

Ziel des NFP 27 ist die Erprobung und Verbesserung von Methoden zur Erfassung und Beurteilung der Wirksamkeit staatlicher Massnahmen. Es werden Evaluationen, Kurzevaluationen und Studien über Lernprozesse in Verwaltungen durchgeführt, um

- ein Know How über Evaluationsmethoden aufzubauen
- Evaluationsmethoden zu erproben und daraus Erfahrungen zu gewinnen
- Anhaltspunkte für die Zeitgerechtigkeit, Zuverlässigkeit und Relevanz von Evaluationen zu gewinnen
- zu überprüfen, inwieweit durch Evaluationen Lernprozesse in Gang gesetzt werden können.

Werner Bussmann **Evaluationen erfolgreich begleiten**

Werner Bussmann

Evaluationen staatlicher Mass- nahmen erfolg- reich begleiten und nutzen

Ein Leitfaden

E f f e t s
des mesures étatiques
Wirksamkeit
staatlicher Massnahmen
E f f e t t i
delle misure statali

E f f e t s
des mesures étatiques
Wirksamkeit
staatlicher Massnahmen
E f f e t t i
delle misure statali

E f f e t s
des mesures étatiques
Wirksamkeit
staatlicher Massnahmen
E f f e t t i
delle misure statali

Werner Bussmann

Evaluationen staatlicher Massnahmen erfolgreich begleiten und nutzen:

Ein Leitfaden

© Werner Bussmann, Bern

1. und 2. Auflage 1995, Verlag Rüegger AG Chur/Zürich

Für die Veröffentlichung auf dem Internet leicht bearbeitete Auflage

Der vorliegende Leitfaden ist im Rahmen des Nationalen Forschungsprogramms "Wirksamkeit staatlicher Massnahmen" (NFP 27) erarbeitet worden.

Werner Bussmann

**Evaluationen staatlicher Massnahmen erfolgreich
begleiten und nutzen:**

Ein Leitfaden

© Werner Bussmann, Bern

2. und 2. Auflage 1995, Verlag Rüegger AG Chur/Zürich
Für die Veröffentlichung auf dem Internet leicht bearbeitete Auflage

Ziel des NFP 27 ist die Erprobung und Verbesserung von Methoden zur Erfassung und Beurteilung der Wirksamkeit staatlicher Massnahmen. Es werden Evaluationen, Kurzevaluationen und Studien über Lernprozesse in Verwaltungen durchgeführt, um

- ein Know How über Evaluationsmethoden aufzubauen
- Evaluationsmethoden zu erproben und daraus Erfahrungen zu gewinnen
- Anhaltspunkte für die Zeitgerechtigkeit, Zuverlässigkeit und Relevanz von Evaluationen zu gewinnen
- zu überprüfen, inwieweit durch Evaluationen Lernprozesse in Gang gesetzt werden können.

Inhaltsverzeichnis

Inhaltsverzeichnis	5
Vorwort	9
0 Überblick	11
1 Sechs Schritte, um ein Evaluationsprojekt erfolgreich in Auftrag zu geben, zu begleiten und zu nutzen	13
1.1 Das Problem erkennen	13
1.2 Die Evaluation planen.	15
1.21 Die Forschungsfrage festlegen	15
1.22 Den Kreis der Beteiligten grob abgrenzen	19
1.23 Interne oder externe Evaluation?	19
1.24 Den Verwendungszweck der Evaluation bestimmen ...	22
1.25 Den zeitlichen Rahmen festlegen und die erforderlichen finanziellen und personellen Mittel abschätzen	23
1.26 Über die Weiterbearbeitung der Evaluation entscheiden	23
1.3 Offerten einholen und den Auftrag vergeben	24
1.31 Soll eine Evaluierbarkeitsprüfung in Auftrag gegeben werden?	24
1.32 Ein Pflichtenheft erarbeiten und Offerten einholen.	25
1.33 Die Offerten überprüfen	26
1.34 Auswählen	28
1.35 Auftragsvergabe als Verhandlung	32
1.4 Die Evaluation sinnvoll begleiten	33
1.41 Wer begleitet die Evaluation?	33
1.42 Grundsätze für die Begleitung	34
1.5 Die Evaluationsresultate optimal nutzen	36
1.51 Rechenschaft ablegen	37
1.52 Den Vollzug verbessern	37

1.53	Kosten einsparen	38
1.54	Massnahmen aufheben	39
1.55	Massnahmen ändern oder neugestalten.	40
1.56	Lernprozesse als Ziel von Evaluationen	42
1.57	Exkurs: Nutzung und Missbrauch von Evaluationen	43
1.6	Vorkehrern treffen, um künftig die Wirksamkeit von Massnahmen besser zu erfassen	45
2	Was sind Evaluationen staatlicher Massnahmen?	51
2.1	Evaluation, Begriff	51
2.2	Was kann alles evaluiert werden?	51
2.3	Welche Aspekte und Kriterien können Evaluationen beinhalten?	53
2.4	Welchen Zeithorizont decken Evaluationen ab?	55
2.5	Was kosten Evaluationen und wie lange dauern sie?	56
3	Welcher Strategien und Forschungstechniken bedienen sich Evaluationen?	59
3.1	Möglichkeiten zur Klassifikation von Evaluationsstrategien	59
3.2	Begriffe: Evaluationsstrategie und Forschungstechnik	60
3.3	Überblick über die Evaluationsstrategien	60
3.31	Deskriptive, normative und kausalerorientierte Fragen	60
3.32	Fragestellung und Evaluationsstrategie.	61
3.4	Evaluationsstrategien für deskriptive und normative Fragen	62
3.5	Evaluationsstrategien für kausalerorientierte Fragestellungen	65
3.51	Explizit und implizit vergleichende Evaluationsstrategien	65
3.52	Kontrollgruppen-Vergleich.	69
3.53	Querschnitt-Vergleiche.	71
3.54	Längsschnitt-Vergleich.	74
3.6	Forschungstechniken	76

3.61	Überblick/Darstellung ausgewählter Forschungs- techniken	76
3.62	Fallstudien	78
3.63	Umfragen	80
3.64	Analyse vorhandener Daten.	83
3.7	Exkurs: Modelle als Ausgangspunkt und Resultat von Evalu- ationen	84
3.8	Methodenmix und Triangulation	89
3.9	Hinweise zur Wahl der Evaluationsstrategie und der Forschungstechnik	91
Anhang 1	Erläuterung der wichtigsten Begriffe	95
Anhang 2	Glossar der wichtigsten Begriffe in vier Sprachen	100
Anhang 3	Literatur	101
Anhang 4	Stichwortverzeichnis	106

Vorwort

Die in der Nachkriegszeit stark angewachsene Staatstätigkeit hat seit den 70er Jahren zu wachsenden Zweifeln darüber geführt, ob staatliche Massnahmen korrekt vollzogen werden, ob sie die erwünschten Wirkungen zeitigen und ob sie wirtschaftlich gerechtfertigt werden können. Evaluationen sind eine der Antworten auf diese nagenden Fragen. Seit den 60er Jahren in den USA entwickelt, haben sie im europäischen Raum zunehmend Verbreitung gefunden und werden in der Schweiz auf Stufe des Bundes, der Kantone und Gemeinden immer häufiger eingesetzt.

Der vorliegende Leitfaden soll den Umgang mit dem vielfältig und flexibel einsetzbaren Instrument "Evaluationen" erleichtern. Er richtet sich an Personen, welche Evaluationen in Auftrag geben, (mit) begleiten und nutzen möchten. Erfahrungsgemäss wird die Verwaltung in vielen Fällen Auftraggeber von Evaluationen sein, weshalb viele Beispiele aus diesem Umfeld gewählt werden. Auch Parteien, Verbände und weitere Gruppen werden indessen immer häufiger zu diesem Instrument greifen. Ihnen allen soll der Leitfaden eine nützliche Hilfe sein.

Der Leitfaden bildet einen ersten Teil des Schlussergebnisses des Nationalen Forschungsprogramms "Wirksamkeit staatlicher Massnahmen" (NFP 27). Ein zweiter Teil wird in Form eines "Handbuchs über Methoden der Evaluationsforschung" in rund einem Jahr erscheinen und sich vor allem an die Wissenschaft und an die Evaluationspraxis richten.

Der Impuls zur Erarbeitung des Leitfadens erfolgte im April 1993 während eines dreiwöchigen Forschungsaufenthalts in den USA. Dabei konnte ich namentlich aus den Kontakten mit dem GAO (General Accounting Office) und den verschiedenen Gesprächen, die ich mit Eleanor Chelimsky und Ray Rist führte, Nutzen ziehen. Die anschliessenden Arbeiten konnten sich auf zahlreiche Gespräche im Kreis der Expertengruppe und der Projektbearbeiter des NFP 27 stützen. Auf erste Entwürfe des Leitfadens gaben mir namentlich Andreas Balthasar, Frohmut Gerheuser, Leo Meyer und Ruedi Meier ausführliche und hilfreiche Kommentare. Frau Madeleine Rieke brachte die Texte in druckbare Form und bewies dabei viel Beharrungsvermögen und Geduld. Heinrich Koller als Direktor des Bundesamts für Justiz stellte die Büroinfrastruktur zur Verfügung. Ernst Buschor als Präsident der Expertengruppe des NFP 27 gab mir viel Vertrauen und Freiraum bei der Erarbeitung der Schlussberichte. Allen

erwähnten Personen möchte ich herzlich danken. Ein spezieller Dank geht ausserdem an meine Frau und die beiden Kinder, deren Vitalität mithalf, meine gelegentlichen Arbeitsschübe unbeschadet zu überstehen.

Bern, Oktober 1994

0 Überblick

Der Leitfaden richtet sich an Personen, welche eine Evaluation in Auftrag geben, begleiten und nutzen wollen bzw. darüber Abklärungen durchführen. Er ist keine Anleitung zur Selbstevaluation, aber kann Ideen geben, welche Aspekte bei der eigenen Beurteilung des Vollzugs oder der Wirksamkeit beachtet werden sollten.

Der Leitfaden umfasst drei Teile. Teil 1 behandelt die Abwicklung einer Evaluation - aus der Sicht des Auftraggebers - von der Problemformulierung bis zur Nutzung. Teil 2 informiert darüber, was Evaluationen sind, und Teil 3 darüber, welche Strategien und Forschungstechniken eingesetzt werden, um aussagekräftige Resultate zu erzielen.

Teil 1 erörtert in sechs Schritten das konkrete Vorgehen bei der Vergabe, Begleitung und Nutzung von Evaluationen. Kapitel 1.1 behandelt anhand von Beispielen die möglichen Auslöser für Evaluationen. Kapitel 1.2 zeigt Ihnen, welche Punkte Sie beachten sollten, wenn Sie den Grobrahmen für die Evaluation bestimmen. Kapitel 1.3 gibt Ihnen Hinweise zum Einholen von Offerten, zur Auswahl und zum Vertragsabschluss. Kapitel 1.4 informiert über die Begleitung der Evaluation und den dafür einzusetzenden Zeitaufwand. Kapitel 1.5 zeigt Ihnen, für welche Zwecke Sie die Evaluationsresultate nutzen können. Evaluationen können vor allem dazu beitragen, den Vollzug zu verbessern, die Wirkungen der in Frage stehenden Massnahmen zu überprüfen und fundiert darzustellen sowie die Änderung oder Neukonzipierung einer Massnahme an die Hand zu nehmen. Kapitel 1.6 gibt Ihnen Hinweise darüber, wie Sie bei der Neugestaltung einer Massnahme sicherstellen können, dass die Wirksamkeit später überprüft und belegt werden kann.

Teile 2 und 3 vermitteln Ihnen grundlegende Kenntnisse über Evaluationen. Teil 2 behandelt die wichtigsten Arten von Evaluationen, die ihnen zugrundegelegten Kriterien und erwähnt auch den Zeitbedarf und die Kosten von Evaluationen. Teil 3 behandelt die in der Evaluationsforschung hauptsächlich verwendeten Strategien und Forschungsmethoden. Besonderes Gewicht wird auf die Evaluationsstrategien zur Wirkungsermittlung gelegt.

In Anhang 1 werden die wichtigsten verwendeten Begriffe kurz erläutert, Anhang 2 enthält ein Glossar von wichtigen Evaluationsbegriffen in vier Sprachen, Anhang 3 umfasst das Literaturverzeichnis und Anhang 4 das Stichwortverzeichnis.

Vielleicht noch ein paar Hinweise darauf, was Sie in den nachfolgenden Kapiteln **nicht** finden werden. Dies ist kein Kurs für erfolgreiche Verwaltungsführung. Dieses Thema wäre zu breit. Wir konzentrieren uns auf die Frage, wie bessere Informationen über den Vollzug und die Wirkungen beschafft werden

können. Wir wissen, dass diese Anstrengungen zur Ermittlung der Wirksamkeit und Wirtschaftlichkeit staatlicher Massnahmen ab und zu nur ungern gesehen werden, sind aber überzeugt, dass sie im Interesse der Bürgerinnen und Bürger sind und längerfristig honoriert werden. Im Folgenden beschränken wir uns auch auf sozialwissenschaftlich abgestützte Evaluationen staatlicher Massnahmen. Dies ist einer von vielen Wegen, die begangen werden müssen. Eine innovative und kooperative Verwaltung hat sehr viele weitere Möglichkeiten mit je spezifischen Vor- und Nachteilen, um zu Informationen über die Wirksamkeit staatlicher Massnahmen zu gelangen: Verwaltungskontakte, Rückmeldungen von Bürgerinnen und Bürgern, Kontakt mit Verbänden, politische Vorstösse, Zeitungsmeldungen, Leistungsindikatoren, amtliche Statistik u.a.m.. Diese Möglichkeiten werden hier nicht behandelt. Wir möchten uns auf das Instrument "Evaluationen" konzentrieren, um Sie aussagekräftig darüber zu informieren. Mit allem Nachdruck möchten wir indessen erwähnen, dass Evaluationen Ergänzung und nicht Ersatz von anderen bestehenden Mechanismen der Erfolgskontrolle sein sollen und einen Beitrag dazu leisten können, dass staatliches Handeln den Bedürfnissen der Bürgerinnen und Bürger noch besser als bisher entspricht.

1 Sechs Schritte, um ein Evaluationsprojekt erfolgreich in Auftrag zu geben, zu begleiten und zu nutzen

1.1 Das Problem erkennen

Am Anfang jeder Evaluation steht ein Problem, zu dessen sinnvoller Lösung oder Bewältigung wesentliche Informationen fehlen. Es kann sich beispielsweise um eines der folgenden Probleme handeln:

Beispiel 1: In der Baudirektion eines kleinen Zentralschweizer Bergkantons überlegen sich der Regierungsrat und die Beamten, zu welchem Zeitpunkt die drei grossen Pässe am besten vom Schnee geräumt werden. Würde der Aufwand bei einer etwas späteren Schneeräumung sinken? Aber ergäben sich dann nicht negative Auswirkungen auf den Tourismus und das Gastgewerbe? (Ausgangssituation für die Studie Ecoplan 1993b im Kanton Uri)

Beispiel 2: Das Bundesamt für Umwelt, Wald und Landschaft macht sich mehr und mehr Gedanken darüber, ob die Vorschriften des Natur- und Landschaftsschutzes, welche für Skipisten gelten, tatsächlich auch befolgt werden. Es ist zunehmend skeptisch, ob seine Richtlinien und Wegleitungen in der Praxis überhaupt beachtet werden, und möchte Hinweise darüber, wie es zu einer besseren Umsetzung seiner Vorschriften beitragen kann. (Ausgangssituation für die Studie Hirsch et al. 1995)

Beispiel 3: Eine grosse Schweizer Stadt ist gezwungen, die Ausgaben auch im Sozialbereich zu kürzen. Die Gemeinschaftszentren der Stadt, die durch eine gemeinnützige Organisation geführt werden, sollen mehr als eine Million Franken einsparen. Aber wo soll gespart werden? Linear oder gezielt in Bereichen, wo der Schaden nicht klein ist? Die Vorsteherin des Sozialamts wünscht, dass zusammen mit den LeiterInnen der Gemeinschaftszentren eine sinnvolle Sparstrategie ausgearbeitet wird. (Ausgangssituation für die Studie Price Waterhouse 1993)

Beispiel 4: Ein Vizedirektor im Bundesamt für Justiz möchte den Vorentwurf für ein Bundesgesetz über die Gleichstellung zwischen Frau und Mann auf seine mutmasslichen Folgen im wirtschaftlichen Bereich überprüfen lassen. Welche wirtschaftlichen Kosten wird ein Diskriminierungsverbot zur Folge haben? Wird die Bilanz möglicherweise sogar positiv ausfallen? (Ausgangslage für die Studie Littmann-Wernli 1991)

Beispiel 5: Als Pionier hat der Kanton Tessin 1984 ein kantonales Gesetz über Treuhänder erlassen. Nach zehn Jahren ist nun Zeit für eine Bilanz. Hat

sich das Gesetz bewährt? Soll es beibehalten, ausgebaut oder aufgehoben werden? (Ausgangslage für eine 1994 noch laufende Studie)

Beispiel 6: Die für die Hotellerieförderung zuständige Abteilung des BIGA macht sich seit längerer Zeit Gedanken über die Wirksamkeit und Effizienz der Förderung. Bekannt ist, wie viele Hotels Bürgschaften und Darlehen erhalten. Im Dunkeln liegt aber, ob die geförderten Hotels im Konkurrenzkampf nachher wirklich die Nase vorn haben und ob die Investitionen ihre Investitionskraft nachhaltig stärken. (Ausgangslage für die Studie Hanser/Kuster/Cavelti 1994)

Evaluationen sind ein flexibel und vielfältig einsetzbares Instrument. Sie können genutzt werden, um über staatliche Massnahmen Rechenschaft abzulegen (Beispiele 5 und 6), um den Vollzug zu verbessern (Beispiel 2), um Kosten zu sparen bzw. die Wirtschaftlichkeit zu verbessern (Beispiele 1 und 3), um Massnahmen aufzuheben (ev. Beispiel 5) oder zu ändern (Beispiele 2 und 5) oder um neue Massnahmen auszuarbeiten (Beispiel 4). Evaluationen werden je länger desto mehr zu einem unentbehrlichen Instrument, um staatliches Handeln transparenter, wirksamer und wirtschaftlicher zu gestalten. Sie werden benutzt von der Regierung im Rahmen der Aufsicht, von der Verwaltung im Rahmen ihrer Verantwortung für die Mitgestaltung und den Vollzug staatlicher Massnahmen, vom Parlament im Rahmen der Oberaufsicht, aber auch von Parteien, Verbänden und Bürgergruppen.

Zur Erarbeitung aussagekräftiger Evaluationen werden alle gängigen (sozial-)wissenschaftlichen Methoden der Datensammlung und -auswertung sowie Entscheid- und Problemlösungsverfahren verwendet. In neuester Zeit wird zudem vermehrtes Gewicht auf die sachgerechte Umsetzung von Evaluationen gelegt.

Der Hauptbeitrag der Evaluationsforschung im Vergleich zu anderen Instrumenten der Informationsbeschaffung (betriebswirtschaftliche Instrumente, Controlling u.a.m.) liegt vor allem in der Erhellung der Möglichkeiten und Grenzen bei der Ermittlung der Wirkungen staatlicher Massnahmen im sozialen und wirtschaftlichen Bereich. Das Hauptproblem liegt hier in der Isolierung des Einflusses staatlicher Massnahmen von anderen Einflussfaktoren (Konjunktur, Wertewandel u.a.m.) auf das komplexe soziale Geschehen.

→Die Anfangskapitel von Teil 2 geben Ihnen weiteren Aufschluss darüber, was Evaluationen sind, was sie leisten können und kosten. Sie finden aber auch in den nachfolgenden Beispielen Hinweise über die Einsatzmöglichkeiten von Evaluationen. Über die Evaluationsstrategien bei der Ermittlung der Wirkungen staatlicher Massnahmen gibt Ihnen Kap. 3.5 Aufschluss.

Evaluationen sind somit kein Zaubermittel. Sie kosten Geld und brauchen Zeit (vgl. Kap. 2.5). Sie setzen deshalb eine sorgfältige Zeit- und Kostenplanung

voraus. Sie erfordern beim Auftraggeber bzw. bei den Adressaten Sachverstand, tatkräftigen Einsatz, die Fähigkeit zur Konzentration auf das Wesentliche, Lernfähigkeit und Kooperationsbereitschaft. In den nachfolgenden Kapiteln wird gezeigt, welche konkreten Schritte die Voraussetzung für die Durchführung und Nutzung aussagekräftiger Evaluationen bilden¹.

1.2 Die Evaluation planen

Der erste Schritt besteht darin, die Fragestellung zu klären und den Kontext einer möglichen Evaluation näher zu umschreiben. Anders gesagt: der Grobrahmen der Untersuchung ist abzustecken. Konkret soll entschieden werden,

- ob das Problem weiterverfolgt werden soll
- wer in die weiteren Arbeiten einzubeziehen ist
- ob das Problem intern oder durch einen externen Auftrag weiterbearbeitet werden soll
- wozu ein allfälliger Auftrag dienen soll
- welches der zeitliche Rahmen ist
- welche finanziellen und personellen Mittel eingesetzt werden sollen.

1.21 Die Forschungsfrage festlegen

Wenn Sie daran denken, eine (interne oder externe) Evaluation in Auftrag zu geben, dann ist dies meist mit der Hoffnung verbunden, recht umfassende Fragen beantworten zu können, z.B.:

- Zeitigen bestimmte Subventionen nachhaltige Wirkungen?
- Welches sind die Schwachpunkte im Vollzug des Gesetzes?
- Erreichen wir mit der Informationskampagne auch wirklich diejenigen, welche die Informationen am dringendsten benötigen?

Evaluationen können in der Regel nur auf gezielte Fragen fundierte Antworten geben. Sie sind nicht in der Lage, mit vertretbaren Kosten alle Informationsbedürfnisse zu erfüllen. Je nach Wissensbedürfnissen und Fragestellung ergibt sich eine Konzentration auf eine oder mehrere der folgenden Blickrichtungen von Evaluationen:

¹Bei der Formulierung der nachfolgenden Kapitel habe ich bei von Andreas Balthasar erarbeiteten Unterlagen für den Kurs "Einsatz externer Expertinnen und Experten" wesentliche Anleihen gemacht.

1. Evaluation des **Vollzugs** (Vollzugsanalysen, Vollzugskontrollen, Vollzugsstudien, Prozessanalysen): Eine Evaluation des Vollzugs befasst sich mit den Aktivitäten der Verwaltung sowie mit dem damit in Verbindung stehenden Verhalten der Privaten. Oft haben die Vollzugsinstanzen gute Kenntnisse über Aspekte des Vollzugs (persönliche Kontakte, Fachkonferenzen, statistische Daten u.a.m.). Eine Überprüfung des Vollzugs durch ein Evaluationsteam kann indessen in verschiedenen Fällen sinnvoll sein, z.B. im Hinblick auf eine umfassende und zugleich kritische Bestandesaufnahme, zur Vertiefung einzelner bedeutsamer Vollzugsfragen (z.B. Prioritätensetzung beim Einsatz knapper Mittel) oder für einen Quervergleich kantonaler oder kommunaler Vollzugsorganisationen.

Beispiel: Eine schon fast klassische Vollzugsuntersuchung ist die Studie von Delley u.a.(1982) über den Vollzug der damaligen Lex Furgler. Das erstaunliche Resultat der Untersuchungen war, das die Kantone dieses Gesetz sehr innovativ für ihre jeweiligen Bedürfnisse nutzten.

2. Evaluation der **Wirkungen** (Wirkungsanalysen, Wirkungskontrollen, Impact-Analysen): Die Ermittlung der Wirkungen von staatlichen Massnahmen steht heute im Zentrum vieler Bestrebungen, weil sie die Voraussetzung für eine fundierte Bewertung darstellt. Die Frage lautet: Was wäre anders, wenn die Massnahme nicht vollzogen würde? Es geht somit um die Isolierung des Einflusses der Massnahme auf das komplexe soziale und wirtschaftliche Geschehen. Zu beachten ist, dass die Zuverlässigkeit der Evaluation von Wirkungen z.T. von der Möglichkeit sinnvoller Vergleiche abhängt (Kontrollgruppen-, Querschnitt- und Längsschnitt-Vergleiche, vgl. Kap. 3.5). Ausserdem ist es nötig, sich auf bestimmte Wirkungsaspekte zu konzentrieren, um aussagefähige Resultate zu erzielen.

Beispiel: Hanser/Kuster/Cavelti 1994 untersuchten, welche Auswirkungen die Hotellerieförderung auf die Investitionstätigkeit und das Betriebsergebnis der geförderten Hotels hatte. Sie rekonstruierten den jeweiligen Investitionsentscheid, werteten Daten über die Situation der Hotels vor und nach der Investition aus und verglichen auch mit Hotels, welche die Investition ohne staatliche Hilfe finanzierten bzw. lange keine Investitionen vornahmen.

3. Evaluation der **Wirtschaftlichkeit** (Wirtschaftlichkeitsanalysen, Effizienzanalysen): Solche Studien dienen der Beantwortung der Frage, ob der Output oder die Wirkungen auch mit weniger Ressourcen erreicht werden können bzw. ob der Aufwand (Kosten) für eine bestimmte Massnahme durch die Wirkungen gerechtfertigt wird. Wirtschaftlichkeitsanalysen, ob impressionistisch oder systematisch durchgeführt, entsprechen damit einem wachsenden Bedürfnis der Öffentlichkeit nach Transparenz über Nutzen und

Kosten staatlichen Handelns. Die Durchführung eigentlicher Kosten-Nutzen-Analysen erfordert eine monetäre Bewertung von Wirkungen und setzt eingehende Sachkenntnisse voraus. Etwas geringere Anforderungen an die monetäre Quantifizierung stellen Kosten-Effektivitäts-Analysen. Alle diese Studien können prospektiv vor der Ergreifung einer Massnahme bzw. der Tatigung einer Investition oder retrospektiv zur Einleitung allfalliger Korrekturen vorgenommen werden.

Beispiel: In einer Untersuchung ber die Schneeraumung der Urner Passstrassen (Ecoplan 1993b) wurden die Kostenvorteile einer zeitlich verspateten ffnung der Passstrassen den daraus resultierenden Nachteilen (Auswirkungen auf das Tourismusgewerbe) gegenbergestellt. Es wurde empfohlen, die Passffnung zeitlich leicht nach hinten zu verschieben.

4. **Monitoring (Vollzugskontrolle und -steuerung):** Das Monitoring dient dazu, permanente Informationen ber Aspekte des Vollzugs und der Wirkungen bereitzustellen. Die ermittelten Daten knnen der berwachung des Vollzugs dienen, z.B. durch Vergleich der Ist- und der Soll-Werte oder durch den Vergleich mit Zahlen aus anderen Kantonen. Das Monitoring kann sich auch auf zielrelevante Grssen beziehen und Einblick in deren Entwicklung geben. Hierbei sind die Bemhungen zur Erarbeitung von **Leistungsindikatoren** zu erwahnen, welche Aufschluss ber den Output der Verwaltung und die erzielten Wirkungen geben sollen. Meist kann aufgrund von Leistungsindikatoren aber nicht nachgewiesen werden, dass das Resultat ausschliesslich auf Grund der staatlichen Aktivitaten erzielt worden ist. Beim Monitoring und bei Leistungsindikatoren handelt es sich um punktuelle, aber permanente Instrumente, wahrend Evaluationen der Wirkungen vertiefte, aber oft auf einen bestimmten Zeitpunkt bezogene Ergebnisse bringen. Beides sind komplementare Instrumente.

Beispiel: Die im Rahmen der sog. Praventivmassnahmen gegen die Arbeitslosigkeit angebotenen Mglichkeiten (v.a. Weiterbildungskurse) dienen dazu, die Arbeitschancen von Arbeitslosen zu verbessern. Inwieweit erreichen die Kurse ihre Ziel kurz-, mittel- und langfristig? In einer eingehenden Studie zeigte Freiburghaus (1987), wie eine permanente Erfolgskontrolle der Praventivmassnahmen aufgezogen werden kann. Fur den Erfolg solcher Weiterbildungsmassnahmen wichtige Indikatoren (Besuch der Kurse, Einschatzung nach Abschluss der Kurse u.a.m.) sollten erfasst werden.

5. Evaluation der **mutmasslichen knftigen Wirkungen** einer vorgesehenen Massnahme (ex ante Evaluation, prospektive Evaluation): Diese Perspektive steht bei der Ergreifung neuer oder der Revision bestehender Massnahmen im Zentrum. Es gibt sehr unterschiedliche Methoden zur Beantwortung prospektiver Fragen: impressionistische Abschatzung mglicher Wir

kungsabläufe, Erarbeitung von Szenarien (z.B. unter Mithilfe der Delphi-Methode), Beurteilung der Massnahmenentwürfe unter unterschiedlichen Rahmenbedingungen, Erarbeitung quantitativer Modelle und Simulation der Wirkungen. Die Mithilfe eines Evaluationsteams kann verschiedenen Zwecken dienen: Hilfestellung bei der Zukunftsexploration (z.B. Durchführung einer Delphi-Umfrage über mögliche Szenarien der Zukunftsentwicklung), Erarbeitung fundierter Hypothesen über das Wirkungsgefüge von Massnahmen, Erarbeiten von Vorschlägen zur besseren Evaluierbarkeit vorgesehener Massnahmen u.a.m.

Beispiel: Nach dem Reaktorunfall in Tschernobyl mehrten sich die Stimmen, welche einen Ausstieg aus der Kernenergie forderten. Der Bundesrat beauftragte eine Arbeitsgruppe mit der Ausarbeitung von Energieszenarien, um die Möglichkeiten und Konsequenzen eines Ausstiegs der Schweiz aus der Kernenergie zu untersuchen. Die Arbeitsgruppe prüfte insgesamt 18 Szenarien, wovon 5 besondere Bedeutung hatten. Meinungsverschiedenheiten innerhalb der Arbeitsgruppe ergaben sich namentlich im Hinblick auf die Auswirkungen einer drastischen Energiesparpolitik auf den Komfort und auf die volkswirtschaftlichen Auswirkungen eines Ausstiegs (Bericht der Expertengruppe Energieszenarien 1988).

Die Festlegung der Forschungsfrage ist ein **Angelpunkt** jeder Evaluation. Dabei wird ein politisches oder administratives Problem in eine bearbeitbare Forschungsfrage umformuliert.

Je besser Sie Ihre Frage einengen und spezifizieren können und je klarer Ihre Vorstellungen über den Beitrag des Evaluationsteams bei der Lösung Ihrer Probleme sind, desto präziser und kostengünstiger wird die Evaluation ausfallen. Obgleich es Aufgabe eines kompetenten Evaluationsteams ist, mit Ihnen gemeinsam eine sinnvolle Fragestellung für die Evaluation zu entwickeln, wird die Vorbereitung erleichtert, wenn Sie vorgängig möglichst realistische Vorstellungen entwickeln. Wie eine anfänglich recht breite Fragestellung eingengt und präzisiert werden kann, wird am nachfolgenden Beispiel gezeigt.

Der für die Wirtschaftsförderung zuständige Regierungsrat des Kantons X möchte wissen, ob die kantonale Wirtschaftsförderungspolitik "gut funktioniert". Nach mehreren Besprechungen mit den für den Vollzug zuständigen Beamten und einem informell beigezogenen Evaluationsexperten werden folgende mögliche Untersuchungsfragen herauskristallisiert:

- ◆ *Sind die Dienstleistungen der kantonalen Wirtschaftsförderungsstelle bei den Klein- und Mittelbetrieben ausreichend bekannt?*
- ◆ *Wie kann die Zusammenarbeit unter den mit der Wirtschaftsförderung befassten kantonalen Stellen verbessert werden?*

High Tech Ist es sinnvoll, prioritär High Tech-Betriebe zu unterstützen?

Nach Gesprächen mit dem Regierungsrat stellt sich heraus, dass die beiden letzten Fragen als prioritär erachtet werden. Die Frage der Koordination wird in Zusammenarbeit mit Organisationsexperten des Personalamts intern bearbeitet; für die Frage der High Tech-Betriebe werden anschliessend Offerten von in Regional- und Wirtschaftsfragen bewanderten Evaluationsteams eingeholt.

Welche der ursprünglich gestellten und breit gefassten Fragen durch eine Evaluation sinnvoll beantwortet werden können, muss in jedem einzelnen Fall abgeklärt und später auch in die Verhandlungen mit dem Evaluationsteam einbezogen werden.

1.22 Den Kreis der Beteiligten grob abgrenzen

An einer Evaluation sind meist mehrere Dienststellen und Vorgesetzte interessiert. Es gilt deshalb, den Kreis der von der Evaluation Betroffenen grob abzugrenzen. Ist nur Ihre Dienststelle impliziert? Haben andere Dienststellen Interesse an den Fragestellungen oder ist gar deren Beteiligung im Hinblick auf eine erfolgreiche Evaluation nötig? Ist es sinnvoll, weitere Betroffene (Verbände, Parteien u.a.m.) bereits jetzt einzubeziehen? Wer muss einen allfälligen Evaluationsauftrag formell unterzeichnen (eine Vorabklärung über die Vorstellungen dieser Person ist in der Regel sinnvoll)? Welche Personen und Institutionen haben ein Interesse an einer Evaluation? Von wem ist Unterstützung bzw. sind allenfalls Widerstände zu erwarten?

1.23 Interne oder externe Evaluation?

Wir gehen in diesem Leitfaden davon aus, dass Sie eine Evaluation ausserhalb ihrer eigenen Dienststelle in Auftrag geben, sei es an eine verwaltungsnahe Anstalt (z.B. landwirtschaftliche Forschungsanstalt), sei es an eine unabhängige Stelle (Universitätsinstitut, privates Büro). Das bedeutet nicht, dass wir die Präferenz auf eine externe Evaluation legen. Wie aus den nachfolgenden Bemerkungen ersichtlich sein wird, gibt es viele gute Gründe, eine Evaluation an eine verwaltungsinterne Stelle zu vergeben oder sie gar selbst durchzuführen. In letzteren beiden Fällen verfügt die Dienststelle in der Regel aber bereits über etablierte Forschungskontakte bzw. allenfalls sogar über ein Know How in Evaluationsfragen. Vor allem Personen und Dienststellen, welche keine regelmässigen Forschungskontakte haben, dürften den grössten Be

darf nach einem Leitfaden wie dem vorliegenden haben, weshalb wir den Inhalt auf diesen hypothetischen Fall ausgerichtet haben.

Bei der Frage nach der Zweckmässigkeit interner oder externer Evaluation geht es nicht um ein Entweder-Oder, sondern um die optimale Kombination beider Möglichkeiten. Es sind komplementäre und nicht sich ausschliessende Vorgehensweisen. Dies gilt es im Auge zu behalten, wenn nachfolgend einige Vor- und Nachteile interner und externer Evaluation beschrieben werden.

Interne Evaluation wird in gewissem Umfang von jeder Dienststelle betrieben. Es geht um die Darstellung und Beurteilung der eigenen Tätigkeiten und deren Wirkungen. Die Wirkungserfassung erfolgt meist in intuitiver Form, ohne dass staatliches Handeln und die Veränderung der Zielgrössen in einem gesicherten Zusammenhang zueinander stehen. Das Schwergewicht interner Evaluation liegt meist bei der Evaluation des Vollzugs. Sofern in der Dienststelle selbst oder in einer Forschungsabteilung der entsprechenden Institution wissenschaftlicher Sachverstand vorhanden ist, ist allenfalls auch eine interne Evaluation der Wirkungen möglich.

- Vorteile

Eine interne Evaluation kann in der Regel auf gutem Sachverstand und auf der Kenntnis des Politikbereichs und der relevanten Wissenschaftsdisziplin aufbauen. Es bestehen gute persönliche Beziehungen zu Verwaltungsstellen, Vollzugsinstanzen und Adressaten der Politik. Die verwaltungsinternen Daten und Statistiken sind bekannt. Während der Evaluation können laufend Zwischenergebnisse übermittelt werden. Interne Evaluationsteams verfügen innerhalb der Verwaltung meist über grosse Glaubwürdigkeit, und durch die Ergebnisse und Empfehlungen anvisierte Dienststellen werden recht gerne geneigt sein, diese umzusetzen.

- Nachteile

Interne Evaluation kann mit Betriebsblindheit verbunden sein. In methodischer Hinsicht bringen nicht alle internen Forschungsstellen die für eine Evaluation der Wirkungen notwendigen Kenntnisse mit. Die Glaubwürdigkeit der Ergebnisse und der Empfehlungen für verwaltungsexterne Institutionen und Personen (Politiker, Verbände, Parteien u.a.m.) ist in der Regel nicht sehr gross.

Bei einer **externen Evaluation** haben Sie die Möglichkeit, zwischen Universitätsinstituten und privaten Büros auszuwählen. Trotz Unterschieden im Profil solcher Institutionen (vgl. AGREB 1990, S. 13) stellen wir die Vor- und Nachteile gemeinsam dar.

- Vorteile

Externe Evaluationsstellen verfügen in der Regel über gute Kenntnisse des

Vorgehens (Evaluationsstrategien, Forschungstechniken). Sie haben Erfahrung aus dem gleichen oder weiteren Politikbereichen. Z.T. bringen sie eine breite Palette von Fähigkeiten (wissenschaftliche Kenntnisse, Beratungspraxis) mit. Sie verfügen über Distanz zum Forschungsgegenstand und können neue Perspektiven und Ideen einbringen. Dies erhöht die Glaubwürdigkeit der Resultate gegenüber aussenstehenden Personen.

- Nachteile

Externe Evaluationsteams kennen den Untersuchungsgegenstand oft weniger gut als die Verwaltungspraktiker und brauchen z.T. eine gewisse Anlaufzeit. Sie sind mit den konkreten Anliegen der Verwaltung nicht unbedingt vertraut, und es bedarf vertiefter Anstrengungen, damit die Resultate bedürfnisgerecht ausfallen.

Keine ausgeprägten Unterschiede bestehen nach unserer Meinung bezüglich Kosten und Objektivität der Untersuchung. Externe Evaluationen sind in der Regel mit vergleichsweise höheren ausgewiesenen Budgetkosten verbunden, da verschiedene Kosten interner Evaluation (Personalkosten, Büroaufwand, Ausbildungskosten) verdeckt bleiben. Sie sind von den Gesamtkosten her betrachtet aber nicht notwendigerweise teurer. Umgekehrt sind interne Evaluationen nicht notwendigerweise schönfärberischer bzw. dem Auftraggeber gegenüber unkritisch. Es gibt externe Evaluationsteams, die auf ihren Ruf als unabhängige und seriöse Auftragnehmer bedacht sind, aber auch solche, die im Hinblick auf die Vergabe weiterer Forschungsaufträge an einem guten Einvernehmen mit dem Auftraggeber interessiert sind.

Interne und externe Evaluation sind von Fall zu Fall in geeigneter Weise zu kombinieren. Zur Illustration zwei Beispiele:

Beispiel 1: Eine Dienststelle möchte wissen, wie sie geeignete Vollzugsdaten bereitstellen kann, um die Wirkungen einer bestimmten Massnahme zu erfassen. Sie gibt bei einem in Evaluationsfragen bewanderten Wissenschaftler eine Evaluierbarkeitsprüfung in Auftrag. Im Gutachten, welches dem Amt erstattet wird, werden die Ziele und die Wirkungsmechanismen der untersuchten Massnahme detailliert behandelt, und es werden Wege aufgezeigt, um die Wirkungen permanent zu erfassen (Monitoring, Leistungsindikatoren, Hinweise für spätere Evaluationen der Wirkung). Das Amt beschliesst, bei den Vollzugsinstanzen und den Adressaten der Massnahme verschiedene Daten zu erheben, die für die Beurteilung der Wirksamkeit nützlich sind. Verschiedene weiter gehende Empfehlungen zur Datenerhebung können auf Grund der Personalknappheit nicht verwirklicht werden (Studie Freiburghaus 1987).

Beispiel 2: Eine Evaluation des Vollzugs und der Wirkungen wird von einem privaten Büro durchgeführt. Die Evaluation betrifft auch eine halbprivate Institution, welche den Grossteil des Vollzugs sicherstellt. Die für die

Wirkungsanalyse notwendigen statistischen Grundlagen werden von dieser Institution selbst bereitgestellt. Sie wird über die Zwischenergebnisse laufend orientiert und setzt verschiedene Vorschläge bereits um, noch bevor der Schlussbericht der Öffentlichkeit vorgestellt wird.

1.24 Den Verwendungszweck der Evaluation bestimmen

Es lohnt sich, früh Überlegungen über die mögliche Verwendung bzw. Nutzung der Evaluation anzustellen. Gewisse Hinweise hat Ihnen vielleicht bereits die Abgrenzung des Kreises der Beteiligten gegeben. Sie erkennen, wer Interesse an der Evaluation und deren Ergebnissen hat und wie die entsprechende Person oder Institution allfällige Ergebnisse verwenden könnte.

Die sachgerechte Nutzung ist der Angelpunkt jeder Evaluation. Die Verantwortlichen für eine Massnahme (Leiter der Dienststelle, zuständiger Departementsvorsteher/in, interessierte Parlamentskommissionen u.a.m.) sind möglichst früh über die bevorstehende Evaluation zu informieren, und es ist angezeigt, dass sie sich über für die Entscheidung relevanten Faktoren äussern. Falls dies nicht schon geschehen ist, sind bei der Planung des finanziellen und personellen Aufwands die für eine sachgerechte Nutzung sinnvollen Schritte (z.B. adressatenspezifische Kurzfassungen oder mündliche Präsentationen der Ergebnisse) zu berücksichtigen.

Bereits jetzt ist möglicherweise abschätzbar, ob die Evaluation "bloss" im Sinne der Rechenschaftslegung oder gezielt, beispielsweise für Kostensenkungen oder Massnahmenänderungen eingesetzt werden soll.

→Hinweise über die möglichen Verwendungszwecke von Evaluationen gibt Ihnen Kap. 1.5.

Im Zusammenhang mit dem Verwendungszweck stellt sich auch die Frage nach der Publikation der Evaluationsstudie. Wem sollen die Resultate zur Kenntnis gebracht werden? Der Dienststelle und ihren Vorgesetzten, weiteren Beteiligten, einer breiten Öffentlichkeit? Namentlich um Lernprozesse zu ermöglichen (vgl. Kap. 1.56) plädieren wir für grösstmögliche Transparenz und Offenheit im Bereich von Evaluationen. Dennoch wird es manchmal angebracht sein, Evaluationen als engen Beratungsprozess auszugestalten und weitere Beteiligte sowie die Öffentlichkeit nicht über die Resultate zu informieren. Dieses Vorgehen ist insbesondere dann angebracht, wenn eine Evaluation auch die Organisationsentwicklung sowie die Verbesserung der Führungs- und Personalstrukturen zum Inhalt hat.

1.25 Den zeitlichen Rahmen festlegen und die erforderlichen finanziellen und personellen Mittel abschätzen

Aus dem Verwendungszweck ergeben sich meist auch die zeitlichen Anforderungen an den Evaluationsprozess. Es gilt festzulegen, wann die Zwischen- und die Endergebnisse vorliegen sollen.

Ausserdem ist der finanzielle und personelle Grobrahmen zu bestimmen. Wenn Sie die Evaluation intern durchführen, müssen Sie das für die Durchführung der Evaluation erforderliche Personal freistellen. Bei einer externen Evaluation müssen Sie Budgetmittel reservieren und Personal für die Begleitung und Nutzung der Evaluationsstudie vorsehen.

→Es ist wichtig, dass Sie von realistischen Annahmen bezüglich Zeit und Kosten ausgehen. Kap. 2.5 gibt Ihnen entsprechende Hinweise. Angaben über den Zeitaufwand für die Begleitung und Nutzung finden Sie in Kap. 1.4.

1.26 Über die Weiterbearbeitung der Evaluation entscheiden

Eine Evaluation ist nur dann sinnvoll, wenn der davon erwartete Nutzen die dafür voraussichtlich einzusetzenden personellen und finanziellen Ressourcen übersteigt.

In vielen Fällen sind auch Varianten des Vorgehens möglich und es gilt, die voraussichtlich beste Variante auszuwählen.

Beispiel: Das BUWAL wollte wissen, ob die Vorschriften des Natur- und Landschaftsschutzes bei skisportbedingten Landschaftseingriffen eingehalten werden. Es fasste als erste Variante ein Umwelt-Monitoring ins Auge und zwar in Form einer Aufnahme und Interpretation von Luftbildern zahlreicher Skigebiete, um die Entwicklung im Zeitablauf zu verfolgen. Die zweite Variante bestand darin, eine bestimmte Anzahl Skigebiete sehr intensiv und aus unterschiedlichen Perspektiven (naturwissenschaftlich, juristisch, soziologisch und ökonomisch) zu untersuchen, um das Gewicht von Umweltschutz-Auflagen im Geflecht aller wichtiger Interessen zu beobachten (Das BUWAL entschied sich für die erste Variante; das NFP 27 finanzierte später die zweite Variante, vgl. Hirsch et al. 1995).

Weiterführende Literatur

Weitere Hinweise geben vor allem der Leitfaden für das Management von Forschungsprojekten der Bundesverwaltung (AGREB 1990) und die dazugehörigen Anhänge.

1.3 Offerten einholen und den Auftrag vergeben

Wenn Sie sich über den Grobrahmen und die Hauptfragen der Evaluation klar sind, können Sie die Auftragsvergabe vorbereiten². Auch wenn Sie mit einem Universitätsinstitut oder einem Evaluationsteam/Beratungsbüro eine gute und enge Forschungszusammenarbeit pflegen, empfehlen wir Ihnen, durch ein offenes oder ein Einladungsverfahren Konkurrenzofferten einzuholen, und zwar aus folgenden Gründen:

- Aus Konkurrenzofferten können sich Hinweise auf neue und vielversprechende Fragestellungen und Evaluationsstrategien ergeben.
- Sie haben v.a. im Hinblick auf die Kosten eine bessere Verhandlungsposition.
- Sie können Ihre Vergabe im Hinblick auf spätere Kontrollen fundierter rechtfertigen.
- Möglicherweise kann das Projekt am besten durch Zusammenarbeit verschiedener Evaluationsteams bzw. entsprechende Teilaufträge bearbeitet werden.

Nach einer groben Vorauswahl auf zwei bis drei mögliche Auftragnehmer ist es meist sinnvoll, die in den Offerten unterbreiteten Vorschläge gemeinsam mit den entsprechenden Evaluationsteams zu besprechen. Dabei können Verständnisfragen beantwortet sowie Fragestellung, Evaluationsstrategie und Forschungstechnik weiter konkretisiert werden.

1.31 Soll eine Evaluierbarkeitsprüfung in Auftrag gegeben werden?

Bei grösseren Evaluationsvorhaben (über 200'000 Franken) kann es sinnvoll sein, eine Evaluierbarkeitsprüfung in Auftrag zu geben (vgl. auch Balthasar 1993, 6f.). Dies sind Abklärungen und Überlegungen zur Planung und Durchführbarkeit der Untersuchung, folglich Machbarkeitsstudien für Evaluationen. Sie haben namentlich Aufschluss zu geben über die zu evaluierende

²Wir gehen zwar vom Fall einer externen Evaluation aus; einige der nachfolgend erwähnten Schritte können aber auch bei einer internen Evaluation sinnvoll sein.

Massnahme, über die der Evaluation zugrundezulegenden Kriterien, über mögliche Evaluationsstrategien, über mögliche Forschungstechniken und über die Datenlage. Im Rahmen einer Evaluierbarkeitsprüfung können beispielsweise folgende Fragen angegangen werden: Ist die zu evaluierende Massnahme genügend abgrenzbar? Auf welche Ziele (gesetzlich formulierte Ziele, Verlautbarungen von Verbänden, Bedürfnisse der Zielgruppen u.a.m.) könnte sich die Evaluation abstützen? Gibt es geeignete Informationsquellen über den Vollzug oder die Wirkungen? Gibt es Widerstände gegen eine Evaluation? Welches soll die Rolle des Evaluationsteams sein (Informationsbeschaffer, Berater/Ratgeber, Schiedsrichter u.ä.)?

Evaluierbarkeitsprüfungen erhöhen die Chance, dass die Evaluation zu einem Erfolg wird. Sie können auch zum Ergebnis führen, dass die gestellten Fragen nicht oder nur mit sehr hohem finanziellem und zeitlichem Aufwand beantwortet werden können. Eine Evaluierbarkeitsprüfung kann einer Dienststelle allenfalls auch Hinweise geben, wie sie geeignete Informationen über die Wirksamkeit der von ihr durchgeführten Massnahmen **selbst** ermitteln kann, und eine Alternative zu einer eigentlichen Evaluation bilden.

1.32 Ein Pflichtenheft erarbeiten und Offerten einholen

Die Offertstellung wird erleichtert und das Ergebnis verbessert, wenn Sie möglichst präzise Angaben über den Auftrag liefern. Sie müssen das Ziel und Ecksteine vorgeben; an den Offertstellern liegt es, den Weg anzugeben, auf welchem die Vorgaben erreicht werden sollen.

Deshalb könnte ein **Pflichtenheft**, das Sie den Interessierten zustellen, Angaben zu folgenden Punkten enthalten:

- Anlass, Problem
- Fragestellung
- Verwendungszweck der Ergebnisse
- Zu evaluierender Sachverhalt, zu untersuchende staatliche Massnahme
- Spezifische Fragen, die zu bearbeiten sind
- Allfällige Vorarbeiten, Vorstudien
- Auftraggeber, Zuständige Person/Auskunftsperson
- Weitere an der Evaluation beteiligte Dienststellen
- Qualifikationsanforderungen an das Evaluationsteam
- Zeitliche Vorgaben für Zwischen- und Schlussergebnisse
- Form der Zwischen- und Schlussprodukte (mündliche, schriftliche Orientierungen)
- Kostenrahmen, allfällige Vorgaben für die Honoraransätze
- Datum für die Einreichung der Offerten, erwarteter Umfang

1.33 Die Offerten überprüfen

Bevor Sie unter den Offerten eine Auswahl treffen, werden Sie diese auf allfällige Lücken überprüfen. Daraus können Sie ermitteln, ob die Offerte auf alle bedeutsamen Fragen Antwort gibt. Die nachfolgende Checklist enthält Punkte, die auf Grund der Erfahrungen des NFP 27 vor der Inangriffnahme der Evaluation geklärt werden sollten. Je nach Fragestellung werden indessen weitere Punkte zu regeln sein bzw. sind einzelne der in der Checklist erwähnten Fragen nicht relevant.

Checklist für die Überprüfung der Offerten auf Vollständigkeit

Besteht Klarheit darüber,

- welche Massnahme evaluiert wird (z.B. ganzer Erlass, bestimmte rechtliche Regelung, bestimmte Beiträge u.a.m.)?
- auf welche Fragen (Vollzug, Wirkungen, Monitoring, weiteres) sich die Evaluation konzentriert?
- nach welchen Kriterien (Effektivität, Wirksamkeit, Wirtschaftlichkeit) die Massnahme beurteilt werden soll?
- worauf sich die Zielanalyse abstützen wird (rechtlich formulierte Ziele, informelle Ziele/Verlautbarungen von Aktoren, Bedürfnisse der Zielgruppe u.a.m.)?
- welches Wirkungsmodell der Evaluation zugrundegelegt wird?
- welche Evaluationsstrategien verwendet werden?
- welche Forschungstechniken verwendet werden?
- welche Personen die Evaluation durchführen werden?
- welche Arbeitsschritte wann (ev. durch wen) unternommen werden?
- wer die Arbeit des Evaluationsteams begleiten wird?
- wann von Ihnen oder anderen Dienststellen Inputs (z.B. statistische Unterlagen, Kommentare, Schreiben an von der Evaluation Betroffene u.a.m.) erwartet werden?
- ob der Datenzugang gesichert ist?
- woraus die Zwischen- und Endprodukte bestehen (Berichte und deren Umfang, mündliche Präsentationen, Übersetzungen von Berichten u.a.m.)?
- wann Sie diese Zwischen- und Endprodukte erhalten werden?
- ob zusätzliche Umsetzungsmassnahmen (z.B. Durchführung von Seminaren, Präsentationen bei Entscheidungsinstanzen, Zeitschriftenartikel) vorgesehen sind?
- Wie viel die Evaluation (zu welchem Preisstand) kosten wird und aus welchen Bestandteilen sich die Kosten zusammensetzen?

1.34 Auswählen

Sie sind nun in der Lage, eine fundierte Entscheidung zu treffen. Die dafür massgebenden Kriterien hängen vom konkreten Evaluationsprojekt ab. Wir geben zuerst einige Hinweise auf nach unseren Erfahrungen wichtige Faktoren für die Auswahl und bieten anschliessend eine Checklist für die Beurteilung der Offerten an.

Inhaltliche Faktoren

- ◆ **Robuste Evaluationsstrategie:** Hier gilt es sicherzustellen, dass Schwierigkeiten bei der Realisierung einzelner Forschungsschritte nicht das ganze Projekt gefährden. Dies kann am besten durch eine kombinierte Evaluationsstrategie und durch Methodenmix geschehen.
- ◆ **Flexibilität des Vorgehens:** Wenn im Rahmen einer Evaluation neues Territorium begangen wird, sind Überraschungen (z.B. bezüglich der Verfügbarkeit von Informationen) kaum auszuschliessen. In solchen Fällen ist es von Vorteil, wenn die Untersuchungsanlage noch Änderungen ermöglicht. Ein flexibles Vorgehen ist beispielsweise bei Fallstudien und bei mündlichen Interviews eher möglich als bei einer breit angelegten schriftlichen Befragung oder bei sehr aufwändigen statistischen Auswertungen.
- ◆ **Abklärung der Datenlage:** Es gilt sorgfältig abzuklären, ob die erforderlichen Daten vorhanden sind und (z.B. EDV-gestützt) ausgewertet werden können. Die Beschränkung auf eine Evaluationsstrategie oder auf bestimmte Forschungstechniken setzt eine vergleichsweise eingehendere Abklärung der Datenlage voraus. Dies kann es z.B. erforderlich machen, dass bei einer Umfrage in heiklen Gebieten (z.B. finanzielle Verhältnisse) Pilottests durchgeführt, dass bei sekundärstatistischen Auswertungen die Voraussetzungen für eine EDV-gestützte Auswertung geprüft oder bei einem Querschnitt-Vergleich die Gegebenheiten (Grobanalyse der Massnahmen, Vergleichbarkeit des Umfelds) grob abgeklärt werden.
- ◆ **Umsetzungsmassnahmen:** Sie können ein massgeblicher Kostenfaktor sein, dessen Bedeutung bei der Auftragsvergabe z.T. unterschätzt wird. Eine Offerte, welche keine Umsetzungsmassnahmen enthält, kann auf den ersten Blick vorteilhafter erscheinen als ein Angebot, welches die entsprechenden Massnahmen (samt Kosten!) einplant; sie wird im Endergebnis aber möglicherweise teurer sein. Vor allem wenn Sie in der Verwaltung Lernprozesse auslösen wollen (vgl. Kap. 1.56), ist der Umsetzung als einer Daueraufgabe ausreichend Beachtung zu schenken. Ein partizipatives Vorgehen, welches Anliegen der Auftragnehmer laufend in den Evaluationsprozess einzubeziehen versucht und gleichzeitig ständige Rückmeldungen ermöglicht, kann Lernprozessen durchaus förderlich sein (braucht in

dessen meist auch Zeit und verursacht Kosten). Die Umsetzungsmassnahmen können bereits vor der Inangriffnahme festgelegt werden. Wenn über ihre Ausgestaltung noch zu viele Unsicherheiten bestehen, können sie aber auch zu einem späteren Zeitpunkt bestimmt werden. In letzterem Falle sollte für die Umsetzung eine nicht zu knapp bemessene Reserve eingestellt werden, damit Ihnen später nicht die Budgetmittel fehlen, um die Erkenntnisse aus der Evaluationsstudie in- und ausserhalb der Verwaltung auf sinnvolle Art und Weise zu verbreiten.

Personelle Faktoren

- **Qualität des Evaluationsteams:** Sie kann für den Erfolg der Evaluation von ausschlaggebender Bedeutung sein. Die Qualität des Evaluationsteams beruht auf vier Faktoren, deren Gewichtung Ihnen überlassen bleibt.
 - Das Evaluationsteam sollte mit der entsprechenden Evaluationsstrategie und den anzuwendenden Forschungstechniken vertraut sein. Dies betrifft nicht nur für den Laien schwer verständliche Techniken wie beispielsweise multivariate Auswertungen oder ökonomische Verfahren. Auch scheinbar (!) einfachere Techniken wie Umfragen oder Tiefeninterviews setzen Sachverstand und Erfahrung voraus, wenn sie zu aussagekräftigen Resultaten führen sollen.
 - Das Evaluationsteam sollte gute Kenntnisse des zu untersuchenden Sachgebiets (z.B. Landwirtschaftspolitik, Soziale Sicherheit, Technologiepolitik) bzw. der massgebenden Disziplin (Ökonomie, Soziologie, Politikwissenschaft u.a.m.) aufweisen. Dies erlaubt ein rasches und zielgerichtetes Vorgehen und erleichtert den Zugang des Evaluationsteams zu Daten (z.B. Bereitschaft von Schlüsselpersonen für Interviews). Ausserdem sind die Resultate eines Evaluationsteams, das zu den Kennern eines Sachgebiets zählt, in der Regel glaubwürdiger.
 - Wenn mehrere selbständige Personen bzw. Evaluationsteams am Projekt teilnehmen, sollte für eine gute Zusammenarbeit (z.B. durch bereits bestehende Formen der Zusammenarbeit) Gewähr bestehen.
 - Das Evaluationsteam sollte, last but not least, über gute kommunikative Fähigkeiten verfügen (Kontakte mit Ihnen, mit den durch die Evaluation betroffenen Personen, schriftliche und mündliche Präsentationen).

Für die Beurteilung der Offerten im Hinblick auf die Auswahl bietet Ihnen die nachfolgende Checklist Hilfe. Es wird an Ihnen liegen, einzelne Faktoren stärker bzw. weniger stark zu gewichten. Wir empfehlen Ihnen, die Offerten nüchtern zu beurteilen. Lassen Sie sich nicht durch schöne Fachausdrücke oder

ausgefeilte Graphiken blenden. Es ist wichtig, dass Sie das vorgeschlagene Vorgehen verstehen. Seien Sie vorsichtig, wenn Ihnen das Evaluationsteam auf alle Fragen Antworten verspricht. Es könnte sein, dass es sich der Tragweite Ihrer Fragen gar nicht bewusst ist und beispielsweise Fragen nach den Wirkungen einer Massnahme durch ein blosses Auflisten der Zielindikatoren beantworten zu können glaubt. Seien Sie ebenfalls vorsichtig, wenn Ihnen das Evaluationsteam Antworten auf Fragen, die Sie gar nicht gestellt haben, versprechen möchte. Es ist möglich, dass das Team mit anderen Sachfragen oder mit anderen methodischen Instrumenten vertraut ist, als im gegebenen Fall nötig wäre.

→ Wenn Sie eine Evaluation der Wirkungen in Auftrag geben möchten und Sie die eingereichten Offerten nur teilweise überzeugen, empfehlen wir Ihnen, Kap. 3.5 anzuschauen.

Checklist für die Offertenbeurteilung

OffertstellerIn:.....

Beurteilungskriterien	Ja/Nein	Kommentar
Inhaltliche Kriterien		
Die Aufgabenstellung ist gut erfasst worden.	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
Auf alle Fragen wird eingegangen.	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
Es wird nachvollziehbar dargelegt, wie die Fragen angegangen werden.	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
Das Vorgehen und der gewählte methodische Ansatz überzeugen.	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
Bei Wirkungsanalysen: Evaluationsstrategie und Forschungsmethoden sind geeignet zur Erfassung der Wirkungen der Massnahmen.	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
Der Zeitplan scheint realistisch, es gibt keine Flaschenhälse.	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
Das Projekt ist im Hinblick auf Zwischen- und Schlussergebnisse sinnvoll etappiert.	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
Die Vorschläge zur Umsetzung sind sinnvoll.	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
Personell		
Das Evaluationsteam hat ausreichende methodische Kenntnisse (über Evaluationsstrategien, Forschungstechniken) sowie ausreichendes Fach-(z.B. ökonomische Kenntnisse) und Sachwissen.	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
Die vorgeschlagene Projektorganisation scheint zur Erledigung des Auftrags geeignet.	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
Die Form der Zusammenarbeit zwischen Auftraggeber und -nehmer ist eindeutig definiert.	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
Gegenüber dem Evaluationsteam sind keine gewichtigen Widerstände zu erwarten.	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
Finanzielle Kriterien		
Der vorgeschlagene Kostenrahmen ist für die vorgesehenen Arbeiten angemessen.	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
Die vorgeschlagenen Abrechnungsmodalitäten erlauben eine laufende Kontrolle des Preis- und Leistungsverhältnisses.	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
Allfällige Vorgaben betreffend die Ansätze werden eingehalten.	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	

Weitere Bemerkungen

Quelle: Andreas Balthasar, Unterlagen für den Kurs "Einsatz externer Expertinnen und Experten", im Rahmen des Weiterbildungsprogramms "Gemeinde und Stadtverwaltung im Wandel" des Forschungszentrums für schweizerische Politik (mit kleineren Ergänzungen des Autors)

1.35 Auftragsvergabe als Verhandlung

Es ist selten, dass Sie unter den Offerten, welche Ihnen unterbreitet werden, das perfekt auf Sie zugeschnittene Angebot finden. In der Regel ist vor der Auftragsvergabe eine Aushandlungsphase erforderlich, in welcher Sie Ihre Wünsche und Vorstellungen weiter konkretisieren und umgekehrt das Evaluationsteam seine methodischen Optionen prazisiert. Diese Phase kann fur den Erfolg der Evaluation recht wichtig sein. Es gelingt moglicherweise, das Vorgehen, den Zeitplan und die Kosten noch besser auf Ihre Bedurfnisse zuzuschneiden. Indem Sie bewusst bestimmte methodische Vorgehensweisen ausscheiden (z.B. aus Grunden der Zuverlassigkeit, aus Zeit- oder Kostengrunden), werden Sie das gewahlte Vorgehen und die Ergebnisse umso besser akzeptieren konnen. Die Verhandlungsphase erlaubt Ihnen auch abzuschatzen, ob und inwieweit das Evaluationsteam methodisch versiert ist und Ihre kritischen Fragen uber die Evaluationsstrategie und die Forschungsmethoden beantworten kann.

Sie werden am Ende des Aushandlungsprozesses wahrscheinlich feststellen, dass es bei Evaluationen keine "idealen" oder "einzig richtigen" Wege gibt. Die Festlegung eines Evaluationsauftrags ist vielmehr ein Optimierungsprozess, in welchem zwischen Informations- und Lernbedurfnissen auf der einen Seite sowie Zeit und Kosten auf der andern Seite ein ausgewogenes Verhaltnis bestehen muss.

Es scheint uns wichtig, dass Sie vor der Vertragsunterzeichnung kurz innehalten und die Projektskizze nuchtern betrachten. Eine Evaluation wird in der Regel nur so gute Resultate bringen, wie es aufgrund des Forschungsplans erwartet werden kann. Sie durfen deshalb nicht mit Wundern rechnen. Vor der Vergabe des Auftrags kann es sich lohnen, die folgenden Fragen ein letztes Mal durchzugehen:

- Verstehe ich die Evaluationsstrategie? Macht sie Sinn? Ist sie geeignet, Antwort auf die gestellten Fragen zu geben? Konnen die Resultate voraussichtlich genutzt und umgesetzt werden?³ Sind keine weiteren Verbesserungen am Forschungsplan mehr moglich?
- Verdient das Evaluationsteam Vertrauen im Hinblick auf die Bearbeitung der anspruchsvollen Fragestellungen?
- Ist der Zeitplan realistisch? Sind keine Flaschenhalse (zu kurz bemessene Fristen) feststellbar? Kommen die Resultate rechtzeitig fur anstehende Entschiede?
- Konnen die Kosten der Evaluation durch den damit voraussichtlich erzielbaren Nutzen gerechtfertigt werden?

³Wenn Sie diesen Punkt vertiefen mochten, finden Sie in Kap. 1.5 weitere Angaben.

Bei allfälligen Zweifeln lohnt es sich, noch einmal über die Bücher zu gehen und die entsprechenden Punkte mit dem Evaluationsteam zu diskutieren bzw. von einer Evaluation abzusehen. Sie haben ein Recht darauf, Klarheit über den Forschungsplan zu erhalten. Wenn Sie hingegen alle obigen Fragen mit Ja beantworten können, sollte einem Vertragsabschluss nichts mehr im Wege stehen.

Der letzte Schritt der Auftragsvergabe ist die Gestaltung und Unterzeichnung des Vertragstexts. Dabei kann in der Regel von der Offerte, die ergänzt und auf der Basis des in den Verhandlungen erzielten Ergebnisses bereinigt wird, ausgegangen werden.

Weiterführende Literatur

Als weitere Arbeitsinstrumente allgemeiner Art können Sie AGREB 1990 und Eidgenössisches Personalamt 1993 verwenden. Eine gute Einführung in das Vorgehen bei der Evaluation gibt ferner Wholey 1983.

1.4 Die Evaluation sinnvoll begleiten

Im Hinblick auf die Begleitung und Steuerung des Evaluationsprojekts stehen zwei Fragen im Vordergrund: Erstens **wer** die Evaluation begleiten soll und zweitens **wie**, d.h. nach welchen Grundsätzen die Begleitung erfolgen soll.

1.41 Wer begleitet die Evaluation?

Die Bestimmung der Projektorganisation im Hinblick auf die Begleitung hat Auswirkungen auf die Umsetzung, aber auch auf die Kosten. Folgende Fragen sind zu beantworten: (1) Welche Person(en) aus der Dienststelle, welche die Evaluation in Auftrag gibt (und finanziert), soll(en) gegenüber dem Evaluationsteam die Verantwortung wahrnehmen? (2) Soll für die Begleitung und Steuerung der Evaluation allenfalls eine Begleitgruppe gebildet werden? Welche Aufgabe hat die Begleitgruppe zu erfüllen? Welche Personen sollen darin Einsitz nehmen?

Nach den Erfahrungen im Rahmen des NFP 27 können **Begleitgruppen** ein geeignetes Instrument zur Begleitung und Mitgestaltung des Evaluationsprozesses sein. Sie sollten die wichtigsten, am Evaluationsprozess und an den Evaluationsresultaten interessierten Personen einbeziehen und gleichzeitig eine bestimmte Grösse (8-12 Personen) nicht überschreiten. Die Begleitgruppe kann einerseits dem Evaluationsteam wertvolle Informationen vermitteln. Sie ist andererseits dadurch, dass sie privilegierten Zugang zu den Evaluationsre-

sultaten hat, sehr oft ein sinnvolles Gefäss für die Umsetzung der Resultate. Folgende Aufgaben können beispielsweise der Begleitgruppe übertragen werden:

- Mitwirkung bei der Festlegung der Forschungsfrage und des Forschungsplans
- Festlegung der bei der Evaluation (z.B. in Form von Interviews) einzubeziehenden Personen oder Gruppen
- Beurteilung der Evaluationsstrategien und Forschungstechniken (z.B. im Hinblick auf Plausibilität und Verständlichkeit)
- Beurteilung der Evaluationsresultate (mündliche Präsentationen, Zwischenberichte, Entwurf Schlussbericht)
- Mitwirkung bei der Bestimmung der im Anschluss an die Evaluation zu treffenden Änderungen

Negativ ins Gewicht fällt der Zeitaufwand für die Begleitgruppe. Termine müssen frühzeitig festgesetzt werden und sollten eine Reserve für bei der Evaluation allenfalls auftretende Schwierigkeiten enthalten. Begleitgruppen verursachen auch finanzielle Aufwendungen und zwar im Hinblick auf administrative Kosten (z.B. Saalmieten, Mahlzeiten, ev. Reiseentschädigungen für TeilnehmerInnen) und im Hinblick auf den Zeitaufwand für die Vorbereitung der Sitzungen. Da Zeit ein kostbares Gut ist, müssen die Sitzungen der Begleitgruppe vom Evaluationsteam sehr seriös vorbereitet werden (was sich im Budget niederschlagen wird).

Wir raten davon ab, Begleitgruppen als blosses Alibi-Instrument einzusetzen. Sie sollen vielmehr eine echte Mitwirkung anderer Personen am Evaluationsprozess ermöglichen. Der formelle Auftraggeber braucht sich vor dieser Mitwirkung nicht zu fürchten, wenn er seinerseits genügenden Sachverstand in die Begleitgruppe einbringen kann.

1.42 Grundsätze für die Begleitung

Für die Begleitung der Evaluation ist eine gute Dosierung von Vertrauen und Kontrolle geeignet. Evaluationsteams verdienen in der Regel das Vertrauen, das ihnen entgegengebracht wird. Sie bringen professionelles Know-How und Erfahrungen im Management von Forschungsprojekten mit und haben ein Interesse daran, dass die Evaluation zu einem Erfolg wird.

Gleichzeitig ist es wichtig, dass Sie dem Evaluationsteam zeigen, dass Sie an der pünktlichen Erarbeitung der Zwischen- und Endresultate interessiert sind. Je besser Sie Präsenz markieren und die Wichtigkeit ziel- und sachgerechter Ergebnisse kommunizieren können, desto besser wird das Ergebnis Ihren Vorstellungen entsprechen. Eine eingehende Begleitung der Evaluation

kann sich im Hinblick auf die Umsetzung sehr auszahlen. Je sachkundiger Sie die Evaluation begleiten, desto besser können die während des Evaluationsprozesses noch zu fällenden Entscheide (z.B. Stichprobenumfang der Umfrage, Bestimmung der in der zweiten Runde vorzunehmenden Interviews, Umfang und Gestaltung des Zwischenberichts) auf Ihre Bedürfnisse zugeschnitten werden. Der Aufwand für die Begleitung wird Ihnen im Hinblick auf die Umsetzung wieder zugute kommen.

Planen Sie deshalb für die Begleitung der Evaluation genügend Zeit ein. Aufgrund unserer Erfahrungen sollten für die Begleitung einer grösseren - z.B. zehnmonatigen - Evaluation pro Monat mindestens 1-2 Tage vorgesehen werden. Diese Zeit werden Sie brauchen für:

- Telephonische Kontakte (z.B. je alle 2 Wochen); Besprechung auftretender Detailfragen
- Treffen mit dem Evaluationsteam (ca. einmal pro Monat) bzw. einer allfälligen Begleitgruppe (alle 2-4 Monate); Vorbereitung grösserer Entscheide im Hinblick auf das Forschungsprojekt
- Orientierung anderer Beteiligter in- und ausserhalb Ihrer Dienststelle über den Projektfortschritt und (Zwischen)-Ergebnisse
- Analyse der Zwischen- und Schlussresultate
- Umsetzung

Es ist durchaus möglich, dass das Evaluationsteam während der Evaluation den Forschungsplan anders als Sie interpretiert oder dass beispielsweise aufgrund der Datenlage Änderungen vorgenommen werden müssen. In diesem Fall ist es wichtig, die Probleme aufs Tapet zu bringen und eingehend zu besprechen. Oft wird es möglich sein, eine geeignete Lösung zu finden, welche sowohl Ihre Informationsbedürfnisse als auch die durch das Budget gesetzten Grenzen berücksichtigt.

Wir möchten nicht verschweigen, dass es in seltenen Fällen bei Evaluationen zu ernsthaften Schwierigkeiten kommen kann. Die Gründe dafür können sehr vielfältig sein:

- Widerstände bei beteiligten Dienststellen und/oder betroffenen Gruppen
- Änderung der Rahmenbedingungen (Ablehnung einer Vorlage in einer Volksabstimmung, neue richtungweisende Entscheide der politischen Behörden, Budgetprobleme u.a.m.)
- Nicht oder nur schlecht durchführbarer Forschungsplan (mangelnde Eignung einer Kontrollgruppe für den Vergleich, ungenügende Zahlenreihen bei einem Längsschnitt-Vergleich, Probleme bei der Datenbeschaffung, Schwierigkeiten bei der EDV-gestützten Auswertung)
- Personelle Probleme innerhalb des Evaluationsteams (Betrachtung einer ungeeigneten Person mit Forschungsaufgaben, schlechte Zusammenarbeit innerhalb des Evaluationsteams, längere Krankheit oder Unpässlichkeit von

Schlüsselpersonen im Evaluationsteam, Absprung einer sehr fähigen Person im Evaluationsteam)

Es gibt keine Patentrezepte für solch heikle Situationen. Sie können versuchen, die Schwierigkeiten einigermaßen zu beheben und Ersatzlösungen zu finden (z.B. Ersatz eines ungeeigneten Projektmitarbeiters, etwas anders gelagerte Stichprobe), um dennoch zu brauchbaren Resultaten zu kommen. Bei Problemen, die nicht durch das Evaluationsteam zu verantworten sind (z.B. mangelnde Eignung der durch die Verwaltung bereitgestellten Daten) wird ev. ein ergänzender Auftrag nötig sein. Wenn ein erfolversprechender Abschluss der Evaluation nicht abzusehen ist, muss eine Beschränkung oder gar ein Abbruch des Projekts in Betracht gezogen werden. Wir raten Ihnen jedenfalls davon ab, bei ungenügenden (Zwischen-)Resultaten, die durch eine unsorgfältige Planung und ein ungenügendes Projektmanagement des Evaluationsteams bedingt sind, weitere zusätzliche Mittel in das Evaluationsprojekt zu stecken. Umgekehrt kann es sich durchaus lohnen, Zusatzaufträge zu vergeben, wenn das Evaluationsteam gute Arbeit geleistet hat und sich interessante Zusatzfragen stellen.

1.5 Die Evaluationsresultate optimal nutzen

Sie sind mit einem Problem konfrontiert worden und haben eine Evaluation in Auftrag gegeben und begleitet, um darauf Antworten zu erhalten. Bereits bei der Vorbereitung des Evaluationsprojekts sind Sie möglicherweise auf Fragen gestossen, die Sie sich bisher noch nie in dieser Form gestellt haben. Beispielsweise: Welches sind unsere Hauptziele? Welche Personen, Betriebe, Gegenden sollen durch unsere Massnahmen erreicht werden? Hat die Massnahme zu den befürchteten Nebenwirkungen geführt? Besteht ein Potential an Einsparungen, ohne dass die Hauptwirkungen gefährdet werden?

Durch den Kontakt mit dem Auftragnehmer haben Sie Kenntnis von den Zwischenergebnissen erhalten. Vielleicht hat dieser enge Kontakt bereits **Lernprozesse** ausgelöst (oft spielen sich die wichtigsten Lernprozesse während der Evaluation ab). Nun liegen die Ergebnisse vor (schriftlich, mündliche Präsentation). Sie können von Ihnen oder von weiteren Personen für verschiedene Zwecke genutzt werden.

1.51 Rechenschaft ablegen

Vielleicht sind die Ergebnisse nicht neu. Sie zeigen Sachverhalte auf, die Ihnen schon bekannt waren. Vielleicht wissen Sie nun fundierter über den

Vollzug, die Wirksamkeit, oder die Wirtschaftlichkeit Bescheid. Sie können belegen, dass das anvisierte Zielpublikum erreicht wird. Es hat sich ergeben, dass die Nebenwirkungen weniger gravierend sind, als behauptet wurde. Gut so. Nicht immer müssen aus einer Evaluation Änderungen und Verbesserungen resultieren. Die Öffentlichkeit hat ein Anrecht zu wissen, ob Massnahmen und Gesetze, die kosten bzw. den Privaten bestimmte Pflichten auferlegen, die gesetzten Ziele erreichen. Transparenz über staatliche Massnahmen hilft, das Vertrauen der Bürgerinnen und Bürger in den Staat zu erhalten oder zu vergrössern. Durch eine Evaluation können Sie allenfalls auch einem "falschen Alarm" zuvorkommen und sich Argumente beschaffen, um dem Ruf nach einer Beitragskürzung oder nach einem Abbau von Vorschriften (oder je dem Gegenteil) entgegenzutreten zu können. Die Demokratie braucht fundierte Argumente.

Wir möchten mit Nachdruck davon abraten, Evaluationen als reines Alibi-Instrument zur Legitimationsbeschaffung zu verwenden. Dies mag Ihnen kurzfristigen Gewinn verschaffen. Wir sind aber überzeugt, dass die Kontrollorgane auf Ebene des Bundes und der Kantone gelegentlich überprüfen werden, ob die Vergabe von Evaluationen einem wirklichen Bedürfnis entspricht und sorgfältig abgewickelt wird.

1.52 Den Vollzug verbessern

Sehr oft können Evaluationen Hinweise geben, wie Sie den Vollzug verbessern können. Insbesondere Evaluationen des Vollzugs dienen diesem Ziel. Auch Evaluationen der Wirkungen können Ihnen Hinweise über mögliche Verbesserungen des Vollzugs geben, wenn sie nicht als "Black box-Evaluation" konzipiert sind und den Wirkungspfad sorgfältig rekonstruieren.

Beispiel 1: Sie haben aus der Evaluation erfahren, dass die Dienstleistungen, die Sie anbieten, 90 Prozent der Grossbetriebe, aber nur 40 Prozent der Kleinbetriebe bekannt sind. Sie werden in den kommenden Jahren Ihre Informationsanstrengungen zugunsten der Kleinbetriebe wesentlich verstärken und dann überprüfen, ob sich bei diesen Betrieben der Kenntnisstand und die Inanspruchnahme der Dienstleistungen verbessert hat.

Beispiel 2: Eine Evaluation des Vollzugs hat vor allem zu ermitteln versucht, worauf die in Ihrem Kanton bestehenden Vollzugsunterschiede zwischen den Gemeinden zurückzuführen sind. Die Untersuchung hat ergeben, dass dafür weniger die politische Struktur oder die Finanzlage der Gemeinden verantwortlich ist, sondern die Problemsensibilität und der Ausbildungsstand des Vollzugspersonals. Deshalb werden Sie in den kommenden Monaten und

Jahren Ihre Anstrengungen vornehmlich auf die Aus- und Weiterbildung richten (vgl. Studie Balthasar/Knöpfel 1994).

Beispiel 3: Die in Auftrag gegebene Studie zeigt, dass die mit der Massnahme angestrebte Verbesserung der Umweltqualität durch die Zunahme der Emittenten (z.B. Zunahme der Fahrzeuge, Zunahme der Gebäude) zunichte gemacht wird. Gleichzeitig wird aber noch ein erhebliches Potential im Hinblick auf eine konsequente Durchsetzung der Vorschriften ermittelt. Deshalb werden Sie Ihre Anstrengungen vorderhand auf einen normgerechten Vollzug konzentrieren und mit einer Verschärfung der Rechtsnormen zuwarten.

1.53 Kosten einsparen

Als vielfältig einsetzbares Instrument können Evaluationen dazu beitragen, Kosten einzusparen. Vorschläge für Einsparungsmöglichkeiten können aus Evaluationen des Vollzugs und der Wirkungen resultieren. Evaluationen können aber auch spezifisch auf die Ermittlung von Möglichkeiten zur Kostensenkung ausgerichtet werden. Dem Kriterium der Wirtschaftlichkeit wird in diesem Falle die grösste Priorität zugemessen. Es geht darum zu ermitteln, welche Einsparungen den geringsten Einfluss auf die mit der Massnahme anvisierten Hauptwirkungen haben. Spezifisch zu erwähnen sind Kosten-Nutzen-Analysen und Kosten-Effektivitäts-Analysen, welche in der Regel prospektiv zur Bewertung von Handlungsalternativen eingesetzt werden.

Beispiel 1: Eine Rekonstruktion der Investitionsentscheide der durch die Subvention anvisierten Zielgruppe ergibt, dass 70% der Beitragnehmer die Investition auch ohne staatliche Hilfe vorgenommen hätten. Dies ist ein recht hoher Mitnahmeeffekt. Durch eine Einengung des Kreises der Beitragsberechtigten und durch eine bessere Instruktion der mit der Gesuchsprüfung betrauten Personen wird versucht, den Mitnahmeeffekt zu vermindern.

Beispiel 2: Eine Zielgruppenbefragung von Empfängern staatlicher Beiträge ergibt, dass einzelne Personen und Betriebe auf diese Leistungen nicht angewiesen und gar erstaunt waren, dass sie in den Genuss der Beiträge kamen. Es wird nun versucht, durch eine Änderung der rechtlichen Grundlagen solche Fälle von der Subventionsberechtigung auszuschliessen.

Beispiel 3: Der finanzielle Engpass einer Grossstadt macht die Kürzung der Ausgaben einer im Sozialbereich tätigen Dienststelle um 1,4 Millionen Franken notwendig. Zuerst wird eine lineare Kürzung der Ausgaben erwogen. Zugunsten einer Evaluation wird darauf verzichtet. Zum ersten Mal wird dabei Transparenz über die Vollkosten (Personalkosten, Materialkosten, kalkulatorische Raumkosten) der über die ganze Stadt verteilten Dienststelle erstellt, werden zusammen mit dem Vollzugspersonal die erbrachten Dienst

leistungen nach ihrer Wichtigkeit rangiert und werden anschliessend Prioritäten und Posterioritäten für Kürzungen festgelegt. Durch die Evaluation wird bei der Dienststelle zum ersten Mal das Bewusstsein über die mit der Sozialarbeit anvisierten Ziele und über die Kosten geschaffen. Die Kürzungen können dort vorgenommen werden, wo sie für die Betroffenen am ehesten verkraftbar sind. Ein Quervergleich der unterschiedlichen Dienststellen verbessert zudem das Bewusstsein für deren Besonderheiten und für Verbesserungspotentiale (vgl. Studie Price Waterhouse 1993).

Es ist legitim, dass Evaluationen implizit oder explizit für Kosteneinsparungen eingesetzt werden. Evaluationen können vertiefte Grundlagen bereitstellen, wie bestimmte Kürzungen am besten auf spezifische Ausgabenposten verteilt werden können.

1.54 Massnahmen aufheben

Gegebenenfalls kann eine Evaluation auch Hinweise darauf geben, dass eine Massnahme ohne gravierende Nachteile aufgehoben werden kann.

Beispiel: Die Evaluation des im Heimarbeitsgesetz enthaltenen Gleichstellungsgebots zwischen Heimarbeitern und Fabrikarbeitern (Gerheuser/Schmid 1993) zeigte, dass erstens die Zahl der dem Erlass unterstellten Personen dauernd zurückgeht, dass zweitens mit grosser Wahrscheinlichkeit keine Fälle rücksichtsloser Ausbeutung mehr vorkommen und dass drittens eine konsequente Durchsetzung des Gleichstellungsgebots zum Verschwinden bestimmter Heimarbeitsformen führen würde, was von den betroffenen Heimarbeitern indessen nicht gewünscht wird. In ihren Schlussfolgerungen erwägen die Autoren die Aufhebung der Vorschrift als ein mögliches Szenario, wobei sie davon ausgehen, dass die hauptsächlich von Frauen ausgeübte Heimarbeit durch das Gleichstellungsgebot gemäss (künftigem) Bundesgesetz über die Gleichstellung von Frau und Mann ausreichend geschützt sein wird.

Vor allem Evaluationen der Wirtschaftlichkeit vermögen Hinweise darauf zu geben, ob die Kosten einer Massnahme durch den Nutzen der dadurch erzielten Wirkungen aufgewogen werden. Sie können gegebenenfalls auch den Verzicht auf von politischen Gruppen zusätzlich geforderte Massnahmen rechtfertigen helfen.

1.55 Massnahmen ändern oder neugestalten

Evaluationen als empirische Untersuchungen sind vornehmlich auf die Erfassung des Ist-Zustands ausgerichtet (vgl. Kap. 2.4). Ihr wichtigstes Ziel besteht indessen oft in der Vorbereitung künftiger Entscheide. Um zu wissen, wohin man geht, braucht es fundierte Kenntnisse darüber, wo man steht. Erst wenn bekannt ist, wie der Vollzug abläuft und welche Wirkungen die Massnahme zeitigt, kann der Bedarf an Änderungen einigermaßen verlässlich bestimmt werden.

Wäre die Zukunft eine einfache Verlängerung der Vergangenheit, dann wäre die Neugestaltung einer Massnahme eine sehr einfache Sache. Es wären bloss die Mängel der bestehenden Massnahme auszumerzen. Die Änderungen würden sich sozusagen automatisch aus einer (ex post/retrospektiven) Evaluation ergeben.

Doch die Wirklichkeit ist komplex und oft unberechenbar. Neben einer Evaluation des Vollzugs und der Wirkungen braucht es deshalb weitere drei Schritte:

Analyse der Problementwicklung:

Es gilt, die bisherigen Entwicklungen zu analysieren und möglichst fundierte Mutmassungen anzustellen, wie sich die Rahmenbedingungen und die Problemlage in Zukunft entwickeln werden. Es sind beispielsweise folgende Fragen zu beantworten: Wie hat sich das Problem entwickelt? Ist eine Verbesserung, Verschlechterung, Stabilisierung eingetreten? Wie wird sich die Situation voraussichtlich entwickeln? Welche künftigen neuen Faktoren müssen in Rechnung gestellt werden? Während namentlich bei Wirtschafts-, Verkehrs-, Energie- und Umweltfragen z.T. eine quantitative Modellierung der bisher abgelaufenen bzw. künftig erwarteten Entwicklungen erfolgt, wird in anderen Politikbereichen, wenn überhaupt explizite Überlegungen angestellt werden, mit Szenarien gearbeitet. Denkbar ist auch, die Zukunftsvorstellungen von Experten des Fachgebiets zu ermitteln, wofür sich die Technik der Delphi-Umfrage eignet. Bei anspruchsvollen Techniken der Zukunftsexploration kann es durchaus sinnvoll sein, wissenschaftliche Beratung in Anspruch zu nehmen.

Analyse der Handlungsmöglichkeiten:

Auf der Basis der Problemanalyse, der Evaluation des Vollzugs und der Wirkungen, unter Berücksichtigung von (positiven und negativen) Beispielen in verwandten Politikbereichen, in anderen Kantonen oder anderen Ländern und nicht zuletzt auf der Grundlage Ihrer und der Kreativität weiterer an der Neugestaltung beteiligter Partner gilt es, die künftigen Handlungsalternativen zu bestimmen. Diese können aus **Kopien** von bereits bestehenden Massnahmen

(aus anderen Politikbereichen, anderen Kantonen oder Ländern), aus der **Adaptation** bereits bestehender Massnahmen an die veränderten Verhältnisse, aus einem **Amalgam** bereits bestehender Massnahmen-Elemente, aus einer **Synthese** solcher Massnahmen-Elemente zu etwas Neuem oder gar aus einer **Erfindung** bestehen.

Auswahl der geeigneten Massnahme:

In einer Grobanalyse gilt es vorerst, ungeeignete Handlungsalternativen auszuschneiden. Anschliessend sind die Handlungsalternativen (auf der Basis Ihrer bisherigen Erfahrungen und Kenntnisse) aus der Perspektive der Kriterien "Effektivität/Vollzugstauglichkeit", "Wirksamkeit" und "Wirtschaftlichkeit" approximativ zu beurteilen. Es gilt, Fragen zu beantworten wie: Welche Vollzugsprobleme könnten bei den einzelnen Handlungsalternativen auftreten? Wie gross wird der Zeitbedarf für den Vollzug sein? Welche Wirkungen und Nebenwirkungen könnten eintreten? Welcher finanzielle Aufwand und welche weiteren Belastungen Privater sind zu erwarten? Wie in Kapitel 3.7 erwähnt sind in einigen wenigen Politikbereichen Modellierungen und Simulationen der Auswirkungen von künftigen Massnahmen möglich. In anderen Bereichen wird, wenn überhaupt mit spezifischen Methoden, mit Szenarien gearbeitet. In der Regel handelt es sich somit um eine versuchsweise Wirkungsabschätzung. Sie bildet die Grundlage für die Handlungsauswahl. Letztere erfordert eine Aufreihung der Ziele nach ihrer Bedeutung, eine Bestimmung der Wirksamkeit der einzelnen Massnahmen und eine Auswahl nach dem zu erwartenden Nutzen. An sich gibt es Modelle und Software für eine möglichst systematische Handlungsauswahl, die indessen für die allermeisten Entscheidungsprozesse im öffentlichen Bereich nicht anwendbar sind. Entscheidend ist, dass eine bewusste und argumentativ begründbare Auswahl getroffen wird. Neue und unkonventionelle Handlungsmöglichkeiten sollten nicht vorschnell aufgrund emotionaler Argumente ausgeschieden werden; ihr Beitrag zur Lösung des Problems sollte seriös abgeschätzt werden.

In der Regel werden die oben aufgezeigten Schritte nicht von einer einzigen Person, sondern im Rahmen von Kommissionsarbeiten durchgeführt. Die Zusammenarbeit mehrerer Personen ermöglicht Kreativität und erlaubt die argumentative Durchdringung möglicher Handlungsalternativen. Die Gefahr besteht indessen, dass sich eine Gruppenmeinung rasch und ohne seriöse Prüfung einzelner Handlungsalternativen durchsetzt. Soweit Sie Ihre Verantwortung in den Prozess einbringen können, sollten Sie darauf hinwirken, dass alle sinnvollen Alternativen seriös geprüft und die Auswahl fundiert getroffen wird.

Die systematische und bewusste Durchführung der aufgezeigten Schritte gibt Ihnen noch nicht die Garantie, dass die Massnahmen ohne Schwierigkeiten vollzogen werden können und im erwarteten Ausmass wirken. Die Zukunft ist

immer gut für Überraschungen. Das Ziel sollte sein, auf der Basis der vorhandenen Kenntnisse die Massnahme so zu gestalten, dass sie den möglichen absehbaren Entwicklungen Rechnung trägt. Mehr ist nicht möglich. Wichtig ist indessen, dass Sie bei der Gestaltung der Massnahme sicherstellen, dass Sie inskünftig periodisch die geeigneten Informationen darüber erhalten, ob der Vollzug im geplanten Rahmen abläuft und ob anzunehmen ist, dass die vorgesehenen Wirkungen erzielt werden (vgl. Kap. 1.6).

1.56 Lernprozesse als Ziel von Evaluationen

Das Hauptziel von Evaluationen besteht zusammenfassend wohl darin, Lernprozesse zu ermöglichen und zu fördern. Dank besserem Wissen soll besseres Handeln möglich sein. Lernprozesse haben eine individuelle und eine kollektive Dimension. Nur Individuen nehmen Wissen auf. Aber die Veränderung der Wirklichkeit durch Handeln erfolgt in kollektivem Zusammenhang, in Organisationen oder gar in Netzwerken. Oft ist zu Beginn nur schwierig absehbar, wer in diesen Lernprozess einbezogen werden soll: Ist es eine bestimmte Dienststelle, sind es die Beamten auf Ebene Bund, Kantone und Gemeinden, ist es ein grösseres Netzwerk unter Einschluss der Verbände der Betroffenen oder ist es gar die breite Öffentlichkeit? Die sich aus Evaluationen ergebenden Lernprozesse sind nicht völlig planbar. Durch Transparenz, insbesondere die Zugänglichmachung von Resultaten wird indessen die Möglichkeit geschaffen, dass der Kreis der in den Lernprozess einbezogenen Personen und Institutionen möglichst breit gezogen wird.

Dem Bedarf nach lernfördernder Evaluation hat die Evaluationsforschung bisher dadurch Rechnung getragen, dass sie Vorgehensweisen entwickelt hat, welche auf die Bedürfnisse der Auftragnehmer sehr konkret eingehen. Seit einiger Zeit werden vermehrt **interaktive Evaluationsmethoden** eingesetzt. Hierbei wird die Evaluation als offener, gestaltbarer Prozess verstanden, in welchen sowohl Anliegen des Evaluationsteams als auch Anliegen der Betroffenen einfließen können (vgl. Kap. 3.62). Interaktive Evaluationsmethoden können dazu beitragen, Lernprozesse in einem Netzwerk von Personen bzw. Institutionen in Gang zu bringen. Evaluation nähert sich damit der **Mediation**, d.h. der Nutzung sozialwissenschaftlichen Sachverstands zur besseren Bewältigung von politischen und gesellschaftlichen Konflikten. Daneben werden allerdings weiterhin und in recht grossem Umfang Evaluationen aus unabhängiger und freier Warte nötig sein. Solche Evaluationen können in dem Ausmass, in dem sie auf zuverlässiger Basis neues Wissen schaffen, Lernprozesse in Gang setzen, ohne dass dieses Ziel explizit auf die Fahne geschrieben wird.

1.57 Exkurs: Möglicher Missbrauch von Evaluationen

Im Zusammenhang mit Lernprozessen und der Nutzung von Evaluationen muss ein Wort über den möglichen Missbrauch von Evaluationen gesagt werden. Dabei ist vorauszuschicken, dass zwischen legitimer Verwendung von Evaluationen und Missbrauch verschiedene Schattentöne bestehen und somit die Grenzen fließend sind.

Missbrauchsgefahren lauern bereits bei der **Auftragsvergabe** von Evaluationen. Die Forschungsfrage kann so formuliert werden, dass nur die positiven Auswirkungen einer Massnahme erfasst und beispielsweise mögliche negative Nebenwirkungen gar nicht untersucht werden. Diese Einseitigkeit kann sich im Evaluationsdesign fortsetzen: Als Vergleichseinheiten werden beispielsweise nur Kantone oder andere Länder ausgewählt, bei denen man sicher ist, dass sie vergleichsweise schlechter abschneiden. Auch bei der Anwendung einzelner Forschungstechniken kann ein positives Resultat vorgespurt werden: Bei Umfragen werden beispielsweise der Massnahme kritisch gegenüberstehende Gruppen nicht einbezogen. Statistiken, welche ein ungünstiges Bild der Lage zeichnen, werden als unzuverlässig eingestuft und gar nicht erst ausgewertet. Das Resultat solchen Vorgehen ist eine Alibi-Evaluation - dazu angetan, ein positives Bild der Massnahme zu zeichnen. Mit umgekehrten Vorzeichen kann eine Evaluation - als Killer-Evaluation - auch benutzt werden, um eine bestimmte Massnahme "abzuschiesse". Auch hier handelt es sich um eine Alibi-Evaluation, die nur dazu herhalten muss, einen bereits gefassten Entscheid zu legitimieren.

Gegen solche Missbräuche von Evaluationen gibt es keine Patentrezepte. Wir setzen Hoffnungen in eine kritische Öffentlichkeit, in hinterfragende und sachkundige Medien, in eine Wissenschaft, welche sorgfältig über ihre Standards wacht, und in Kontrollorgane der Exekutive und des Parlaments⁴, welche stichprobenartig Evaluationsstudien auf ihre methodische Fundiertheit überprüfen. Dadurch werden nicht alle Missbräuche verhindert werden können. Aber wir sind überzeugt, dass mit der weiteren Verbreitung von Evaluationen auch Vorkehrungen getroffen werden, um allfälligen Auswüchsen zu begegnen. Universitätsinstitute und private Büros werden in Zukunft noch mehr als heute gezwungen sein, impliziten oder expliziten ethischen Richtlinien nachzuleben, wenn sie ihre wissenschaftliche Glaubwürdigkeit längerfristig erhalten wollen.

Auch **nach dem Vorliegen der Forschungsergebnisse** können Missbräuche vorkommen. Ganz klar als Missbrauch einzustufen sind die Fälschung

⁴So hat in den USA das General Accounting Office als Kontrollorgan des Parlaments viel Sachverstand aufgebaut, um die von der Exekutive veranlassten Evaluationen wissenschaftlich zu durchleuchten.

unliebsamer Daten, das Herausstreichen kritischer Passagen und die Nichtveröffentlichung seriös erarbeiteter Evaluationsresultate, obgleich dies vereinbart worden war. Daneben gibt es zahlreiche Fälle, die nicht eindeutig als Missbrauch einzustufen sind. Dazu gehört die einseitige Verwendung von Evaluationsresultaten durch Medien und Politiker. Es kann durchaus vorkommen, dass in den Medien aus einem Bericht, der zu einem überwiegend positiven Urteil der Massnahme gelangt, die wenigen kritischen Stellen herausgegriffen und als Gesamtaussage des Berichts dargestellt werden (und umgekehrt). Soweit nicht Falschzitationen vorkommen, kann gegen eine derart einseitige Verwendung rechtlich (z.B. in Form von Gegendarstellungen) kaum vorgegangen werden. Bekannt ist auch, dass Politiker Argumente aus wissenschaftlichen Analysen herauspicken und als Munition in politischen Debatten verwenden. Auftraggeber und Auftragnehmer von Evaluationen müssen sich damit abfinden, dass die Auseinandersetzung um staatliche Massnahmen in der politischen Arena erfolgt und dort die Spielregeln gesetzt werden.

Sehr heikle Fragen im Spannungsfeld zwischen legitimer Nutzung und Missbrauch tauchen namentlich bei der Formulierung von Schlussberichten auf, die für die Öffentlichkeit bestimmt sind. Aus der Sicht des Evaluationsteams reagiert die Verwaltung oft völlig mimosenhaft auf vereinzelte kritische Bemerkungen im Schlussberichtsentwurf. Die Auftragnehmer umgekehrt können sich bereits ausmalen, wie diese Stellen aus dem Zusammenhang gerissen, aufgebauscht und im politischen Kampf eingesetzt werden. Wir möchten hier einerseits den Auftraggebern (v.a. in der Verwaltung) zu bedenken geben, dass die Kontrollfunktion von Presse und Politik ein gelegentliches Aufzeigen von Schwächen erfordert und dass die daraus entstehenden Konsequenzen für die Verwaltung selten gravierend sind. Wir glauben, dass Regierung und Verwaltung einen besseren Umgang mit Kritik entwickeln müssen. Den Auftragnehmern möchten wir gleichzeitig zu bedenken geben, dass Evaluationen in politischen Auseinandersetzungen recht oft als Munition verwendet werden und dass nach dem Motto "if anything can go wrong, it will go wrong" eine grosse Chance besteht, dass bestimmte Aussagen aus dem Zusammenhang gerissen und aufgebauscht werden. Wir würden dem Evaluationsteam demnach empfehlen, bei der Redaktion von Berichten, die für die Öffentlichkeit bestimmt sind, den Grundsatz zu befolgen: Hart in der Sache, moderat im Ton.

Weiterführende Literatur

Über die möglichen Verwendungszwecke von Evaluationen gibt vor allem Chelimsky 1987 Auskunft. Weiterführende Hinweise zur prospektiven Wirkungsabschätzung finden sich in Bussmann 1992. Hinweise über Entscheidungsmodelle geben Böhret/Hugger 1980. Rose 1993 behandelt die Möglichkeiten, aus den Erfahrungen anderer Gebietseinheiten und aus früheren Erfahrungen mit Massnahmen Lehren zu ziehen. Im Hinblick auf die Förderung von Lernprozessen bieten unter den Projekten des NFP 27 vor allem Rey et al. 1993, Bierter/Binder 1993 und Hirsch et al. 1995 Einblick in die entsprechenden interaktiven Vorgehensweisen. Die Beiträge in Bussmann 1994 und insbesondere derjenige von Binder 1994 behandeln die Lernthematik bzw. die Nutzung von Evaluationen zu diesem Zweck.

1.6 Vorkehren treffen, um künftig die Wirksamkeit von Massnahmen besser zu erfassen

Wenn eine neue oder geänderte Massnahme von den politischen Entscheidungsinstanzen angenommen worden ist, könnte man sagen: Ende gut, alles gut. Staatliches Handeln kennt aber kein Ende, sondern ist ein dauernder Prozess. Früher oder später werden Fragen über die Effektivität, die Wirksamkeit und die Wirtschaftlichkeit der ergriffenen Massnahmen auftauchen. Wenn Sie daran interessiert sind, später die Wirksamkeit der Massnahmen belegen zu können, was wir hier annehmen, dann müssen Sie **jetzt** die geeigneten Vorkehren dazu treffen. Andernfalls wird eine Evaluation nur mit hohem Aufwand oder unter Umständen nur mit fragwürdigen Methoden durchgeführt werden können. Wenn Sie die Evaluierbarkeit verbessern möchten, können Sie in den folgenden Bereichen ansetzen:

Erhebung von Vollzugsdaten

Vollzugsdaten fallen bei der Verwaltungstätigkeit an. Sie werfen ein Licht sowohl auf die Leistungen der Verwaltung (z.B. für bestimmte Tätigkeiten eingesetztes Personal, getroffene Massnahmen, erteilte Bewilligungen, geleistete Beiträge) als auch auf das Verhalten der Adressaten der Massnahme (z.B. Anzahl der Anfragen bzw. der eingereichten Gesuche, Zahl der sanierungsbedürftigen Anlagen, Umfang der durchgeführten Bauarbeiten). Vollzugsdaten können in der Regel recht leicht beschafft werden. Die Vollzugsinstanzen können gebeten oder verpflichtet werden, periodisch bestimmte Daten, welche für die Beurteilung des Vollzugs oder der Wirksamkeit wichtig sind, zu liefern. Eine besonders günstige Lage, zu Vollzugsdaten zu gelangen, besteht

dann, wenn eine staatliche Leistung erbracht wird (Beitrag, Genehmigung eines Baugesuchs usw.).

Bei der Bestimmung der zu ermittelnden Daten gilt es, nicht nur gegenwärtige, sondern auch künftige Bedürfnisse möglichst gut abzuschätzen. Welche Probleme könnten auf der politischen Ebene auftauchen? Über welche Wirkungsaspekte sollte in Zukunft fundiert informiert werden? Welche Teilmassnahmen könnten in Bezug auf ihre Wirtschaftlichkeit zu Diskussionen Anlass geben? Über solche und weitere Fragen lohnt es sich frühzeitig nachzudenken. Selbstverständlich ist es nicht möglich, alle Informationsbedürfnisse vorzusehen. Ebenso ist es auch nicht zweckmässig, für alle denkbaren Fälle Informationspflichten vorzusehen. Hier wie überall gilt der Grundsatz der Verhältnismässigkeit. Zum Teil wird nicht zu umgehen sein, dass im Falle rasch auftauchender Informationsbedürfnisse kurzfristig neue Daten erhoben werden müssen. Gleichzeitig ist daran zu erinnern, dass längerfristige Zahlenreihen für die Beurteilung der Wirksamkeit einer Massnahme sehr wichtig sein können. Bei jeder Massnahmenänderung gilt es auch abzuklären, ob bisher Daten erfasst worden sind, welche nur noch von geringem Nutzen sind oder welche auf andere Weise einfacher und besser erfasst werden können. Letzteres kann beispielsweise dann zutreffen, wenn Sie bisher grosse Datenmengen erhoben haben und sich nunmehr auf eine Stichprobe der Grundgesamtheit beschränken können, bei der wesentlich detailliertere Angaben erfasst werden.

Verarbeitung und Auswertung der Vollzugsdaten

Die Erfassung von Vollzugsdaten lohnt sich nur, wenn deren Verarbeitung und Auswertung sichergestellt ist. Nehmen Sie mit den für statistische Aufgaben verantwortlichen Stellen Kontakt auf und benutzen Sie gegebenenfalls die methodische Beratung durch das Bundesamt für Statistik. Denken Sie frühzeitig an die Bereitstellung von EDV-Apparaten und Personal. Je besser Sie die Verarbeitung und Auswertung planen, desto weniger Personal werden Sie anschliessend dafür einsetzen müssen (Verwendung maschinenlesbarer Informationen). Achten Sie darauf, dass die Daten im Hinblick auf unterschiedliche Aspekte ausgewertet können. Es gibt heute einfach handhabbare Datenbanksysteme, welche solche Auswertungen ermöglichen. Nehmen Sie auch frühzeitig mit der Datenschutzstelle Kontakt auf, um sicherzustellen, dass durch Auswertungen keine Rückschlüsse auf einzelne Personen vorgenommen werden können. Aussagekräftige Auswertung und Beachtung der Datenschutzbelange können durchaus Hand in Hand gehen.

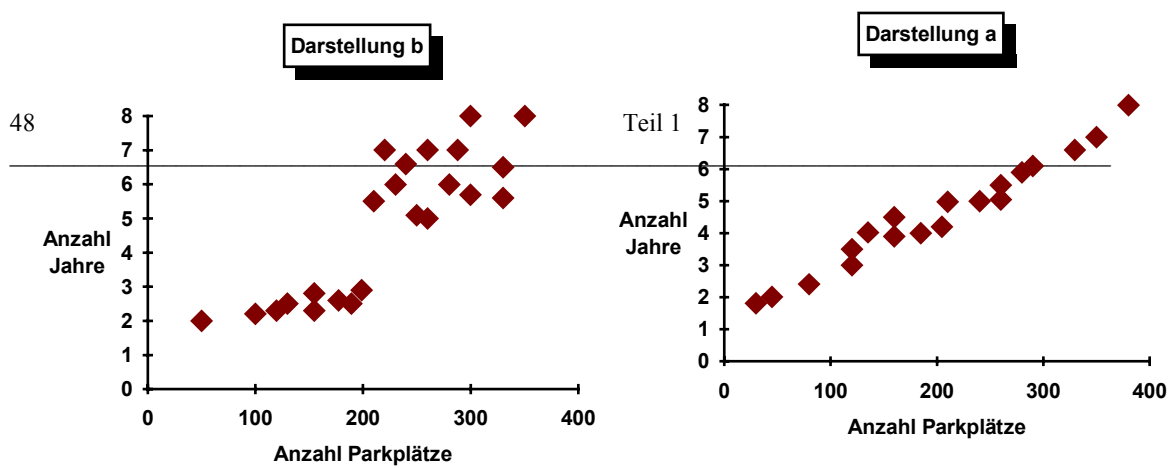
Vergleiche zur besseren Evaluierbarkeit der Massnahme möglich machen

Wirkungsermittlungen sind dann am zuverlässigsten, wenn sie auf der Grundlage sinnvoller Vergleiche erfolgen (vgl. Kap. 3.5).

Als erstes erwähnen wir den Längsschnitt-Vergleich, bei welchem der Zustand vor und nach Inkrafttreten einer Massnahme verglichen wird. Dabei ist allerdings wichtig, dass der Einfluss von Faktoren, die im Zeitablauf schwanken (z.B. Konjunkturlage), konstant gehalten wird. Ein Längsschnitt-Vergleich ist in der Regel desto sinnvoller, je markanter die Massnahmenänderung ausfällt. Dies widerspricht z.T. der Tradition kleiner, schrittweiser Änderungen. Wir empfehlen Ihnen auf jeden Fall, den Ist-Zustand **vor** Inkrafttreten der Massnahme möglichst eingehend zu erfassen, um später einen Längsschnitt-Vergleich durchführen zu können.

Als zweite und dritte Möglichkeit erwähnen wir den Kontrollgruppen-Vergleich und den Querschnitt-Vergleich. Die Ermittlung der Wirkungen wird erleichtert, wenn die in Aussicht genommene Massnahme nicht auf alle möglichen Adressaten oder auf alle möglichen Gebiete ausgedehnt wird. Bestimmte Personengruppen, Branchen oder Gebiete werden somit von der Regelung ausgenommen. Dadurch wird es möglich, die Situation bei den Adressaten der Regelung (Zielgruppe) mit der Situation einer Personengruppe zu vergleichen, welche der Regelung nicht unterliegt (Kontrollgruppen-Vergleich) bzw. die Situation zwischen mehreren Gebieten (Querschnitt-Vergleich) zu vergleichen.

Beispiel 1: Die Umweltverträglichkeitsprüfung wird nur für Bauprojekte ab einer gewissen Grössenordnung vorgeschrieben, z.B. für Parkhäuser ab 200 Parkplätzen. Dank diesem Grenzwert kann eine statistische Auswertung der Wirkungen der Umweltverträglichkeit auf die Dauer des Bauverfahrens vorgenommen werden. Es kann beispielsweise untersucht werden, ob der Grenzwert von 200 Parkplätzen einen markanten Einfluss auf die Dauer des Bauverfahrens hat. Die Analyse einer gewissen Zahl von UVP-pflichtigen und nicht-UVP-pflichtigen Bauverfahren kann z.B. folgende Bilder ergeben:



**Abbildung 1 Länge des Bauverfahrens in Abhängigkeit der im Bau-
projekt vorgesehenen Anzahl Parkplätze (hypothetisches
Beispiel)**

In Darstellung a hängt die Dauer des Bauverfahrens von der Höhe der Anzahl Parkplätze ab; die UVP-Pflicht hat keinen Einfluss darauf. Ein anderes Bild zeigt Darstellung b: oberhalb des Grenzwertes von 200 Parkplätzen ist eine deutlich grössere Dauer des Bauverfahrens festzustellen; die UVP-Pflicht hat somit signifikanten Einfluss darauf.

Beispiel 2: Die durch das Bundesgesetz über die Förderung des Hotel- und Kurortskredites bereitgestellten Hilfen werden nicht von allen Hotels, die in den Genuss der entsprechenden Leistungen gelangen könnten, in Anspruch genommen. Dafür sind verschiedene Gründe verantwortlich (mangelnde Kenntnis der Hilfsmöglichkeiten, genügende Eigenkapitalbasis, ordnungspolitische Grundsätze). Der Vergleich zwischen Hotels, welche die Leistungen (Bürgschaften, Zinskostenbeiträge) in Anspruch genommen haben und solchen, die dies nicht taten, erlaubt Rückschlüsse auf die Beurteilung der Hotellerieförderung durch die Hoteliers und auf die Wirkungen der Massnahmen (Studie Hanser/Kuster/Cavelti 1994).

Beispiel 3: Während die Form der Halbgefängenschaft für kurze Freiheitsstrafen bis zu 3 Monaten bereits seit dem 1. Januar 1974 möglich ist, können die Kantone seit dem 1. Januar 1986 auch Einschliessungs- und Gefängnisstrafen von 3 - 6 Monaten versuchsweise in der Form der Halbgefängenschaft vollziehen. Für Freiheitsstrafen bis zu einer Dauer von höchstens 30 Tagen steht den Kantonen seit dem 1. Mai 1990 ferner die Vollzugsform der gemeinnützigen Arbeit zur Verfügung. Die Kantone sind verpflichtet, ihre Erfahrungen mit diesen neuen Vollzugsformen auszuwerten. Die Tatsache, dass letztere nicht durchgehend angewendet werden, ermöglicht den Vergleich mit herkömmlichen Formen des Strafvollzugs bezüglich Kosten, Vollzugstauglichkeit und Resozialisierung.

Dass Adressaten "durch die Maschen des Gesetzes schlüpfen" oder dass die Kantone einen Erlass nicht oder sehr ungleichmässig vollziehen, ist aus ethischer Sicht vielleicht fragwürdig, weil es dem Postulat der Gleichbehandlung widerspricht. Ungleichbehandlung, Ausnahmeregelungen und ungleicher

Vollzug sind jedoch Möglichkeiten, um mehr über die Wirkungen des Vollzugs zu erfahren. Sie sind namentlich bei neuen Massnahmen oder unsicheren Rahmenbedingungen eine geeignete Möglichkeit, um für spätere - definitiv gedachte - Entscheide bessere Grundlagen zu schaffen. Gegebenenfalls können solche Massnahmen rechtlich in die Form einer Versuchsregelung gekleidet werden: sie werden befristet und nur dann weitergeführt, wenn sie sich bewähren.

Beispiele: In der Öffentlichkeit bekannt und umstritten ist die Heroinabgabe an Drogensüchtige. Gestützt auf das Betäubungsmittelgesetz können 700 Drogenabhängige in Schweizer Städten unter ärztlicher Kontrolle Heroin, Morphin oder Methadon beziehen. Daneben gibt es aber verschiedene weitere Versuchsregelungen, v.a. im Verkehr (z.B. Tempo 50), in den Medien (z.B. lokale Rundfunkversuche), bei der Schaffung und Nutzung von elektronischen Datenbanken und im Schulbereich (z.B. Fünftagewoche, Tagesschulen).

Wir legen Ihnen nahe, die Frage der Evaluierbarkeit bereits bei der Massnahmengestaltung zu berücksichtigen und auf die Agenda der zu behandelnden Fragen zu setzen. Sollten in Ihrer Dienststelle oder in der Kommission, welche die Neugestaltung der Massnahme behandelt, zu wenig fundierte Kenntnisse über Evaluationsstrategien vorhanden sein, dann kann Ihnen eine Person, welche mit solchen Fragen vertraut ist, relativ leicht (und für wenig Geld) Hinweise für eine bessere Evaluierbarkeit der Massnahme geben.

Weiterführende Literatur

Aus juristischer Sicht äussern sich Mader 1988 und 1993 sowie Mastronardi 1991 zu Versuchsregelungen.

2 Was sind Evaluationen?

2.1 Evaluation, Begriff

Evaluation bedeutet dem Wortsinn nach die Beurteilung eines Gegenstands oder Sachverhalts. Der Begriff Evaluation kann in einem sehr breiten Sinne gebraucht werden, der weder den zu evaluierenden Sachverhalt (Produkt, Institution, Massnahme, grösserer Politikbereich), noch das Ziel der Evaluation (Kaufentscheid, Reorganisation, Massnahmenverbesserung), noch die Evaluationsstrategie (sozialwissenschaftliche Untersuchungsmethoden, impressionistische Bewertung), noch die evaluierende Instanz (Vollzugsbehörde, Regierung, Aufsichtsbehörde, freie wissenschaftliche Forschung u.a.m.) einschränkt.

In unserem Zusammenhang beschränken wir uns indessen erstens auf den **öffentlichen Bereich** (wobei Evaluationsstrategien durchaus auf bestimmte Fragen in Firmen, z.B. auf interne Ausbildungsprogramme oder auf firmeneigene Massnahmen der Kulturförderung angewendet werden können).

Zweitens konzentrieren wir uns auf **systematische oder wissenschaftliche Evaluationen**. Das sind Evaluationen, welche

- nachvollziehbare Überlegungen und Aussagen enthalten,
- auf einem systematischen Vorgehen beruhen, welches eine Ermittlung der wichtigen Wirkungszusammenhänge anstrebt,
- dem Zweck dienen, möglichst objektive Aussagen zu gewinnen,
- und in einer Form vorliegen, die ihren methodischen Charakter erkennen lässt.

Drittens konzentrieren wir uns auf die Evaluation staatlicher **Massnahmen**⁵. Dies sind analytisch isolierbare Tätigkeiten des Staates im Hinblick auf bestimmte, allerdings oft vage formulierte Ziele, die sich auf Veränderungen in Wirtschaft und Gesellschaft beziehen. Sie beinhalten in der Regel personelle und finanzielle Ressourcen.

2.2 Was kann alles evaluiert werden?

Die Evaluation **staatlicher Massnahmen** deckt ein weites Feld ab. Gegenstand von Evaluationen können sein:

⁵Wir verwenden durchgehend den Begriff "Massnahme". Die Unterschiede zum Begriffs des Programms, der in der angelsächsischen Evaluationsliteratur dominiert, sind gradueller Natur. Das Programm ist in den USA nicht bloss ein analytischer Begriff, sondern auch eine Rechtsfigur, weshalb uns die wörtliche Übernahme problematisch schien.

- ganze Gesetze oder einzelne Gesetzesbestimmungen
- Subventionen (z.B. Beiträge an Waldwege, an Lawinenverbauungen)
- Institutionen und Einrichtungen (z.B. Schulen, Verkehrsbetriebe, vom Staat unterstützte Institutionen wie Verkehrszentrale, gemeinnützige Vereine)
- Anlagen (z.B. Strassen, Bahnlinien, Buslinien)

Für all diese Evaluationsobjekte ist die Methodik recht ähnlich, desgleichen auch für **Programme** - das sind auf bestimmte Ziele orientierte Massnahmenbündel. Wir werden auf die entsprechenden Evaluationsstrategien und Forschungsmethoden in der Folge (Teil 3) noch detailliert eingehen und beschränken uns im Moment auf Hinweise zur Evaluation **ganzer Politikbereiche** (z.B. schweizerisches Ausbildungssystem, Landwirtschaftspolitik u.a.m.). Die für staatliche Massnahmen und Programme gebräuchlichen vergleichenden bzw. nicht vergleichenden Evaluationsstrategien sind hier nur beschränkt tauglich, da verschiedene Massnahmen so zusammenwirken, dass ihr Beitrag an das Gesamtergebnis nicht präzise ermittelt werden kann. Die Vorbedingungen für eine Evaluation sind ähnlich wie bei einem Patienten, der mehrere Medikamente zugleich einnimmt; es wird nur schwer möglich sein, den Einfluss eines einzelnen Medikaments auf seinen Gesundheitszustand abzuschätzen. Ein globaler Gegenstand - ein ganzer Politikbereich - ruft auch nach umfassenden Evaluationsmethoden. Dabei kann man sich u.a. folgender Mittel bedienen: Analyse vorhandener Leistungsindikatoren (im Erziehungssystem z.B. Anzahl der Absolventen der Primär-, Sekundär- und Tertiärstufe) und der Kosten; Beurteilung durch ausländische, z.T. in internationalen Organisationen (z.B. OECD, Weltbank, WHO) tätige Experten (peer review).

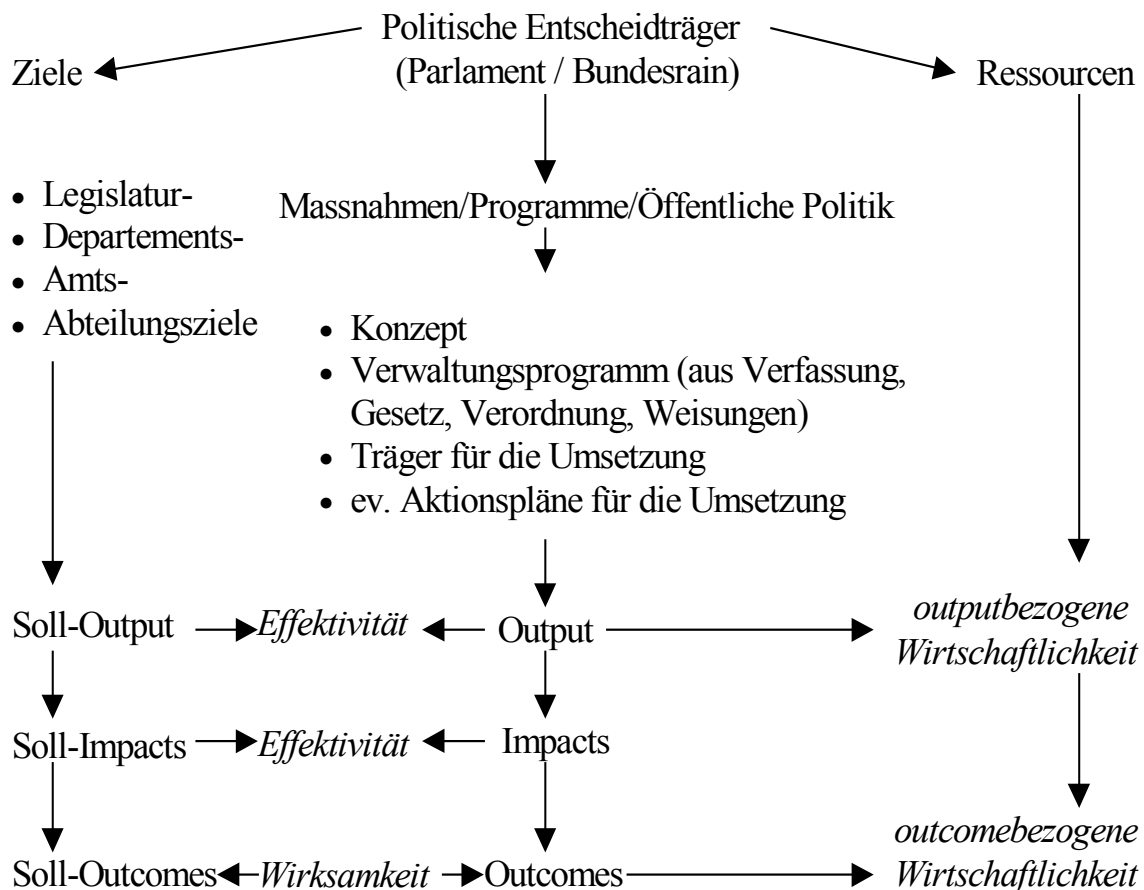
Abbildung 2 Untersuchungsmethodik für staatliche Massnahmen, Programme und Politikbereiche

Gegenstand	Hauptsächliche Untersuchungsmethodik
staatliche Massnahmen	vergleichende/nicht vergleichende Evaluationsstrategien
Programme	dito
Politikbereiche	quantitative Analysen, globale Beurteilungen, Peer review

2.3 Welche Aspekte und Kriterien können Evaluationen beinhalten?

Zwischen dem, was öffentliche Bedienstet tun, und dem, was in der gesellschaftlichen und sozialen Wirklichkeit passiert, bestehen komplexe Wirkungsverläufe. Es gibt mehrere Möglichkeiten, die verschiedenen Tätigkeiten und Sachverhalte zusammenzufassen und zu klassifizieren; die Begriffe werden (noch) nicht einheitlich verwendet. Wir gehen von folgender Darstellung aus:

Abbildung 3 Staatliches Handeln und Wirkungsverläufe



Die **Ressourcen** der Verwaltung umfassen die rechtlichen Kompetenzen, die Sachmittel, die Finanzmittel und das Personal mit all seinen Fähigkeiten (Sachwissen, Problemlösungsmethoden, Konsensfindungsstrategien).

Der **Output** sind die von der Verwaltung hergestellten Güter bzw. erbrachten Dienstleistungen (Drucksachen, Verordnungen, Bewilligungen, Busen, Beiträge u.a.m.).

Die **Impacts** sind die bei den Adressaten der Massnahmen bewirkten Änderungen.

Der **Vollzug** (=Output und Impacts) umfasst das Zusammenspiel der Aktivitäten der Verwaltung einerseits und der durch die staatlichen Massnahmen anvisierten Adressaten andererseits. Massgebendes Kriterium für die Qualität des Vollzugs ist die *Effektivität*. Es geht dabei um den Grad der Übereinstimmung zwischen dem rechtlich geforderten bzw. dem geplanten Verhalten von Vollzugsträgern (*Output*) und/oder Adressaten und deren beobachtbarem Verhalten (*Impacts*). Je nach Massnahmentyp stellt sich die Frage der Effektivität in besonderer Weise: *Befolgungsgrad* bei Geboten und Verboten, *Vollzugsgrad* bei öffentlichrechtlichen Normen, die der Umsetzung durch Verwaltungsinstanzen bedürfen (z.B. Erteilung von Bewilligungen), *Grad der Inanspruchnahme* bei den privatrechtlichen Normen mit fakultativem Charakter und bei Finanzhilfen, *Beachtungsgrad* bei Informationen und Empfehlungen. Die Effektivität ist eine notwendige, aber nicht hinreichende Voraussetzung der Wirksamkeit.

Die **Outcomes/Wirkungen** sind die auf eine Massnahme zurückführbaren Veränderungen in Wirtschaft und Gesellschaft. Zwei Kriterien sind hier massgebend. Das erste ist die *Wirksamkeit*: Dieses Kriterium bezieht sich auf das Verhältnis zwischen den beabsichtigten und den tatsächlich eingetretenen Wirkungen. Die Feststellung und Beurteilung der Wirksamkeit setzt die Bestimmung der Ziele der staatlichen Massnahmen voraus.

Das zweite Kriterium ist dasjenige der *Wirtschaftlichkeit* oder *Effizienz*: Es bezieht sich auf das Verhältnis von Output bzw. erzielten Wirkungen und den dafür eingesetzten Ressourcen (outputbezogene bzw. outcomebezogene Wirtschaftlichkeit). Eine bessere Wirtschaftlichkeit lässt sich entweder dadurch erreichen, dass zur Erzielung eines bestimmten Outputs bzw. bestimmter zielbezogener Wirkungen/Outcomes weniger Ressourcen eingesetzt werden oder dass mit den eingesetzten Ressourcen ein Mehr an Output bzw. zielbezogenen Wirkungen/Outcomes erreicht wird. Unter Wirtschaftlichkeit wird in der Evaluationsforschung oft auch das in monetäre Grössen ausgedrückte Verhältnis zwischen Nutzen/Wirkungen und den dafür eingesetzten Ressourcen/Mitteln verstanden.

Es ist zu beachten, dass sich das Kriterium der *Wirksamkeit* auf Wirkungen bezieht, die aus der staatlichen Massnahme resultieren. Idealerweise gibt das Kriterium der Wirksamkeit den Beitrag zur Erreichung eines oder mehrerer Ziele an, der auf die Massnahme zurückzuführen ist. Es setzt deshalb eine kausale Zuordnung zwischen staatlicher Massnahme und Wirkungen voraus (vgl. Kap. 3.3 - 3.5). Ein blosser Vergleich zwischen gesetzten Zielen und erreichten Werten - ohne kausale Beziehungen - wird in der Regel als *Zielerreichung* bezeichnet. Eine guter Zielerreichungsgrad kann z.B. aufgrund positiver Entwicklungen (in Wirtschaft oder Gesellschaft) erzielt werden, ohne dass die Massnahme überhaupt Wirkungen zeitigt. Eine Analyse der Zielerreichung, z.B.

auf Grund von Outcome-Indikatoren ist meist sehr viel einfacher vorzunehmen als eine Analyse der Wirksamkeit.

2.4 Welchen Zeithorizont decken Evaluationen ab?

Nach dem herrschenden Sprachgebrauch wird unter Evaluation oft die Auswahl eines Produkts im Hinblick auf eine **spätere** Verwendung verstanden. Eine Dienststelle evaluiert beispielsweise Workstations auf ihre Verwendbarkeit oder das Eidgenössische Militärdepartement evaluiert, welches Militärvelo für die nächsten Jahrzehnte truppentauglich ist. In der Regel erfolgt eine solche Evaluation auf der Basis **bestehender** Modelle oder allenfalls Prototypen. Diese werden entweder anhand von Produktbeschreibungen (Workstations) miteinander verglichen, oder es werden Prototypen einem Feldtest unterworfen (Militärvelos).

Im Unterschied zu EDV-Anlagen oder Militärvelos wissen wir bei staatlichen Massnahmen oft nur ungenau, wie das "Produkt" beschaffen ist, nämlich wie der Vollzug abläuft und welche Wirkungen die Massnahme zeitigt. Diese Informationen können nur durch Erfassung der bisher abgelaufenen Prozesse gewonnen werden. Evaluationen haben deshalb sehr oft **retrospektiven Charakter** (= ex post Evaluationen).

Es handelt sich dabei um **empirische** Untersuchungen, d.h. Untersuchungen, welche Sachverhalte in ihren konkreten Ausprägungen zu erfassen versuchen. Evaluationen unterscheiden sich damit von Beurteilungen staatlicher Massnahmen, die aufgrund blosser theoretischer Überlegungen getroffen wurden, z.B. durch Ökonomen mit Bezug auf volkswirtschaftliche Theorien. Evaluationen sind in der Regel nicht auf die Bildung von Theorien oder auf das Herausarbeiten eines breit anwendbaren Erfahrungswissens (z.B. über Instrumente staatlichen Handelns) ausgerichtet, aber können zu beidem Beiträge leisten. Während Evaluationen lange Zeit primär als Instrument zur Feststellung von Wirkungen verstanden wurden, wird in der Evaluationsforschung seit Ende der achtziger Jahre der Theoriebezug wieder stärker hervorgehoben.

Der retrospektive Charakter von Evaluationen ist kein Hindernis für deren Verwendung bei der Änderung und/oder Neukonzipierung einer Massnahme, im Gegenteil. Die Kenntnis der Effektivität, Wirksamkeit und Wirtschaftlichkeit der in Kraft stehenden Massnahme erlaubt, Anhaltspunkte über die Notwendigkeit und die Richtung einer Änderung bzw. Neugestaltung zu erhalten. Diese Kenntnisse erleichtern zudem die Abschätzung der mutmasslichen Wirkungen einer in Auswahl befindlichen neuen Massnahme. Viele

Evaluationen haben zwar retrospektiven Charakter, dienen aber als **Grundlage für künftiges Handeln**.

Prospektive (oder ex ante) **Evaluationen** bezwecken die Abschätzung der Wirkungen oder der Wirtschaftlichkeit möglicher oder vorgesehener Massnahmen. Es handelt sich nicht mehr um empirische Untersuchungen, sondern um Projektionen bestehender Kenntnisse in die Zukunft. Solche Evaluationen können eher impressionistischen Charakter haben. Sie können sich bestimmter Techniken bedienen wie beispielsweise der Szenariotechnik. Sie können aber auch auf ausgefeilten Wirkungsmodellen beruhen. In Sachgebieten, in welchen erprobte Theorien über das Wirkungsgefüge bestehen (v.a. in der Wirtschafts-, Landwirtschafts-, Energie- und Umweltpolitik) können die entsprechenden Modelle sowohl für die retrospektive Darstellung als auch (bei Annahme bestimmter Rahmenbedingungen) für die Simulation der Wirkungen von geänderten Massnahmen verwendet werden können. Ganz allgemein wird eine engere Verknüpfung retrospektiver und prospektiver Evaluationen angestrebt. Evaluationen der Wirtschaftlichkeit werden recht häufig prospektiv angewendet, um die Zweckmässigkeit grosser Investitionsvorhaben (z.B. Strassenbauten, Eisenbahnstrecken) abzuschätzen.

→ Näheres über den Einsatz von Evaluationen zur Änderung oder Neugestaltung von Massnahmen finden Sie in Kapitel 1.5.

2.5 Was kosten Evaluationen und wie lange dauern sie?

Evaluationen sind in der Regel zeitlich und finanziell aufwendig. Für eine eingehende Studie (Evaluation der Wirkungen oder der Wirtschaftlichkeit) muss mit einem Zeitaufwand von 6 Monaten bis 2 Jahren und mit Aufwendungen von 50'000 bis 300'000 Franken gerechnet werden. Bei einem klar eingegrenzten und nicht zu komplexen Untersuchungsgegenstand können mit einem Zeitaufwand von rund 6 Monaten und einem Budget von unter 100'000 Franken durchaus brauchbare Ergebnisse erreicht werden. Noch bescheidenere Fragestellungen, z.B. einfachere Vollzugsabklärungen kosten z.T. unter 50'000 Franken. Ca. je ein Drittel der Kosten fällt für die Vorbereitung und Feinplanung des Projekts, für die eigentliche Durchführung und für die Umsetzung (schriftliche, mündliche Präsentationen) an.

Sozialwissenschaftliche Forschung ist nicht billig, weil sie in der Regel einen recht hohen personellen Einsatz erfordert. Was die Forschungstechniken (vgl. Kap. 3.6) angeht, sind persönliche Interviews (z.B. Tiefeninterviews), breit angelegte Umfragen, statistische Auswertungen mit aufwendiger Daten

aufbereitung bzw. neu zu erarbeitenden ökonometrischen Modellen vergleichsweise teuer; Dokumentenanalysen, Literaturrecherchen, telephonische oder schriftliche Umfragen (z.B. Zielgruppenbefragung) oder einfachere statistische Auswertungen sind umgekehrt vergleichsweise billig. Nicht zu unterschätzen ist der finanzielle Aufwand für Umsetzungsmassnahmen (z.B. in Form von Weiterbildungsveranstaltungen oder in Form von ansprechend gestalteten Broschüren). Eine finanziell aufwendigere Strategie kann sich durchaus rechtfertigen, wenn sie bessere und zuverlässigere Antworten auf Ihre Fragen verspricht.

Evaluationsteams sind durchaus in der Lage, Zeit- und Kostenpläne innerhalb einer 10%-Fehlermarge einzuhalten. Oft zu wenig eingeplant werden Kosten für die Betreuung von Begleitgruppen, für Übersetzungen sowie generell für die Umsetzung.

Weiterführende Literatur

Die beste Einführung in den Gegenstand und die Methodik von Evaluationen geben Rossi/Freeman 1993; eine frühere und z.T. noch mit Unzulänglichkeiten behaftete Auflage dieses Werkes wurde auch ins Deutsche übersetzt (Rossi/Freeman/Hofmann 1988). Aus der Perspektive des Verwaltungsmanagements nach wie vor sehr lesenswert ist Wholey 1983. Zu den Vor- und Nachteilen interner bzw. externer Forschung äussert sich AGREB 1990, S. 11-14.

3. Welcher Strategien und Forschungstechniken bedienen sich Evaluationen?

3.1 Möglichkeiten zur Klassifikation von Evaluationsstrategien

Es ist im Rahmen dieses Leitfadens nicht möglich, die Methodik von Evaluationen eingehend vorzustellen. Hierfür wird im Rahmen des NFP 27 im Jahre 1995/96 ein spezielles Handbuch erscheinen; zudem gibt es verschiedene Lehrbücher (vgl. Literaturangaben zu den einzelnen nachfolgenden Kapiteln).

Ein Grossteil der Literatur über Evaluationen ist indessen detaillierter formuliert, als dies für Auftraggeber von Evaluationen nötig wäre. Zudem sind nur wenige Werke auf Länder wie die Schweiz zugeschnitten. Deshalb soll nachfolgend auf knappem Platz ein Grundverständnis für die wichtigsten Evaluationsstrategien und Forschungsmethoden geschaffen werden.

Es gibt verschiedene Möglichkeiten der Darstellung. Jede geht von impliziten Wertvorstellungen über geeignete und weniger geeignete Methoden aus. Es wäre falsch, die nachfolgenden Ausführungen als "Kochbuch" über Evaluationsstrategien zu verstehen und sklavisch zu befolgen. Jede Darstellungsweise - auch die vorliegende - hat ihre Vor- und Nachteile, ihre Möglichkeiten und Grenzen. Wir verstehen sie als Ausgangs-, nicht aber als Endpunkt für ein vertieftes Verständnis der Methodik von Evaluationen.

In der Evaluationsforschung gibt es eine Vielzahl von Klassifikationsversuchen. In den USA ist namentlich die Unterscheidung in experimentelle, quasi-experimentelle und "weitere" Methoden vorherrschend (z.B. Rossi/Freeman 1993). Einer Übertragung dieser Darstellung auf europäische Verhältnisse steht indessen im Wege, dass hier seltener als in den USA eigentliche Experimente durchgeführt werden und dass der Grossteil der amerikanischen Evaluationsmethodik somit nur selten praktisch angewendet werden kann. In den USA ebenfalls bedeutsam sind die von der Evaluation Research Society entwickelten Standards (ERS Standards Committee 1982), welche von der Unterscheidung in sechs verschiedene Evaluationstypen (prospektive Evaluationen, Evaluierbarkeitsprüfung, Vollzugsstudien, Wirkungsstudien, Monitoring, Metaevaluationen) ausgehen. Wir lehnen uns hier an diese sowie zusätzlich an die vom US General Accounting Office (1991a) entwickelte Systematik an, die durch praktische Erfahrungen mit Evaluationen geprägt ist und sich gut auf die Situation von Ländern wie der Schweiz übertragen lässt.

3.2 Begriffe: Evaluationsstrategie und Forschungstechnik

Wir unterscheiden zwischen **Evaluationsstrategie**, **Forschungstechnik** (oder **Forschungsmethode**) und **Evaluationsdesign**.

- ♦ Eine **Evaluationsstrategie** ist der Hauptansatz der Untersuchung zur Beantwortung der gestellten Fragen im Hinblick auf den Vollzug oder die Wirkungen der staatlichen Massnahme. Sie ist die konzeptionelle Antwort auf die Untersuchungsfrage.
- ♦ Die **Forschungstechnik** ist die gewählte sozial- (allenfalls auch naturwissenschaftliche) Methode zur Gewinnung und Auswertung der Daten im Hinblick auf die Beantwortung der gestellten Untersuchungsfrage.
- ♦ Wie im nächsten Kapitel gezeigt wird, kann bei normativen und deskriptiven Fragen die Evaluationsstrategie weitgehend aus der Verwendung bestimmter Forschungstechniken bestehen. Bei kausalerorientierten Fragen bedarf es in der Regel einer impliziten oder expliziten Vergleichsbasis, auf der die Evaluationsstrategie basiert.
- ♦ Das **Evaluationsdesign** ist die konkrete Untersuchungsanlage zur Beantwortung der gestellten Fragen. Es enthält Angaben über die Evaluationsstrategie und die gewählten Forschungstechniken.

3.3 Überblick über die Evaluationsstrategien

3.31 Deskriptive, normative und kausalerorientierte Fragen

Evaluationen des Vollzugs oder der Wirkungen können deskriptive, normative oder kausalerorientierte Fragestellungen beinhalten.

Bei *deskriptiven* Fragen geht es, wie der Name sagt, um die Beschreibung von Tätigkeiten, Objekten oder Sachverhalten. Die von der Verwaltung vorgenommene Vollzugsberichterstattung beispielsweise hat in der Regel beschreibenden Charakter.

Bei *normativen* Fragen geht es um Vergleiche zwischen Sein und Sollen. Die von Kontrollorganen vorgenommenen Überprüfungen gehen in der Regel von normativen Fragestellungen aus.

Bei *kausalerorientierten* Fragen geht es darum, zwischen Tätigkeiten und Zuständen (Ursache-Wirkungs-)Beziehungen herzustellen. Wissenschaftliche Untersuchungen sind oft von solchen Fragen geleitet.

Am Beispiel der Gewässerschutzpolitik werden Beispiele für diese drei Arten von Fragen gegeben:

Abbildung 4 Deskriptive, normative und kausalerorientierte Fragen

Fragestellung	Vollzug	Ziele bzw. Wirkungen
deskriptiv	Wie viele Kläranlagen sind in der Schweiz in Betrieb?	Wie hat sich die Gewässergüte in den letzten 10 Jahren entwickelt? ⁶
normativ	Werden im Kanton X im Rahmen des Baubewilligungsverfahrens die Gewässerschutzvorschriften eingehalten? ⁷	In wie vielen Gewässern wird die vorgeschriebene Gewässergüte erreicht? ⁸
kausalerorientiert	Inwieweit hat sich der Abbau von Subventionen auf den Bau von Kläranlagen ausgewirkt?	Inwieweit hat die Gewässerschutzpolitik zur Verbesserung der Gewässergüte beigetragen? ⁹ Welche Nebenwirkungen auf die wirtschaftliche Entwicklung haben sich ergeben?
normativ und kausalerorientiert	Auf welche Ursachen sind die festgestellten Vollzugsdefizite (bzw. Vollzugserfolge) im Gewässerschutz zurückzuführen?	Weshalb konnte in den dicht besiedelten Gebieten die vorgeschriebene Gewässergüte noch nicht erreicht werden?

3.32 Fragestellung und Evaluationsstrategie

Die Art der Fragestellung - deskriptiv, normativ oder kausalerorientiert - hat massgebenden Einfluss auf die Evaluationsstrategie. Die Zusammenhänge sind in Abbildung 5 dargestellt und werden in den nachfolgenden Kapiteln erläutert.

Abbildung 5 Fragestellungen und Evaluationsstrategien

Fragestellung	Evaluationsstrategie	Kombinationen
---------------	----------------------	---------------

⁶Es handelt sich um die Beschreibung der Entwicklung anhand eines wichtigen Indikators. Inwieweit Veränderungen der Gewässergüte auf die Gewässerschutzpolitik oder auf andere Faktoren (z.B. Konjunkturlage) zurückzuführen sind, bleibt offen.

⁷Diese Frage betrifft das Kriterium der Effektivität.

⁸Auch hier wird nur die Zielgrösse beleuchtet, und es wird offengelassen, ob und inwieweit die Gewässerschutzpolitik zum Resultat beigetragen hat. Die Frage betrifft das Kriterium der Zielerreichung.

⁹Diese Frage betrifft das Kriterium der Wirksamkeit.

<i>Deskriptive Fragen</i>	keine spezifische Evaluationsstrategie (siehe unter Forschungstechniken)	Kombination mit kausalerorientierten und/oder normativen Fragestellungen
<i>Normative Fragen</i>	Soll-Ist Vergleich (Messung der Abweichung)	Kombination mit kausalerorientierten und/oder deskriptiven Fragestellungen
<i>Kausalerorientierte Fragen</i>	<i>Explizit vergleichende Strategien</i>	
	Kontrollgruppen Experimente unechte Kontrollgruppen Regressions-Diskontinuitäts-Analysen	Kombination mit Längsschnitt-Vergleich
	Querschnitt-Vergleich	Kombination mit Längsschnitt-Vergleich
	Längsschnitt-Vergleich (Vorher-nachher Vergleich)	Kombination mit Querschnitt-Vergleich
	<i>Implizit vergleichende Strategien, z.B.</i> Beurteilung durch Adressaten Befragung von Schlüsselpersonen Lückenlose Rekonstruktion des Ablaufs Rückgriff auf Schattenpreise	

3.4 Evaluationsstrategien für deskriptive und normative Fragen

Obwohl bei Evaluationen normative und kausalerorientierte Fragestellungen die hauptsächliche Herausforderung bilden, sind in jeder Evaluation zumeist auch **deskriptive Fragen** zu beantworten. Sie können die unterschiedlichsten Formen annehmen: Wie lange braucht ein Sachbearbeiter durchschnittlich bis zum Abschluss eines Dossiers? Wie läuft ein Beratungsgespräch für Arbeitslose ab? Wie ist die Einkommenssituation der Empfänger landwirtschaftlicher Direktzahlungen? Welche Personen besuchen städtische Freizeitzentren?

Deskriptive Fragestellungen können durch die Beschreibung der relevanten Sachverhalte und Tätigkeiten beantwortet werden. Eine spezielle Evaluationsstrategie ist nicht nötig. Eingesetzt wird das ganze Spektrum sozialwissenschaftlicher Forschungstechniken (vgl. Kap. 3.6). Universitätsinstitute und private Büros haben ausreichende Erfahrungen in der Bearbeitung deskriptiver

Fragestellungen. Sie bearbeiten solche Fragen kompetent, auch wenn sich hin und wieder schwierige Fragen stellen wie beispielsweise die folgenden: Geben kantonale Steuerdaten zuverlässige Auskunft über die Einkommenssituation der AHV-Rentner? Wie lässt sich die ökonomische und soziale Situation Alleinerziehender am besten ermitteln? Können die im Kanton X ermittelten Erkenntnisse über das Verfahren der Umweltverträglichkeitsprüfung auf die ganze Schweiz übertragen werden? Wie viele ausländischen Personen arbeiten illegal in der Schweiz? Die aus unserer Sicht heikelsten Fragen betreffen somit die Repräsentativität (und damit auch die Übertragbarkeit) von Daten und Erkenntnissen. Eine Knacknuss für sozialwissenschaftliche Forschung sind zudem "verborgene" Bevölkerungsteile (z.B. illegal Eingereiste) und verbotenes oder gesellschaftlich geächtetes Verhalten (z.B. Kindsmisbrauch, illegale Beschäftigung, Steuerhinterziehung, Drogenkonsum).

Die Beantwortung **normativer Fragen** gehört zum Inhalt jeder Evaluation als bewertender Tätigkeit. Bei einer solchen Frage geht es um den Vergleich zwischen Soll- und Ist-Werten und darin liegt eigentlich bereits die geeignete Evaluationsstrategie. Die Sozialwissenschaften haben im Hinblick auf die Überprüfung von Theorien und generell im Bereich von Messungen ausreichende Erfahrungen, um normative Fragen kompetent zu beantworten.

Die Beantwortung normativer Fragen setzt idealtypisch folgende Schritte voraus (die im Rahmen konkreter Evaluationen nicht in jedem einzelnen Fall und in dieser Reihenfolge durchgeführt werden):

- Ermittlung der Vorgaben oder Ziele: Die Ziele können an sich durch Rückgriff auf rechtliche Erlasse und die dazugehörigen Dokumente (Botschaften, parlamentarische Debatten) ermittelt werden. Diese sind in der Regel aber recht vage gefasst, so dass auf weitere Materialien (politische Verlautbarungen, Wertvorstellungen der Betroffenen u.a.m.) zurückgegriffen werden muss. Hilfestellung bei der Zielermittlung können namentlich die Regeln rechtlicher Auslegung bieten. Gegebenenfalls kann das Evaluationsteam hypothetische Vorgaben/Ziele zur Überprüfung einsetzen. Denkbar ist auch, Sachverhalte anhand mehrerer Ziele zu überprüfen.
- Operationalisierung der Vorgaben/Ziele: Letztere werden in bestimmte Leistungsziele umgesetzt, damit sie mit konkreten Sachverhalten verglichen werden können. Das Ziel der Preisniveaustabilität wird beispielsweise mit einer Teuerungsrate von maximal 3 Prozent oder das Ziel der Deckung des Existenzminimums mit einem Jahreseinkommen von Fr. 30'000 für eine vierköpfige Familie in ländlichen Verhältnissen festgelegt.
- Messung: Die Leistungsziele als Soll-Werte werden mit konkreten Sachverhalten (Ist-Werte) verglichen.

- Gesamtbilanz: Die Ergebnisse von mehreren Messungen in Bezug auf unterschiedliche Leistungsziele sind in einer Gesamtbeurteilung zusammenzufassen.

Eine typische normative Frage ist diejenige nach der Effektivität (Vollzugskonformität): Haben die Behörden und die Adressaten der Massnahmen die rechtlichen Bestimmungen und die weiteren Vorgaben befolgt?

Eine weitere normative Frage ist diejenige nach der Zielerreichung. Sie setzt einen Vergleich zwischen Zielen und Ist-Zustand voraus.

Die Frage der Wirksamkeit ist ebenfalls eine normative Frage, die indessen zusätzlich einen kausalen Aspekt aufweist (s. unten). Der kausalen Aspekt normativer Fragestellungen in Bezug auf die Wirksamkeit und Wirtschaftlichkeit wird bei der Festlegung der Evaluationsstrategie oft vernachlässigt. Mit anderen Worten: Es wird oft nur die Zielerreichung und nicht die Wirksamkeit ermittelt, wie an folgendem hypothetischem Beispiel gezeigt werden soll.

Beispiel: Zur Senkung des Energiekonsums sollen spezielle Eierkocher eingesetzt werden, die im Rahmen einer Kampagne verbilligt abgegeben werden. Ziel der Kampagne ist es, 1 Million Eierkocher abzusetzen (Kriterium der Effektivität) und damit den Energieverbrauch um 0.1 Prozent zu senken (Kriterium der Wirksamkeit). Nach einem Jahr zeigt sich, dass beide Kriterien erfüllt werden konnten. Eine vertiefte (kausale) Evaluation ergibt dann aber, dass nur jeder zehnte verkaufte Eierkocher überhaupt eingesetzt wird, da die Bestimmung der erforderlichen Kochdauer Probleme bereitet und dass nur eine Senkung des Energieverbrauchs um 0.01 Prozent erreicht werden konnte. Die effektive Reduktion des Energieverbrauchs war auf andere Faktoren (milder Winter) zurückzuführen.

Evaluationsstrategien zur Beantwortung kausaler Fragen (s. unten) berühren implizit immer auch normative Fragen, ohne dass sich alle Evaluationsteams dieser Tatsache bewusst sind. Implizite und explizite Vergleiche zur Ermittlung der Wirkungen setzen durch den Gegenstand des Vergleichs (Kontrollgruppe, zu vergleichendes Gebiet, Zeitpunkt des Vergleichs) die Messlatte unterschiedlich hoch an.

Beispiel: Die auf einem Experiment beruhende Evaluation einer neuen Therapieform für Gefangene kann sehr unterschiedliche Resultate je nach Ausgangssituation der Gefangenen ergeben. Werden bei den Gefangenen bereits andere Therapien eingesetzt, dann ist möglich, dass nur ein geringer Erfolg resultiert. Bestehen hingegen keine anderen Therapiemöglichkeiten, ist die Chance höher, dass sich die Einführung der Therapie als Erfolg erweist.

Umgekehrt tauchen bei primär deskriptiven oder kausalen Untersuchungen ab und zu heikle Fragen in Bezug auf normative Aspekte auf. Bei zwei im Rahmen des NFP 27 durchgeführten Evaluationen trat klar zutage, dass

im Vollzug von den Behörden und/oder den Normadressaten rechtliche Vorschriften nicht eingehalten wurden. Z.T. basierten diese Erkenntnisse auf vertraulichen Informationen. In beiden Projekten wurden im Schlussbericht diese Tatsachen nur in verklausulierter Form erwähnt, weil nicht das Aufdecken individueller Fehler, sondern die Darstellung der Erfolgsbedingungen der betreffenden Massnahme und z.T. auch Lernprozesse im Vordergrund standen. Evaluationsteams müssen sich namentlich bei rechtlich stark geregelten Fragen (z.B. Baubewilligungsverfahren) bewusst sein, dass rechtliche Vorgaben und effektiver Vollzug auseinanderklaffen können und dass diese Tatsache den Zugang zu Informationen erschweren kann.

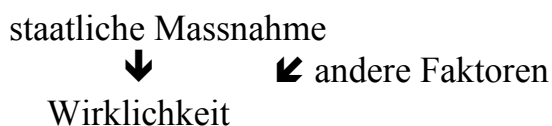
3.5 Evaluationsstrategien für kausalorientierte Fragestellungen

3.51 Explizit und implizit vergleichende Evaluationsstrategien

Kausaloorientierte Fragen betreffen Ursache-Wirkungs-Verhältnisse. Wie hat sich die Pflicht zur Umweltverträglichkeitsprüfung auf die Beteiligung von Umweltverbänden am Vollzug des Umweltschutzrechts ausgewirkt? Welche Auswirkungen haben die den Hoteliers gewährten Bürgschaften und Beiträge auf den Strukturwandel in der Hotellerie? Wie haben sich die strengen Grenzwerte für Hausfeuerungen auf die Wettbewerbssituation der schweizerischen Heizungsindustrie ausgewirkt? Dies sind typische Fragen, welche auf die Erhellung von Ursache-Wirkungs-Beziehungen abzielen.

Bei Evaluationen der Wirkungen und teilweise auch bei Evaluationen des Vollzugs handelt es sich um kausaloorientierte Fragen. Ihre Beantwortung ist in der Regel sehr schwierig, weil zwischen dem Beschluss und Inkraftsetzen einer Massnahme, ihrem Vollzug und den Wirkungen in der Gesellschaft meist ein sehr verschlungenes, der direkten Beobachtung selten zugängliches Wirkungsgefüge besteht. Idealerweise geht es darum, staatliche Massnahme und darauf zurückführbare Veränderungen (der Gesellschaft, der Wirtschaft, der Umwelt usw.) in eine zuverlässige Beziehung zu bringen und dabei andere Einflussfaktoren zu isolieren. Die Frage lautet: Was wäre anders, wenn die staatliche Massnahme nicht vollzogen würde (Vergleich "policy on"- "policy off")?

Abbildung 6 Zuordnung von staatlichen Massnahmen und gesellschaftlicher Wirklichkeit



Wie lassen sich in einer verschlungenen, komplexen Wirklichkeit die oft langsamen, oft mit Verzögerung auftretenden, z.T. schleichenden Veränderungen erfassen? Wie können bestimmte Ursachen von anderen Einflussfaktoren isoliert werden?

Die Antwort der Evaluationsforschung liegt vor allem in der Entwicklung vergleichender Evaluationsstrategien. Rückschlüsse über Ursache-Wirkungs-Beziehungen werden durch Variationen in Bezug auf die Untersuchungseinheiten (Kontrollgruppen-Vergleich), in Bezug auf räumliche Einheiten (Querschnitt-Vergleich) und durch Variation in zeitlicher Hinsicht (Längsschnitt-Vergleich) zu ziehen versucht. Diese auf einer expliziten Vergleichsbasis beruhenden Evaluationsstrategien werden in den drei nachfolgenden Kapiteln eingehend dargestellt.

Freilich können solche Strategien nicht bei allen Massnahmen angewendet werden. Wenn Massnahmen seit langer Zeit unverändert und in allen Kantonen und Gemeinden auf gleiche Weise vollzogen werden und zudem für alle in Betracht fallenden Adressaten gelten, sind Kontrollgruppen, Querschnitt- (mit Ausnahme internationaler Vergleiche) und Längsschnitt-Vergleiche nicht zweckmässig bzw. möglich.

Sofern in solchen Fällen nicht vollständig auf Hinweise über Wirkungen von Entscheiden verzichtet werden soll, gilt es, aus der schwierigen Situation das Beste zu machen. Als zweitbeste Lösung können **implizit vergleichende Evaluationsstrategien** eingesetzt werden. Solche Evaluationsstrategien bedienen sich nicht empirisch bestehender zeitlicher (Längsschnitt-Vergleiche) und anderer (Kontrollgruppen- und Querschnitt-Vergleiche) Variationen, sondern greifen auf behelfsmässige Vergleiche in Form von hypothetischen Gedankenkonstruktionen, von induktiven Überprüfungen von Wirkungsmodellen und von Schattenpreisen zurück. Sie können unter gewissen Umständen durchaus vertretbare Ergebnisse bringen. Folgende Strategien zur Wirkungsermittlung bieten sich an:

- *Beurteilung durch Adressaten der Massnahme*

Die Adressaten (sowie allenfalls weitere Betroffene) der Massnahme können (durch schriftliche Umfragen, Tiefeninterviews, Hearings u.ä.) über ihre Einschätzung der Wirkungen befragt werden. Sie werden ihre Beurteilung auf der Grundlage früherer Erfahrungen (impliziter Längsschnitt-Vergleich), von Quervergleichen (v.a. mit Nachbarländern) oder einer hypothetischen Gedankenkonstruktion ("Wie wäre die Situation, wenn die Massnahme nicht bestehen würde?") vornehmen.

Es ist klar, dass die Antworten durch taktisches Verhalten (Antworten zu eigenem Vorteil) und durch Fehleinschätzungen beeinflusst werden können. Letztere können daraus entstehen, dass beispielsweise gegenintuitive Wirkungsverläufe (z.B. die Möglichkeit, dass Sozialhilfe die Eigeninitiative und die Selbstachtung der Empfänger vermindert) und Nebenwirkungen nicht in Betracht gezogen werden. Versierte Evaluationsteams können solche Fehlerquellen in gewisser Masse kontrollieren, indem sie entsprechende Argumente und Fragen in den Gesprächsprozess einbringen. Es wird aber nie völlig auszuschließen sein, dass Unwissen, Vorurteile und momentane Stimmungen die Antworten der befragten Adressaten prägen.

- *Befragung von Schlüsselpersonen*

Schlüsselpersonen wie Experten des Sachgebiets, Fachpersonen aus der Verwaltung und Vertreter von Interessengruppen können eine weitere Quelle von Informationen über Ursache-Wirkungs-Beziehungen bilden. Ihre Kenntnis des Prozesses soll dazu beitragen, die Wirkungen zu rekonstruieren. Insbesondere Schlüsselpersonen im Vollzug der Massnahmen können wertvolle Kenntnisse über zentrale Aspekte des Wirkungspfades haben. Auch hier erfolgt die Wirkungsermittlung z.T. aufgrund von hypothetischen Gedankenkonstruktionen.

Wie die Adressaten (s.o.) können aber auch Experten und Schlüsselpersonen des Vollzugs Fehleinschätzungen unterliegen. Ihre Stellungnahme kann durch momentane Stimmungen (z.B. Euphorie oder Skepsis über bestimmte Entwicklungen) beeinflusst werden. Gerade in eng miteinander kommunizierenden Netzwerken können einzelne Akteure kraft ihrer institutionellen Macht oder kraft ihrer Persönlichkeit die Meinungen prägen, so dass auch breiter abgestützte Befragungen eine recht einheitliche Meinung zutage fördern.

Versierte Evaluationsteams sind sich dieser Gefahren bewusst. Sie suchen gezielt nach alternativen Erklärungen und konfrontieren die Befragten damit. In iterativen Verfahren werden die in einer ersten Runde ermittelten Ergebnisse den Schlüsselpersonen nochmals unterbreitet, um sie allenfalls zu einem Überdenken ihrer Meinung anzuregen. Gegebenenfalls kann der Evaluationsprozess auch als gemeinsamer Lernprozess ausgestaltet werden, bei welchem darauf geachtet wird, dass jede Meinung, ungeachtet der hierarchischen Position der jeweiligen Person, gleich viel gilt. Freilich wird es auch im Rahmen solcher sorgfältig ausgestalteter Evaluationsprozesse kaum gelingen, aus gegenläufigen und subtilen Ansichten über das Wirkungsgefüge ein zuverlässiges Gesamtergebnis zu rekonstruieren.

- *Lückenlose Rekonstruktion des Ablaufs/eingehende Fallstudie(n)*

Bereits die Befragung von Schlüsselpersonen kann durch das Bestreben geprägt sein, das Wirkungsgefüge möglichst präzise zu rekonstruieren.

Letzteres ist indessen eine eigenständige Option, welche sich verschiedener Datenbeschaffungstechniken (Dokumentenanalysen, Umfragen, Tiefeninterviews, quantitative Datenanalysen u.a.m.) bedient. Es geht um die lückenlose Aufarbeitung eines Falls (vgl. Kap. 3.62). Diese umfassenden Kenntnisse erlauben es, gewisse Wirkungen auf induktivem Wege auszu-schliessen, sofern sich im Ablauf nicht die entsprechenden Indizien, die für das Eintreten der Wirkungen Voraussetzung sind, finden lassen. Freilich bleibt dabei noch ein recht breiter Raum von Möglichkeiten, und es ist nur schwierig abzuschätzen, wie sich die einzelnen Glieder der Wirkungskette zu einem Ganzen, der Wirkung, zusammenfügen.

- *Rückgriff auf Schattenpreise*

Als Behelf für eine Ermittlung der Wirkungen dienen in vom ökonomischen Denken geprägten Evaluationen Marktpreise bzw. Schattenpreise als Schätzungen für theoretisch auf dem Markt erzielbare Leistungen. So können beispielsweise die ökonomischen Wirkungen bestimmter Strassenbauten anhand der Wertsteigerungen der von ihnen direkt oder indirekt erschlossenen Grundstücke approximativ bestimmt werden. Oder es kann untersucht werden, ob und inwieweit sich eine Reduktion der Lärmbelastung in der Umgebung eines Flugplatzes auf die Bodenpreise der betroffenen Gemeinden niederschlägt.

Freilich ist der Markt nur ein sehr unvollkommenes Instrument zur Wirkungsermittlung. Die ökonomischen Stimmzettel sind sehr ungleich verteilt. So kann sich beispielsweise die Zerschlagung der innerstädtischen Drogenszene ökonomisch positiv auf die Umsätze der Geschäfte der unmittelbaren Nachbarschaft niederschlagen, während die aus der Verlagerung der Drogenszene bei den Drogensüchtigen oder ihren Familien anfallenden Kosten unberücksichtigt bleiben. An sich kann die Analyse sukzessive verfeinert werden, um auch solchen Einwänden Rechnung zu tragen. Dennoch bleiben Unwägbarkeiten bestehen, die damit zusammenhängen, dass "Marktpreise" von bestimmten institutionellen Gegebenheiten (z.B. Weltmarktpreise, Kartellpreise, Monopolpreise) abhängig sind und dass diese Zufälligkeiten das Ergebnis der Evaluation mitprägen. Ausserdem ist festzuhalten, dass eine monetäre Quantifizierung von bestimmten Sachverhalten meist sehr schwierig ist. Vor allem Evaluationen der (mutmasslichen) Wirkungen ökonomischer Deregulierung bedienen sich solcher Preiskonzepte.

Implizit-vergleichende Evaluationsstrategien haben somit ihre Gefahren und ihre Chancen im Hinblick auf die Beantwortung kausaloientierter Fragen. In Kap.3.6 finden Sie vereinzelt weitere Hinweise über die Eignung von Forschungsmethoden, Aufschlüsse über Ursache-Wirkungs-Beziehungen zu geben.

3.52 Kontrollgruppen-Vergleich

Die Evaluationsstrategie des Kontrollgruppen-Vergleichs (auch Vergleich mit/ohne Massnahme) dient der Beantwortung kausaloientierter Fragen. Die für eine Evaluation der Wirkungen typische Fragestellung lautet in der Regel: Was wäre in der Wirklichkeit anders, wenn die Massnahme nicht durchgeführt würde? Genau auf diese Frage zielt der Kontrollgruppen-Vergleich. Er wird in der Evaluationsforschung unter der Rubrik "experimentelle und quasi-experimentelle Methoden" behandelt. Wir erwähnen hier folgende Möglichkeiten des Kontrollgruppen-Vergleichs:

Experiment (mit Zufallsauswahl): Aus einer Grundgesamtheit wird eine zufällig ausgewählte Stichprobe von Personen einer bestimmten Massnahme (im Evaluationsjargon: "Behandlung") ausgesetzt. Die bei der Zielgruppe eingetretene Situation wird mit der Situation einer (wiederum zufällig ausgewählten) Kontrollgruppe verglichen, welche der Massnahme nicht ausgesetzt wird. Aus dem Vergleich der Situationen beider Gruppen wird auf die Wirkungen der Massnahme geschlossen.

Beispiel: Ziel eines Experiments ist die Erprobung einer neuen Therapieform für Gefangene. Unter den Gefangenen wird eine Umfrage durchgeführt, wer sich für die Behandlung interessiert. Unter den Interessenten wird eine Zufallsstichprobe von Personen ausgewählt, welche die Therapie beanspruchen können; die übrigen Angemeldeten dienen als Kontrollgruppe. Die Situation für beide Gruppen wird durch Fragebogen, Beobachtungen des Gefängnispersonals sowie mittelfristig durch Daten über die Rückfälligkeitsraten u.a.m. erfasst. Die Ergebnisse werden einander gegenübergestellt, was Rückschlüsse auf den Erfolg bzw. Misserfolg der neuen Therapie erlaubt.

Unechte Kontrollgruppen: Bei diesem Vorgehen werden Kontrollgruppen nicht durch Zufallsauswahl geschaffen, sondern es werden mittels statistischer Methoden aus "natürlich" vorkommenden zeitlichen, örtlichen und personalen Differenzen in der Anwendung der Massnahme Schlüsse auf die Wirkungen gezogen. Um die der Massnahme ausgesetzten Einheiten möglichst gut mit denjenigen zu vergleichen, welche der Massnahme nicht ausgesetzt sind, werden statistische Verfahren wie die Bildung von "Zwillingspaaren" ("matching"), welche sich in möglichst vielen wichtigen Aspekten gleichen, angewendet.

Beispiel: In einer vom NFP 27 mitfinanzierten Kurzevaluation wurde untersucht, wie sich die Unterstützung des landwirtschaftlichen Hochbaus

durch Investitionskredite und Subventionen im Kanton Zürich ausgewirkt hat. Hierzu wurde eine Gruppe stark subventionierter Landwirtschaftsbetriebe einer Gruppe wenig subventionierter Betriebe gegenübergestellt, und es wurden deren landwirtschaftliches Einkommen und deren Verschuldung miteinander verglichen. Zwischen beiden Gruppen ergaben sich deutliche Unterschiede. Die Untersuchung profitierte von der guten Datenlage im Landwirtschaftsbereich (Studie Schnyder 1993).

Oft wird der Kontrollgruppen-Vergleich durch einen Längsschnitt-Vergleich (für Ziel- und/oder Kontrollgruppe) ergänzt. Namentlich bei unechten Kontrollgruppen, welche methodisch anfechtbare Resultate liefern, können dann zuverlässigere Rückschlüsse auf die Wirkungen gezogen werden.

Beispiel: In der soeben erwähnten Untersuchung wurde auch ein Längsschnitt-Vergleich für die Gruppe von Betrieben, welche Investitionen in Stallbauten vornahmen, durchgeführt. Es zeigte sich, dass rund 35% der investierenden Betriebe ihre betriebswirtschaftliche Situation verbessern konnten; eine ähnlich grosse Zahl von Betrieben mit ähnlicher wirtschaftlicher Ausgangslage wurde indessen durch die Investition stark belastet, so dass sich ihr Einkommen kaum verbesserte oder gar verschlechterte.

Regressions-Diskontinuitäts-Analysen: Es handelt sich um ein neues Verfahren, um die Wirkungen einer Massnahme zu ermitteln, die nur für bestimmte Gruppen von Personen, die durch bestimmte Schwellenwerte von den übrigen Personen abgegrenzt sind, gilt. Das Verfahren setzt beim Evaluations-team eingehende statistische Kenntnisse voraus.

Beispiel: Bei der Umweltverträglichkeitsprüfung besteht bei Bauvorhaben, welche mindestens 200 Parkplätze vorsehen, eine Pflicht zur Erstellung eines Umweltverträglichkeitsberichts. Um die Auswirkungen der Umweltverträglichkeitsprüfung auf die Länge des Bauverfahrens zu ermitteln, kann davon ausgegangen werden, dass mit zunehmender Anzahl Parkplätze auch die Dauer des Bauverfahrens zunimmt. Ob die Pflicht zur Erstellung eines Umweltverträglichkeitsberichts massgeblichen Einfluss auf die Länge des Bauverfahrens hat, lässt sich daraus ablesen, ob in der Regressionsgeraden im Verhältnis Parkplatzgrösse/Dauer des Bauverfahrens bei 200 Parkplätzen eine Bruchstelle besteht oder nicht (vgl. Abb. 1 in Kap. 1.6).

Obgleich Kontrollgruppen-Vergleiche für die Evaluation der Wirkungen an sich eine sehr geeignete Evaluationsstrategie darstellen, ist ihre Anwendbarkeit in mehrerer Hinsicht beschränkt. Der Durchführung von Experimenten stehen, wie in Kap.1.6 erwähnt, gewisse ethische und rechtliche Prinzipien (Gleichheit vor dem Gesetz) entgegen. Wenn Massnahmen für die ganze Bevölkerung gelten und landesweit einheitlich angewendet werden, ist es oft schwierig, geeignete Kontrollgruppen zu finden, welche Rückschlüsse auf die

Wirkungen der Massnahme ermöglichen. In Verbindung mit anderen Evaluationsstrategien (v.a. Längsschnitt-Vergleich) kann indessen der (unechte) Kontrollgruppen-Vergleich aussagekräftigere Resultate ermöglichen.

Weiterführende Literatur

Rossi/Freeman 1993 und Mohr 1993 behandeln den Kontrollgruppen-Vergleich sehr eingehend.

3.53 Querschnitt-Vergleiche

Mit Querschnitt-Vergleichen wird eine Evaluation des Vollzugs und der Wirkungen einer staatlichen Massnahme durch den Quervergleich über räumliche Einheiten (Gemeinden, Kantone, Länder) hinweg angestrebt. In der Evaluationsforschung finden sich sehr wenig Angaben über Querschnitt-Vergleiche: Sie werden in der Regel unter dem Kontrollgruppen-Vergleich behandelt. Bei Kontrollgruppen-Vergleichen werden indessen primär Unterschiede zwischen Untersuchungseinheiten (z.B. Adressaten der Massnahmen wie Einzelpersonen und Betriebe) innerhalb eines homogenen Gebiets untersucht, während beim Querschnitt-Vergleich Ausprägungen im Vollzug und den Wirkungen von Massnahmen zwischen nicht notwendigerweise homogenen Gebieten (Ländern, Kantonen, Gemeinden) untersucht werden.

Die realitätsgerechte Berücksichtigung der von Gebiet zu Gebiet unterschiedlichen Voraussetzungen für die Durchführung einer Massnahme ist die hauptsächliche Herausforderung bei Querschnitt-Vergleichen. Es gilt, aus der je nach Gebiet unterschiedlichen Ausgestaltung einer bestimmten Massnahme Schlüsse auf deren Wirkungsmechanismen zu ziehen, wobei der Einfluss der jeweiligen Ausgangs- und Rahmenbedingungen (z.B. Wirtschaftskraft, demographische Faktoren, Familienstrukturen, föderalistische/zentralstaatliche Strukturen) auf das Funktionieren der Massnahme in die Beurteilung einbezogen werden muss. Es gibt verschiedene Vorgehensweisen, um dem Einfluss von Rahmenbedingungen auf die Wirkungsweise von Massnahmen Rechnung zu tragen:

- Es werden die in unterschiedlichen Gebieten (Ländern, Kantone, Gemeinden) in einem bestimmten Politikbereich getroffenen Massnahmen (gesetzliche Grundlagen, Programme, Vollzugstätigkeiten und -strukturen, Entwicklungstrends u.a.m.) eingehend dargestellt. Es handelt sich hier eigentlich um Fallstudien je nach Gebiet (zu den Einzelfallstudien und zu den multiplen Fallstudien, vgl. Kap. 3.62). Vereinzelt bleibt es bei der Darstellung der einzelnen Fälle, oft werden jedoch Querbeziehungen (v.a. im Hinblick auf Vollzugsaspekte) aufgezeigt. Z.T. wird auch versucht, aus dem

Quervergleich der Massnahmen und der damit gemachten Erfahrungen Schlüsse auf grundlegende Wirkungsmechanismen zu ziehen. Auf eine quantitative Auswertung des Verhältnisses Ausgestaltung der Massnahme - Ergebnisse wird indessen verzichtet. Dieses Vorgehen wurde in der Schweiz insbesondere im Rahmen von Vollzugsstudien (z.B. Delley et al. 1982, Linder u.a. 1990) gewählt.

Beispiel: Balthasar/Knöpfel ermittelten drei grundlegenden Modelle von Feuerungskontrollen in den Kantonen. Sie illustrierten das erste Modell (Feuerungskontrolle durch beamtete Personen) anhand des Kantons Luzern, das zweite Modell (Feuerungskontrolle durch Kaminfeger) anhand des Kantons Waadt und das dritte Modell (Feuerungskontrolle durch Revisionsfirmen) anhand des Kantons Basel-Stadt. Eine Beurteilung, ob eines dieser Modelle den andern gesamthaft überlegen ist, wurde nicht angestrebt (Studie Balthasar/Knöpfel 1994).

- Wird eine grosse Zahl von Gebieten in die Untersuchung einbezogen, so können auch quantitative Analysen des Verhältnisses der Ausgestaltung der Massnahme und wichtiger Wirkungsgrössen vorgenommen werden. Dabei gilt es sicherzustellen, dass die Analyse nicht durch andere Einflussfaktoren (z.B. Unterschiede Stadt-Land, wirtschaftsstarke/wirtschaftsschwache Gebiete) systematisch verzerrt wird.

Beispiel 1: Hausheer untersuchte im Rahmen einer Querschnitt-Untersuchung in mehr als 100 Gemeinden, ob und inwieweit sich die Lösung mit Sackgebühren im Vergleich zu herkömmlichen Lösungen auf die Menge des gemischten Kehrrechts und der Ergebnisse der Separatsammlungen (Papier, Metall) auswirkt. Der Autor kam zum Schluss, dass die Sackgebühr zu einer signifikanten Reduktion der Menge des gemischten Kehrrechts und zu einer signifikanten Erhöhung der Mengen bei Separatsammlungen führt. Er war sogar in der Lage nachzuweisen, dass die erwähnten Grössen auch durch die Höhe der Sackgebühr beeinflusst werden (Hausheer 1991).

Beispiel 2: Balthasar/Knöpfel untersuchten 419 Gemeinden in den Kantonen Luzern, Bern und Basel-Landschaft eingehend auf Faktoren, welche die Verbreitung neuer Heizanlagen beeinflussen. Es zeigte sich, dass namentlich im Kanton Luzern der Anteil moderner Heizungen recht hoch ist und dass dort ferner der Qualifikationsgrad der Heizungskontrolleure in den einzelnen Gemeinden einen signifikanten Einfluss auf die Erneuerung von Ölfeuerungen hat (Balthasar/Knöpfel 1994).

- Namentlich bei zentral gesteuerten Pilotversuchen wird angestrebt, Gebiete, in denen eine bestimmte Massnahme eingeführt wird, mit möglichst gleichartigen Gebieten, in denen dies nicht geschieht, zu vergleichen. Die Untersuchungsanlage ist hierbei ähnlich wie bei Experimenten (vgl. Kap. 3.52).

Beispiel: Zu Beginn der 80er Jahre galt in der Schweiz Tempo 60. Um die Auswirkungen einer Einführung von Tempo 50 abzuklären, wurde in vier Testgebieten Tempo 50 eingeführt. Als Kontrollgruppe dienten vier möglichst ähnliche Gebiete, in welchen Tempo 60 weiterhin galt. Die Auswirkungen der beiden Tempolimiten (auf die Unfallhäufigkeit, die Schwere der Unfälle, den Verkehrsfluss u.a.m.) wurden sorgfältig beobachtet und systematisch miteinander verglichen (Arbeitsgruppe Verkehrssicherheit 1983).

Querschnitt-Vergleiche können eine Mehrzahl sinnvoller Funktionen erfüllen. Bereits die deskriptive Darstellung unterschiedlicher Lösungen kann das Sensorium für neue Möglichkeiten schärfen. So wird beispielsweise in der Drogenpolitik sehr intensiv nach Wegen gesucht, um die Beschaffungskriminalität zu vermindern. Auch aus der Darstellung einzelner Fälle (Länder, Kantone, Gemeinden), die durch markante Erfolge oder Misserfolge gekennzeichnet sind, können Lehren über die Erfolgsbedingungen einer Massnahme gezogen werden. In solchen Fällen können Querschnitt-Vergleiche auch auf anderen Formen des Vergleichs (Kontrollgruppen-Vergleich, Längsschnitt-Vergleich), die in einzelnen Gebieten angewendet wurden, aufbauen. Am stärksten kann das Instrument der Querschnitt-Vergleiche dann zur Beantwortung kausalorientierter Fragen eingesetzt werden, wenn eine Mehrzahl von Gebieten auf bestimmte Ursache-Wirkungs-Beziehungen untersucht wird. Eine solche Untersuchungsanlage ist dann identisch mit derjenigen des Kontrollgruppen-Vergleichs. Freilich wird es nicht in jedem Fall möglich sein, solche gesicherten Erfahrungen auf ein anderes Gebiet zu übertragen, weil dort unter Umständen Rahmenbedingungen bestehen, die mit denjenigen, die in der Mehrzahl der untersuchten Gebietseinheiten bestehen, nicht übereinstimmen. Vor einer mechanistischen Übertragung fremder Erfahrungen ist somit zu warnen.

Weiterführende Literatur

Über die Eignung von Querschnitt-Vergleichen zur Übertragung von Erfahrungen von einem Gebiet auf das andere gibt Rose 1993 einen ausgezeichneten Einblick. In den meisten Projekten des NFP 27 wurden auch gewisse Querschnittsvergleiche zwischen Kantonen oder zwischen Gemeinden vorgenommen, so namentlich in Sangra 1992, Bierter/Binder 1993, Gerheuser/Schmid 1993, Balthasar/Knöpfel 1994, Hanser/Kuster/Cavelti 1994 und Hirsch et al. 1995.

3.54 Längsschnitt-Vergleich

Mit dem Längsschnitt-Vergleich werden die Wirkungen von Massnahmen im Zeitablauf zu ermitteln versucht. Der Vergleich setzt voraus, dass eine Massnahme neu in Kraft gesetzt (z.B. Verbot von Kaskadenverkäufen bei Liegenschaften), geändert (z.B. Änderungen im Eherecht), in der Dosierung variiert (z.B. Geldmengenpolitik der Nationalbank, Höhe der Forstsubventionen) oder aufgehoben (z.B. Bedürfnisklausel für das Gastgewerbe) wird. Es wird untersucht, welche Änderungen der zielrelevanten Grössen auf die Massnahmenänderung folgen. Ob diese kausal auf letztere zurückzuführen sind, bedarf in der Regel eingehender Prüfung (Möglichkeit zufälliger Parallelität; Möglichkeit, dass Änderung der Massnahme und der Zielvariablen auf eine dritte Grösse zurückgeführt werden kann u.a.m.). Wenn die Massnahmenänderung von weiteren Entwicklungen überlagert wird (Konjunkturentwicklung, Wertewandel, Sparmassnahmen der öffentlichen Hand u.a.m.), ist es in der Regel sehr schwer, die Nettowirkung der Änderung zu bestimmen. Deshalb wurde der Längsschnitt-Vergleich in der Evaluationsforschung z.T. nur als zweit- oder drittrangige Strategie betrachtet. Es gibt jedoch eine ganze Anzahl von Voraussetzungen, unter denen dieser Vergleich sehr geeignet ist. Wir erwähnen als Beispiele folgende Fälle:

- Die Massnahme wird **in mehreren Gemeinden, Kantonen oder Ländern neu eingesetzt**. Sofern überall die gleiche Entwicklung der zielbezogenen Grössen zu beobachten ist, kann der Zusammenhang zwischen Massnahme und Wirkungen als gut gesichert gelten.

Beispiel: In einer Studie der Arbeitsgruppe "Gesetzesevaluation" wurde untersucht, welche Auswirkungen die Einführung der Kehrriechtsackgebühr auf die Menge des Gemischtkehrriechts hat. In 11 der untersuchten 21 Gemeinden verminderte sich die Menge des Gemischtkehrriechts stark, in weiteren 4 Gemeinden verminderte sie sich schwach; die übrigen Gemeinden verfügten über keine detaillierten Zahlenangaben. Dass in fast allen Gemeinden ähnliche Ergebnisse beobachtet wurden, erhöhte die Zuverlässigkeit der Aussagen (Studie INFRAS 1990).

- Das **Ausmass** der Änderungen nach Inkrafttreten einer Massnahme ist sehr **gross**, und es gibt keine anderen Gründe, welche plausibel für diese Änderungen verantwortlich gemacht werden können.

Beispiel: Die Evaluation der AIDS-Kampagne basiert auf einem Längsschnitt-Vergleich. Angesichts der markanten Zunahme des Gebrauchs von Präservativen im Zeitablauf scheint es nicht nötig, den Netto-Effekt der AIDS-Kampagne auf das Verhalten im Vergleich zu andern bewusstseinsbildenden Faktoren (Krankheit prominenter Personen, medizinische Berichte, Gespräche im Bekanntenkreis) herauszufiltern.

- Es bestehen **theoretische Grundlagen** über den Bezug zwischen Massnahme und zielbezogenen Variablen sowie allfälligen anderen Einflussfak

toren. Die theoretischen Annahmen sind empirisch gut abgesichert. In diesem Falle werden die im Längsschnitt empirisch ermittelten Änderungen als zuverlässiges Bild der Wirklichkeit betrachtet, sofern sie mit der Theorie übereinstimmen.

Beispiel: Im Gewässerschutz sind die Zusammenhänge zwischen Gewässerreinigung und Wassergüte naturwissenschaftlich dermassen gut bekannt, dass der Bezug zwischen der Entwicklung der Zahl und des Reinigungsgrads der Kläranlagen einerseits und der Entwicklung der Wassergüte im Zeitablauf andererseits aufgrund der Stoffflüsse ohne weiteres rekonstruiert und die Netto-Wirkung der Gewässerschutzpolitik ermittelt werden können.

- Es ist möglich, den Einfluss der Massnahmenänderung auf bestimmte Einheiten (Teile der Bevölkerung, Betriebe u.a.m.) statistisch zu erfassen und damit zu **isolieren** von allfälligen Überlagerungen durch andere Entwicklungen. Der Längsschnitt-Vergleich kann in Kombination mit einer unechten Kontrollgruppe (s. vorangehender Abschnitt) recht aussagekräftige Ergebnisse zeitigen.

Beispiel: In einer Untersuchung der Wirkungen der Bürgschaften und Beiträge nach dem Bundesgesetz über die Förderung des Hotel- und Kurortskredites wurden die Betriebsdaten von unterstützten 48 Hotels im Längsschnitt ausgewertet. Diese Werte wurden mit dem Branchendurchschnitt gewichtet und somit implizit mit der Grundgesamtheit der Hotels verglichen, um den Einfluss von Konjunktur, Schneeverhältnissen u.a.m. herauszufiltern (Studie Hanser/Kuster/Cavelti 1994).

Eine Variante des Längsschnitt-Vergleichs ist der **Trend/Ist-Vergleich**. Sofern bestimmte zielrelevante Grössen einem Trend (z.B. Zunahme der Luftverschmutzung aufgrund der Verkehrszunahme) folgen, ergibt ein blosser Längsschnitt-Vergleich ein falsches Bild. Der Vergleich beinhaltet sinnvollerweise die Gegenüberstellung der Entwicklung, wie sie ohne Massnahmenänderung erfolgt wäre, und der Entwicklung im Gefolge der Massnahme. Nötig ist somit eine Prognose der Entwicklung ohne Massnahme (Status quo-Prognose), welche auf einer Trendextrapolation oder einer anderen Methode der Zukunftsabschätzung beruhen kann.

Weiterführende Literatur

Einen Einblick in die Längsschnittuntersuchungen geben Schmitz 1989 und Widmer 1991. Unter den Studien des NFP 27 sind Hanser/Kuster/Cavelti 1994 sowie Balthasar/Knöpfel 1994 zu erwähnen.

3.6 Forschungstechniken

3.61 Überblick/Darstellung ausgewählter Forschungstechniken

Verschiedene Autoren benützen die Forschungstechnik als zentrales Merkmal zur Klassifikation von Evaluationsstrategien. Aus unserer Sicht ist indessen bei kausalorientierten Fragestellungen der Entscheid über eine sinnvolle Art des Vergleichs (Kontrollgruppen- Querschnitt-, Längsschnitt-Vergleich, implizit vergleichende Evaluationsstrategien) von zentraler Bedeutung, und die Forschungstechnik muss optimal darauf ausgerichtet werden.

Es gibt eine kaum mehr zu überblickende Zahl von Forschungstechniken, im deutschen Sprachraum in der Regel als Methoden der empirischen Sozialforschung bezeichnet. Dabei kann grob zwischen Methoden der **Datenerhebung** und solchen der **Datenauswertung** unterschieden werden.

Wir erwähnen stichwortartig folgende Methoden der **Datenerhebung**:

- Dokumentenanalyse
- offenes/halboffenes Interview
- schriftliche/telephonische/persönliche Umfrage
- statistische Datensammlung (Aufbereitung von Verwaltungsvollzugsdaten, Aufbereiten eigener Statistiken)
- audiovisuelle Aufzeichnung
- Gruppengespräch (interaktiv)
- teilnehmende Beobachtung

Unter den Methoden der **Datenauswertung** sind zu erwähnen:

- Literaturstudien/Dokumentenanalyse (qualitative Gewichtung)
- hermeneutische Interpretationsverfahren (z.B. von mündlichen Interviews)
- Inhaltsanalysen (quantitativ)
- einfachere statistische Auswertungen (Durchschnitt, Median, Standardabweichung, u.a.m.)
- Regression, multiple Regression
- Faktoranalysen, Clusteranalysen, multidimensionale Skalierung
- lineare Programmierung
- kausale Modellierung mit mathematischen Modellen, ökonometrische Modelle, Simulationen

Forschungstechniken werden in **qualitative** und **quantitative** Methoden unterteilt. Im Rahmen von qualitativen Methoden¹⁰ werden Daten in Form von Worten dargestellt. Bei der Verwendung von quantitativen Methoden werden die Daten in Form von Zahlen dargestellt; vorgängig ist eine "Messung", d.h. ein Verfahren der Zuteilung von Ereignissen oder Sachverhalten auf Zahlen notwendig.

Im folgenden werden drei Forschungstechniken näher vorgestellt, welche im Rahmen von Evaluationen oft verwendet werden.

Abbildung 7 Wichtige Forschungstechniken, Kombinationen

Forschungstechnik	Untergruppen	Spezialformen, Kombinationen mit anderen Forschungstechniken
Fallstudien	einzelner Fall ausgewählte Fälle multiple Fallstudien	interaktive, aktorzentrierte Untersuchungen Interkommunale, -kantonale, -nationale Fallstudien
Umfragen	Panel-Umfragen	
Analysen vorhandener Daten	Sekundärstatistische Analysen Evaluationssynthesen	ökonometrische Studien

Weiterführende Literatur

Als Einführungen in die Methoden der empirischen Sozialforschung bzw. in qualitative Methoden sind Atteslander 1985 und Lamnek 1988 dienlich.

3.62 Fallstudien

Wir verstehen unter einer Fallstudie die analytische Beschreibung eines Ereignisses, einer Institution, eines (Vollzugs)prozesses oder einer Massnahme. Untersuchungsgegenstand einer Fallstudie können beispielsweise sein:

- Ausgewählte Baugenehmigungsverfahren im Kanton X
- Erfahrungen von 5 Arbeitslosen mit der staatlichen Arbeitsvermittlung

¹⁰Im angelsächsischen Sprachraum werden zahlenmässige Informationen, bei welchen die Zahlen nicht (im Sinne ordinaler oder kardinaler Skalierung) zueinander in Bezug gesetzt werden können, ebenfalls den qualitativen Methoden zugerechnet. Als Beispiel ist die EDV-gestützte Inhaltsanalyse von Texten zu erwähnen.

- Behördliche Strategien im Zusammenhang mit der Errichtung einer Sondermülldeponie
- Der Vollzug des Invalidenversicherungsgesetzes in einzelnen Kantonen

Der Anlass für Fallstudien liegt in der Regel in der Komplexität des zu untersuchenden Gegenstands, welche eine mehr in die Tiefe als in die Breite gehende Analyse erfordert. Die ermittelte Information wird dementsprechend reichhaltig und komplex sein.

Bei der Bearbeitung von Fallstudien bedient man sich einer oder mehrerer Datenerhebungs- und -auswertungsmethoden (Dokumentenanalysen, Interviews, teilnehmende Beobachtung, hermeneutische Interpretationsverfahren, einfachere statistische Auswertungen u.a.m.). Fallstudien sind daher weniger eine Forschungstechnik als ein Forschungszugang.

Wir unterscheiden zwischen Einzelfallstudien, ausgewählten Fällen und multiplen Fallstudien. **Einzelfallstudien** dienen in der Regel der Beschreibung eines einzelnen Ereignisses, Prozesses oder einer einzelnen Massnahme. Sie können der Illustration dienen (z.B. Darstellung von mehreren parallel laufenden Bewilligungsverfahren bei einem grossen Bauprojekt), wobei fließende Grenzen zum "investigative journalism" bestehen. Sie können zweitens zur Exploration neuer und bisher unbearbeiteter Fragen benutzt werden. Drittens kann sich die Auswahl auf einen kritischen Fall (z.B. Verhalten von Dienststellen in einem Katastrophenfall) beziehen, um daraus Lehren ziehen und Verbesserungen vornehmen zu können. Bei letzterem handelt es sich um die **Untersuchung ausgewählter Fälle nach bestimmten Kriterien**. Es geht um normative Fragen wie: Wurden die durch die Massnahme anvisierten Leistungsziele (z.B. Beschäftigungsquote der Teilnehmer von Weiterbildungsprogrammen für Arbeitslose) erreicht? Verlieh der Vollzug wie geplant? Solche normativen Fragen - Kontrollfragen - können gut anhand ausgewählten Fällen beantwortet werden. Im Vergleich zu einer umfassenden Analyse wird eine solche Falluntersuchung die Frage nicht beantworten können, wie häufig Abweichungen zwischen Soll-Vorgaben und Ist-Werten sind. Bei einer klugen Auswahl (z.B. sowohl "gute" wie "schlechte" Fälle) wird aber ermittelt werden können, ob Abweichungen vorkommen und warum sie vorkommen. Bei **multiplen Fallstudien** werden mehrere Fälle untersucht, sei es um die Aussagen glaubwürdiger zu machen, sei es um gewisse Hypothesen besser zu testen oder um Unterschiede (zwischen Kantonen, Zeitpunkten u.a.m.) hervortreten zu lassen. Zwischen multiplen Fallstudien, die unterschiedliche Gebiete betreffen, und Querschnitt-Vergleichen bestehen fließende Grenzen.

Eine Kombination zwischen Fallstudien und Umfragen (s. nächstes Kapitel) sind **aktorzentrierte, interaktive Evaluationsstrategien**. Sie beziehen (oft

mit offenen oder halboffenen Interviews) die für eine Massnahme wichtigsten **Aktoren** (Vollzugsinstanzen, Adressaten, weitere Betroffene) in die Untersuchung ein. Dabei wird oft auf deren Einschätzungen der Wirksamkeit abgestellt. Das Ziel ist nicht wie bei vielen Umfragen die Ermittlung einer repräsentativen Meinung, sondern vielmehr eine möglichst vollständige Erfassung der verschiedenen möglichen Argumentationsformen; dabei wird der Forschungsprozess in der Regel dann abgeschlossen, wenn bei zusätzlichen Interviews keine neuen Argumente mehr zutage treten und eine "Sättigung" an Argumentationsformen erreicht wird. Aus dem vorgefundenen Material kann das Evaluationsteam gewisse Typisierungen vornehmen. Es wird in der Regel nicht versuchen, aus den vorgefundenen Meinungen **die** Wirklichkeit zu rekonstruieren. Soweit nicht bereits die Argumente **interaktiv** ermittelt worden sind (z.B. durch Hearings, Diskussionsgruppen), wird das Evaluationsteam Ergebnisse und Deutungen an alle oder zumindest die wichtigsten Aktoren weiterleiten, damit Aussagen und Vorschläge gemeinsam diskutiert werden können. Das Evaluationsteam kann bei der Steuerung der Evaluation auf die Wünsche und Fragen wichtiger Betroffener abstellen, z.B. im Rahmen einer Begleitgruppe. Der Evaluationsprozess ist ferner in der Regel auf eine **Verbesserung der Massnahmen** und nicht bloss auf eine Bilanz ausgerichtet. Das aktorzentrierte und interaktive Element wird genutzt, um Fragen und Probleme auf den Tisch zu stellen, gemeinsam nach Lösungen zu suchen und diese zu diskutieren, wobei das Evaluationsteam die Funktion eines Moderators und eines Experten übernimmt. Diese Evaluationsstrategie eignet sich vor allem bei Massnahmen, deren erfolgreicher Vollzug und deren Wirksamkeit von der guten Zusammenarbeit zwischen Behörden und Betroffenen abhängen.

Fallstudien sind in der Regel geeignet zur Beantwortung deskriptiver Fragen. Bei geeigneter Fallauswahl (Untersuchung ausgewählter Fälle nach bestimmten Kriterien) können sie auch dazu dienen, normative Fragen (z.B. gibt es Vollzugsdefizite und wenn ja warum?) zu beantworten. Fallstudien können ausnahmsweise auch zur Beantwortung kausaloientierter Fragen eingesetzt werden, beispielsweise indem an einem einzelnen Fall überprüft wird, ob die angenommene Wirkungskette auch in der Wirklichkeit zu beobachten ist. Ob aus Fallstudien im Hinblick auf eine grössere Grundgesamtheit Aufschluss über Wirkungen erhalten werden kann, ist umstritten.

Fallstudien können mit anderen Forschungstechniken kombiniert werden, z.B. mit Umfragen, um sowohl in die Tiefe wie auch in die Breite gehende Daten zu erhalten.

Weiterführende Literatur

Einen guten Einblick in Fallstudien bieten Yin 1989 und General Accounting Office 1990. Rey et al. 1993 und Bierter/Binder 1993, beides

Schlussberichte von im Rahmen des NFP 27 durchgeführten Projekten mit einem aktorzentrierten, interaktiven Ansatz, geben praxisgerechte Hinweise zu diesem Vorgehen. Ein weiteres Beispiel ist Gerheuser 1991.

3.63 Umfragen

Bei einer Umfrage werden in der Regel aus der Stichprobe einer Grundgesamtheit der Bevölkerung Daten erfasst, um das Auftreten, das Verteilungsmuster und die gegenseitige Beziehung zwischen bestimmten natürlich auftretenden Ereignissen und Zuständen zu ermitteln¹¹. Ziel ist das Sammeln von Information, aus welchen auch Schlüsse auf Teile der Bevölkerung, welche in der Stichprobe nicht enthalten waren, gezogen werden können. Es werden Zufallsstichproben benutzt, bei welchen jedes Segment der zu untersuchenden Bevölkerung die Chance hat, mit einer bestimmten Wahrscheinlichkeit bei der Stichprobe berücksichtigt zu werden. Innerhalb der Grenzen statistischer Zuverlässigkeit können die Ergebnisse auf die zu untersuchende Grundgesamtheit übertragen werden. Im Hinblick auf die Aggregation der Daten wird grosser Wert auf die Einheitlichkeit der von den Stichprobeneinheiten zu sammelnden Daten gelegt. Die Informationen werden in der Regel aus geschlossenen Fragen ermittelt, d.h. die Befragten wählen aus den möglichen Antworten die zutreffende Antwort aus¹². Offene Fragen sind möglich, erfordern aber anschliessend bei der Auswertung eine Codierung der Antworten.

Umfragen können schriftlich/postalisch, telephonisch oder durch persönliche Befragung durchgeführt werden.

Von zentraler Bedeutung ist die Definition der Grundgesamtheit, aus der die Stichprobe gezogen wird. Bei einer umfassenden Studie über die Auswirkungen der AHV wird dies beispielsweise die Bevölkerung über 62 bzw. 65 Jahre sein. Bei einer Studie über die Zielgruppe einer bestimmten Massnahme (z.B. PTT, Wohnbauförderung) wird jeweils sorgfältig abzuklären sein, welche sozioökonomischen und demographischen Schichten dazugehören. Denkbar ist auch die Konzentration einer Umfrage auf Bevölkerungsschichten, bei denen bestimmte Probleme zu vermuten sind (z.B. alleinerziehende Mütter oder Väter, Langarbeitslose, Betriebe im Konkurs u.a.m.).

Beispiel: In einer Studie über die Wohneigentumsförderung des Bundes wurden mit Bundesbürgschaften geförderte Hauseigentümer, Personen mit

¹¹Auf Vollerhebungen wie beispielsweise die Volkszählungen oder die Rekrutenbefragungen wird hier nicht speziell eingegangen.

¹²Soweit nicht im voraus alle möglichen Antworten abgeschätzt werden können, wird im Fragebogen eine Restkategorie ("andere Faktoren:.....") gebildet.

abgelehntem Gesuch um Bundesförderung, Hauseigentümer ohne Bundesförderung sowie Mieter über ihre Situation sowie die Einschätzung der Bundeshilfe befragt. Daraus ergab sich, dass die mit Bundesbürgschaften geförderten Hauseigentümer ein von den nicht geförderten Hauseigentümern deutlich verschiedenes Profil (geringeres Einkommen, kleiner dimensionierte Häuser u.a.m.) aufwiesen. Auch aus der Mieterbefragung ergaben sich aufschlussreiche Informationen, z.B. die Tatsache, dass sich 62 Prozent der befragten Mieter noch nie Gedanken über den Erwerb von Wohneigentum gemacht hatten (Studie Schulz u.a. 1993).

Eine Kombination von Umfragen mit Längsschnitt-Vergleichen sind **Panel-Umfragen**. Dabei wird dieselbe Stichprobe von Personen im Zeitablauf mehrmals befragt. Daraus können Änderungen in Sachverhalten und Meinungen ermittelt werden. Für Entscheide über staatliche Massnahmen sind Informationen über Trends oft nützlicher als rein statische Informationen. Solche Informationen können auch Hinweise auf Wirkungsbeziehungen geben, indem ermittelt werden kann, welcher Sachverhalt vor dem andern auftrat.

Die Durchführung von Umfragen setzt in der Regel eingehende Kenntnisse der Tatbestände voraus. Diese können auf verschiedene Art (Literaturstudium, Tiefeninterviews u.a.m.) beschafft werden. Umfragen setzen eine eingehende Vorbereitung voraus. Teilweise sind fundierte Abklärungen darüber nötig, aus welcher Grundgesamtheit die Stichprobe sinnvollerweise gezogen werden soll und wie gross die Stichprobe für die je unterschiedlichen Segmente ausfallen soll. Danach kommen Pretests (versuchsweise Prüfungen) des Fragebogens, um sicherzustellen, dass die Fragen klar und neutral formuliert sind. Nach der Durchführung der Umfrage ist ein oft mehrfaches Nachhaken erforderlich, um fehlende Antworten zu erhalten oder um die nicht antwortenden Personen durch geeignete andere Personen zu ersetzen. Es gilt sicherzustellen, dass die Umfrageergebnisse nicht durch fehlende Antwortgänge verfälscht werden.

Umfragen erlauben in der Regel die Beantwortung deskriptiv orientierter Fragen. Es können Tatbestände (z.B. Gesundheitszustand, Ausgabenverhalten älterer oder invalider Personen) und Meinungen (z.B. Grad der Bedürftigkeit) erfasst werden. Innerhalb relativ kurzer Zeit und mit vertretbaren Kosten kann aus Umfragen die Einschätzung staatlicher Leistungen (z.B. Postzustellung, Gestaltung der Steuerformulare, Auskunft über AHV-Leistungen) durch die Benutzerinnen und Benutzer ermittelt werden. Umfragen sind hingegen nur bedingt geeignet zur Beantwortung kausalerorientierter Fragen. Z.T. wird versucht, aus Korrelationen, d.h. aus positiven Beziehungen zwischen verschiedenen Grössen, Aufschluss über Wirkungszusammenhänge zu gewinnen. Umfragen können insbesondere dann Ursache-Wirkungs-Beziehungen erhellen, wenn sie Auskunft über die Motive von Personen im Hinblick auf bestimmtes

(beobachtbares) Verhalten geben können (z.B. Gründe für einen Rückgang der Benützung öffentlicher Verkehrsbetriebe; diese können etwa in der Erhöhung der Tarife, in Sicherheitsproblemen, in langen Fahrzeiten oder ungünstigen Verbindungen liegen).

Weiterführende Literatur

Eine Einführung in Umfragetechniken bietet Atteslander 1985; General Accounting Office 1991b behandelt dieses Instrument im Kontext von Evaluationen und Wholey 1983 gibt Hinweise über die Eignung von Telephoninterviews zur Beurteilung der Qualität staatlicher Dienstleistungen. Unter den Projekten des NFP 27 haben v.a. Schulz u.a. 1993 und auch Sangra 1992 diese Forschungstechnik verwendet.

3.64 Analyse vorhandener Daten

Die beiden vorangehenden Forschungstechniken erfordern in der Regel die Erhebung neuer Informationen, sei es durch Umfragen, sei es durch Interviews oder andere Forschungstechniken (im Falle von Fallstudien). Da die eigene Datenerhebung in der Regel mit Kosten verbunden ist, lohnt es sich auf jeden Fall, vorhandene Daten auf ihre Eignung zu prüfen; dieser Schritt erleichtert auch die Vorbereitung zur Gewinnung neuer Daten.

Sekundärstatistische Analysen

Viele Daten (z.B. Vollzugsdaten, Volkszählungen, Mikrozensus, Umfragen) werden für bestimmte Zwecke erhoben. Im Rahmen sekundärstatistischer Analysen werden sie nun für die Beantwortung von Fragen genutzt, für welche sie ursprünglich nicht aufbereitet wurden. In den meisten Fällen können sekundärstatistische Analysen zu deskriptiven Zwecken genutzt werden.

Beispiel: Die in der Volkszählung ermittelten Anteile der Hauseigentümer und Mieter sowie die in der Baustatistik ermittelte Zahl der Neubauten können im Zeitablauf nachgezeichnet, nach Kantonen aufgeschlüsselt und in Bezug zur Wohneigentumsförderung des Bundes oder der Kantone gesetzt werden (vgl. Studie Schulz u.a. 1993).

Unter Umständen kann möglicherweise durch eine Kombination von für deskriptive Zwecke erhobenen Daten Aufschluss über gewisse Ursache-Wirkungs-Beziehungen im Hinblick auf bestimmte staatliche Massnahmen erhalten werden.

Beispiel: In den 60er und 70er Jahren wurden verschiedene Untersuchungen über die Schichtzugehörigkeit der Studierenden durchgeführt. Sie ergaben, dass Studierende aus der Ober- und Mittelschicht übervertreten

waren. Ein neuerer Vergleich der verschiedenen Statistiken sowie der in den letzten 30 Jahren durchgeführten Umfragen über die soziale Herkunft der Studierenden ergab, dass sich trotz ausgebauter kantonaler Stipendienregelungen an der Übervertretung von Studierenden aus der Ober- und Mittelschicht wenig geändert hat (unveröffentlichte Seminararbeit Univ. Bern).

Die Eignung sekundärstatistischer Analysen hängt von der jeweiligen Datenlage ab. Im Wirtschafts- und im Landwirtschaftsbereich beispielsweise sind die Voraussetzungen recht gut. Weitergehende allgemeine Aussagen sind indessen recht schwierig und die Datenlage muss in jedem konkreten Fall eingehend abgeklärt werden.

Evaluationssynthesen

Über bestimmte Fragen bzw. staatliche Massnahmen liegen bereits Untersuchungen vor. Im Rahmen von Evaluationssynthesen werden die Ergebnisse verschiedener Studien aggregiert (zusammengeführt), um Schlussfolgerungen zu ziehen, die verlässlicher sind als diejenigen der Einzelstudien.

Beispiel: Es gibt in der Schweiz eine ganze Reihe von Informationskampagnen, und in verschiedenen Fällen (z.B. AIDS-Kampagne, Zivilschutz-Informationen) wurden Erfolgskontrollen durchgeführt. Im Rahmen einer Evaluationssynthese können nun die Ergebnisse dieser Einzeluntersuchungen zusammengeführt und die Erfolgsvoraussetzungen des Instruments "Informationskampagnen" näher umrissen werden (Projektbeschreibung Balthasar zuhanden des Nationalfonds).

Evaluationssynthesen können qualitativen Charakter haben, aber über typische Literaturstudien hinausgehen. Die Ergebnisse werden gewogen und qualitativ kombiniert, ohne dass eine statistische Aggregation der Einzelstudien angestrebt wird. Daneben gibt es auch die Möglichkeit der Kombination von Einzelstudien. In den USA wurde hierzu die Methode der "Meta-Analyse" entwickelt. Dabei wird aus dem in je verschiedenen Studien quantitativ erhobenen Ausmass an Wirkungen¹³ der Durchschnitt ermittelt.

Weiterführende Literatur

General Accounting Office 1992a und 1992b geben Aufschluss über sekundärstatistische Analysen bzw. Evaluationssynthesen. Unter den Projekten des NFP 27 haben Rieder (Evaluationen in der Agrarpolitik I und II) sowie Stalder et al. 1994 sekundärstatistische Analysen durchgeführt und sich dabei der Linearen Programmierung sowie der Ökonometrie bedient.

¹³Gemessen am Unterschied der Situation zwischen Ziel- und Kontrollgruppe

3.7 Exkurs: Modelle als Ausgangspunkt und Resultat von Evaluationen

Modelle haben für Evaluationen einen wichtigen Stellenwert und zwar als Ausgangspunkt und Resultat von Evaluationen sowie im Zusammenhang mit statistischen Forschungstechniken.

Modelle sind Versuche zur Abbildung der Wirklichkeit. Sie können dingliche (z.B. Uhr als Modell der Erdumdrehung in analoger oder digitaler Form), sprachliche (z.B. Theorien über den Wertewandel) oder numerisch/mathematische (z.B. ökonomisches Konjunkturmodell) Form annehmen. Bei Evaluationen sind letztere beiden Formen vorherrschend.

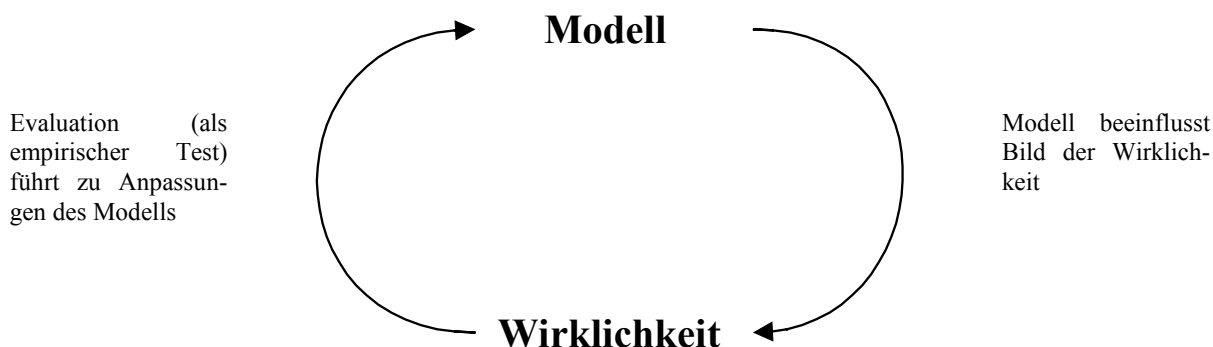
Jede kausalorientierte Evaluation setzt ein **Wirkungsmodell** voraus. Ein Modell ermöglicht eine Fokussierung der Evaluation und erlaubt gleichzeitig, wichtige Annahmen zu spezifizieren, die in der empirischen Untersuchung überprüft werden sollen.

Beispiel: Im Heimarbeitsgesetz besteht die Verpflichtung, HeimarbeiterInnen und BetriebsarbeiterInnen gleich zu entlohnen. In Interviews mit Arbeitgebern und HeimarbeiterInnen wurde ermittelt, ob die erwähnte Vorschrift des Heimarbeitsgesetzes bei den dem Gesetz unterstellten HeimarbeiterInnen eingehalten wird, wobei namentlich auch die Erfolgsvoraussetzungen bei den kantonalen Vollzugsinstanzen näher untersucht wurden. Die Untersuchung beruhte auf dem Wirkungsmodell, dass die Verwirklichung des Gleichstellungsgebotes von den Aktivitäten der Vollzugsinstanzen und den je spezifischen Interessen der Arbeitgeber und HeimarbeiterInnen abhängt: ein durchaus plausibles Modell. (Studie Gerheuser/Schmid 1993). Im Rahmen einer anders fokussierten Untersuchung hätten aber auch andere Wirkungsaspekte ins Zentrum gerückt werden können, beispielsweise die Frage, ob das Gleichstellungsgebot auch auf nicht dem Heimarbeitsgesetz unterstellte ArbeitnehmerInnen ausgestrahlt habe oder ob dieses Gebot einen Einfluss auf die Auslagerung von Arbeitsplätzen in Billiglohnländer gehabt habe.

Modelle sind für die wissenschaftliche Forschung und für kausalorientierte Evaluationen unentbehrlich, indem sie eine Eingrenzung und Präzisierung der Forschungsfrage ermöglichen; sie bergen gleichzeitig aber die Gefahr, dass sie die Evaluation durch das Bild der Wirklichkeit, das sie vorprägen, in eine einseitige oder gar falsche Richtung lenken. Dies kann beispielsweise dann geschehen, wenn in einem Modell nur die intendierten Wirkungen einer Massnahme einbezogen und bedeutsame Nebenwirkungen ausgeblendet werden.

Modelle bilden zugleich auch den Endpunkt einer Evaluation, da jedes Evaluationsresultat ein Versuch ist, Aspekte der Wirklichkeit abzubilden.

Abbildung 8 Modelle als Ausgangspunkt und Resultat von Evaluationen



Modelle können sehr unterschiedliche Dimensionen aufweisen.

- Sie können **statischen** ("Eine Erhöhung des Benzinpreises um 5 Prozent führt zu einer Reduktion des Verbrauchs um 0.5 Prozent") oder **dynamischen** ("Eine Erhöhung des Benzinpreises führt zuerst zu einer Verminderung der Autofahrten, dann zur Entwicklung sparsamerer Autos, anschließend zu einem Gewöhnungseffekt".) **Charakter** haben.
- Sie können **partiellen** (z.B. Einfluss der Preise öffentlicher Verkehrsmittel auf deren Benützung) oder **globalen** (Benützung der öffentlichen und privaten Verkehrsmittel in Abhängigkeit des Wertewandels, der Entwicklung der Siedlungsstruktur, der Preisentwicklung, des Angebots an öffentlichen Verkehrsleistungen u.a.m.) **Charakter** aufweisen.
- Sie können **raumzeitlich beschränkte** ("die im Jahre 1993 erfolgte Erhöhung der Parkbussen hat zu einer Abnahme des Trottoirparkierens um 21 Prozent geführt) oder **umfassende** ("Der Einfluss von Strafen auf das Deliktverhalten hängt von der Höhe der Strafe und der Chance, entdeckt zu werden, ab") **Geltung** beanspruchen.

Qualitative Modelle sind in Worte gefasst, **quantitative** Modelle in Zahlen. Letzteres sind mathematische Gleichungssysteme, in welchen das Ausmass des Einflusses zwischen Grössen (Variablen) in numerischen Werten (sog. Parameter) ausgedrückt wird. Grundsätzlich sind beide Formen von Modellen in die jeweils andere "Sprache" übersetzbar. Namentlich bei einer grossen Zahl von Variablen bietet eine mathematisch/numerische Darstellung den Vorteil grösserer Überschaubarkeit.

Namentlich quantitative Modelle spielen in Bezug auf die Erwartungen an die Leistungsfähigkeit von Evaluationen eine beträchtliche Rolle. Solche Modelle erlauben unter Umständen, die künftigen Auswirkungen bestehender bzw. geplanter Massnahmen unter Annahme bestimmter Rahmenbedingungen abzuschätzen und die möglichen Auswirkungen alternativer Massnahmen zu simulieren.

Beispiel: Eine ökonometrische Untersuchung der Auswirkungen der Ausländerpolitik auf den Strukturwandel in der Schweiz schloss alle bedeutenden Märkte (Gütermarkt und Arbeitsmarkt) ein. Das entwickelte Modell konnte die Vergangenheitsentwicklung auf zufriedenstellende Weise nachzeichnen. In einer Simulation wurden die möglichen Auswirkungen einer Arbeitsmarkoliberalisierung (stärkere Einwanderung qualifizierter, schwächere Einwanderung niedrigqualifizierter Arbeitskräfte) abgeschätzt. Es ergaben sich gemäss Modell markante Auswirkungen auf die wichtigen wirtschaftlichen Sektoren, aber erstaunlicherweise nur recht geringfügige Auswirkungen auf die Entwicklung der Produktivität (Studie Stalder et al. 1994).

Die Anforderungen an quantitative Modelle sind erheblich. Es müssen ausreichende Daten vorhanden sein und die Beziehungen zwischen wichtigen Variablen sollten über längere Zeitabschnitte stabil bleiben. Wichtig ist ferner eine langfristige wissenschaftliche Durchdringung des Sachgebiets. Gute Voraussetzungen für quantitative Modelle bestehen beispielsweise in Wirtschafts-, Landwirtschafts-, Energie- und z.T. in Umweltfragen.

Quantitative Modelle erwecken aufgrund ihrer "Sprache" bei Laien oft Erwartungen an ihre Präzision (im Sinne der "Übereinstimmung" mit der Wirklichkeit sowie der Fähigkeit, künftige Entwicklungen zu prognostizieren). Auch wenn quantitative Modelle eine empirische Überprüfung erleichtern, bieten sie keineswegs Gewähr dafür, dass künftige Entwicklungen zuverlässig(er) abgeschätzt werden können.

Eine zentrale Fragestellung quantitativer Modelle ist die gegenseitige Beziehung zwischen den Variablen, namentlich inwieweit sich Veränderungen der (durch staatliche Massnahmen beeinflussbaren, sog. exogenen) Grössen in Veränderungen der davon abhängigen (sog. endogenen) Grössen auswirken. Diese Verhaltenskoeffizienten oder Parameter können auf dreierlei Art ermittelt werden und dementsprechend können auch drei Arten von Modellen unterschieden werden:

Technische Parameter bzw. technisch/naturwissenschaftlich vorgegebene Modelle: Bei diesen Modellen sind die Parameter durch technische Gegebenheiten vorgegeben.

Beispiel: In der Umweltpolitik spielen quantitative Modelle, die auf technischen Parametern beruhen, eine wichtige Rolle. Beruhend auf Verkehrsprognosen und auf Kenntnissen der Reinigungswirkungen von Katalysatoren (=technischer Parameter) sowie auf Annahmen über die Erneuerung der Fahrzeugflotte konnten Mitte der achtziger Jahre die mutmasslichen Auswirkungen einer Einführung des Katalysators auf die Schadstoffbelastung der Luft abgeschätzt werden.

Statistisch (empirisch) ermittelte Parameter bzw. positive statistische Modelle: Die Verhaltenskoeffizienten werden (oft unter Zuhilfenahme

theoretischer Kenntnisse) durch statistische Methoden (wie Regressionsrechnungen, Pfadmodelle, Faktor- Varianz- und Clusteranalysen) ermittelt bzw. geschätzt, wobei die Ergebnisse oft anhand theoretischer Kenntnisse auf ihre Plausibilität überprüft und notfalls angepasst werden.

Beispiel: Die im vorletzten Beispiel erwähnte ökonometrische Analyse (Studie Stalder et al. 1994) bediente sich statistischer Verfahren, namentlich der Regressionsrechnung, um die Beziehung zwischen wichtigen Grössen, die in der Schnittmenge von Ausländer- und Strukturpolitik bedeutsam sind, abzuschätzen.

Normativ gesetzte Parameter bzw. normative Modelle: Aufgrund normativer Theorien werden Parameter festgelegt (mathematisch spezifiziert und mit Normdaten quantifiziert). Die Parameter werden nicht empirisch überprüft, sondern als plausible Annahmen gesetzt. Eine statistische Überprüfung der Validität (d.h. der "Richtigkeit" der Ursache-Wirkungs-Beziehungen) des Modells wird nicht angestrebt. Soweit das Modell prognostische Zwecke erfüllt, kann indessen später ermittelt werden, ob die aufgrund des Modells abgebildeten Entwicklungen in der Wirklichkeit tatsächlich eintreffen. Falls dies zutrifft, ist dies ein Indiz für die Glaubwürdigkeit des Modells auch für prognostische Zwecke; die Gefahr besteht aber, dass sich in der Zwischenzeit Beziehungen zwischen wichtigen Grössen verändert haben.

Beispiel: Ein auf der Linearen Programmierung beruhendes Modell der Landwirtschaft zeichnete namentlich die Zusammenhänge zwischen der Tierhaltung und dem Getreideanbau auf. Es beruhte auf der Annahme, dass sich die Bauern wirtschaftlich rational verhalten. Das Modell konnte Hinweise über den künftigen bäuerlichen Strukturwandel und dessen regionale und sektorale Auswirkungen geben. In Simulationen konnte namentlich auch das Szenario einer Liberalisierung aufgrund des GATT durchgespielt werden (Studie Felder/Rieder in: Evaluationen in der Agrarpolitik I 1994).

Oft sind Modelle auch Kombinationen dieser drei Typen. Ein auf linearer Programmierung beruhendes normatives ökonomisches Modell kann sich beispielsweise auf technisch-naturwissenschaftliche Zusammenhänge (z.B. Betriebsgrösse und Ernteerträge) abstützen.

Weiterführende Literatur

Eine kurze Einführung in quantitative Modelle als Evaluationsinstrument bieten Rieder/Meyer 1994. Im NFP 27 wurden quantitative Modelle vor allem in zwei Studien verwendet (Evaluationen in der Agrarpolitik I/II 1994 sowie Stalder et al. 1994).

3.8 Methodenmix und Triangulation

Um sinnvolle Aussagen über den Vollzug und die Wirkungen staatlichen Handelns zu machen, ist im Rahmen von Evaluationen die Verwendung unterschiedlicher Zugänge und unterschiedlicher Forschungstechniken sinnvoll. Dadurch ist es in der Regel möglich, die Facetten der Wirklichkeit einigermaßen zufriedenstellend zu erfassen. Ein **Methodenmix** ist im Hinblick auf die Erzielung aussagekräftiger Resultate oft auch kostengünstiger als die Ausrichtung auf eine einzige Forschungstechnik. Evaluationsteams bedienen sich beispielsweise Dokumentenanalysen, Literaturstudien, grober sekundärstatistischer Analysen sowie Interviews, um einen Überblick über das Untersuchungsfeld zu erhalten, bevor sie weitere Techniken der Datenerhebung (z.B. Umfragen) und -auswertung (z.B. multiple Regression) einsetzen.

Beispiel: Balthasar/Knöpfel 1994 erfassten die Auswirkungen von Umweltvorschriften auf Erfindungen im Bereich der Hausfeuerungsanlagen, auf die technische Umsetzung dieser Erfindungen in verbesserte Hausfeuerungen und auf deren Verbreitung. Sie untersuchten in Form einer Fallstudie das Verhältnis zwischen Umweltschutz-Grenzwerten und Erfindungen, im Rahmen einer Marktstudie den Einfluss der Typenprüfung auf Innovationen bei Hausfeuerungen und mit Umfragen und statistischen Analysen den Einfluss der Feuerungskontrolle auf die Modernisierung von Heizungsanlagen. Durch Zeitreihenanalysen versuchten sie schliesslich, die Gesamtwirkungen behördlicher Massnahmen zu erfassen.

Die **Triangulation** bedient sich überlappender Zugänge, um die Zuverlässigkeit von Evaluationsresultaten zu verbessern. Sie kann eine oder mehrere der folgenden Formen umfassen:

1. Kombination von Evaluationsstrategien: Bei kausalorientierten Fragestellungen kann eine Kombination von Vergleichen (Kontrollgruppen- Querschnitt-, Längsschnitt-Vergleich) oder deren Kombination mit einer implizit-vergleichenden Evaluationsstrategie die Zuverlässigkeit schwacher Ergebnisse verbessern bzw. die Zuverlässigkeit bereits recht guter Ergebnisse zusätzlich erhärten.

Beispiel : Einen interessanten Vergleich unterschiedlicher Evaluationsstrategien und Forschungstechniken ermöglichen zwei von der Arbeitsgruppe "Gesetzesevaluation" (AGEVAL) in Auftrag gegebene Studien über die Berufliche Vorsorge. Im Rahmen einer sekundärstatistischen (ökonomischen) Untersuchung im Längsschnitt wurden die Wirkungen des Bundesgesetzes über die Berufliche Vorsorge auf den Arbeitsmarkt ermittelt (Studie Schätti 1990). Gerheuser 1991 untersuchte dieselbe Fra

gestellung durch eine (nicht repräsentative) Befragung von Personalchefs aus Betrieben.

2. Überlappung von Forschungstechniken: Damit wird versucht, Aussagen über Sachverhalte gut abzusichern. Es kann darum gehen, die Aussagen einzelner Personen auf mehrfache Art zu überprüfen (z.B. durch Interviews und durch teilnehmende Beobachtung), einen Personenkreis auf verschiedene Weise bei der Ermittlung von Sachverhalten einzubeziehen (z.B. durch Tiefeninterviews ergänzt durch eine repräsentative Umfrage) oder mehrere Datenquellen zu ähnlichen Sachverhalten auszuwerten (z.B. sekundärstatische Analysen und repräsentative Umfragen).

vgl. Studie Balthasar/Knöpfel weiter oben

3. Aufstellen unterschiedlicher (rivalisierender) Wirkungshypothesen: Es ist bekannt, dass Wirkungshypothesen dazu verleiten können, die Evaluationsstrategien und die Forschungstechniken so zu wählen, dass die Wirkungshypothesen (z.B. Wirkungslosigkeit eines bestimmten Handlungsinstruments) bestätigt werden. Das Aufstellen rivalisierender Wirkungshypothesen, kann geeignet sein, diese Gefahr zu vermindern.
4. Bearbeitung derselben Fragestellung durch unterschiedliche Evaluationsteams: dieses Vorgehen ist meist mit einer der drei obigen Triangulationsmöglichkeiten verbunden.

Für ein Beispiel sei auf Ziffer 1 verwiesen.

5. Zuteilung unterschiedlicher Aufgaben und Rollen innerhalb eines Evaluationsteams: auch dieses Vorgehen ist meist mit einer der drei oberen Triangulationsmöglichkeiten verbunden.

Die Evaluation der Walliser Wirtschaftsförderung erfolgte interaktiv, d.h. unter Einbezug der Betroffenen. Zuerst wurden mit Interviews deren Einschätzungen erhoben. Die Zwischenresultate wurden anschliessend durch Gruppenarbeit vertieft. Dabei wurden innerhalb des Evaluationsteams die Funktionen der Animation, der Expertise und der Methodik möglichst auf verschiedene Personen verteilt (Studie Rey et al. 1993, S. 31).

Weiterführende Literatur

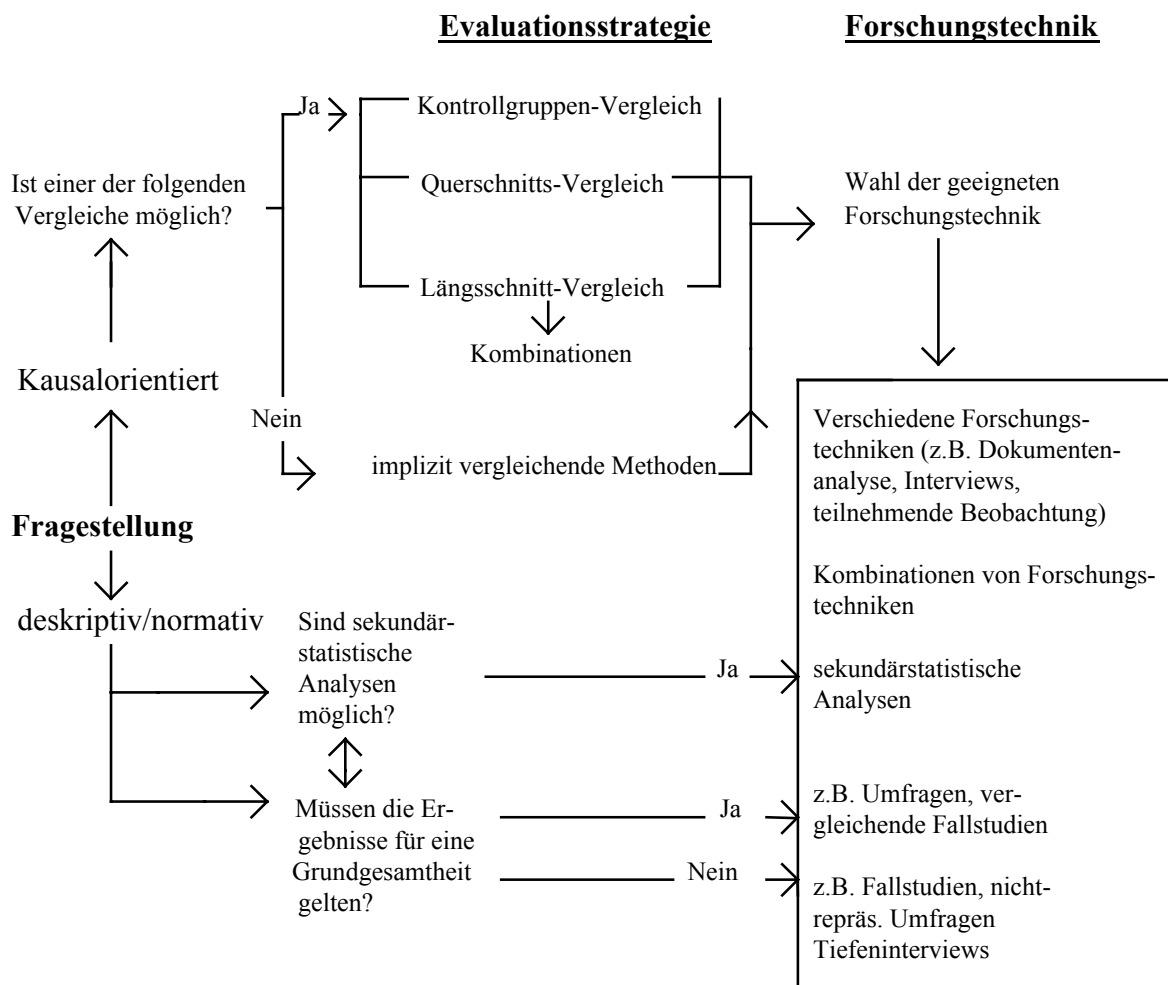
Ein Verfechter von Methodenmix und Triangulation ist Patton 1990 (S. 186-198, 464-470). Vor allem Balthasar/Knöpfel 1994 haben sich unter den Projekten des NFP 27 eines Methodenmixes und der Triangulation bedient und geben hierzu Hinweise.

3.9 Hinweise zur Wahl der Evaluationsstrategie und der Forschungstechnik

Die Wahl der Evaluationsstrategie und der Forschungstechnik ist eine gemeinsame Aufgabe von Auftraggeber und Evaluationsteam. In der Regel wird ein Evaluationsteam bestimmte Evaluationsstrategien und Forschungstechniken auch deshalb vorschlagen, weil sie mit dem entsprechenden Analyseinstrumentarium vertraut ist. Die Einholung mehrerer Offerten und deren kritische Beurteilung kann helfen, das für die Fragestellung optimale Vorgehen zu finden.

Evaluationen setzen sehr gute Kenntnisse des sozial- (und allenfalls auch des natur-)wissenschaftlichen Instrumentariums voraus. Auch wenn in Abbildung 9 versucht wird, gewisse Hinweise zum Vorgehen zu geben, muss betont werden, dass die Bestimmung der Evaluationsstrategie und der Forschungstechnik eine Kunst und demnach auch ein kreativer Prozess ist, für den es keine pfannenfertigen Rezepte gibt.

Abbildung 9 Wahl von Evaluationsstrategien und Forschungstechniken



Die Wahl der Evaluationsstrategie und der Forschungstechniken ist ein Optimierungsprozess, bei welchem es gilt, die Fragestellung, die Evaluationsstrategien, die Forschungstechniken sowie die finanziellen und zeitlichen Vorgaben in eine bestmögliche Beziehung zueinander zu bringen. Dies ist eine höchst anspruchsvolle Aufgabe, die indessen von einigen Evaluationsteams im Rahmen des NFP 27 gemeistert wurde. Welche Fragen und Aspekte dabei angesprochen werden, kann vielleicht am besten anhand einiger Schwächen von Evaluationen, denen wir im Laufe der letzten Jahre in Gesuchen und Projekten in- und ausserhalb des NFP 27 begegnet sind, dargestellt werden.

- Verwendung deskriptiver Ansätze zur Beantwortung kausalorientierter Fragen. Dies ist der weitaus häufigste Fehler, der nach unseren Erfahrungen bei der Wahl des Evaluationsdesigns gemacht wird. Die Beantwortung kausalorientierter Fragen geschieht mit Vorteil im Rahmen explizit vergleichender Evaluationsstrategien: als zweitbeste Lösung und für den Fall, dass keine Vergleiche möglich sind, können auch implizit-vergleichende Eva

lationsstrategien eingesetzt werden. In letzterem Falle ist es notwendig, dass sich das Evaluationsteam ihrer beschränkten Aussagekraft bewusst ist, um nicht unzulässige Schlussfolgerungen zu ziehen.

- Zu komplexes Evaluationsdesign. Die Erarbeitung eines differenzierten Wirkungsmodells ist ein sinnvoller Schritt im Rahmen jeder Evaluation. Vor deren Inangriffnahme (d.h. bei der Offertstellung) muss sich das Evaluationsteam bewusst sein, dass auf ein Wirkungsmodell mit vielen Variablen nur dann kausalorientierte Antworten möglich sind, wenn eine sehr grosse Zahl von Beobachtungsdaten im Längs- oder Querschnitt (Kontrollgruppen- und Querschnittvergleiche) ausgewertet werden kann. Andernfalls ist es nicht möglich, den Einfluss der einzelnen Variablen auf das Wirkungsfeld zu ermitteln. Verschiedene Evaluationsteams sind sich dieser Voraussetzungen nicht bewusst.
- Ungenügend grosse Stichproben für interkantonale Auswertungen von Umfragen. Ziel einzelner Evaluationen ist es, aufgrund von Umfragen Aussagen über den kantonalen Vollzug machen zu können. Damit dies möglich ist, braucht es aber pro Kanton eine Stichprobe von mindestens 100 Personen pro Untersuchungssegment. Auf gesamtschweizerischer Ebene ist somit eine sehr hohe Stichprobe erforderlich.¹⁴
- Übertragung der Ergebnisse aus einem Segment auf eine Grundgesamtheit ohne vorangegangene statistische Absicherung. Oft sind sich Evaluationsteams zu wenig bewusst, dass sie bestimmte Daten nur bei einem Teilsegment (z.B. in bestimmten Kantonen, bei bestimmten Firmen) erhoben haben, und ziehen Schlussfolgerungen auf eine grössere Grundgesamtheit (z.B. die Schweiz als Ganzes, alle Firmen), als aus statistischen Gründen vertretbar wäre.
- Sehr aufwendige Datenauswertungstechniken bei ungenügender Datenqualität. Für bestimmte Auswertungen werden teilweise sehr aufwendige (ökonomische) Verfahren eingesetzt. Der entsprechende Aufwand lohnt sich aber nur dann, wenn die Daten den entsprechenden theoretischen Variablen entsprechen und die Datenqualität genügend ist.
- "Umfunktionieren" einer Kontroll-Evaluation in eine Verbesserungs-Evaluation. Wir möchten diesen Punkt erwähnen, obwohl er die Evaluationsmethodik nur indirekt berührt. In verschiedenen Evaluationen, die von Kontrollinstanzen in Auftrag gegeben wurden, versuchte das Evaluationsteam nachträglich, den Betroffenen Lernimpulse zu vermitteln. Dazu ist es in der Regel zu diesem Zeitpunkt zu spät. Sofern eine Evaluation Lernimpulse

¹⁴Durch kantonal geschichtete Stichproben könnte die Stichprobengrösse vergleichsweise vermindert werden; dadurch wird aber möglicherweise für kleinere gesellschaftliche Teilsegmente die Aussagekraft auf gesamtschweizerischer Ebene vermindert.

auslösen soll, müssen "vertrauensbildende Massnahmen" am Anfang und nicht am Ende einer Evaluation stehen.

Anhang 1 Erläuterung der wichtigsten Begriffe

Bias: Ausmass, in welchen eine Messung, Stichprobe oder analytische Methode einen bestimmten Wert unter- oder überschätzt bzw. von spezifischen theoretischen oder inhaltlichen Voraussetzungen ausgeht.

Black box-Evaluation: Untersuchung und Beurteilung der Wirkungen, ohne zu ermitteln, wie sie zustandekommen. Verzichtet wird namentlich auf eine Untersuchung des Vollzugs.

Controlling: Betriebswirtschaftlich inspiriertes Konzept einer laufenden Erfolgskontrolle. Während Controlling in der Privatwirtschaft als Ergänzung und Weiterentwicklung des betrieblichen Rechnungswesens entstanden ist, wird der Begriff im öffentlichen Sektor z.T. viel weiter gefasst und umfasst auch die Wirkungskontrolle. Der Begriff Controlling wird in dieser Form nur im deutschsprachigen Raum verwendet; im angelsächsischen Bereich werden primär die Begriffe "evaluation", "performance review" u.a.m. verwendet.

Delphi-Umfrage: Befragung von Fachexperten zu Zukunftsfragen. Die Befragung erfolgt über mehrere Runden. Zuerst erfolgt in der Regel eine konventionelle Befragung (schriftlich, grösstenteils mit standardisierten Fragen). In der zweiten Runde werden die Experten mit den Antworten aller Experten konfrontiert, um die Weiterentwicklung ihrer Einschätzungen anzuregen. Es können auch neue Argumente in die Befragung aufgenommen werden.

diachron: bezeichnet ein Evaluationskonzept, welches die Entwicklung bestimmter Grössen im Zeitablauf untersucht (s. Längsschnitt-Vergleiche).

Effektivität: Dieses Kriterium bezieht sich auf die Umsetzung/den Vollzug der Massnahme. Es geht um den Grad der Übereinstimmung zwischen dem rechtlich geforderten bzw. dem geplanten Verhalten von Vollzugsträgern und/oder Adressaten und deren beobachtbarem Verhalten (Outputs und Impacts). Je nach Massnahmentyp stellt sich die Frage der Effektivität in besonderer Weise: Befolgungsgrad bei Geboten und Verboten, Vollzugsgrad bei öffentlichrechtlichen Normen, die der Umsetzung durch Verwaltungsinstanzen bedürfen (z.B. Erteilung von Bewilligungen), Grad der Inanspruchnahme bei den privatrechtlichen Normen mit fakultativem Charakter, aber auch z.B. bei Finanzhilfen, Beachtungsgrad bei Informationen und Empfehlungen etc. Die Effektivität ist eine notwendige, aber nicht ausreichende Voraussetzung der Wirksamkeit.

Effizienz: (→Wirtschaftlichkeit)

Evaluation: Der Begriff Evaluation bedeutet dem Wortsinn nach Beurteilung eines Gegenstandes oder Sachverhalts. Er kann in einem sehr breiten Sinne gebraucht werden, der weder den zu evaluierenden Sachverhalt (Produkt, Institution, Massnahme, grösserer Politikbereich), noch das Ziel der Evaluation (Kaufentscheid, Reorganisation, Massnahmenverbesserung), noch die Evaluationsstrategie (sozialwissenschaftliche Untersuchungsmethoden, impressionistische Bewertung), noch die evaluierende Instanz (für Massnahmen verantwortliche Behörde, Aufsichtsbehörde, freie wissenschaftliche Forschung u.a.m.) einschränkt. Im Leitfaden beschränken wir uns indessen erstens auf den öffentlichen Bereich (wobei Evaluationsstrategien durchaus auf bestimmte Fragen in Firmen, z.B. auf interne Ausbildungsprogramme, Beiträge an die Kulturförderung u.a.m. angewendet werden können). Zweitens konzentrieren wir uns auf methodische Evaluationen. Das sind Evaluationen, welche a) nachvollziehbare Überlegungen und Aussagen enthalten, b) auf einem systematischen Vorgehen beruhen, welches eine Ermittlung der wichtigen Wirkungszusammenhänge anstrebt, c) dem Zweck dient, möglichst objektive Aussagen über Vollzug und Wirkungen zu gewinnen, und d) in einer Form vorliegt, die ihren methodischen Charakter erkennen lässt. Drittens konzentrieren wir uns auf die Evaluation von Massnahmen. Dies sind analytisch isolierbare Tätigkeiten des Staates im Hinblick auf bestimmte, allerdings oft vage formulierte Ziele, die sich auf Veränderungen in Wirtschaft und Gesellschaft beziehen. Sie beinhalten in der Regel personelle und finanzielle Ressourcen.

Evaluationsdesign: Konkrete Untersuchungsanlage zur Beantwortung bestimmter Forschungsfragen. Das Evaluationsdesign enthält Angaben zur Evaluationsstrategie und zu den gewählten Forschungstechniken.

Evaluationsforschung: wird weitgehend identisch mit dem Begriff "Evaluation" verwendet, wobei namentlich die Verwendung von Methoden der angewandten Sozialwissenschaften im Vordergrund steht.

Evaluationsstrategie: Hauptansatz der Untersuchung zur Beantwortung der gestellten Fragen im Hinblick auf den Vollzug, die Wirkungen oder die Wirtschaftlichkeit der staatlichen Massnahme. Konzeptionelle Antwort auf die Untersuchungsfrage.

Evaluierbarkeitsprüfung: Eine Evaluierbarkeitsprüfung (oder Vor-Evaluation) dient der Vorbereitung und Planung einer Evaluation. Sie soll namentlich über die Ziele der staatlichen Massnahme, das Wirkungsmodell, die Evaluationsstrategie, die Forschungstechniken sowie die Datenlage Aufschluss geben. Eine Evaluierbarkeitsprüfung kann auch ohne spätere Evaluation sinn

voll sein und Hinweise für eine bessere Selbstevaluation der Massnahmen geben.

ex ante Evaluation: (→prospektive Evaluation)

ex post Evaluation: (→retrospektive Evaluation)

externe Validität: (→Übertragbarkeit)

Forschungstechnik: die gewählte sozial- (allenfalls auch naturwissenschaftliche) Methode zur Gewinnung und Auswertung der Daten im Hinblick auf die Beantwortung der gestellten Untersuchungsfrage.

Generalisierbarkeit von Evaluationsresultaten (→Übertragbarkeit)

Impacts: Änderungen des Verhaltens der Adressaten der Massnahme

Interne Validität: Ausmass, in welchem in einer Untersuchung/Evaluation die Ursachen bzw. Wirkungen bestimmter Phänomene ermittelt werden können.

Kontrollgruppen (-Vergleich): Ermittlung der Wirkungen einer staatlichen Massnahme durch Vergleich des Zustands, der bei der Zielgruppe eingetreten ist, und des Zustands, der bei einer (der Massnahme nicht ausgesetzten) Kontrollgruppe eingetreten ist.

Kosten-Effektivitäts-Analysen: Ermittlung der Kosten unterschiedlicher Massnahmen (z.B. Reduktion Geschwindigkeit, Ausbau von Strassen, Gurten-tragpflicht auf den Rücksitzen) zur Erzielung einer bestimmten Wirkung (z.B. Reduktion der Strassentoten um eine bestimmte Anzahl Personen). Sie bilden eine Alternative zu Kosten-Nutzen-Analysen, wenn der Nutzen nicht in Geldwert bestimmbar ist.

Kosten-Nutzen-Analysen: Verfahren zur Bestimmung der Wirtschaftlichkeit einer Massnahme, wobei Kosten und Wirkungen(=Nutzen) in Geldwert quantifiziert werden.

Leistungsindikatoren: Statistiken, welche Rückschlüsse über den Output von Verwaltungen und die daraus entstehenden Wirkungen ermöglichen (sollen) und sowohl monetäre wie nichtmonetäre Aspekte beinhalten.

Längsschnitt(-Vergleich): Untersuchungen, welche durch Beobachtung von bestimmten Einheiten im Zeitablauf Rückschlüsse auf die Wirkungen von Massnahmen zu ziehen versuchen.

Massnahme, staatliche: Analytisch isolierbare Tätigkeiten des Staates im Hinblick auf bestimmte, allerdings oft vage formulierte Ziele, die sich auf Veränderungen in Wirtschaft und Gesellschaft beziehen. Sie beinhalten in der Regel personelle und finanzielle Ressourcen.

Monitoring: Laufende Beobachtung, Bereitstellung ständiger Daten über Vollzugs- und Wirkungsaspekte.

Outcomes/Wirkungen: Auf eine bestimmte Massnahme zurückführbare Veränderungen in Wirtschaft, Gesellschaft oder der Umwelt.

Output: Die von der Verwaltung hergestellten Güter bzw. erbrachten Dienstleistungen (Drucksachen, Verordnungen, Bewilligungen, Bussen, erteilte Beiträge u.a.m.).

Prospektive Evaluation: Abschätzung der mutmasslichen Wirkungen vorgesehener Massnahmen.

Qualitative Daten: Daten, welche in Form von Worten ausgedrückt werden.

Quantitative Daten: Information, welche in Form von Zahlen ausgedrückt wird (erfordert "Messung" als Verfahren der Zuteilung von Zahlen auf beobachtete Sachverhalte oder Ereignisse).

Querschnitt (-Vergleich): Vergleich von Gebieten (Länder, Kantone, Gemeinde) im Hinblick auf den Vollzug und die Wirkungen bestimmter staatlicher Massnahmen.

Repräsentative Stichprobe: Eine Stichprobe, welche dieselbe Verteilung von Merkmalen aufweist wie die Bevölkerung, aus welcher die Stichprobe gezogen worden ist.

Retrospektive Evaluation: Empirische Erfassung der Wirkungen von Massnahmen, welche in Kraft gesetzt und vollzogen worden sind.

synchron: bezeichnet ein Evaluationskonzept, welches nicht den Zeitpunkt variiert (→Längsschnitt-Vergleich), sondern verschiedene Gruppen, Gebiete o.ä. zu einem selben Zeitpunkt untersucht (s. Kontrollgruppen-Vergleiche, Querschnitt-Vergleiche).

Übertragbarkeit von Evaluationsresultaten: Ausmass, in welchem die Resultate einer Wirkungsanalyse über die untersuchten Einheiten hinaus auf andere Gegebenheiten (andere Gruppen, andere räumliche Einheiten, unterschiedliche Zeitpunkte) übertragen (oder verallgemeinert) werden können.

Vollzug: Zusammenspiel der Aktivitäten der Verwaltung einerseits und der durch die staatlichen Massnahmen anvisierten Adressaten andererseits.

Wirksamkeit: Verhältnis zwischen den beabsichtigten und den tatsächlich eingetretenen Outcomes/Wirkungen, die auf die staatliche Massnahme zurückzuführen sind.

Wirkungen: Auf eine bestimmte Massnahme zurückführbare Veränderungen in Wirtschaft und Gesellschaft.

Wirkungsanalysen/Wirkungskontrollen: Untersuchung der Wirkungen staatlicher Massnahmen. Dabei werden sowohl die zielkonformen wie auch die zielkonträren Wirkungen untersucht. Besondere Beachtung wird oft Mitnahmeeffekten, Verdrängungseffekten und anderen Nebenwirkungen geschenkt.

Wirkungsmodell: Annahmen (welche ev. auf wissenschaftlichen Theorien beruhen können) darüber, wie eine staatliche Massnahme wirkt. Ein Wirkungsmodell dient der Konzeption oder Änderung einer Massnahme und ist Grundlage für eine Evaluation der Wirkungen.

Wirkungspfad: Rekonstruktion der Wirkungsketten ausgehend von einer bestimmten staatlichen Massnahme, ev. bezogen auf bestimmte, spezifische Wirkungsaspekte.

Wirtschaftlichkeit: Verhältnis zwischen dem vorgesehenen oder tatsächlichen Output und den vorgesehenen oder tatsächlichen Ressourcen (Wirtschaftlichkeit der Outputproduktion) bzw. Verhältnis zwischen den voraussichtlichen oder tatsächlichen Outcomes/Wirkungen und den vorgesehenen oder tatsächlichen Ressourcen (=outcomebezogene Wirtschaftlichkeit). Eine bessere Wirtschaftlichkeit lässt sich entweder dadurch erreichen, dass zur Erzielung bestimmter zielbezogener Outcomes/Wirkungen weniger Ressourcen eingesetzt werden oder dass mit den eingesetzten Ressourcen ein Mehr an Output bzw. zielbezogenen Outcomes/Wirkungen erzielt wird. Unter Wirtschaftlichkeit wird in der Evaluationsforschung oft auch das in monetären Grössen ausgedrückte Verhältnis zwischen Nutzen/Wirkungen und den dafür eingesetzten Ressourcen/Mitteln verstanden (→Kosten-Nutzen Analysen).

Zeitreihenuntersuchungen: Untersuchungen, welche durch Beobachtung der Veränderungen bestimmter Einheiten im Zeitablauf Rückschlüsse auf die Wirkungen von Massnahmen zu treffen versuchen.

Zielerreichung: Verhältnis zwischen vorgegebenen Zielen und tatsächlich eingetretener Situation (ohne kausale Zuordnung zu bestimmten staatlichen Massnahmen).

Anhang 2 Glossar der wichtigsten Begriffe in vier Sprachen

Deutsch	Französisch	Italienisch	Englisch
Beteiligte	partie prenante, protagoniste	detentori di interessi	stakeholder
Bilanz-Evaluation	évaluation récapitulative	valutazione di sintesi	summative evaluation
Effektivität	effectivité	effettività	implementation, compliance
Erfolgskontrolle	analyse des résultats	analisi dei risultati	results analysis
Ergebnis	résultats, réalisations, extrants	risultati, rendimenti	outputs, results, products
Evaluation	évaluation	valutazione	evaluation
Evaludierbarkeitsprüfung, Vorstudie	étude d'évaluabilité	pre-valutazione	evaluability assessment
Kontrollgruppe	groupe de contrôle, groupe témoin	gruppo di controllo	control group
Kosten-Nutzen-Analyse	analyse coût-avantage	analisi costi-benefici	cost-benefit analysis
Kosten-Wirksamkeits-Analyse	analyse coût-efficacité	analisi costi-efficacia	cost-effectiveness-analysis
Massnahme	mesure	metodo	measure, program
Mitnahmeeffekt	effet d'aubaine	impatto lordi	deadweights
Politik, öffentliche	politique publique	politica	policy
Sparsamkeit	emploi ménager des fonds	economicità	economy
Verbesserungs-Evaluation	évaluation endo-formative	valutazione formativa	formative evaluation
Vollzug	mise en oeuvre, exécution	esecuzione	implementation
Vollzugskontrolle, Monitoring	pilotage, suivi	monitoraggio	monitoring
Wirksamkeit,	efficacité	efficacia	effectiveness
Wirkung, Effekt	effet, impact	impatto	outcome, impact, effect
Wirtschaftlichkeit, Effizienz	efficience	efficienza	efficiency
Ziel	objectif, but, finalité	obiettivo	goal, objective, target, aim
Zielgruppe	groupe cible, public cible	gruppo obiettivo	target group

Anhang 3 Literatur

(⁺=Kurzberichte von Projekten des NFP 27; * = Schlussberichte von Projekten des NFP 27; ** vom NFP 27 mitfinanzierte Kurzevaluationen)

- AGREB 1990, Leitfaden für das Management von Forschungsprojekten der Bundesverwaltung, Bern, Bundesamt für Bildung und Wissenschaft (Bezugsquelle)
- Arbeitsgruppe Verkehrssicherheit 1983, Versuch Tempo 50, Schlussbericht, EDMZ (A 22315), Bern
- Atteslander Peter 1985, Methoden der empirischen Sozialforschung, Berlin/New York
- Balthasar Andreas (Hrsg.) 1993, Einführung in die Evaluationsforschung für Politikwissenschaftlerinnen und Politikwissenschaftler, Forschungszentrum für Schweizerische Politik der Universität Bern
- Balthasar Andreas/Knöpfel Carlo 1994*, Umweltpolitik und technische Entwicklung, Eine politikwissenschaftliche Evaluation am Beispiel der Heizungen, Basel, Helbing & Lichtenhahn
- Balthasar Andreas/Knöpfel Carlo 1994⁺, Umweltvorschriften und technische Entwicklung, NFP 27, Bern
- Bericht der Expertengruppe Energieszenarien 1988, Energiepolitische Entscheidungsmöglichkeiten, Herausgeber: Eidg. Verkehrs- und Energiewirtschaftsdepartement (Bezugsquelle EDMZ, 9.88 3600)
- Bierter Willy/Binder Hans-Martin 1993*, Unternehmerische Innovationsprozesse, Die Wirksamkeit staatlicher Förderungsmassnahmen - eine Evaluation, Chur/Zürich, Rüegger
- Bierter Willy/Binder Hans-Martin 1994⁺, Unternehmerische Innovationsprozesse, Die Wirksamkeit staatlicher Förderungsmassnahmen - eine Evaluation, NFP 27, Bern
- Binder Hans Martin 1994, Lernprozess als Ziel einer interaktiven Evaluationsstrategie, in: Bussmann Werner (Hrsg.), Lernen in Verwaltungen und Policy-Netzwerken, Chur/Zürich, 75-84
- Böhret Carl/Hugger Werner 1980, Test und Prüfung von Gesetzesentwürfen, Köln
- Bussmann Werner (Hrsg.) 1994, Lernen in Verwaltungen und Policy-Netzwerken, Chur/Zürich, Rüegger Verlag
- Bussmann Werner 1992, Rechtsetzung unter Unsicherheit, in: Gesetzgebung heute, 2, 39-52

-
- Chelimsky Eleanor 1987, Linking Program Evaluation to User Needs, in: Palumbo Dennis J. (Hrsg.), *The Politics of Program Evaluation*, Newbury Park, Sage, 72-99
- Delley Jean-Daniel et al. 1982, *Le droit en action*, St. Saphorin
- Delley Jean-Daniel 1994* (avec la collaboration de Torricelli Fischer Claire et Rincon José-M.), *Impact et limites des mesures de protection de l'air, le cas des installation industrielles et artisanales*, travaux CETEL n° 42
- ECOPLAN 1993a**, *Gestaltungsmöglichkeiten und Auswirkungen verschiedener Submissionsordnungen*, Evaluation im Auftrag der Bau-, Verkehrs- und Energiedirektion des Kantons Bern, (veröffentlichter Bericht)
- ECOPLAN 1993b**, *Schneeräumung von Urner Passstrassen*, Evaluation im Auftrag der Baudirektion des Kantons Uri, (veröffentlichter Bericht)
- Eidgenössisches Personalamt (Hrsg.) 1993, *Leitfaden für den Einsatz externer Beraterinnen und Berater*, Abteilung Personal- und Organisationsentwicklung (Bezugsort), Bern
- ERS Standards Committee 1982, *Evaluation Research Society standards for program evaluation*, in: Rossi Peter H. (Ed.), *Standards of evaluation practice*, *New Directions in Program Evaluation*, Vol. 15, San Francisco, Jossey-Bass
- Evaluationen in der Agrarpolitik I 1994*, Nationales Forschungsprogramm 27, Reihe Schlussberichte, Bern, mit folgendem Beitrag: Felder Stefan/Rieder Peter, *Analyse des Angebotsverhaltens und der Produktionsstruktur im Agrarsektor mit Hilfe eines LP Strukturmodells*
- Evaluationen in der Agrarpolitik II 1994*, Nationales Forschungsprogramm 27, Reihe Schlussberichte, Bern, mit folgenden Beiträgen: Baur Priska/Meyer Leo/Rieder Peter, *Politökonomische Analyse der Getreidepolitik*; Felder Stefan/Meyer Leo/Rieder Peter, *Vorzeitige Stilllegung von Höchsttierbeständen in der Fleisch- und Eierproduktion 1980-1984*; Cavagn Gion/Meyer Leo/Rieder Peter, *Analyse des Angebotsverhaltens der Getreideproduzenten*
- Forster Sophie/Jeanrenaud Claude 1993**, *La promotion économique neuchâteloise, bilan 1993*, Neuchâtel
- Freiburghaus Dieter 1987, *Präventivmassnahmen gegen die Arbeitslosigkeit in der Schweiz*, Bern, Haupt
- General Accounting Office 1990a, *Case Study Evaluations*
- General Accounting Office 1990b, *Prospective Evaluation Methods: The Prospective Evaluation Synthesis*
- General Accounting Office 1991a, *Designing Evaluations*
- General Accounting Office 1991b, *Using Structured Interviewing Techniques*
- General Accounting Office 1992a, *Quantitative Data Analysis: An Introduction*

- General Accounting Office 1992b, The Evaluation Synthesis
- Gerheuser Frohmüt W. 1991, Die Wirkungen der beruflichen Vorsorge auf den Arbeitsmarkt, Chur/Zürich, Rüegger Verlag
- Gerheuser Frohmüt W./Schmid André 1993*, Bedingt wirksam: Das Lohn-gleichheitspostulat des Heimarbeitsgesetzes, Bern, NFP 27, Serie Schlussberichte
- Gerheuser Frohmüt W./Schmid André 1993+, Bedingt wirksam: Das Lohn-gleichheitspostulat des Heimarbeitsgesetzes, NFP 27, Bern
- Hanser Christian/Kuster Jürg/Cavelti Guido 1994*, Auswirkungen der staatli-chen Hotellerieförderung, Bern, EDMZ
- Hanser Christian/Kuster Jürg/Cavelti Guido 1994+, Auswirkungen der staatli-chen Hotellerieförderung, Bern, NFP 27, Reihe Kurzberichte
- Hausheer Othmar 1991, Die Kehrachtsackgebühr - ein wirkungsvolles um-weltpolitisches Instrument? Eine empirische Untersuchung, Chur/Zürich, Rüegger Verlag
- Hirsch Madeleine et al. 1995, Zukunft Skitourismus
- INFRAS 1990, Evaluation de la taxe sur les sacs à déchets urbains, cahier de l'environnement, no 136, OFEFP, Berne
- INFRAS 1990, Evaluation Kehrachtsackgebühr, Schriftenreihe Umwelt Nr. 136, BUWAL, Bern
- Lamnek S. 1988, Qualitative Sozialforschung, Band I und II, München
- Linder Wolf u.a. 1990, Verbrauchsabhängige Heizkostenabrechnung, Förde-rung erneuerbarer Energien, Bern, Forschungszentrum für schweizerische Politik
- Littmann-Wernli Sabina 1991, Auswirkungen eines Diskriminierungsverbotes und eines verstärkten Kündigungsschutzes, überarbeitete Fassung eines Gutachtens im Auftrag des Bundesamtes für Justiz, in: Gesetzgebung heute 1991/2, 31-82
- Lüscher Claude/Wyss Christian 1994*, Geringere Umweltbelastung dank Umweltverträglichkeitsprüfung? Bern, NFP 27, Reihe Schlussberichte
- Lüscher Claude/Wyss Christian 1994+ Geringere Umweltbelastung dank Umweltverträglichkeitsprüfung? Bern, NFP 27, Reihe Kurzberichte
- Mader Luzius 1988, Experimentelle Gesetzgebung, in: Grimm Dieter/Maihofer Werner (Hrsg.), Gesetzgebungstheorie und Rechtspolitik, Opladen, Westdeutscher Verlag, 211 - 221
- Mader Luzius 1993, La législation expérimentale en Suisse, in: Morand Char-les-Albert, Evaluation législative et lois expérimentales, Presses universi-taires d'Aix-Marseille, 221-242
- Mastronardi Philippe 1991, Experimentelle Rechtsetzung im Bund, in: Zeit-schrift für Schweizerisches Recht, NF Bd. 110, 1. Halbband, Heft 4/5, 449-469

- Meyer Leo/Rieder Peter 1994⁺, Evaluationen in der Agrarpolitik - Fallbeispiele im Getreidebau und der Tierhaltung, NFP 27, Bern
- Mohr Lawrence B. 1993, Impact Analysis for Program Evaluation, Newbury Park, Sage
- Monnier Eric 1987, Evaluations de l'action des pouvoirs publics, Paris
- Morand Charles-Albert (ed.) 1993, Evaluation législative et lois expérimentales, Presses universitaires d'Aix-Marseille
- Patton Michael Quinn 1990, Qualitative evaluation and research methods, 2nd edition, Newbury Park, Sage,
- Price Waterhouse 1993**, Evaluation der Zürcher Gemeinschaftszentren (ZGZ), Evaluation im Auftrag des Sozialamtes der Stadt Zürich, unveröffentlichtes Typoskript
- Rey Michel et al. 1993*, La politique valaisanne de promotion économique: une évaluation privilégiant le point de vue des acteurs, Programme national de recherche 27, Berne
- Rey Michel/Vodoz Luc 1993⁺, La promotion économique valaisanne, PNR 27, Berne
- Rieder Peter et al. 1990, Direktzahlungen an die Berglandschaft; Eine Gesetzesevaluation, Institut für Agrarwirtschaft, Zürich, VdF
- Rose Richard 1993, Lesson-drawing in Public Policy, A guide to Learning across Time and Space, Chatham House Publishers Inc.
- Rossi Peter H./Freeman Howard E., 1993, Evaluation: a Systematic Approach. Fifth Edition, Newbury Park, Sage
- Rossi Peter H./Freeman Howard/Hofmann Gerhard 1988, Programm-Evaluation: Einführung in die Methoden angewandter Sozialforschung, Stuttgart, Ferdinand Enke Verlag
- Sangra Emmanuel 1992*, Evaluation des actions d'information de la protection civile sur le comportement de la population en cas d'alarme, Editions J-M Blanc SA, Le Mont-sur-Lausanne
- Schaetti Stephan 1990, Berufliche Vorsorge und Arbeitsmarkt, WWZ-Studien, Nr. 18, Basel
- Schmitz B. 1989, Einführung in die Zeitreihenanalyse: Modelle, Softwarebeschreibung, Anwendungen, Bern
- Schnyder Stefan 1993**, Über die Verteilung und Wirkung der Investitionshilfen im landwirtschaftlichen Hochbau des Kantons Zürich, Studie im Auftrag des Meliorations- und Vermessungsamtes der Stadt Zürich, unveröffentlichtes Typoskript
- Schulz Hans-Rudolf/Muggli Christoph/Hübschle Jörg 1993*, Wohneigentumsförderung durch den Bund, Bern, Schriftenreihe Wohnungswesen, Band 55, EDMZ (Bezugsquelle)

-
- Schulz Hans-Rudolf/Muggli Christoph/Hübschle Jörg 1993⁺, Die Wirksamkeit der staatlichen Wohneigentumsförderung, NFP 27, Bern
- Stalder Peter/Frick Andres/Gaillard Serge/Salzgeber Renate 1994*, Die Auswirkungen der Ausländerpolitik auf den Strukturwandel der Schweiz, Bern, NFP 27, Reihe Schlussberichte
- Stalder Peter/Frick Andres/Gaillard Serge/Salzgeber Renate⁺, Die Auswirkungen der Ausländerpolitik auf den Strukturwandel der Schweiz, Bern, NFP 27, Reihe Kurzberichte
- Wholey Joseph S. 1983, Evaluation and Effective Public Management, Boston, Little Brown
- Widmer Thomas 1991, Evaluation von Massnahmen zur Luftreinhaltepolitik in der Schweiz, Eine quasi-experimentelle Interventionsanalyse nach dem Ansatz Box/Tiao, Zürich, Rüegger Verlag
- Yin Robert K. 1985, Case study research, Newbury Park, Sage

Anhang 4 Stichwortverzeichnis

- Begleitgruppen, 33-34
Betroffene, 67
Bias, 95
Black box-Evaluation, 37, 95
- Controlling, 95
- Delphi-Umfrage, 18, 95
diachron, 95
- Effektivität, 53-54, 64, 95
Effizienz, s. Wirtschaftlichkeit
Evaluation, Begriff, 51, 96
Evaluationsdesign, 60, 96
Evaluationsforschung, 59, 96
Evaluationsstrategie, 60-69, 92, 96
Evaluationssynthesen, 84
Evaluierbarkeitsprüfung, 24-25, 96
ex ante Evaluation, s. prospektive Evaluation
Experiment, 69-70
Experten, Schlüsselpersonen, 67
externe Evaluation, 21
- Fallstudien, 68, 78-80
Forschungstechnik, 60, 76-78, 92, 97
- Generalisierbarkeit, s. Übertragbarkeit
- interaktive Evaluationsmethoden, 42
interne Evaluation, 19-20
interne Validität, 97
- Kausalität, 61-62, 64f.
Kontrollgruppen(-Vergleich), 47-49, 69-71, 97
- Kosten von Evaluationen, 56-57
Kosten, s. Wirtschaftlichkeit
Kosten-Effektivitäts-Analysen, 17, 38, 97
Kosten-Nutzen-Analysen, 17, 38, 97
- Längsschnitt-Vergleich, 47, 74-76, 97
Leistungsindikatoren, 17, 97
Lernprozesse, 42-43
- Mediation, 42
Methodenmix, 89
Monitoring, 17, 98
- Offerten für Evaluationen, 25, 26-31
Output, 53-54, 98
- Panel-Umfragen, 82
Programm, 52
prospektive Evaluation, 17-18, 40-42, 56, 98
- Qualitativ, 77, 86, 98
Quantitativ, 77, 86, 98
Querschnitt-Vergleiche, 71-74, 98
- Regressions-Diskontinuitäts-Analysen, 47-48, 70-71
Repräsentativität, 63, 98
retrospektive Evaluation, 55, 98
- Sekundärstatistische Analysen, 83
staatliche Massnahmen, 51-52, 97
synchron, 98
- Trend-/Ist-Vergleich, 76
Triangulation, 89-90

-
- Übertragbarkeit (von Evaluations-
resultaten), 98
Umfragen, 80-83
Umsetzung, 28-29, 36-42
- Versuchsregelung, 49
Vollzug, 16, 37-38, 45-46, 53-54, 98
Vollzugssteuerung, s. Monitoring
- Wirksamkeit, 45-49, 53-54, 64, 98
- Wirkungen, 16, 53-54, 98
Wirkungsanalysen/-kontrollen, 16,
99
Wirkungsmodell, 56, 85, 99
Wirkungspfad, 99
Wirtschaftlichkeit, 16, 38-39, 53-54,
99
- Zeitreihen, 99
Zeitbedarf für Evaluationen, 56-57
Zielerreichung, 54-55, 64, 99

Veröffentlichungen des NFP 27

Schlussberichte

Rey Michel, et al., La politique valaisanne de promotion économique: une évaluation privilégiant le point de vue des acteurs, Berne 1993, 210 p., 20 francs

Gerheuser Frohmüt W., Schmid André, Bedingt wirksam: Das Lohngleichheitspostulat des Heimarbeitsgesetzes, Bern 1993, 154 Seiten, Fr. 20.–

Evaluationen in der Agrarpolitik I, Bern 1994, 192 Seiten, Fr. 20.–

Stefan Felder, Peter Rieder, Analyse des Angebotsverhaltens und der Produktionsstruktur im Agrarsektor mit Hilfe eines LP Strukturmodells

Evaluationen in der Agrarpolitik II, Bern 1994, 154 Seiten, Fr. 10.–

Priska Baur, Leo Meyer, Peter Rieder; Politökonomische Analyse der Getreidepolitik

Stefan Felder; Leo Meyer, Peter Rieder; Vorzeitige Stilllegung von Höchsttierbeständen in der Fleisch- und Eierproduktion 1980-1984

Gion Cavegn, Leo Meyer, Peter Rieder; Analyse des Angebotsverhaltens der Getreideproduzenten

Stalder Peter, Frick Andres, Gaillard Serge, Salzgeber Renate, Die Auswirkungen der Ausländerpolitik auf den Strukturwandel der Schweiz, Bern 1994, 97 Seiten, Fr. 20.–

Lüscher Claude, Wyss Christian, Geringere Umweltbelastung dank Umweltverträglichkeitsprüfung? Bern 1994, 85 Seiten, Fr. 20.–

Kurzberichte

Rey Michel, Vodoz Luc, La promotion économique valaisanne, Berne 1993

Schulz Hans-Rudolf, Muggli Christoph, Hübschle Jörg, Die Wirksamkeit der staatlichen Wohneigentumsförderung, Bern 1993

Gerheuser Frohmüt W., Schmid André, Bedingt wirksam: Das Lohngleichheitspostulat des Heimarbeitsgesetzes, Bern 1993

Bierter Willy, Binder Hans-Martin, Unternehmerische Innovationsprozesse, Die Wirksamkeit staatlicher Förderungsmassnahmen-eine Evaluation, Bern 1994

Balthasar Andreas, Knöpfel Carlo, Umweltvorschriften und technische Entwicklung, Bern 1994

Meyer Leo, Rieder Peter, Evaluationen in der Agrarpolitik - Fallbeispiele im Getreidebau und der Tierhaltung, Bern 1994

Hanser Christian, Kuster Jürg, Cavelti Guido, Auswirkungen der staatlichen Hotellerieförderung, Bern 1994

Stalder Peter, Frick Andres, Gaillard Serge, Salzgeber Renate, Die Auswirkungen der Ausländerpolitik auf den Strukturwandel der Schweiz, Bern 1994

Lüscher Claude, Wyss Christian, Geringere Umweltbelastung dank Umweltverträglichkeitsprüfung? Bern 1994

Delley Jean-Daniel, Impact et limites des mesures de protection de l'air. Le cas de l'industrie et de l'artisanat, Berne 1994

Diskussionsbeiträge

Evaluationen: Wozu und für wen? Tagung vom 14. August 1991 in Bern, mit Beiträgen von Eleanor Chelimsky, Patrick Viveret, Dieter Freiburghaus und Werner Bussmann, 52 Seiten, gratis, lieferbar solange Vorrat

Die Veröffentlichungen können bezogen werden:

Schweizerischer Nationalfonds, Abt. IV/NFP 27, Postfach 2338, 3001 Bern, Tel. 031 308 23 49

Weitere Veröffentlichungen aus dem NFP 27

Bussmann Werner (Hrsg.), Lernen in staatlichen Verwaltungen und Policy-Netzwerken, Chur/Zürich, Rüegger 1994, 144 Seiten, Fr. 34.--

Morand Charles-Albert (ed.), Evaluation législative et lois expérimentales, Presses universitaires d'Aix-Marseille 1993

Sangra Emmanuel, Evaluation des actions d'information de la protection civile sur le comportement de la population au cas d'alarme, Editions J-M Blanc SA, Le Mont-sur-Lausanne 1992

Schulz Hans-Rudolf, Muggli Chrisoph, Hübschle Jörg, Die Wirksamkeit der staatlichen Wohneigentumsförderung, Schriftenreihe Wohnungswesen (Hrsg. Bundesamt für Wohnungswesen), Band 55, Bern 1993, Bezug bei der EDMZ

Bierter Willy, Binder Hans-Martin, Unternehmerische Innovationsprozesse, Die Wirksamkeit staatlicher Förderungsmassnahmen - eine Evaluation, Rüegger, Chur/ Zürich 1993

Balthasar Andreas, Knöpfel Carlo, Umweltschutz und technische Entwicklung, Helbing & Lichtenhahn, Basel 1994

Hanser Christian, Kuster Jürg, Cavelti, Guido, Hotellerieförderung durch Bund und Kantone, Evaluationen der Auswirkungen in der Hotellerie, Schriftenreihe BIGA, Beiträge zur Tourismuspolitik Nr. 3, Bern 1994, EDMZ (Form. 710.502 d)

Delley Jean-Daniel (avec la collaboration de Fischer-Torricelli Claire et Rincon José-M.), Impact et limites des mesures de protection de l'air. Le cas de l'industrie et de l'artisanat, Travaux CETEL, juin 1994

Veröffentlichungen von Kurzevaluationen, welche vom NFP 27 mitfinanziert worden sind

ECOPLAN, Gestaltungsmöglichkeiten und Auswirkungen verschiedener Submissionsordnungen, Evaluation im Auftrag der Bau-, Verkehrs- und Energiedirektion des Kantons Bern, 1993, 200 Seiten, Fr. 30.-- (BVE, Reiterstrasse 11, 3011 Bern, Tel. 031 633 31 63)

ECOPLAN, Schneeräumung von Urner Passstrassen, Evaluation im Auftrag der Baudirektion des Kantons Uri, 1993, 58 Seiten, Fr. 25.-- (Bauamt des Kantons Uri, 6460 Altdorf, Tel. 044 4 26 11)

Forster Sophie/Jeanrenaud Claude, La promotion économique neuchâteloise, bilan 1993, Neuchâtel (Institut de recherches économiques et régionales, Tel. 038 21 10 85)

Diese Veröffentlichungen sind im Buchhandel bzw. bei der erwähnten Bezugsquelle erhältlich.