schwäche besteht jedoch darin, dass sich die Rahmenbedingungen im Zeitablauf ändern können. So ist es oftmals schwierig, die Wirkungen von Massnahmen herauszukristallisieren. Ein weiteres Problem besteht darin, dass politische Massnahmen selten als einfache Interventionen erfolgen. Viel häufiger sind es lang andauernde, vielschichtige Prozesse beruhend auf kleinen Schritten, differenzierten Fristsetzungen und häufigen Programmkorrekturen. Oft ist es nicht möglich, aufzuschlüsseln, welcher Schritt zu welcher Wirkung geführt hat.

Als Möglichkeit, um die Unzulänglichkeiten der einzelnen Vergleichsebenen etwas zu vermindern, bietet sich die *konzeptionelle Triangulation* an. Dabei werden die Vergleichsebenen kombiniert. Die Evaluation einer Massnahme zur Förderung erneuerbarer Energien würde beispielsweise nicht allein auf einem Zeitreihenvergleich aufbauen. Zusätzlich könnte ein Quervergleich der Politiken verschiedener Kantone angelegt und beide Teiluntersuchungen verknüpft werden. Triangulation führt allerdings zu recht komplexen Untersuchungsanlagen. Wir kommen darauf in Abschnitt 3.2 zurück.

Neben der Vergleichsebene ist der Forschungsansatz das zweite Element eines Untersuchungsdesigns. Die empirische Sozialforschung unterscheidet in diesem Zusammenhang zwei Möglichkeiten (Reinhold 1991, S. 391 f.). Der *Einzelfall-approach* kennzeichnet sich dadurch, dass ein oder mehrere Einzelfälle in ihrer Komplexität herausgearbeitet werden. Der *Survey-approach* versucht dagegen, Hypothesen mit Hilfe von Daten aus einer Vielzahl von Untersuchungseinheiten systematisch zu falsifizieren. Dabei liegt ein besonderes Augenmerk auf der Vergleichbarkeit der Erhebungssituation für alle "Fälle".

Abschliessend ist es wichtig darauf hinzuweisen, dass die Auswahl eines geeigneten Evaluationsdesigns nicht nur durch die For-

schungslogik und durch theoretische Überlegungen bestimmt wird. Ebenso wichtig sind die Rahmenbedingungen, welche durch die zeitlichen, finanziellen und qualifikatorischen Ressourcen sowie durch den geplanten Verwendungszweck der Untersuchung vorgegeben werden. Die Wahl des geeigneten Untersuchungsdesigns ist eine anspruchsvolle Aufgabe. Die Erfahrung lehrt nämlich, dass vielfach ohne oder ohne klare Vergleichsebene gearbeitet wird. Dadurch wird die Aussagekraft der Evaluation erheblich beeinträchtigt.

3.2 Wichtige Designs

Durch die unterschiedliche Kombination der Elemente eines Evaluationsdesigns ergibt sich eine Vielzahl von möglichen Untersuchungsanlagen. Im Folgenden gehen wir auf einzelne davon vertieft ein.

3.2.1 Fallstudie

Eine Fallstudie ist ein Forschungsansatz, bei welchem "besonders interessante Fälle hinsichtlich möglichst vieler Dimensionen beobachtet (...) beschrieben und analysiert" werden. Zentrale Bedeutung bei der Fallstudie hat aber nicht die Tatsache möglichst vieler 'Dimensionen' oder Variablen, sondern entscheidend ist vielmehr, dass die Untersuchungsobjekte nicht auf einige wenige Variablen reduziert werden. Dies würde dem Untersuchungsobjekt nicht gerecht werden können, weil das Herausgreifen einzelner Merkmale einigermassen willkürlich erscheint. Die Komplexität des Gegenstandes verkürzt und die Individualität und Identität der zu Untersuchenden verletzen könnte. Tatsächlich geht es () besonders darum, ein ganzheitliches und realistisches Bild der sozialen Welt zu zeichnen. Mithin sind möglichst alle für das Untersuchungsobjekt relevanten Dimensionen in die Analyse einzubeziehen' (Lamnek 1989, S. 5)

Die Zielsetzung einer Fallstudie ist wesentlich abhängig davon, ob ein eher qualitativer oder ein quantitativer Zugang zur Fragestellung vorhanden ist. Aus der Perspektive der *quantitativen Forschungstradition* lässt sich unterscheiden zwischen Fallstudien, die eher der Hypothesenbildung dienen und solchen, welche in erster Linie die Hypothesenprüfung vor Augen haben (vgl. u.a. Eckstein 1975). *Hypothesenbildende Fallstudien* haben die Aufgabe, kom-

plexe Zusammenhänge aufzuhellen und so zur Entwicklung empirisch begründbarer theoretischer Konzepte, Theorien und Hypothesen beizutragen (Lamnek 1989, S. 11). Sie dienen somit dem Kennenlernen und der Interpretation von Problemzusammenhängen und der Formulierung von empirisch relevanten Hypothesen. Diese lassen sich in der nachfolgenden Hauptuntersuchung testen. Bei hypothesenbildenden Fallstudien werden meist typische Fälle untersucht, d.h. solche, welche den "Normalfall" repräsentieren. Hypothesenprüfende Fallstudien dagegen analysieren eher kritische Fälle. Damit sollen Hypothesen bestätigt, relativiert oder verworfen werden. Ausgangspunkt bilden Annahmen, welche das Eintreten eines bestimmten Ereignisses gewissermassen zwingend voraussagen. Zeigt eine Fallstudie nun ein Ausbleiben des unter den angegebenen Ausgangsbedingungen prognostizierten Ereignisses, so ist eine Modifikation oder eine Verwerfung der verwendeten Hypothese notwendig. Andernfalls gilt die Hypothese als bestätigt.

In der *quantitativen Forschungstradition* kommen Fallstudien im allgemeinen ergänzende Funktionen zu. Den Hauptteil der Studie bilden dort meist Breitenuntersuchungen. Beispiele für Fallstudien in diesem Kontext finden sich in der Evaluation der Hausfeuerungskontrolle, einer Massnahme aus dem Umweltschutzbereich. In einem Projekt haben A. Balihar und C. Knöpfel (1994, S. 84 ff.) mit drei Fallstudien, die dem Vollzug der Massnahme in je einem Kanton gewidmet waren, diejenigen Faktoren herausgearbeitet, welche die Wirksamkeit der Hausfeuerungskontrolle beeinflussen können. In einem zweiten Schritt wurde der Einfluss der einzelnen Faktoren in einem Quervergleich zwischen 500 Gemeinden regressionsanalytisch ermittelt.

Im Rahmen der *qualitativen Forschungstradition* wird den Fallstudien eine eigenständige Bedeutung zugemessen. "Die Einzelfallstudie im qualitativen Paradigma strebt eine wissenschaftliche Rekonstruktion von Handlungsmustern auf der Grundlage von alltagsweltlichen, realen Handlungsfiguren an. Dabei versucht der Forscher, nicht nur als alltagsweltlicher Handlungspartner die Figuren nachzuvollziehen, sondern diese in den wissenschaftlichen Diskurs zu überführen und Handlungsmuster zu identifizieren, indem er allgemeinere Regelmässigkeiten vermutet" (Lamnek 1989, S. 16). Lamnek beschreibt die Erarbeitung einer Fallstudie als zweistufigen Prozess: In der ersten Phase ist der Forscher bemüht, die erfassten Gegebenheiten zu verstehen. In der zweiten Phase versucht er auf der Basis des einen Falles oder im Vergleich mit anderen Fällen, typische Mu-

ster herauszukristallisieren. Die Offenheit des Fallstudienzugangs, sei es bezüglich des theoretischen Konzepts gegenüber den involvierten Personen, in der Wahl der Erhebungstechniken oder in der Gestaltung der Erhebungssituation, ermöglicht es, Zusammenhängen auf die Spur zu kommen, welche weder vom Programm beabsichtigt, noch vom Forscher vorausgesehen waren. Einzelfallstudien sind in der Evaluationsforschung sehr häufig. Als Beispiele seien die Evaluation der Tätigkeit des Bundesamt für Umwelt, Wald und Landschaft (Knoepfel 1991) oder die Evaluation der Wirtschaftsförderung des Kanton Wallis (Rey et al. 1993) genannt.

Neben Einzelfallstudien werden häufig auch vergleichende Fallstudien angefertigt. So wurden beispielsweise im Rahmen des NFP 27 die Wirtschaftsförderungen der Kantone Solothurn und St. Gallen miteinander verglichen (Blerter & Binder 1994). Vergleichende Fallstudien haben den Vorteil, dass der Einfluss unterschiedlicher Situationen auf das Programm bestimmt und unterschiedliche Wirkungsfaktoren identifiziert werden können. Häufig gehen jedoch die Vorteile der vergleichenden Analyse angesichts unvollständiger oder inkompatibler Daten bei der Auswertung wieder verloren.

Ob sich mit einer Fallstudie die angestrebten Ziele auch erreichen lassen, ist in erster Linie von der Auswahl der Fälle abhängig. In der Anleitung des General Accounting Office zur Durchführung von Fallstudien (GAO 1990) nimmt daher dieser Arbeitsschritt breiten Raum ein. Als Oberkriterien werden Bequemlichkeit, Zufallsauswahl und erwarteter Nutzen unterschieden. Das Bequemlichkeitskriterium kommt dann zum Zug, wenn jener Fall untersucht werden soll, über den am leichtesten Informationen erhältlich sind. Diesem pragmatischen Vorgehen steht die reine Zufallsauswahl gegenüber. Eine solche Auswahl bietet die beste Voraussetzung für Repräsentativität und Verallgemeinbarkeit der Ergebnisse. Im allgemeinen schlägt das GAO jedoch vor, die Auswahl streng an den Zielen der Evaluation zu orientieren.

Das Office unterscheidet dabei sieben verschiedene Selektionskriterien. Erstens ist es möglich, Extreme auszuwählen, um dadurch Hinweise auf die Ursachen für grosse Varianzen zu bekommen. Als zweites sind Musterbeispiele denkbar. Sie legen dar, was zum Erfolg in einzelnen Fällen beigetragen hat. Das Gegenbeispiel stellen die kritischen Fälle dar. Sie weisen auf spezifische Probleme hin. Weiter ist es möglich, typische Fälle auszuwählen. Sie geben Aufschluss darüber, wie die Wirkungskette im "Normalfall" zusammenhängt.

Durch die Beschreibung von Fällen, welche ein spezielles Cluster vertreten lassen sich verschiedene Programmtypen vergleichen. *Illustrative Fälle* dienen dazu, Probleme oder Chancen von Programmen besonders eindrücklich darzulegen. Als letzten Typ bezeichnet das GAO *Spezialfälle*. Dabei wird ein herausragendes Ereignis untersucht. Dabei kann es sich beispielsweise um einen Fall handeln, der besonders viel Aufsehen erregt hat.

Zusammenfassend ist hervorzuheben, dass Fallstudien den Vorteil bieten, dass die Untersuchung nicht auf isolierte Variablenzusammenhänge beschränkt werden muss. Vielmehr ist es möglich, eine Vielfalt von Einflussfaktoren im Auge zu behalten. Kein anderes Analysekonzept besitzt eine derartige Flexibilität und erlaubt es, zahlreichen unterschiedlichen Wirkungen gleichzeitig nachzugehen, weitere Verknüpfungen herzustellen und Erhebungsräume kurzfristig zu verändern (Hellstern/Wollmann 1983, S. 71). Fallstudien sind darum dann besonders angebracht, wenn die Wirkungszusammenhänge noch wenig bekannt sind und grosser Wert auf Flexibilität gelegt werden soll. Allerdings haben Fallstudien auch wichtige Schwächen. Mit Fallstudien lassen sich zwar gewisse Wirkungszusammenhänge auf induktivem Weg ausschliessen, dennoch bleibt aber ein recht breiter Raum von Erklärungsmöglichkeiten. Die Verlässlichkeit der aufgezeigten Zusammenhänge ist vielfach ungesichert. Sieht man einmal von der grundsätzlichen Opposition gegen eine letztlich hermeneutische Vorgehensweise ab, so hängt die Qualität der Resultate weitgehend von der sozialwissenschaftlichen Kompetenz der ForscherInnen und von weichen qualitativen Expertinnenurteilen ab. Wieweit die aus Untersuchungen dieser Art gewonnenen Aussagen auf andere Fälle übertragbar sind, ist zudem auch eine Frage der Einbindung der Untersuchung in einen übergreifenden Bezugsrahmen.

3.2.2 Breitenuntersuchungen

Das Gegenstück zum Fallstudienansatz bildet der *Survey-approach* (vgl. Reinhold 1991, S. 391). Sein Kernstück bildet die Erhebung und Auswertung von Daten für eine Vielzahl, eventuell sogar für alle Fälle. Ausgehend von Kenntnissen über die soziale Wirklichkeit werden dazu zuerst Hypothesen entwickelt. Diese werden dann operationalisiert und anhand von Daten überprüft, die aus Befragungen, Beobachtungen oder Sekundäranalysen stammen.

Breitenuntersuchungen arbeiten mit quantitativen Methoden der empirischen Sozialforschung. Dabei liegt ein besonderes Augenmerk auf der Vergleichbarkeit der Erhebungssituationen für alle Untersuchungseinheiten. Man versucht damit störende Einflussfaktoren auszuschliessen oder - wenn das nicht gelingt - ihren Einfluss auf alle Untersuchungsobjekte konstant zu halten. Durch die Interpretation der quantitativen Verteilung der Merkmalsausprägungen der Resultate versucht der Evaluator oder die Evaluatorin dann die Ausgangshypothesen zu falsifizieren. Gelingt die Falsifizierung nicht, so ist die Hypothese vorläufig als bestätigt anzusehen. Ein Beispiel für einen solchen Untersuchungsansatz bildet die Evaluation der staatlichen Hotellerieförderung, die durch Hanser et al. (1994) im Rahmen des NFP 27 bearbeitet wurde. Dabei wurden alle rund 1800 Unterstützungsgesuche an die Schweizerische Gesellschaft für Hotellerieförderung in den Jahren 1984 bis 1989 ausgewertet, um Häufigkeiten verschiedener Kategorien von Innovationsfinanzierung und Unterstützungsleistung zu erfassen und zu interpretieren.

3.2.3 Experimentelle und quasi-experimentelle Designs

Sozialexperimente mit Zufallsauswahl sind die strengste Methode zur Bestimmung der Wirkungen von politischen Massnahmen. Der Ansatz geht auf die in der Psychologie und der Biologie entwickelte Methode des kontrollierten Experiments zurück. Dort werden nach dem Zufallsprinzip Gruppen gebildet. Die Experimentiergruppe wird danach einer Massnahme ausgesetzt, die Kontrollgruppe nicht. Bei der Experimentiergruppe wird aufgrund des Massnahmen Einsatzes mit einer bestimmten Wirkung gerechnet. Die Wirkung wird mit Messwerten zu belegen versucht (Attelander 1984 S. 190 ff.). In ihrer Einführung in die Programmevaluation schildern Rossi et al. (1988) verschiedene Beispiele für solche Experimente. Um zu prüfen, ob die Gewährung einer begrenzten finanziellen Ueberbrückungshilfe ehemaliger Strafgefangenen bei ihrer sozialen Eingliederung hilft, wurden in den USA 400 entlassene Strafgefangene nach dem Zufallsprinzip drei Versuchsgruppen und einer Kontrollgruppe zugeteilt. Die erste Versuchsgruppe erhielt Arbeitslosengeld für die Dauer von 13 Wochen und besondere Hilfe zur beruflichen Eingliederung. Die zweite erhielt nur das Arbeitslosengeld und die dritte nur Hilfe zur beruflichen Eingliederung. Die Kontrollgruppe profitierte weder von Arbeitslosengeld noch von Hilfe bei der Arbeitssuche. Die 400 Personen wurden ein Jahr lang nach ihrer Entlassung in bestimmten Ab-

ständen interviewt. Auf diese Weise wurden Rückschlüsse auf die Wirksamkeit der einzelnen Massnahmen gezogen (Rossi et al 1980)

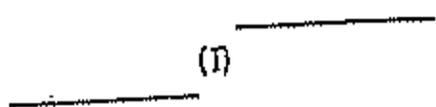
Es zeigten sich aber verschiedene Grenzen des experimentellen Ansatzes. Ethische oder rechtliche Gründe standen einer willkürlichen Zuordnung der Menschen zu Experimentiergruppen und Kontrollgruppen im Wege. Wenn ein Programm sinnvoll ist und es den Menschen wahrscheinlich hilft, bedeutet die Verweigerung bestimmter Dienstleistungen faktisch eine Benachteiligung oder gar eine Schädigung der Kontrollgruppe. Weiter sind Experimente teuer und zeitraubend. Die verschiedenen Gruppen müssen nämlich eine statistisch relevante Grösse aufweisen. Zudem sind Ergebnisse im allgemeinen erst nach langer Zeit zu erwarten. Vielfach ist es dann zu spät, um noch Anpassungen an der Massnahme vornehmen zu können. Schliesslich sind auch Experimente nicht frei von methodischen Problemen. Ein Beispiel dafür ist der Hawthorne-Effekt, der zeigt, dass ein Messvorgang das Verhalten der Untersuchten beeinflusst (dazu Rossi et al 1988 S. 96 ff.)

Wenn echte Experimente nicht durchführbar sind, müssen die Evaluatoren oder Evaluatorinnen andere Wege wählen, um die durch die Zufallsauswahl angestrebte Äquivalenz von Kontroll- und Versuchsgruppen zumindest näherungsweise zu erreichen. Eine Möglichkeit bilden dabei *quasi-experimentelle Untersuchungsanlagen*, wie sie vor allem von Cook & Campbell (1976) beschrieben wurden. Die wichtigsten sind in der folgenden Übersicht zusammengestellt (Darstellung 2: Quasi-experimentelle Designs).

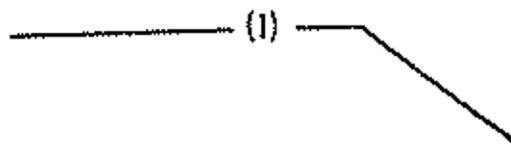
Die Vergleichsgruppen werden dabei aus natürlich vorkommenden zeitlich, örtlich oder personalen Differenzen in der Anwendung der Massnahmen gebildet und so Schlüsse auf die Wirkungen gezogen. Die Schwierigkeit liegt in der geschickten Auswahl einer vergleichbaren Kontrollgruppe. Quasi-Experimente sind Realexperimenten methodisch unterlegen, da es meist nicht gelingt, alle störenden Einflüsse zu kontrollieren. Ein Beispiel für eine quasi-experimentell angelegte Evaluation im Rahmen des NFP 27 bildet eine Studie, welche sich mit der Unterstützung des landwirtschaftlichen Hochbaus durch Investitionskredite und Subventionen im Kanton Zürich befasst hat. Hierzu wurde eine Gruppe stark subventionierter Landwirtschaftsbetriebe einer Gruppe wenig subventionierter Betriebe gegenübergestellt. Weiter wurden deren landwirtschaftliches Einkommen und deren Verschuldung miteinander verglichen. Zwischen den Gruppen gab es klare Unterschiede (Schnyder 1993).

Darstellung 2: Zeitpfade einer Wirkung
 (Quelle: Hellstern 1991)

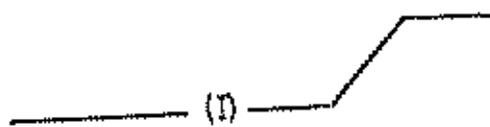
A Abrupt change in level.



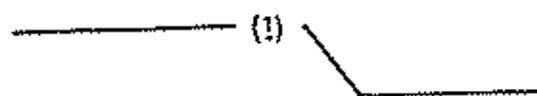
F. Delayed change in direction.



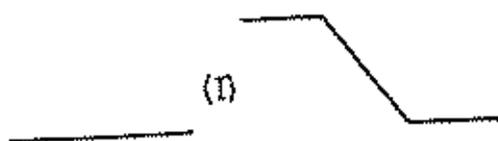
B Delayed change in level.



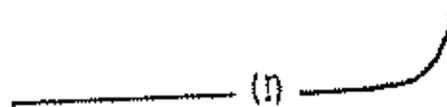
G. Temporary change in direction



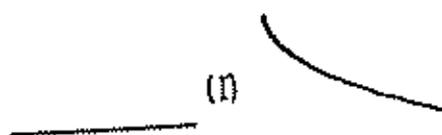
C. Temporary change in level..



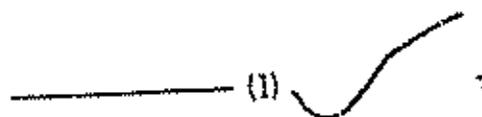
H. Accelerated change in direction



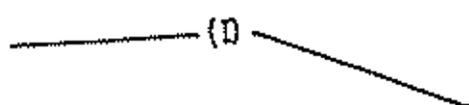
D. Decaying change in level.



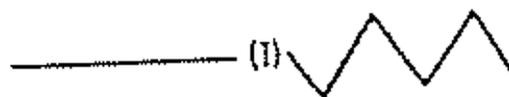
I. "Evolutionary operations" effect



E Abrupt change in direction.



J. Change in variability



3.2.4 Zeitreihenuntersuchung

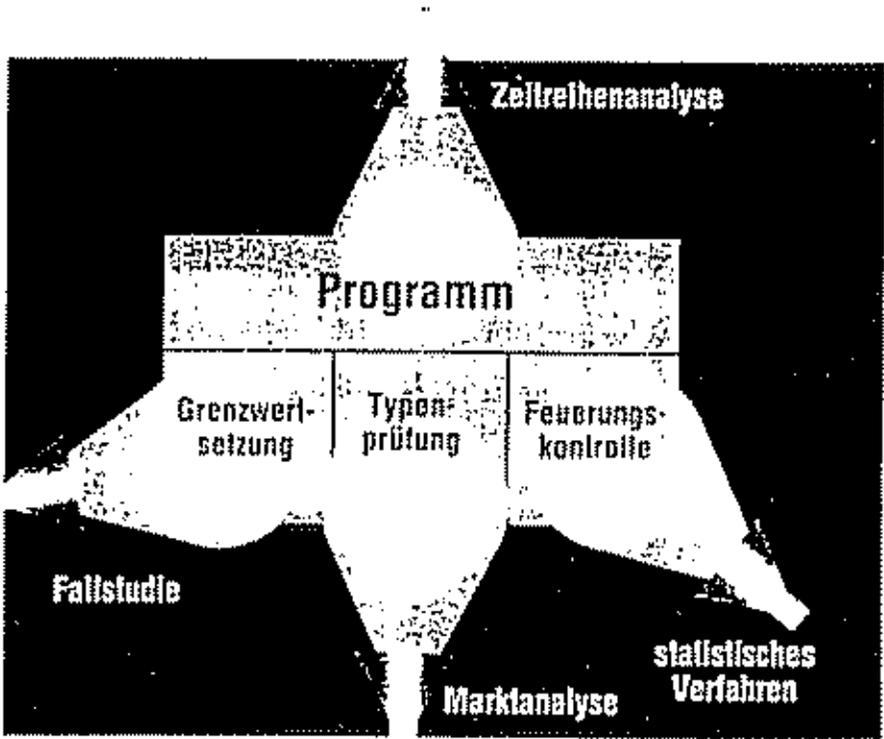
Die konzeptionelle Absicht von Zeitreihenuntersuchungen besteht darin, den Zustand vor der Einführung einer Massnahme demjenigen danach gegenüberzustellen und daraus auf die Wirkungen von Interventionen zu schliessen. Um trendbedingte sowie zufällige Entwicklungen in den Griff zu bekommen, beschränken sich entsprechende Untersuchungen im allgemeinen nicht nur auf einen Variablenvergleich vorher/nachher sondern beobachten die Entwicklung während eines möglichst langen Zeitraums (Zeitreihenanalysen, Interventionsanalysen). Die Wirksamkeit einer politischen Massnahme zeigt sich darin, wie sich der Verlauf der Reihe nach der Intervention - über die üblichen zufälligen Schwankungen hinaus - gegenüber vorher verändert

Darstellung 3 zeigt verschiedene Zeitpfade einer Wirkung. Sie macht deutlich, dass in Zeitreihenuntersuchungen der Zusammenhang zwischen der abhängigen und der unabhängigen Variable als ein Prozess zu betrachten ist. Die Stärke und sogar die Richtung des Einflusses, den eine Intervention auf die Output-Reihe hat, können sich mit der Zeit ändern. Andere Vergleichsebenen sind demgegenüber nicht in der Lage, den Wirkungsverlauf einer Intervention adäquat zu erfassen. Ein Beispiel für eine *statistisch gestützte Zeitreihenanalyse* ist die Evaluallion von Massnahmen zur Luftreinhaltepolitik in der Schweiz von Thomas Widmer (1991)

Derartige Techniken stellen hohe Anforderungen an die Datenreihen (vgl. Menard 1991). So werden in den meisten Fällen mehr als 15 aufeinanderfolgende Datenpunkte vor und nach der Intervention verlangt. Finden mehrere relevante Interventionen in kurzen Abständen statt, so sind zusätzliche Datenpunkte notwendig. Die letztgenannte Anforderung erschweren die Anwendung von Zeitreihenanalysen in der Evaluationsforschung. In der Praxis erfolgen politische Eingriffe nämlich selten als einfache Interventionen. Viel häufiger sind es Prozesse kleiner Schritte mit sich überlagernden externen Einflüssen.

In jedem Fall ist es zweckmässig, die Interpretation von Zeitreihen mit der Beobachtung chronologischer Entwicklungsverläufe einzelner Variablen in einfachen grafischen Darstellungen zu beginnen. In einer Studie der Arbeitsgruppe Gesetzesevaluation wurde bei

Darstellung 3: Konzeptionelle Triangulation
(Quelle: Balthasar & Knöpfel 1994)



spletsweise untersucht, welche Auswirkungen die Einführung der Kehrichtsackgebühr hat. Dazu wurden gemeindespezifische Daten über die Entwicklung der Menge des Gemischkehrichs gesammelt und interpretiert (INFRAS 1990). Derartige pragmatische Vorgehensweisen sind in der Evaluationsforschung weit verbreitet. Die Ergebnisse basieren jedoch in erster Linie auf Intuition und Erfahrung und entbehren statistischer Absicherung.

3.2.5 Konzeptionelle Triangulation

Im Zusammenhang mit der Frage nach der Gültigkeit qualitativer Daten hat die Triangulation in letzter Zeit neue Aktualität gewonnen. Der Begriff stammt ursprünglich aus dem Bereich der militärischen Navigation und bezeichnet eine Strategie, um ausgehend von verschiedenen Referenzpunkten die exakte Position eines Objekts festzustellen. In die Sozialwissenschaften wurde der Gedanke von D.T. Campbell & D. Fiske (1959) eingeführt. Deren Idee war es, eine Hypothese durch die Konfrontation mit einer Serie komplementärer Testmethoden zu überprüfen. Dabei sollte der Forscher eine Vermutung auf der Basis der ersten Methode durch Befunde einer zweiten Methode absichern. Zudem wurde die Chance in Betracht gezogen, dass bei der Anwendung der zweiten Technik auf überraschende Erkenntnisse geschlossen werden könnte. Durch Methoden-triangulation sollen weiter in den einzelnen Methoden steckende Fehler entdeckt werden. Es wurde erwartet, dass es äusserst unwahrscheinlich sei, mit mehreren Techniken dieselben falschen Resultate zu erhalten.

In den siebziger Jahren wurde die Triangulation von N.K. Denzin (1970) weiterentwickelt. Er schlug vor, vier Typen von Triangulation zu unterscheiden: Als "Daten-Triangulation" bezeichnete er den Einbezug verschiedener Datenquellen in eine Untersuchung. Unter "Investigator-Triangulation" verstand er den Einbezug unterschiedlicher Beobachter oder Forscher, um aktorbedingte Verzerrungen der Ergebnisse zu vermeiden. Die Kombination verschiedener Erhebungstechniken benannte er "methodologische Triangulation". Schliesslich wies er als vierten Typ auf die "Theorien-Triangulation" hin. Hierbei werden die Daten mit verschiedenen Theorien konfrontiert und so die Erklärungskraft der einzelnen Ansätze geprüft (vgl. u.a. Flick 1992). In der Evaluationsforschung hat besonders Michael Quinn Patton (1990) auf die Chancen der Triangulation hingewie-

sen. Er schliesst sich der Unterscheidung von vier Typen von Triangulation, die Denzin getroffen hat, an:

In der Evaluationsforschung kann Triangulation auch die Kombination von Vergleichsebenen bezeichnen. Insbesondere die Verbindung von Quer- und Längsschnittuntersuchungen hat in der Evaluationsforschung eine lange Tradition. Zur Klärung des Sprachgebrauchs bezeichnen wir die Kombination von Vergleichsebenen als *"konzeptionelle Triangulation"* und grenzen diese von den anderen Formen von Triangulation, insbesondere von der Daten- und der methodologischen Triangulation ab. Vor allem zwei Aspekte lassen die konzeptionelle Triangulation als erfolgversprechendes Vorgehen erscheinen: einerseits das gegenseitige Einfließen von Informationen aus den Teiluntersuchungen und andererseits die Vertiefung und Verbreiterung des Erklärungsgehalts durch verschiedene Zugangsweisen. Beim ersten Aspekt profiliert die Gesamtbetrachtung davon, dass durch unterschiedliche analytische Zugangsweisen in Teiluntersuchungen Ergebnisse zu erwarten sind, welche als Fragestellungen, als Datengrundlage oder als Randbedingung in andere Teile einfließen können.

Der zweite Aspekt, der sich als ergänzende Dimension bezeichnen lässt, unterstreicht, dass - ähnlich wie in der Fotografie - auch in der Evaluation zusätzliche "Lampen" einen Gegenstand besser ausleuchten und unerwünschte Schatten zu beseitigen vermögen. Durch unterschiedliche Zugangsweisen wird die Abhängigkeit der Ergebnisse von einzelnen Analysekonzepten relativiert. Siegfried Lamnek hebt aber zurecht hervor, dass eine Triangulation keine "Hauptwirklichkeit" erkennen lässt, sondern zu komplementären Bildern der Realität führt. Durch unterschiedliche Zugangsweisen wird die Abhängigkeit der Ergebnisse von einzelnen Analysekonzepten jedoch relativiert. Triangulation sollte demnach weniger der Konstruktion einer "objektiven" Wahrheit dienen, als bewusst die Breite und die Tiefe der Analyse erweitern. Es sind von den einzelnen Untersuchungsteilen nicht deckungsgleiche, sondern komplementäre Ergebnisse zu erwarten, welche sich ineinander fügen und sich ergänzen können, aber nicht kongruent sein müssen (Lamnek 1988, S. 236). Es bleibt also nach wie vor möglich, dass sich einzelne Teilergebnisse widersprechen und dass Bereiche des Untersuchungsgegenstandes weiterhin im dunkeln bleiben.

Verdeutlichen wir an einem Beispiel, das aus dem Forschungsprojekt von A. Balthasar und C. Knöpfel (1994) heraus entwickelt wurde.

was wir unter konzeptioneller Triangulation verstehen. Stellen wir uns vor, dass die Wirkung einer technischen Prüfung durch eine staatliche Behörde (z.B. Typenprüfung) auf die technische Entwicklung untersucht werden soll. Ein erster Zugang könnte in diesem Fall eine Zeitreihenanalyse der Forschungs- und Entwicklungsinvestitionen von Herstellern im Zeitverlauf mit der Einführung der technischen Prüfung als Intervention darstellen. Zweitens wäre es möglich, das Innovationsverhalten einzelner Firmen in der Schweiz und in Deutschland im Quervergleich mit Blick auf allfällige organisatorische Unterschiede, die sich aus der Massnahme ergeben, zu ergründen. Konzeptionelle Triangulation in der Evaluationsforschung bedeutet, dass die Wirkungen einer einzigen Massnahme mit unterschiedlichen Zugängen analysiert werden. Im geschilderten Beispiel können also Erkenntnisse der Zeitreihenanalyse als Randbedingungen in den Quervergleich einfließen, und der Quervergleich beeinflusst die Auswahl der Datenreihen für die Zeitreihenanalyse. Fragestellungen, welche sich aus dem einen Untersuchungsteil ergeben, lassen sich im Rahmen des anderen Teils abklären. Zusammen führen beide Teile zu gegenseitig sich ergänzenden Bildern der realen Wirkungen der technischen Prüfung. Insgesamt werden so mehr Erkenntnisse über die Wirkungszusammenhänge erarbeitet.

4 Forschungstechniken

Unter dem Titel Forschungstechniken fassen wir die Methoden zur Gewinnung und zur Auswertung empirischer Daten zusammen. Die Literatur dazu ist so ausführlich, dass wir an dieser Stelle nur auf die Literaturliste verweisen wollen. Wie für die ganze sozialwissenschaftliche Praxis nimmt auch in der Evaluationsforschung die Diskussion um den Unterschied zwischen qualitativen und quantitativen Vorgehensweisen breiten Raum ein. Dabei wird deutlich, dass es sich nicht um beliebig austauschbare unterschiedliche Methoden handelt. Dahinter stecken vielmehr zwei grundverschiedene Zugänge zu Untersuchungsgegenständen.

5. Umsetzung

Wissenschaftliche Forschung und politische Praxis stehen in einem vieldiskutierten Spannungsverhältnis. Die Politik stellt den Anspruch an die Evaluation, dass deren Ergebnisse in der Praxis umgesetzt werden können. Wir haben oben darauf hingewiesen, dass im Zu-

sammenhang mit der Entwicklung lernorientierter Evaluationsstrategien neue Ansätze entwickelt und getestet wurden. Solche Zugänge gehen davon aus, dass die Umsetzung nur bei einem kontinuierlichen Informationsaustausch bereits während der laufenden Untersuchung optimal erfolgen kann (vgl. auch Kap. 5.3.2).

In jedem Fall ist es wichtig, die Ergebnisse von Evaluationen adressatengerecht zu gestalten. H. Wottawa und H. Thierau (1991) weisen darauf hin, dass Sozialwissenschaftlerinnen zu einem sehr komplexen durchschnittlichen kognitiven Fähigkeiten der Sprachverarbeitung überfordernden Satzbau neigen. Sie verwenden nahezu selbstverständlich Fachausdrücke. Es ist wichtig, bei der Abfassung von Berichten sowie bei anderen Formen der Präsentation das Zielpublikum im Auge zu haben. Ausser dem bei wissenschaftlichen Projekten üblichen umfassenden Schlussbericht sind spezifische, zielgruppengerechte Formen schriftlicher, mündlicher oder audiovisueller Darstellung denkbar.

Mit der Forderung nach der Nützlichkeit von Evaluationsergebnissen für die politische Praxis geraten Studien oftmals mitten in politische Spannungsfelder. Schon die Frage, welche Massnahmen evaluiert werden sollen, ist vielfach umstritten. EvaluatorInnen sollten sich darum Rechenschaft darüber geben, wessen Kriterien bei der Auswahl massgeblich waren. AuftraggeberInnen sind nämlich aus ganz unterschiedlichen Motiven an Evaluationen interessiert.

Grundsätzlich lassen sich vier Anreize unterscheiden. Die erste Variante bildet ein technisches Interesse: Evaluationen sollen Informationen liefern, welches zur Optimierung staatlicher Interventionen eingesetzt werden kann. Evaluationen erfüllen in diesem Fall eine zulässige *instrumentelle Funktion*. Die Umsetzung erfolgt dadurch, dass sich Entscheidungsinstanzen und andere Interessengruppen direkt und konkret auf Forschungsergebnisse beziehen. Die zweite Möglichkeit besteht in einer eher *konzeptionellen Motivation*. Auch dieser Impuls ist für die Evaluation weitgehend unproblematisch. Die Evaluation reflektiert, stellt Fragen, schafft Begriffe, bringt Dinge zur Sprache und kann Lernprozesse auslösen. Die Untersuchungsergebnisse werden in diesem Falle als Denkanstösse bei der Weiterbearbeitung potentieller Lösungswege für bestimmte politische Probleme verwendet. Der dritte Motivationsstrang bilden *argumentative Absichten*. Evaluationen werden eingesetzt, um politische Argumente zu erarbeiten, welche dann unter strategischen Gesichtspunkten umgesetzt werden. Viertens können Evaluationen auch veranlasst

werden, um zur Erhöhung der Akzeptanz von Massnahmen und zu deren Rechtfertigung beizutragen. In diesem Fall dient Evaluation in erster Linie der *Legitimierung*. Vor allem bei den letzten beiden Formen besteht die Gefahr einer unzulässigen Instrumentalisierung der Evaluation durch die Politik. Nur die genehmten Resultate werden weiterverwendet (vgl. u.a. Rieder & Varone 1991)

Das Thema 'Umsetzung von Evaluationsergebnissen' weist aber noch auf grundsätzlichere Aspekte der Beziehung von Wissenschaft und Politik hin. Im Zentrum steht die Frage, ob wissenschaftliche Beratung tendenziell zu einem Ausschluss demokratischer Kontrolle führt und die Politikumsetzung in die Hand von interessengelenkten Technokraten legt oder ob Wissenschaft Entscheidungsprozesse offenlegt und Wirkungsüberlegungen anstellt, welche für die Rechtfertigung des Einsatzes von Steuermitteln unerlässlich sind (Knoepfel 1989). Die Dynamiken der Interaktionsprozesse zwischen Wissenschaft und Politik sind insbesondere im Anschluss an die systemtheoretischen Überlegungen von Niklas Luhmann Thema einer noch nicht abgeschlossenen politikwissenschaftlichen Debatte (vgl. Freiburghaus & Zimmermann 1985).

Literatur

- Alemann Ulrich von und Fomdran Erhard 1990. Methodik der Politikwissenschaft. Eine Einführung in die Arbeitstechnik und Forschungspraxis. Vierte überarbeitete und erweiterte Fassung. Stuttgart: Kohlhammer
- Arbeitsgruppe Gesetzesevaluation 1991. Die Wirkungen staatlichen Handelns besser ermitteln: Probleme. Möglichkeiten. Vorschläge. Schlussbericht an das Eidg. Justiz und Polizeidepartement. Bern: EDMZ
- Attlesander Peter 1984. Methoden der empirischen Sozialforschung. Berlin: de Gruyter
- Balthasar Andreas und Knöpfel Carlo 1994. Umweltpolitik und technische Entwicklung. Eine politikwissenschaftliche Evaluation am Beispiel der Heizungen. Basel: Helbing & Lichtenhahn
- Beywl Wolfgang 1988. Zur Weiterentwicklung der Evaluationsmethodologie. Grundlegung, Konzeption und Anwendung eines Modells der responsiven Evaluation. Frankfurt/Bern: Lang
- Blerter Willy und Binder Hans-Martin 1994. Unternehmerische Innovationsprozesse. Die Wirksamkeit staatlicher Fördermassnahmen - eine Evaluation. Chur/Zürich: Rüegger
- Busmann Werner 1994. Evaluationen staatlicher Massnahmen erfolgreich begleiten und nutzen - ein Leitfaden. Chur/Zürich: Rüegger
- Campbell Donald T. und Fiske Donald 1959. Convergent and discriminant Validation by the Multifit-Multimethod Matrix. In: Psychological Bulletin 56 S. 81-105
- Cook Thomas D. und Campbell Donald T. 1976. The design and conduct of quasi-experiments and true experiments in field settings. In: Handbook of Industrial and organizational psychology. M. D. Dunnette (Ed.) Chicago: Rand McNally S. 223 ff
- Denzin Norman K. 1970. The Research Act. New York
- Eckstein Harry 1975. Case Study and Theory in Political Science. In: Handbook of Political Science. Bd. 7. F. I. Greenstein and N. W. Polsby (Eds.) Reading, Mass
- Fitz-Gibbon Carol Taylor und Morris Lynn Lyons 1987. How to design a program evaluation. Program evaluation kit 3. Newbury Park: Sage
- Flick Uwe 1992. Entzauberung der Intuition. Systematische Perspektiven-Triangulation als Strategie der Geltungsbegründung qualitativer

Daten und Interpretationen. In: Analyse verbaler Daten. Über den Umgang mit qualitativen Daten. J.H.P. Hoffmeyer-Zlotnik. Opladen
Freiburghaus Diefel und Zimmermann Willy 1985. Wie wird forschung
politisch relevant?. Bern: Haupt.

Friedrichs Jürgen 1973. Methoden empirischer Sozialforschung Rein-
bek

General Accounting Office 1990. Case Study Evaluations. Transfer
Paper 10 1 9

Goetze U 1991. Szenario-Technik In der strategischen Unterneh-
mensplanung Wiesbaden: Deutscher Universitätsverlag

Guba E. G und Lincoln Y. S 1987. The Countenance of Fourth-
Generation Evaluation: Description, Judgment, and Negotiation. In:
Evaluation Studies Review Annual. D. S. Cordray und M. W. Lipsey.
London

Hanser Christian, Kuster Jürg und Cavelli Guido, Hotellerieförderung
durch Bund und Kantone. Evaluation der Auswirkungen in der Hotel-
lerie. Schriftenreihe BIGA. Beiträge zur Tourismuspolitik Nr. 3. Bern:
EDMZ

Hellstern Gerd-Michael 1991, Quasi-experimentelle Designs am Bei-
spiel von Zeitreihen. Materialien zum Seminar "Methodik von Evalua-
tionen" des NFP 27. Kandersteg 1991

Hellstern Gerd-Michael und Wollmann Hellmut 1983. Evaluations-
forschung. Ansätze und Methoden - dargestellt am Beispiel des
Städtebaus. Basel/Boston/Stuttgart: Birkhäuser

INFRA5 1990. Evaluation Kehrichtsackgebühr. Schriftenreihe Umwelt
Nr. 136. Ökonomie. Bern: BUWAL

Knoepfel Peter 1989. Wissenschaftliche Beratung im Politikvollzug: Ge-
fahr für die Demokratie? Einige Thesen. In: Jahrbuch der Schweizeri-
schen Gesellschaft für Politikwissenschaft 29. Bern: Haupt S. 81-102

Knoepfel Peter 1991. Evaluation des BUWAL. Expertenbericht zur
Evaluation der Luftreinhaltung, des ländlichen Gewässerschutzes
und der UVP des Bundes. Bern: EDMZ

Lamnek Siegfried 1989. Qualitative Sozialforschung. Band 2: Metho-
den und Techniken. München: Psychologie Verlags Union

Menard Scott 1991. Longitudinal Research. Newbury Park: Sage

Patton Michael Quinn 1990. Qualitative Evaluation and Research
Methods Newbury Park: Sage (2nd edition)

- Reinhold Gerd unter Mitarbeit von Siegfried Lamnek und Helga Recker 1991. Soziologie-Lexikon. München Wien: Oldenbourg
- Rey Michel, Alberton Siegfried und Vodoz Luc 1993. La politique valaisanne de promotion économique: une évaluation privilégiant le point de vue des acteurs. Berne: FNRS/PNR 27
- Rieder Stefan und Varone Frédéric 1991. Fokus und Methoden sozialwissenschaftlicher Evaluationsforschung im Energiebereich. Eine Literatursauswertung. Bern: Institut für Politikwissenschaft
- Risl Ray C. On the application of program evaluation designs: sorting out their use and abuse. In: Knowledge in Society, The International Journal of knowledge transfer. Winter 1989-90 vol 2 number 4, S. 74-96
- Rossi Peter H. und Freeman Howard E. 1993. Evaluation. A Systematic Approach. 5th edition. London: Sage
- Rossi Peter H., Berk R. A. und Lenihan K. J. 1980. Money, Work and Crime: Some Experimental Evidence. New York: Academic
- Rossi Peter H., Freeman Howard E. und Hofmann Gerhard 1988. Programm-Evaluation. Einführung in die Methoden der angewandten Sozialforschung. Stuttgart: Enke
- Roth Erwin unter Mitarbeit von Klaus Heidenreich (Hrsg.) 1993. Sozialwissenschaftliche Methoden: Lehr- und Handbuch für Forschung und Praxis. München. Wien: Oldenbourg
- Schmitz Bernhard 1989. Einführung in die Zeitreihenanalyse: Modelle, Softwarebeschreibung, Anwendungen. Bern: Huber
- Schnyder Stefan 1993. Ueber die Verteilung und Wirkung der Investitionshilfen im landwirtschaftlichen Hochbau des Kantons Zürich. Studie im Auftrag des Meliorations- und Vermessungsamtes der Stadt Zürich. unveröffentlichtes Typoskript
- Scriven Michael 1980. The logic of evaluation. California: Edgepress
- Seiffert Helmut 1969. Einführung in die Wissenschaftstheorie I. München: Beck
- Widmer Thomas 1991. Evaluation von Massnahmen zur Luftreinhaltepolitik in der Schweiz. Eine quasi-experimentelle Interventionsanalyse nach dem Ansatz Box/Tiao. Chur/Zürich: Rüegger
- Wottawa Heinrich und Thierau Heike 1991. Lehrbuch Evaluation. Bern/Stuttgart/Toronto: Huber