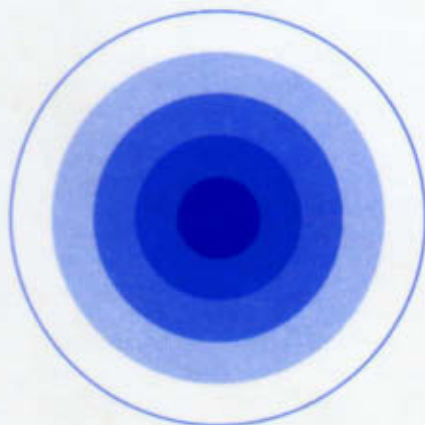

Einführung in die Politikevaluation



Werner Bussmann
Ulrich Klöti
Peter Knoepfel
(Hrsg.)

Dies ist eine redigitalisierte Fassung des Buches,
mit kleinen formalen Abweichungen
© 2004 bei den Herausgebern

Helbing & Lichtenhahn

Werner Bussmann/Ulrich Klöti/ Peter Knoepfel

Einführung in die Politikevaluation

Werner Bussmann
Ulrich Klöti
Peter Knoepfel
(Hrsg.)

Einführung in die Politikevaluation

Helbing & Lichtenhahn
Basel und Frankfurt am Main

Publiziert mit Unterstützung des Schweizerischen
Nationalfonds zur Förderung der wissenschaftlichen Forschung.

Die Deutsche Bibliothek - CIP Einheitsaufnahme

Einführung in die Politikevaluation / Werner Bussmann ... (Hrsg.). - Basel ;
Frankfurt am Main : Helbing und Lichtenhahn, 1997

ISBN 3-7190-1581-5

Alle Rechte vorbehalten. Das Werk und seine Teile sind urheberrechtlich
geschützt.

Jede Verwertung in anderen als in den gesetzlich zugelassenen Fällen bedarf
deshalb der vorherigen Einwilligung des Verlages.

ISBN 3-7190-1581-5

Bestellnummer 21 01581

© 1997 by Helbing & Lichtenhahn Verlag AG, Basel

Printed in Germany

INHALTSVERZEICHNIS

INHALTSVERZEICHNIS	V
---------------------------	----------

AUTORENNACHWEIS	XI
------------------------	-----------

ABKÜRZUNGEN	XIII
--------------------	-------------

VORWORT	XVII
----------------	-------------

EINFÜHRUNG DER HERAUSGEBER	XIX
-----------------------------------	------------

TEIL I: HINTERGRÜNDE UND ENTWICKLUNGEN I

<i>KAPITEL 1: DIE ENTWICKLUNG VON EVALUATIONEN IM INTERNATIONALEN KONTEXT</i>	4
--	----------

1.1. Zwei Wellen der Institutionalisierung	4
1.2. Drei Entwicklungslinien	5
1.3. Ursachen der Entwicklung	6
1.3.1. Veränderung der fiskalischen Situation	6
1.3.2. Politische Konstellationen	7
1.3.3. Verfassungskontext	9
1.4. Evaluation und Finanzkontrolle	10
1.5. Formen der Institutionalisierung	11

<i>KAPITEL 2: EVALUATIONEN IN DER SCHWEIZ</i>	13
--	-----------

2.1. Rahmenbedingungen	13
2.1.1. Geschlossenheit und Offenheit des politischen Systems	13
2.1.2. "Keine Experimente" und "föderalistisches Labor" als gegenläufige Gegebenheiten	15
2.1.3. Sozialwissenschaftlicher Beratungsbedarf	16
2.1.4. Die Erschliessung der peripheren Gebiete als helvetische Variante der "Great society"	17
2.2. Wissenschaftliche Genese der Evaluationsproblematik	18
2.2.1. Vorläufer	18
2.2.2. Von der Implementations- zur Wirkungsforschung	19
2.2.3. Evaluationen und andere Instrumente der Erfolgskontrolle (Controlling und NPM)	21
2.3. Methodik: Gegebenheiten und Tendenzen	23
2.4. Rechtliche Verankerung von Evaluationen bzw. Erfolgskontrollen	25
2.4.1. Evaluationen zum bewussten Gewinnen von Erfahrungen ("experimentelle Evaluationen")	26
2.4.2. Bestimmungen über die Durchführung von Evaluationen bzw. Erfolgskontrollen	27
2.4.2.1. Evaluationsklauseln	27
2.4.2.2. Versuchserlasse	28

2.4.2.3. Budgetmittel für die Evaluation	29
2.4.3. Institutionen mit dem Auftrag, Evaluationen bzw. Erfolgskontrollen durchzuführen	29
2.5. Erfahrungen	31
2.6. Evaluationen und Konkordanzdemokratie	33

TEIL II: CHARAKTERISTIKA, OBJEKTE UND NUTZUNGSZUSAMMENHÄNGE **37**

KAPITEL 3: INHALTLICHE UND METHODISCHE ANFORDERUNGEN AN WISSENSCHAFTLICHE POLITIKEVALUATIONEN	39
3.1. Staatliches Handeln	40
3.2. Bewerten von Wirkungen	45
3.2.1. Explizite Kriterien	45
3.2.2. Ausrichten auf Wirkungen	45
3.3. Wissenschaftliche Untersuchungsanlage	48
3.4. Evaluationsstandards	55
KAPITEL 4: DIE ÖFFENTLICHE POLITIK ALS EVALUATIONSOBJEKT	58
4.1. Institutionelle und substantielle öffentliche Politiken	59
4.2. Evaluation substantieller öffentlicher Politiken und Massnahmen	62
4.2.1. Die (substantielle) "öffentliche Politik" und ihre Akteure	62
4.2.2. Gesamt- und Teilevaluationen (Einzelmassnahmen, Aktionsprogramme und Projekte) öffentlicher Politiken	66
4.3. Gegenstände von Politikevaluationen	68
4.3.1. Traditionelle Evaluationstypisierung	68
4.3.2. Typisierung nach den Stufen des Policy-Cycle	69
4.4. Prospektive und retrospektive Evaluation	74
KAPITEL 5: EVALUATIONSGEGENSTÄNDE UND EVALUATIONSKRITERIEN	78
5.1. Politikkonzept - Empirische Konsistenz und logische Kohärenz des Wirkungsmodells	78
5.2. Verwaltungsprogramm - Kohärenz	83
5.2.1. Einführung	83
5.2.2. Typische Elemente von Verwaltungsprogrammen	84
5.2.2.1. Ziel(e)	86
5.2.2.2. Evaluative Elemente	87
5.2.2.3. Operative Elemente (Instrumente)	87
5.2.2.4. Behördenarrangement, finanzielle Mittel und weitere Ressourcen	88
5.2.2.5. Instrumentelle und prozedurale Elemente	90
5.2.3. Kohärenz und Rechtmässigkeit von Verwaltungsprogrammen	91
5.3. Behördenarrangement - Eignung und Rechtmässigkeit der Vollzugsstrukturen	95
5.4. Aktionsprogramme für den Vollzug - Adäquanz der Planung	98
5.5. Output, Impact und Effektivität	100
5.5.1. Output	100
5.5.2. Impact	103
5.5.3. Effektivität	104
5.6. Outcomes, Wirksamkeit und Wirtschaftlichkeit	109
5.6.1. Outcomes	109

Inhaltsverzeichnis

VII

5.6.2. Wirksamkeit	111
5.6.3. Wirtschaftlichkeit (Effizienz)	113
5.7. Metaevaluationen, Evaluationssynthesen und Metaanalysen - Qualität der Evaluation	115

KAPITEL 6: TYPISCHE NUTZUNGSZUSAMMENHÄNGE VON EVALUATIONEN	119
6.1. Einleitung: Evaluationen als Instrument öffentlicher Politik	119
6.2. Massnahmen, Akteure und politische Konstellationen als Ansatzpunkte	119
6.3. Zu evaluierende Massnahmen oder Politiken	120
6.4. Evaluationen und ihre Akteure	121
6.4.1. Sich einen direkten Informationsvorteil verschaffen	122
6.4.2. Die eigene strategische Position stärken	122
6.4.3. Die Wahrheit herausfinden	125
6.4.4. Durch Studienaufträge Koalitionen bauen	126
6.4.5. Beschwichtigen, indem Studien in Auftrag gegeben werden	127
6.5. Evaluationen und allgemeinpolitischer Kontext	128
6.5.1. Welcher Staat - welche Politikinstrumente?	128
6.5.2. Wieviel Staat?	130

TEIL III: EVALUATION UND VERWANDTE INSTRUMENTE - NEUERE TENDENZEN

133

KAPITEL 7: EVALUATION UND POLITIKANALYSE	134
7.1. Evaluation und klassische Politikanalyse	134
7.2. Evaluation und neuere Politikanalyse (unter Einbezug der Veränderung von Netzwerken öffentlicher Politiken)	139
7.2.1. Zum Begriff des Netzwerks	139
7.2.2. Evaluation als Ressource im Netzwerk	143
7.2.3. Zusammenfassung	145

KAPITEL 8: EVALUATION UND MONITORING	147
8.1. Zum Begriff Monitoring	147
8.2. Integrierte Politikbeobachtung	150
8.3. Geographisches Policy-Monitoring	153

KAPITEL 9: EVALUATION UND	TI	156
9.1. Umfeld		156
9.2. Definition von Mediation		158
9.3. Anwendungsfelder		161
9.4. Beziehungen zwischen Evaluation und Mediation		163
9.5. Zusammenfassung und Ausblick		169

TEIL IV: DURCHFÜHRUNG, KONZEPTE, METHODEN

173

KAPITEL 10: ARBEITSSCHRITTE	175
10.1. Problembezeichnung	175
10.1.1. Eingrenzung der Fragestellung	176
10.1.2. Bestimmung des Evaluationskontextes und der Evaluationsdimension	177
10.2. Machbarkeitsprüfung	179

10.3. Wirkungsmodell	180
10.4. Untersuchungsdesign und Forschungstechnik	182
10.5. Umsetzung	183
KAPITEL 11: UNTERSUCHUNGSDESIGNS	185
11.1. Dimensionen von Evaluationsdesigns	185
11.1.1. Einzelfalluntersuchungen und vergleichende Untersuchungsanlagen	187
11.1.1.1. Einzelfalluntersuchungen	187
11.1.1.2. Vergleichende Untersuchungsanlagen	188
11.1.2. Quer- und Längsschnittanalysen	189
11.1.2.1. Querschnittanalysen	189
11.1.2.2. Längsschnittanalysen	190
11.1.2.3. Kombination und Integration	191
11.1.3. Auswahl und Vollerhebung	192
11.1.3.1. Auswahl	192
11.1.3.2. Vollerhebung	193
11.1.4. Experimentelle, quasi-experimentelle und nicht-experimentelle Designs	194
11.1.4.1. Experimentelle Designs	194
11.1.4.2. Quasi-experimentelle Designs	195
11.1.4.3. Nicht-experimentelle Designs	195
11.1.5. Entscheidungsbaum zur Ausgestaltung eines Evaluationsdesigns	196
11.2. Häufig verwendete Untersuchungsdesigns	198
11.2.1. Die Einzelfallstudie	199
11.2.2. Vergleichende Fallstudien	202
11.2.3. Der quasi-experimentelle Vorher/Nachher-Vergleich	204
11.2.4. Die Zeitreihenanalyse	207
11.2.5. Kombination und Triangulation	210
KAPITEL 12: FORSCHUNGSMETHODEN	214
12.1. Forschungsparadigmen	214
12.2. Qualitative Verfahren	216
12.2.1. Interaktive Evaluation	218
12.2.2. Gütekriterien qualitativer Evaluationsforschung	220
12.2.2.1. Reliabilität	220
12.2.2.2. Validität	221
12.2.3. Möglichkeiten und Grenzen qualitativer Evaluationsforschung	223
12.2.4. Datenerhebung	223
12.2.4.1. Dokumentenanalyse	223
12.2.4.2. Das problemzentrierte Interview	225
12.2.4.3. Gruppendiskussion	228
12.2.5. Datenauswertung	230
12.3. Quantitative Verfahren	233
12.3.1. Datenerhebung	233
12.3.2. Datenauswertung	243
12.3.3. Darstellung der Ergebnisse	253
TEIL V: LERNORIENTIERTE EVALUATIONEN	257
KAPITEL 13: UMSETZUNG VON LERNEN DURCH EVALUATIONEN	257
13.1. Die umsetzungsorientierte Lernforschung	258

Inhaltsverzeichnis	IX
13.2. Lernen als neues Paradigma	261
13.3. Erfahrungen mit lernorientierten Evaluationen	264
13.4. Voraussetzung für Lernprozesse in öffentlichen Politiken	270
13.5. Implikationen für die Umsetzung von Evaluationen	278
LITERATURVERZEICHNIS	281
STICHWORTVERZEICHNIS	317
AUTORENVERZEICHNIS	325
ANHANG: VERZEICHNIS DER VERÖFFENTLICHUNGEN AUS DEM NATIONALEN FORSCHUNGSPROGRAMM "WIRKSAMKEIT STAATLICHER MASSNAHMEN"	331
<i>SCHLÜSSBERICHTE</i>	<i>331</i>
<i>KÜRZBERICHTE</i>	<i>332</i>
<i>SYNTHESEN</i>	<i>333</i>
<i>PROJEKTE</i>	<i>334</i>
<i>KURZEVALUATIONEN</i>	<i>335</i>

AUTORENNACHWEIS

Kapitel 1 Hans-Ulrich Derlien

Kapitel 2 Werner Bussmann

Kapitel 3 Ulrich Klöti

Kapitel 4 Peter Knoepfel, Werner Bussmann

Kapitel 5 Peter Knoepfel, Frédéric Varone, Werner Bussmann,
Luzius Mader

Kapitel 6 Werner Bussmann, Peter Knoepfel

Kapitel 7 Peter Knoepfel, Ingrid Kissling-Näf, Werner Bussmann

Kapitel 8 Ingrid Kissling-Näf, Peter Knoepfel

Kapitel 9 Helmut Weidner, Peter Knoepfel

Kapitel 10 Andreas Balthasar

Kapitel 11 Ulrich Klöti, Thomas Widmer

Kapitel 12 Thomas Widmer, Hans-Martin Binder

Kapitel 13 Ingrid Kissling-Näf, Peter Knoepfel, Werner Bussmann

ABKÜRZUNGEN

AGEVAL	Arbeitsgruppe "Gesetzesevaluation"
AHV	Alters- und Hinterlassenenversicherung
BEI	Bundesblatt
BG	Bundesgesetz
BIGA	Bundesamt für Industrie, Gewerbe und Arbeit
BUWAL	Bundesamt für Umwelt, Wald und Landschaft
CATL	Computer assisted telephone interviewing
CEAT	Communauté d'études pour l'aménagement du territoire
CETEL	Centre d'étude, de technique et d'évaluation législatives
CV	Variationskoeffizient
EAWAG	Eidgenössische Anstalt für Wasserversorgung, Abwasserreinigung und Gewässerschutz
EDMZ	Eidgenössische Druck- und Materialzentrale
EJPD	Eidgenössisches Justiz- und Polizeidepartement
EPA	Eidgenössisches Personalamt
ERS	Evaluation Research Society
ETHZ	Eidgenössische Technische Hochschule Zürich
EVED	Eidgenössisches Verkehrs- und Energiewirtschafts- departement
EWR	Europäischer Wirtschaftsraum
FAC	Forschungsanstalt für Agrikulturchemie und Umwelthygiene, Bern/Liebefeld
FF	Feuille fédérale
GAO	United States General Accounting Office (Rechnungshof)
GIS	Geographisches Informationssystem
HMO	Health maintenance organization
Hrsg	Herausgeber
i E	im Erscheinen

IDHEAP	Institut de hautes études en administration publique
INTOSAI	Internationale Organisation der obersten Rechnungskontrollbehörden
KWF	Kommission zur Förderung der wissenschaftlichen Forschung (heute KTI: Kommission für Technologie und Innovation)
MTV	Motorisierter Individualverkehr
N	Fallzahl, Zahl der Untersuchungseinheiten
NABEL	Nationales Beobachtungsnetz für Luftfremdstoffe
NABO	Nationales Bodenbeobachtungsnetz
NFP	Nationales Forschungsprogramm
NFP 27	Nationales Forschungsprogramm "Wirksamkeit staatlicher Massnahmen"
NIMBY	Not in my backyard
NPM	New Public Management
OECD	Organisation für wirtschaftliche Zusammenarbeit und Entwicklung
ÖLS	Ordinary-least-squares (Methode der kleinsten Quadrate)
PEMD	Program Evaluation and Methodology Division (des US General Accounting Office)
PEMS	Policy and Expenditure Management System
PPBS	Planning, Programming and Budgeting System
PTT	Post-, Telefon- und Telegrafienbetriebe
SAGUF	Schweizerische akademische Gemeinschaft für Umweltforschung
SAKE	Schweizerische Arbeitskräfteerhebung
SANW	Schweizerische Akademie der Naturwissenschaften
SBB	Schweizerische Bundesbahnen
SIDOS	Schweizerischer Informations- und Daten-Archivdienst für die Sozialwissenschaften
SKUB	Schweizerische Kommission für Umweltbeobachtung

SR	Systematische Sammlung des Bundesrecht
TQM	Total Quality Management
UVP	Umweltverträglichkeitsprüfung
WEG	Wohnbau- und Eigentumsförderungsgesetz
WZB	Wissenschaftszentrum Berlin für Sozialforschung

VORWORT

1987 beschloss der Bundesrat die Lancierung des Nationalen Forschungsprogramms 27, "Wirksamkeit staatlicher Massnahmen". Dies war Ausdruck der Tatsache, dass ernsthafte Zweifel an der Wirksamkeit staatlichen Handelns aufgetreten waren und dass diese Fragen bei den sich abzeichnenden finanziellen Problemen voraussichtlich an Bedeutung gewinnen würden. Die Expertengruppe entschied sich, die Evaluation im Bereich der regulativen Politik in den Mittelpunkt zu stellen und finanzierte bzw. unterstützte insgesamt 20 Evaluationen, um die Methodik der Evaluation zu verbessern und um die Erfahrungen zu vertiefen.

Dabei zeigte sich, wie sehr der Evaluationskontext eines Landes prägend wirkt. Die schweizerische Demokratie bezieht ihre Legitimation in erster Linie aus dem (direkt-) demokratischen Entscheidungsverfahren und kaum aus Wirkungsanalysen. Referendumsfähig sind meist Gesetze mit vagen Zielen und offenem Budgetrahmen. Es ist deshalb auch typisch für die Schweiz, dass vor allem auf Ebene des Bundes eine im internationalen Vergleich ausserordentlich starke Delegation von den Gesetzen auf die Verordnungen erfolgt. Die Wirkungsanalyse wird damit zur verwaltungs-internen Angelegenheit, mit der sich das Parlament allenfalls bei budgetären Engpässen befasst.

Die politischen Instanzen zeigten lange Zeit wenig Neigung, die Wirkungen objektiv und extern analysieren zu lassen. Die Folge war, dass Evaluationen nur punktuell eingesetzt wurden. Entsprechend schmal war bis vor kurzem die Basis erfahrener Evaluatorinnen und Evaluatoren. Dazu kommt, dass die Kleinheit der politischen Körperschaften und der Rückstand in der amtlichen Statistik wissenschaftlichen Evaluationen enge Grenzen setzen. "Gute Evaluationen" werden damit häufig zu teuer, oder es werden von den mit Evaluationen betreuten Personen Ergebnisse verlangt, die methodisch nicht ausreichend abgesichert werden können. Die Erfahrungen zeigen, dass Evaluatorinnen und Evaluatoren häufig solchen Sachzwängen nachgeben (müssen) und methodische Defizite bewusst oder unbewusst akzeptieren.

Unterdessen hat sich die finanzielle Situation weiter angespannt. New Public Management hat an politischer Attraktivität gewonnen, wenn auch der Wettbewerb um die besten Wirkungs-Kostenverhältnisse noch zögerlich gefördert wird. Das Anliegen der Wirksamkeit ist in der politischen

Agenda noch weiter nach oben gerückt, auch wenn der Gesetzgeber mit wirkungsorientierten Regelungen immer noch Mühe bekundet.

In diesem Umfeld ist es wertvoll, wenn die Autoren ihr Wissen und ihre Erfahrungen über Evaluation in der Schweiz aus einer internationalen Perspektive zusammenfassen. Denn nicht nur die Evaluatorinnen und Evaluatoren, sondern auch die politischen Auftraggeber müssen die Möglichkeiten und Grenzen dieses Instrumentes kennen. Das Werk verzichtet auf einen für Dritte schwer verständlichen Fachjargon und konzentriert sich auf die wesentlichen politischen und methodischen Fragestellungen der Evaluation in unserem schwierigen Umfeld. Es zeigt die realistischen Grenzen der Evaluation im Kontext eines föderalistischen, direktdemokratischen Kleinstaates. Dennoch ist Evaluation nicht chancenlos oder gar unmöglich. Die Autoren stellen dar, wie angesichts solcher Randbedingungen nach internationalen Standards valable Evaluationen erarbeitet werden können. Das Werk ist damit eine wertvolle Bereicherung namentlich für Evaluatoren, aber auch für Auftraggeber und "Konsumenten" der Evaluationen und stellt gleichzeitig einen wichtigen Baustein zur notwendigen vermehrten Wirkungsorientierung dar. Ich danke den drei Haupt- und allen Mitautoren für diesen wertvollen Beitrag.

Prof. Dr. Ernst Buschor
Präsident der Expertengruppe
des NFP 27

EINFÜHRUNG DER HERAUSGEBER

Das vorliegende Buch ist ein Gemeinschaftswerk: Es ist das Ergebnis eines mit dem Nationalen Forschungsprogramm (NFP) "Wirksamkeit staatlicher Massnahmen" bzw. mit dessen Vorgängerprogrammen eingeleiteten Lernprozesses schweizerischer Politikwissenschaftlerinnen und Politikwissenschaftler zum Thema "Evaluation öffentlicher Politik". Die Gruppe der acht Mit- und drei Hauptautoren dürfte ein für die schweizerische Wissenschaftslandschaft repräsentatives Autorenkollektiv sein, das jüngere und ältere, aus der deutschen und der französischen Schweiz stammende sowie praxis- und theorieverbindende Wissenschaftlerinnen und Wissenschaftler vereint. Diese Autorenschaft bringt auch die von uns anvisierten Adressaten der vorliegenden Einführung zum Ausdruck: Zum einen sind dies Studierende der Politik- und Verwaltungswissenschaft an den deutsch- und französischsprachigen Hochschulen der Schweiz; zum anderen wollen wir damit auch Praktikerinnen und Praktiker aus der Evaluationsforschung oder aus der öffentlichen Verwaltung ansprechen. Diese haben sich kürzlich (1996) im Rahmen der Schweizerischen Evaluationsgesellschaft als Berufsvereinigung zusammengefunden. Die breite Autorenschaft und der Umstand, dass dieses Buch in deutsch und französisch erscheint, sollte es inskünftig ermöglichen, von Genf bis St. Gallen und von Basel bis Chiasso in Sachen Evaluation eine gemeinsame Sprache zu sprechen. In diesem Sinne will das Buch ein Standardwerk für die theoretische und praktische Evaluationsforschung in unserem Lande sein.

Die vorliegende Schrift versteht sich nicht als ein umfassendes Handbuch oder gar als eine Grundlegung für eine neue Evaluationstheorie. Sie will vielmehr eine nach didaktischen Gesichtspunkten konzipierte Einführung in die bestehende schweizerische Evaluationsforschung sein, wie sie bisher in Lehre und Praxis fehlte. Dieses Bestreben kommt zum Ausdruck im stufenweisen Aufbau des Buches, der von den konzeptionellen Grundlagen bis hin zu den methodischen Techniken reicht, in den zahlreichen Querverweisen, im ausgebauten Schlagwortverzeichnis und im bewussten Verzicht auf jene Vollständigkeit, die man etwa in Handbüchern findet.

Trotz seiner Selektivität zeigt das Buch den grossen Reichtum an Evaluationen in unserem Land. Dies wurde dadurch möglich, dass sich alle Beteiligten dazu verpflichteten, ihre bisher nur relativ kleinen Personenzirkeln vermittelten Kenntnisse in die Gesamtschau einzubringen. Sicher wurde dieses Unterfangen durch das NFP 27 erleichtert, das für die Bil-

dung dieser Community einen wichtigen Kristallisationspunkt darstellte. Systematische Bestrebungen, die schweizerische Evaluationsforschung u.a. mit der vorliegenden Einführung in der helvetischen Lehr- und Forschungslandschaft stärker zu verankern, haben ihre Ursache aber auch im zunehmenden Bedarf nach Politikevaluationen in der öffentlichen Verwaltung, im politischen und im gesellschaftlichen Raum. Die allseits geforderte Transformation des öffentlichen Sektors, etwa nach Modellen des New Public Management, der wirkungsorientierten Verwaltungsführung oder des Total Quality Management setzen wissenschaftliche Politikevaluationen geradezu voraus. Dasselbe gilt für Bestrebungen, dem Staat durch eine erhöhte Leistungs- und Sinnstiftungsfähigkeit neue Legitimationsressourcen zuzuführen. Gerade der Vollzugsföderalismus und die markante kulturelle, geographische, politische und administrative Vielfalt unseres Landes bildet für bundeseinheitliche öffentliche Politiken eine anderswo kaum vorhandene Vielfalt an Umsetzungsbedingungen und Wirksamkeitschancen. Was andere in ihren Labors mühsam konstruieren müssen, finden schweizerische Evaluationsforscherinnen und -forscher in der Politikwirklichkeit vor.

Gleichwohl ist Evaluationsforschung keine schweizerische Erfindung. Der erwähnte Reichtum an schweizerischen Evaluationsforschungen muss denn auch eher als eine Art Aufhol- denn als Pioniertat bezeichnet werden, zu der es in den letzten fünf bis zehn Jahren gekommen ist. Denn Evaluationsforschung war in den meisten europäischen Ländern und namentlich in den USA bereits seit den siebziger Jahren ein fest etablierter Zweig der Wissenschaften und ein Bestandteil der politischen und administrativen Praxis. Diesem Umstand trägt die vorliegende Einführung Rechnung, indem sie in wesentlichen konzeptionellen und methodischen Bereichen auf ausländische Literatur abstellt. Das gilt namentlich für die US-amerikanische und, in geringerem Ausmass, für die bundesdeutsche Evaluationsforschung. Beide haben auch die wissenschaftliche Biographie der Hauptautoren in den frühen siebziger Jahren massgeblich mitgeprägt.

An der vorliegenden "Einführung in die Politikevaluation" haben insgesamt elf Personen mitgewirkt. Obwohl die drei Hauptautoren im Interesse der Homogenität als Mitautoren oder Redakteure teilweise in beträchtlichem Ausmass in die einzelnen Kapitel eingegriffen haben, weist der Text gewisse Überschneidungen unter den Kapiteln auf. Dieser "Restbestand" an Überlappungen erscheint uns aus didaktischen Gründen als vertretbar; lernen ohne Redundanz ist kaum vorstellbar und macht auch wenig Spass.

Bei Gemeinschaftswerken ist üblicherweise die Liste der Verdankungen länger als bei gewöhnlichen Monographien. Wir möchten zunächst den Autoren und der Autorin für ihre Mitwirkung an dieser Schrift danken; der wahrhaft iterative Entstehungsprozess hat ihnen einiges an Be- und Überarbeitung abverlangt. Zu danken haben wir aber auch dem Schweizerischen Nationalfonds und den Organen des Nationalen Forschungsprogramms 27, die mit dem Auftrag zu dieser Schrift einen konzeptionellen und methodischen Synthesebericht vorlegen. Zu verdanken haben wir schliesslich auch all jene Forscherinnen und Forscher, die an den in diesem Buch zusammengefassten Evaluationsstudien mitgearbeitet haben. Dasselbe gilt für die zahlreichen Mitarbeiterinnen und Mitarbeiter in den Sekretariaten, Bibliotheken und Verwaltungen, die zum Zustandekommen dieses Buches ebenfalls beigetragen haben. Ein grosser Dank gilt schliesslich Erika Blanc, die, grosszügig unterstützt vom IDHEAP, das Layout, den Text und die definitive Gestaltung dieses Buches produziert hat.

Bern, Zürich, Chavannes-près-Renens

Werner Bussmann

Ulrich Klöti

Peter Knoepfel

TEIL I: HINTERGRÜNDE UND ENTWICKLUNGEN

Evaluation hat vielfältige Hintergründe. Sie hat ihre Wurzeln einerseits in den einzelnen Disziplinen, die sich seit längerer Zeit mit der Ermittlung der Wirksamkeit öffentlicher Politiken befassen, und andererseits in disziplinübergreifenden methodischen Bemühungen, die namentlich in den USA in den sechziger Jahren und seither auch in manchen anderen Ländern einsetzten.

Seit es öffentliche Massnahmen gibt, werden auch *disziplinspezifisch* Überlegungen zu ihrer Wirksamkeit angestellt. In der Schweiz historisch dokumentiert sind beispielsweise die Bemühungen, die Bürger mit polizeirechtlichen Instrumenten vom Tabakrauchen abzuhalten; untersucht wurde auch die Neuorientierung dieser Politik in Richtung einer steuerlichen Belastung des Tabaks nach dem Scheitern polizeilicher Massnahmen (Hengartner und Merki 1993: 378-382). Der systematische Rückgriff auf die Sozialwissenschaften zur Wirkungsermittlung setzte allerdings erst in den vierziger und fünfziger Jahren ein. Bereits in der unmittelbaren Nachkriegszeit wurden in zahlreichen Politikbereichen Überlegungen über die Wirksamkeit bisheriger Massnahmen angestellt. Das gilt namentlich im Zusammenhang mit der Bereinigung der Hinterlassenschaft der Krisen- und Kriegswirtschaft (Landwirtschaft, Preiskontrollen, Emser Werke, Wohnungsmarkt u.a.m.). In zunehmendem Masse stützte man sich dabei auf statistische und wissenschaftliche Befunde ab. Diese Tendenz wurde entscheidend gefördert durch Fortschritte in den einzelnen Disziplinen, welche erlaubten, Aussagen über die Wirkungen spezifischer staatlicher Massnahmen zu machen. So wurden etwa in der Bildungsforschung immer mehr die Wirkungen von pädagogischen Massnahmen, Lehrmitteln und Lehrformen empirisch ermittelt. Auch in den Wirtschaftswissenschaften wurde der Zusammenhang zwischen staatlichen Interventionen und wirtschaftlicher Entwicklung theoretisch erörtert und empirisch erforscht. Die Rechtslehre begann, die Perspektive rein normativ orientierter Fragen vermehrt mit derjenigen der "Rechtstatsachen" zu verbinden (Mader 1985: 18 ff. und 150 ff.). Auch die schweizerische Politikwissenschaft weitete ihren Untersuchungsbereich von politischen Theorien, Prozessen und Institutionen auf konkrete öffentliche Politiken aus (Linder, Hotz und Werder 1979; Germann et al. 1979; Knoepfel 1979).

In verschiedenen weiteren Disziplinen und Spezialzweigen (Betriebswirtschaftslehre, Agrarwirtschaft, Forstwirtschaft u.a.m.) wurden ebenfalls Grundlagen für eine wissenschaftliche Beschäftigung mit öffentlichen Politiken gelegt.

Evaluationen sind damit solide eingebettet in die Bedürfnisse der Einzelpolitiken nach Argumentation über Wirkungen und in die disziplinenbezogenen Bestrebungen zur Wirkungserfassung. Sie sind in diesem Sinne ein neuer Begriff für eine altbekannte Tätigkeit, und man könnte an Monsieur Jourdain denken, der plötzlich mit Erstaunen merkt, dass er Prosa spricht. Doch die Analogie schiesst übers Ziel hinaus, denn es ist wesentlich einfacher Prosa zu sprechen als methodisch saubere Evaluationen durchzuführen. Zudem hat die Evaluation Wurzeln, die über die einzelnen Politikbereiche und Disziplinen hinausreichen und die es rechtfertigen, eine eigene Begrifflichkeit zu entwickeln. In den USA sind namentlich seit den sechziger Jahren im Gefolge des von Präsident Kennedy deklarierten "War on Poverty" umfangreiche Bemühungen unternommen worden, die Ermittlung der Wirkungen staatlicher Programme sowie deren Bewertung und Verbesserung auf einer tragfähigen methodischen Grundlage durchzuführen¹. Nach der Beurteilung amerikanischer Autoren (Chelimsky 1985: 2-6; Rossi und Freeman 1993: 9-27) bediente sich die Evaluationsforschung dabei neuer entscheidungstheoretischer Erkenntnisse und Techniken (PPBS, Kosten-Nutzen-Analysen, Kosten-Effektivitäts-Analysen, Operations Research) sowie neuentwickelter sozialwissenschaftlicher Erhebungs- und Auswertungsmethoden (z.B. Umfragen, computergestützte Auswertung und weiterentwickelte statistische Tests). Zusätzlich anzufügen wäre - nicht zuletzt mit Blick auf europäische Entwicklungen der Evaluationsforschung - der Einbezug von Erkenntnissen der Implementationsforschung (Mayntz 1980; Mayntz 1983) resp. der Politikanalyse (vgl. Bussmann 1993; Knoepfel 1995e) sowie in neuester Zeit (im Zusammenhang mit der Budgetknappheit) Überlegungen der Managementlehre. Daraus ist ein recht kohärentes Gebilde entstanden, welches Überlegungen zum staatlichen Handeln, zur Wissensermittlung, zur Bewertung, zur Forschungspraxis und zur Umsetzung (vgl. Shadish, Cook und Leviton 1991) enthält. Scriven (1991: 1-43, 363 f.) geht dabei so weit, der Evaluation den Status einer *Transdisziplin*, vergleichbar mit der Logik und Statistik, zuzumessen. Die methodische Konsolidierung der

¹ Einen guten Einblick in die Pioniere dieser Entwicklung und den heutigen Stand der methodischen Reflexion geben Shadish, Cook und Leviton 1991.

Evaluationsforschung ist ablesbar an der Gründung von Evaluationsgesellschaften (amerikanische, kanadische, australasiatische, europäische, britische und schweizerische Evaluationsgesellschaft), an den spezifischen Zeitschriften zu diesem Thema (Evaluation and Program Planning, Evaluation Review, Evaluation Practice, Evaluation) sowie, last but not least, an den Standards, welche eine seriöse Qualitätskontrolle von Evaluationen ermöglichen.

Gegenwart und Zukunft der Evaluationsforschung stehen somit in einem fruchtbaren Spannungsverhältnis zwischen politikbereichs- bzw. disziplinspezifischer und übergreifender (trans- und multidisziplinärer) Betrachtungsweise. Zudem ist, wie nachfolgend gezeigt wird, ein Austausch und Dialog über Methodik und Praxis von Evaluationen in Gang gekommen, der Länder mit unterschiedlichem rechtlichem (kontinental-europäische und angelsächsische Rechtstradition) und wissenschaftlichem (unterschiedlicher Ausbau der Sozialwissenschaften) Hintergrund mit einschliesst.

KAPITEL 1: DIE ENTWICKLUNG VON EVALUATIONEN IM INTERNATIONALEN KONTEXT

Hans-Ulrich Derlien

1.1. ZWEI WELLEN DER INSTITUTIONALISIERUNG

Überblickt man den heutigen Stand der Evaluation, soweit sie in den Zentralregierungen in Amerika und Europa institutionalisiert ist, stellt man sehr schnell unterschiedliche Entwicklungsniveaus und Reifegrade fest. Dies gilt sowohl hinsichtlich der praktischen Erfahrung mit diesem Instrument der Rationalisierung von Politik als auch hinsichtlich der Häufigkeit, der Regelmässigkeit und dem methodischen Niveau, auf dem Evaluationen betrieben werden. Die von Ray Rist (1990a) herausgegebenen *Länderberichte* ergeben folgendes Gesamtbild:

- Ohne Zweifel kann das am höchsten entwickelte Evaluationssystem in den *USA* beobachtet werden. Mit rund 30 Jahren Erfahrung in der Programmevaluation hat das amerikanische System ein hohes Mass methodischer Professionalität erreicht. Diese Professionalisierung erlaubt es auch, in den *USA* die Programmevaluation von traditionellen Techniken der Kontrolle und der Rechnungsprüfung deutlich zu unterscheiden. Dazu hat beigetragen, dass das amerikanische Evaluationssystem sowohl innerhalb der Administration als auch im Kongress - und dazu gehört als Hilfsorgan auch der US-Rechnungshof - stark institutionalisiert ist. Die amerikanischen Evaluationsbemühungen haben entsprechende Entwicklungen in anderen Ländern, wenn nicht ausgelöst, so doch zumindest unterstützt.
- In eine zweite Gruppe von Ländern lassen sich *Schweden*, *Kanada* und die *Bundesrepublik Deutschland* einordnen, denen gemeinsam ist, dass sie bis in die sechziger Jahre zurückreichend erste Versuche mit der Programmevaluation unternommen haben. Allerdings blieb diese trotz Institutionalisierung in verschiedenen Politikbereichen relativ fragmentiert, und die Zahl der Evaluationsstudien, die durchgeführt wurden, ist im Vergleich zu den *USA* relativ niedrig.

In Kanada sind ähnlich frühe Evaluationsversuche zunächst gescheitert. Dies gilt in gewisser Weise auch für Grossbritannien, wo das Programme Analysis and Review System in den siebziger Jahren wieder eingestellt wurde. Für Grossbritannien wie auch für Kanada lässt sich jedoch feststellen, dass Ende der siebziger Jahre neue Versuche mit der Programmevaluation unternommen worden sind. Dabei ging Kanada den Weg, die Evaluationsfunktion dem Rechnungshof (Comptroller General) sowie den einzelnen Ministerien zu übertragen, während die Evaluation in Grossbritannien keinen organisatorischen Kristallisationspunkt erhielt, sondern eher Teil der allgemeinen, neuen Managementphilosophie wurde und primär im Rahmen des Budgetprozesses zum Tragen kam.

- Dieser zweite Anlauf, der in Kanada und Grossbritannien unternommen wurde, fällt zeitlich und systematisch zusammen mit Bemühungen zur Institutionalisierung der Programmevaluation in *Norwegen, Dänemark, den Niederlanden, Frankreich, der Schweiz, Spanien* und sogar in der *Europäischen Gemeinschaft*. Auch für diese Gruppe ist kennzeichnend, dass sie die Programmevaluation stärker auf den Budgetprozess bzw. auf die Bewältigung der Budgetengpässe beziehen, als dies während der ersten Generation der Programmevaluation üblich war.

1.2. DREI ENTWICKLUNGSLINIEN

Diese Entwicklung lässt *drei Trends* erkennen:

- In den achtziger Jahren entsteht eine zweite Evaluationsbewegung, die sehr stark an Fragen des *Output des Regierungshandelns*, möglicherweise weniger an den realen Auswirkungen, den Impacts, orientiert ist. Selbst die meisten derjenigen Länder, die die Evaluation schon in den sechziger Jahren etabliert hatten, verstärken ihre Evaluationsbemühungen unter Einschluss der Kostenrechnung (cost-accounting), um die Verwaltungskosten etwa einer Dienstleistung oder eines Verwaltungsaktes zu ermitteln.

- Während die erste Evaluationsbewegung in den sechziger und zu Beginn der siebziger Jahre primär am Prozess der Planung und der Programmentwicklung orientiert und damit letztlich auf die Bedürfnisse der programmverantwortlichen Fachpolitiker in Regierung und Parlament zugeschnitten war, ist die zweite Welle Ende der siebziger Jahre eher zentralen politischen Bemühungen entsprungen, die Berechtigung laufender Reformprogramme zu überprüfen, und ist in erster Linie am *Budgetprozess* ausgerichtet.
- In diesem Zusammenhang gewinnt die externe Kontrollinstanz, der *Rechnungshof*, an Bedeutung. Damit hat sich der Schwerpunkt der Evaluation in einer Reihe von Ländern in den achtziger Jahren von einer vormals primär regierungsinternen Angelegenheit auf die Parlamente verschoben. Die Parlamente gewinnen dabei um so mehr an Bedeutung, wie ihnen die nationale Rechnungsprüfungsbehörde **aus-**schliesslich oder doch primär zuarbeitet.

1.3. URSACHEN DER ENTWICKLUNG

Es lassen sich verschiedene Ursachen für die unterschiedlichen Entwicklungspfade angeben (Derlien 1990).

1.3.1. Veränderung der fiskalischen Situation

Das Entwicklungsmuster der Frühstarter USA, Schweden, Deutschland und Kanada dürfte seine Ursache *ceteris paribus* zunächst in der volkswirtschaftlichen Situation und folglich den fiskalischen Bedingungen haben, unter denen die Regierungen zu wirken hatten. Die erste Gruppe von Ländern erhielt den Impuls zur Programmevaluation in einer Situation *wirtschaftlichen Wachstums* und *wachsender Budgets*, die die Regierungen in die Lage versetzten, kostspielige Sozialprogramme sowie Reformen im Bildungs- und Gesundheitswesen in Angriff zu nehmen. In diesem Kontext sind auch formale *Planungssysteme* entstanden, die sich, wie in Deutschland, auf die mittelfristige, aus dem Einjahresbudget fortgeschriebene Finanzplanung beschränkten (Schmidt und Wille 1970). Einige darunter versuchten sogar, die Budgetseite mit der Programmseite zu integrieren und die Allokation der Haushaltsmittel aus Programmprioritäten

abzuleiten, wie es in den USA, Schweden und Kanada der Fall war. Diese Planungsbemühungen sind zwar teils gescheitert (zum PPBS: Reinermann 1975) oder unbefriedigend geblieben (Wille 1977). In jedem Falle wurde die Evaluation entweder als logisch konsequenter Abschluss des Planungszyklus betrachtet oder aber dadurch stimuliert, dass die neuen Interventionsprogramme neue Informationsbedürfnisse auslösten. Die Unsicherheit, mit der die z.T. innovativen oder zumindest mit innovativem Anspruch verfolgten Programme operierten, liess das Bedürfnis nach Wirkungskontrollen entstehen. Die Evaluation hatte dabei primär die Funktion, bestehende und neue *Programme effektiver zu gestalten*. Nur gelegentlich waren Evaluationen erforderlich, um diese Programme gegen politische Kritik mit dem Nachweis ihrer Wirksamkeit zu legitimieren.

Die Auswirkungen der völlig geänderten ökonomischen Situation Ende der siebziger Jahre waren fast die gleichen: die Notwendigkeit, mit *knappen Ressourcen* zu haushalten, führte zur Entwicklung von Techniken der Haushaltskürzung, des Cutback-Management. Entsprechend ist der Rahmen, in dem während der zweiten Welle der Evaluationsbemühungen die Institutionalisierung betrieben wurde, typischerweise auch der Budgetprozess gewesen: Produktivitätskampagnen oder Financial-Management-Initiativen in Grossbritannien, Kanada, Norwegen und den Niederlanden leiteten entweder die Reaktivierung bestehender Evaluationsversuche ein wie in Kanada und Grossbritannien, oder dieser neue Managerialismus (Aucoin 1988) institutionalisierte das Konzept der Evaluation erstmals mit dem Zweck, die Ressourcenallokation innerhalb des Budgets zu rationalisieren wie in Norwegen und den Niederlanden. Typischerweise sind denn auch die Träger dieser zweiten Bewegung nicht die Programm-Manager in den Ressorts, sondern eher die Finanzminister und die Rechnungshöfe als traditionelle Wächter des Budgets. Damit geht allerdings einher, dass die Erwartungen an die Evaluation und ihre Funktion innerhalb des Politikzyklus sich leicht verlagert haben. Anstatt Programme effektiver zu gestalten, besteht das Anliegen nun eher darin, *ineffektive Programme zu erkennen* und so zu legitimierbaren Budgetkürzungen zu gelangen.

1.3.2. Politische Konstellationen

In den ökonomisch blühenden sechziger Jahren regierten in den Ländern, in denen frühzeitig die Evaluation institutionalisiert worden war, *Reformparteien* wie die Sozialdemokraten in Deutschland und in Schweden, die

Demokraten in den USA und die Liberalen in Kanada; sie initiierten die Reform- und Interventionsprogramme und hatten offenbar auch eine starke Affinität zu den Sozialwissenschaften, deren methodisches Wissen für die Durchführung von Evaluationen benötigt wurde. Ein weiterer Beleg für diesen politischen Zusammenhang dürfte das Aufgreifen des Evaluationsgedankens in Frankreich nach dem Antritt des sozialistischen Präsidenten im Jahre 1981 sein.

Eher *konservative Regierungen* (in Grossbritannien zwischen 1970 und 1974 sowie seit 1979, in Kanada 1979 und wiederum seit 1984, in Norwegen 1981-1986, in Dänemark seit 1982, in Schweden zwischen 1976 und 1982, in den Niederlanden fast ununterbrochen seit 1977) sahen sich nicht nur mit einer veränderten ökonomischen Situation konfrontiert, sondern nutzten auch das z.T. schon verfügbare Evaluationsinstrumentarium, um das Budget neu zuzuschneiden und dabei gelegentlich auch Reformprogramme der Vorgängerregierungen zu stützen, die sie von Anfang an mit Argwohn betrachtet hatten. Wo analytische Kapazitäten vorhanden waren, konnten die ins Amt kommenden konservativen Regierungen diese benutzen, um existierende Programme "analytisch zu delegitimieren" (Wagner und Wollmann, 1986). Die Ausnahme von der Regel sind die konservativen Regierungen in Washington (1980-1992) und Bonn (seit 1982), die die Evaluation nicht intensiviert haben, sondern auf dem schon bestehenden Niveau weiterbetrieben, ja sie in der Exekutive sogar etwas zurückgenommen haben (USA).

Zu den politischen Rahmenbedingungen gehört auch der internationale neo-konservative Konsensus, dass der Staat sich aus der Gesellschaft zurückziehen habe, zumindest aber nicht weiter expandieren solle. In Verbindung mit Privatisierungsprogrammen dienten Deregulierung und Entbürokratisierung sowie das Zurückfahren der Planungssysteme, die zur Entwicklung und Steuerung von Reformpolitiken gedient hatten, dazu, den Kurs zu ändern. Eine zweite Strategie bestand darin, das traditionelle Budgetverfahren zu verbessern, indem neue Kontrolltechniken z.T. aus dem Privatsektor übernommen wurden wie das Controller-Konzept, die zumindest im Ansatz der Evaluationsmethodik sehr nahe kommen. Schliesslich liessen sich modernisierte ökonomische Kontrolltechniken auch gut in die überkommenen Verfahren der Haushaltskontrolle einbinden (Hofmeister 1994).

1.3.3. Verfassungskontext

Eine Erklärung der unterschiedlichen Entwicklungswege, die die verschiedenen Zentralregierungen gegangen sind, kann sich nicht nur auf den Zusammenhang von wachsender Wirtschaft, Interventionsprogrammen und reformistischen Regierungen oder schrumpfender Wirtschaft, Sparpolitik und konservativen Regierungen stützen. Herausragende Bedeutung scheint auch die Verfassungskonstellation zu haben, in der Exekutive und Legislative zueinander stehen. In einem System des *Party Government*, in dem sich üblicherweise Mehrheitsfraktionen und Regierung Oppositionsfraktionen gegenüberstehen (Lijphart 1984; Weaver und Rockman 1993), entwickelt das Parlament als Ganzes kaum das Bedürfnis, sich unabhängig von der Regierung eine Informationsbasis aufzubauen, in der die Evaluation ihren Platz hat. Zwar haben in Kanada, Norwegen und Grossbritannien sehr wohl die Parlamente Initiativen zur Stärkung der Evaluationsfunktion entwickelt; in diesen Fällen haben die Parlamente aber lediglich bekräftigt, was die Regierungen ohnehin wünschten.

Der Fall der USA verweist auf die eminente Bedeutung, die das *Präsidialsystem* für die Evaluationsentwicklung hat. Hier können sich parlamentarische Mehrheiten gegen einen Präsidenten bilden und ihm Evaluationspflichten aufzwingen oder ihre informationelle Unabhängigkeit betonen. So ist zu verstehen, dass der amerikanische Kongress sich einen gewaltigen Mitarbeiterstab als Gegenbürokratie aufgebaut hat und zu einem starken Pfeiler im amerikanischen Evaluationssystem geworden ist, obwohl die Exekutive von sich aus auf dem Gebiet der Evaluation frühzeitig aktiv wurde. Parlamentarisches Interesse vermag dann die Evaluationskapazitäten der Exekutive sogar zu bewahren, wenn die Administration selbst nicht allzu enthusiastisch auf diesem Gebiet ist, geht es doch um die Erhaltung des informationellen Machtgleichgewichts.

Ein weiterer, hiermit zusammenhängender konstitutioneller Faktor, dem man einige Bedeutung beimessen sollte, ist die *Zuordnung der zentralen Rechnungshöfe*. In den meisten Ländern (USA, Kanada, Norwegen, Niederlande, Dänemark) sind diese externen Kontrollinstanzen der Legislative zugeordnet. Während der deutsche Bundesrechnungshof historisch der Exekutive berichtete und erst seit 1969 seine Funktion auch und verstärkt am Bundestag und am Rechnungsprüfungsausschuss ausgerichtet hat (seit 1987 wird sein Präsident vom Bundestag gewählt), hat etwa der französische Cours des Comptes, obwohl ebenfalls dem Parlament berichtend, bis heute eine eindeutig exekutive Orientierung bewahrt, was mit dem

Karrieresystem seiner Mitglieder zusammenhängt. In Schweden schliesslich ist der Rechnungshof ausschliesslich ein Instrument der Zentralregierung, und die parlamentarische Rechnungsprüfung verfügt hier nur über einen kleinen Mitarbeiterstab, um ad hoc Untersuchungen anzustellen. Tendenziell nutzen die Parlamente die Kontrolle durch die Rechnungshöfe um so stärker als Machtbasis, sogar in einem System der Parteienregierung, je unabhängiger die Finanzkontrolle von der Exekutive institutionalisiert ist oder je stärker sie dem Parlament zugeordnet ist. In den USA konnte diese Anbindung des Rechnungshofes in Verbindung mit dem Präsidialsystem natürlich die Evaluationsfunktion im Kongress nur noch weiter stärken.

Andererseits können *Föderalismus*, starke *Dezentralisierung* und Formen *direkter Demokratie* den Evaluationsbedarf geringer erscheinen lassen als in unitarischen Staaten mit längeren Rückkoppelungsschleifen (vgl. Kapitel 2).

1.4. EVALUATION UND FINANZKONTROLLE

Allgemein lässt sich aus dem internationalen Vergleich zunächst folgender Zusammenhang ableiten: Bei exekutiver Orientierung der Rechnungshöfe, prozeduraler Trennung von Politikformulierung und Budgetierung (two-track-system) und Institutionalisierung der Evaluation seitens der Programmentwicklung in den Ministerien ist die Erweiterung des Kontroll-Instrumentariums der Rechnungshöfe um die Evaluation weniger wahrscheinlich als bei einer Konfiguration mit legislativer Anbindung der Rechnungshöfe, Erfahrung mit integrierter Programmentwicklung und Budgetierung und Auftreten der Evaluation unter den Kontextbedingungen der zweiten Welle der Institutionalisierung (Derlien 1994).

Es scheint ein Vorteil der jüngeren neo-konservativen Management-Orientierung gegenüber der älteren Programm-Orientierung der Evaluation zu sein, dass genau dieser Zusammenhang zwischen Evaluation und *Budgetierung* hergestellt wird. Allerdings bleibt abzuwarten, welche Auswirkungen diese Verbindung von Evaluation und Budget auf den Inhalt und die methodische Qualität von Evaluationen hat. Der Zeitdruck der Budgetaufstellung und die Kostenorientierung könnten dazu führen, dass das methodische Anspruchsniveau an die Evaluationen abnimmt, ja dass möglicherweise im Budgetprozess legitimatorische Motive Überhand-

nehmen und methodisch schwer überprüfbare Erfolgsberichte produziert werden (Gray et al. 1993).

Die Verbindung von Budgetprozess, Rechnungsprüfung und Evaluation bleibt auch deshalb prekär, weil diese Funktionen unterschiedliche Professionalisierung erfordern. Zwar führt das US General Accounting Office (GAO) jährlich etwa 1000 Evaluationen durch und verfügt über eine eigene Methoden-Abteilung; bei näherem Hinsehen stellt man jedoch fest, dass die organisatorische Integration der Evaluation eine *Zweispungkeit in der professionellen Zusammensetzung* und Orientierung des GAO-Personals erzeugt hat: Infolge unterschiedlicher Ausbildung und verschiedener Tätigkeitsschwerpunkte koexistieren klassische Rechnungsprüfer mit Evaluatoren, ohne dass diese Dualität bislang in einer sozialen Integration hätte überwunden werden können.

In Kanada, wo im Zuge der zweiten Institutionalisierungswelle (und mehrfacher vom PPBS zum PEMS reichender Budget-Reformen) ebenfalls die Evaluationsfunktion in das traditionelle Auditing integriert wurde, haben sich die Finanzkontrolleure inzwischen - wie der deutsche Bundesrechnungshof - darauf zurückgezogen, die Evaluationsbemühungen der Ministerien zu evaluieren und methodische Richtlinien herauszugeben, also das zu betreiben, was die Spezialisten *Metaevaluation* nennen, die Durchführung von Evaluationen aber den Departements zu überlassen.

1.5. FORMEN DER INSTITUTIONALISIERUNG

Im Ergebnis dieser Entwicklungen lassen sich heute systematisch folgende *fünf institutionalisierte Formen* der Programmevaluation unterscheiden:

- Am weitesten verbreitet ist die *programmspezifische* Institutionalisierung. Häufig, besonders wenn es sich um eine experimentelle Politik handelt, wird bereits bei der Verabschiedung der (gesetzesförmigen) Programme eine Evaluationspflicht festgelegt. Träger der Evaluation ist dann in der Regel eine Fachabteilung in einem Ministerium². Typisch ist diese Institutionalisierungsform bei den Programmverantwortlichen in den Ländern, die zur ersten Welle der Evaluation zu rechnen sind.

² Diese dezentralisierte Institutionalisierung hat im übrigen zur Folge, dass es gelegentlich schwer ist, einen vollständigen Überblick über das Ausmass regelmässig betriebener Evaluationen zu gewinnen.

- *Ressortinterne Evaluationseinheiten* bilden sich in der Regel dort, wo mehrere Abteilungen programmspezifische Evaluationen durchführen müssen, so dass sich eine strukturelle Ausdifferenzierung dieser Funktion wie in den meisten Ministerien in den USA und Kanada lohnt.
- *Zentralisierung der Evaluation in der Exekutive* ist relativ selten. Findet sie sich in der Behörde des Regierungschefs, werden meist nicht einzelne Programme, sondern ganze Politikfelder evaluiert oder hinterfragt. Als zentrale Stelle kommt aber auch das Justizressort oder das für die Haushaltsaufstellung zuständige Ressort³ in Frage.
- *Parlamentarische Evaluation* ist, wie ausgeführt, typisch für das amerikanische Präsidialsystem mit seiner starken Gewaltenteilung, kann sich aber auch anderenorts entwickeln (so in der Schweiz).
- *Evaluation durch die Rechnungshöfe* findet sich vielerorten, insbesondere aber dort, wo die Rechnungshöfe Hilfsorgan der Legislative sind; exemplarisch hierfür ist der amerikanische Bundesrechnungshof.

Schliesslich lässt sich regelmässig eine gewisse Redundanz in den heutigen Evaluationssystemen feststellen, die darin besteht, dass mehrere Stellen im politisch-administrativen System nebeneinander oder funktional aufeinander bezogen die Evaluation betreiben.

³ So spielt in den USA das dem Präsidenten zugeordnete Office of Management and Budget eine gewisse Rolle, und in der Bundesrepublik Deutschland hat der Finanzminister die Möglichkeit, bei der Budgetaufstellung Nutzen-Kosten-Untersuchungen, zu denen auch die Evaluationen gerechnet werden, anzufordern (Derlien 1976).

KAPITEL 2: EVALUATIONEN IN DER SCHWEIZ

Werner Bussmann

2.1. RAHMENBEDINGUNGEN

2.1.1. Geschlossenheit und Offenheit des politischen Systems²

Das schweizerische Regierungssystem ist geprägt durch Eigenheiten, welche auch auf die Modalitäten der Nutzung von Evaluationen Einfluss haben. Diese Spezifika sind vor allem die direktdemokratischen Elemente, die Konkordanz sowie der ausgeprägte Föderalismus.

Es gilt heute unter Politikwissenschaftlern weitgehend als anerkannt, dass die *direktdemokratischen Rechte* massgeblichen Einfluss auf die Herausbildung des schweizerischen Regierungssystems gehabt haben. Vorab das Referendum, und später auch die Initiative, gaben der Opposition starke Waffen in die Hand und erzwangen deren weitestmögliche Einbindung in die Regierungsverantwortung. Diese hat seit 1959 in der sogenannten "Zauberformel" ihren Niederschlag gefunden. Das System der *Konkordanzregierung* als Spielart einer auf Dauer und (minimale) inhaltliche Abstimmung angelegten Koalitionsregierung hat dazu geführt, dass die Entwicklung von Evaluationen sowohl den Anliegen sozialdemokratischer Wohlfahrtspolitik wie auch den konservativen Sparbemühungen entsprechen musste. Dadurch unterscheidet sich die Schweiz von Institutionalisierungsbemühungen anderer Länder, welche meist nur einem der beiden Anliegen entsprachen (vgl. Kapitel 1 sowie Derlien 1990).

Direktdemokratische Beteiligung und Machtteilung im Rahmen der Konkordanzregierungen bilden für die sozialwissenschaftliche Beratung und damit auch für die Evaluation *Hindernisse*, eröffnen aber auch Chancen. Vorab die Hindernisse: Plebiszitäre Entscheide legen im betroffenen Sachgebiet staatliches Handeln für etwelche Zeit fest. So erstickte beispielsweise die Ablehnung der Zuteilung vermehrter Kompetenzen an den Bund im Hochschulwesen Ende der siebziger Jahre zahlreiche andere Zentralisierungsversuche im Keim, und die Ablehnung des EWR-Abkommens Ende 1992 lahmte seither eine offensive schweizerische Aussenpolitik auch gegenüber anderen Organisationen (UNO, NATO-Partner-

¹ Eine gekürzte Version dieses Kapitels erschien in englischer Sprache (Bussmann 1995b).

² Als Einführung in das schweizerische politische System sei u.a. Linder 1995 empfohlen.

schaft für Frieden usw.). Volksabstimmungen prägen die Meinungen darüber, was als "politisch realisierbar" erachtet wird. Aus diesem Grund ist die politische Diskussion über jene Fragen, die durch Volksentscheide oder durch Vetopositionen der Regierungsparteien oder wichtiger Verbände praktisch determiniert sind, sehr stark auf bestimmte Handlungsmöglichkeiten eingeengt und damit faktisch "geschlossen". Sozialwissenschaftliche Beratung, welche mehr als nur marginale Änderungen vorschlägt, wird - dies ist ein grosser Unterschied zu völlig geschlossenen Regierungssystemen - zwar toleriert, aber zugleich mit "benign neglect" bestraft. Mit anderen Worten: Evaluation kann ein sehr frustrierendes Geschäft sein.

Daneben ist aber auch *Offenheit* zu erwähnen. Trotz Konkordanz gibt es ausserhalb und innerhalb der Regierungsparteien Opposition, deren wichtigstes Werkzeug das Referendum und die Initiative ist. Referenden und Initiativen fokussieren die öffentliche Diskussion auf bestimmte Sachfragen, und daraus ergeben sich durchaus auch Gelegenheiten für die Kenntnisnahme und Verbreitung von Evaluationsresultaten. Sozialwissenschaftliche Forschung kann in solchen Momenten durchaus aus dem wissenschaftlichen Subsystem heraufgespült und in der öffentlichen Auseinandersetzung zur Kenntnis genommen werden. Dass solche Phasen politischer Thematisierung mit "explosiven Situationen" und für in der politischen Diskussion unerfahrene Forscherinnen und Forscher mit Gefahren (einseitige, wenn nicht missbräuchliche Verwendung von Forschungsergebnissen) verbunden sind, liegt auf der Hand (vgl. Freiburghaus und Zimmermann 1985: 95 f.).

Trotz der durch Konkordanz politisch oft blockierten Auseinandersetzung haben sich in der Schweiz immer wieder *Phasen grösserer Reformbereitschaft* bemerkbar gemacht, als ob der Rückstau an Änderungen sich plötzlich Bahn brechen müsste. Die Bereinigung der Notrechts- und Wirtschaftseingriffe nach dem 2. Weltkrieg, die Akzeptanz des Planungsthemas seit Ende der sechziger Jahre oder die durch den sog. Finanzskandal im Kanton Bern ausgelösten Veränderungen sind Beispiele hierfür. In solchen Situationen kann dann durchaus ein grösserer *Beratungsbedarf* entstehen, und die Sozialwissenschaftlerinnen und Sozialwissenschaftler und Evaluatorinnen und Evaluatoren, die vorher auf der Wartebank saßen, sind nun plötzlich - wenn auch meist nur für beschränkte Zeit - gefragte Partnerinnen und Partner. Es ist durchaus möglich, ja wahrscheinlich, wenn auch durch empirische Forschung nicht nachgewiesen, dass der konzeptionelle Gebrauch (zum Begriff vgl. Weiss und Bucuvalas 1980

sowie Wingens 1988) von sozialwissenschaftlichem Wissen den Boden für grössere Reformschübe ebnen kann.

Zu erwähnen ist schliesslich die Bereitschaft der politischen Instanzen, Empfehlungen, die in den durch die Konkordanz gelegten Rahmen hineinpassen, anzunehmen und umzusetzen und entsprechende Massnahmen weiterzuentwickeln und zu perfektionieren.

2.1.2. "Keine Experimente" und "föderalistisches Labor" als gegenläufige Gegebenheiten

Schweizerisches Recht zeichnet sich durch die kontinentaleuropäische Figur des Gesetzes aus, wobei die Schweiz dogmatisch namentlich durch die in Deutschland entwickelten Begrifflichkeiten geprägt worden ist. Zentral sind die Maximen der *Gleichheit vor dem Gesetz*, der Rechtsstaatlichkeit, der Konstanz und somit auch der Berechenbarkeit. Im Vergleich zur in den USA verbreiteten Figur des *Programms* ist das Gesetz durch einen recht breiten thematischen Anwendungsbereich (z.T. ganze Politikbereiche), häufig durch mangelnde Zielvorgaben, durch eine enge Normierung des Verwaltungshandelns und durch meist nur beschränkte oder gar fehlende finanzielle Vorgaben gekennzeichnet. Freilich haben im Laufe der letzten 20 Jahre finale Verwaltungsprogramme und das Staatshandeln bündelnde Pläne in der schweizerischen Rechtslandschaft durchaus auch Einzug gehalten (Luftreinhalteprogramm, Massnahmenpläne, Impulsprogramme, Nationale Forschungsprogramme, Energie 2000, Präventionskampagnen u.a.m.).

Was die Reform- und Experimentierbereitschaft der Schweiz angeht, so ist von gegenläufigen Tendenzen auszugehen. Spürbar war und ist (insbesondere auf Bundesebene und generell als Auswirkung des Referendums) ein Konservatismus. "Nur keine Experimente" lautete lange Zeit die Devise. Dies bedeutete zweierlei. Erstens sollte die Schweiz nicht eine *Pionierrolle* übernehmen, sondern neue Massnahmen erst übernehmen, wenn sie sich im Ausland bewährt hatten (Beispiel Mehrwertsteuer). Man wollte die im Ausland gemachten Fehler vermeiden und nur Massnahmen[^] welche sich anderswo bewährt hatten, übernehmen. Zweitens waren *gesellschaftspolitische Experimente*, die durch Einschränkung des personellen und räumlichen Geltungsbereichs von Massnahmen und durch Vergleich von Ziel- und Kontrollgruppe geprägt sind, verpönt. Freilich hat sich inzwischen bei beiden Formen von Experimenten eine Lockerung

der Abwehrreflexe eingestellt: Die Schweiz nahm auf Bundesebene in verschiedenen Sachbereichen (u.a. Umweltschutz) im internationalen Vergleich zeitweise eine Pionierrolle ein, und Experimente in der Form räumlich und/oder zeitlich beschränkter Versuche greifen immer mehr um sich (vgl. Morand 1993 und Mader 1993 sowie Kapitel 2.4.2.).

Daneben verfügt die Schweiz in den Kantonen über ein "*föderatives Labor*". Trotz ihrer Kleinheit sind die Kantone und Gemeinden namentlich im steuerlichen Bereich im internationalen Vergleich bemerkenswert autonom (vgl. Frey 1991; Bussmann 1992b). Trotz zunehmender Politikverflechtung (sei es durch Völkerrecht, Staatsvertragsrecht, geschriebenes und ungeschriebenes Verfassungsrecht sowie durch politikbereichsspezifische Erlasse des Bundes) besitzen die Kantone noch eigenständige Aufgabenbereiche (Bildung, Gesundheit, Polizei, Fürsorge, Kultur u.a.m.) und nehmen Kantone und Gemeinden den Vollzug von Bundespolitiken als Vehikel zur Gestaltung ihrer eigenen spezifischen Interessen wahr (Bussmann 1980 und 1981; Kissling-Näf und Knoepfel 1992; Klöti, Haldemann und Schenkel 1993). Die sich dadurch ergebende Politikdifferenzierung bildete insbesondere seit Ende der siebziger Jahre ein reiches Betätigungsfeld für die Vollzugs- und Wirkungsforschung.

2.1.3. Sozialwissenschaftlicher Beratungsbedarf

Die Überschaubarkeit der kleingekammerten politischen Körperschaften (vgl. Bussmann 1992b) sowie das namentlich bei der Gemeindeverwaltung, z.T. auch bei den Kantonsregierungen nach wie vor bedeutsame Milizprinzip schränkten während langer Zeit den Bedarf nach eingehenden sozialwissenschaftlichen Studien ein, da die Bedürfnisse der Bevölkerung und die Wirkungen staatlichen Handelns für die politischen Instanzen praktisch augenfällig waren. Begünstigt durch den Ausbau des privaten und öffentlichen Verkehrs sind ehemals räumlich getrennte Gebiete in grössere, die Gemeinde- und Kantonsgrenzen überschreitende Agglomerationsgebilde zusammengewachsen. Das Entstehen einer semi-urbanen "Grossagglomeration Schweiz" (Bassand, Joye und Schuler 1988) mit sich in den Zentren kumulierenden Problemen (Verkehr, Fürsorge, Drogen, Kriminalität) hat gleichzeitig auch die politischen Strukturen "unterwandert". Dieser Prozess ruft nach sozialwissenschaftlichem Wissen zur Gestaltung der politischen Strukturen und zur Bearbeitung oder Lösung spezifischer Probleme (Eidgenössisches Personalamt 1996).

Die Sozialwissenschaften sind diesem *erhöhten Beratungsbedarf* indessen noch nicht durchwegs gewachsen. In einer im Jahre 1991/92 in vier Disziplinen durchgeführten Evaluation unter Beizug ausländischer Experten wurde festgestellt, dass die Sozialwissenschaften mit schwierigen Bedingungen konfrontiert seien. Als Stichworte wurden etwa erwähnt, wesentliche Bereiche der Sozialwissenschaften seien unterentwickelt, es bestehe ein Graben zwischen Lehre und Forschung, es fehle an einer genügenden kritischen Masse (namentlich auch aufgrund der sprachlichen Fragmentierung), die Kooperation zwischen Institutionen und Disziplinen im Hinblick auf Forschungsausbildung und -programme sei ungenügend und schliesslich mangle es an einem professionellen Forschungsmanagement, weshalb die Synergien im Bereich der angewandten Forschung nicht ausgeschöpft werden könnten (Schweizerischer Wissenschaftsrat 1993).

2.1.4. Die Erschliessung der peripheren Gebiete als helvetische Variante der "Great society"

Ein letzter Punkt bleibt als Spezifikum in der Entwicklung der Evaluationsforschung zu erwähnen. Während in den USA und beispielsweise Frankreich die Integration unterprivilegierter Bevölkerungsschichten und Rassen in den Hauptstrom gesellschaftlichen und wirtschaftlichen Lebens eine zentrale politische Aufgabe darstellte und weiter darstellt, bildete in der Schweiz die Vermeidung einer Entleerung der peripheren Räume (Berggebiete, Voralpen, Jura, ländliche Regionen des Mittellands) *die* Herausforderung. Pointiert ausgedrückt: Die Integration der peripheren Gebiete ins wirtschaftliche Leben der Schweiz bildete seit Ende der sechziger Jahre die helvetische Variante der Vision einer "Great society". Zur Erreichung dieser Ziele bediente man sich nicht nur sektoraler Massnahmen der Finanz-, Agrar-, Verkehrs-, Medien-, Kultur- und der Regionalpolitik. Die ganze Umbildung des dualen in einen komplexen, kooperativen und von Finanzströmen genährten Föderalismus diente dem Ziel, ohne vollständige Zentralisierung einen interkantonalen und -regionalen Ausgleich zu erzielen. Sozialwissenschaftler wirkten sowohl bei der Konzipierung der Massnahmen (z.B. Prof. Stocker) wie bei deren Erfolgskontrolle mit (vor allem Nationales Forschungsprogramm "Regionalprobleme", siehe weiter hinten). Im Vergleich zu den USA ergaben sich in der Schweiz Impulse für die Entwicklung der Evaluation nicht nur aus der

Sozial- und Bildungspolitik, sondern auch aus der *Regionalpolitik* und den anderen erwähnten Aufgabenfeldern. Die Evaluationsforschung in der Schweiz war und ist deshalb vergleichsweise weniger durch die Disziplinen der Pädagogik und Psychologie geprägt als in den USA, sondern stärker durch die Ökonomie/Soziologie, insbesondere durch die Regionalökonomie, sowie durch die Politikwissenschaft.

Freilich scheint die Regionalpolitik in ihrer heutigen Form ihren Kulminationspunkt überschritten zu haben. In ebenso hohem, wenn nicht höherem Masse wie die Ländernachbarn ist die Schweiz zu einem Einwanderungs- und Asylland geworden mit allen damit verbundenen Schwierigkeiten eines Zusammenlebens von Menschen aus verschiedenen Kulturen. Probleme wie Diskriminierung (der Frauen, aber auch anderer Rassen), Kriminalität, Drogensucht, Gesundheitsvorsorge, Zunahme des Anteils der Rentner, Verlust an Lebensqualität in den grossen Städten u.a.m. schieben sich in die politische und wissenschaftliche Diskussion. Es ist deshalb zu erwarten, dass auch in diesem Punkt ein Anschluss an die im angelsächsischen Bereich diskutierten Themenstellungen erfolgen wird (Schweizerischer Nationalfonds 1996).

2.2. WISSENSCHAFTLICHE GENESE DER EVALUATIONS-PROBLEMATIK

2.2.1. Vorläufer

Evaluationen sind Bestandteile der Beratung von Politik. Sie haben darin ihre Vorläufer. Zu denken ist beispielsweise an Heinrich Pestalozzi, der eigene, z.T. neue Vorstellungen über die Volksschule entwickelte und zu verwirklichen versuchte. Namentlich mit der Entwicklung und Auffächerung der Bundespolitik seit Ende des 19. Jahrhunderts und insbesondere seit Mitte des 20. Jahrhunderts haben natur- und sozialwissenschaftliche Konzepte und Lehrmeinungen immer wieder Einfluss auf politische Entscheide gehabt, sei dies in der frühen Gesundheitspolitik (Epidemiebekämpfung), in der Agrarpolitik (Versuchs- und Forschungsanstalten), in der Wirtschaftspolitik (Krisenbekämpfung, später Konjunktur-, Struktur-, Wettbewerbspolitik u.a.m.) oder in der Umweltpolitik (Forschungen der EAWAG für die Gewässerschutzpolitik).

Eine eigentliche Vorreiterrolle in der Rezeption US-amerikanischer Evaluationsforschung spielte die *Bildungsforschung*. In den siebziger Jah-

ren setzte in zahlreichen Kantonen eine Euphorie im Hinblick auf die Erprobung neuer Unterrichtsmethoden, Curricula und Schulmodelle ein. Das experimentelle Paradigma hielt Einzug, und Schlüsselkonzepte der amerikanischen Evaluationsforschung (wie Bilanz- und Verbesserungsevaluation) wurden in die deutsche und französische Sprache übertragen³.

2.2.2. Von der Implementations- zur Wirkungsforschung

Freilich erfolgten, abgesehen vom Pionier "Bildungsforschung", bis in die siebziger Jahre Evaluationen weitgehend impressionistisch und unreflektiert. In der Regel wurde davon ausgegangen, dass Konzept bzw. rechtliche Verankerung, Vollzug und Wirkungen weitgehend identisch erfolgten d.h. diese Unterscheidungen wurden oft noch gar nicht explizit vorgenommen.

Auch die Ende der sechziger Jahre einsetzende *Planungsdiskussion* (vgl. v.a. Knoepfel 1977; Linder, Hotz und Werder 1979) ging noch stark von der Identität von Wille und Verwirklichung und vom Glauben an die Machbarkeit von Politik (im Sinne von "policy") aus. Bemerkenswert waren indessen zwei Neuerungen: Einerseits die Annahme der Komplexität und Vernetztheit öffentlicher Politik und andererseits das Bestreben, die Wünsche und Anliegen der Betroffenen interaktiv in die Planung einzubeziehen, was eine gesellschaftliche Differenzierung und die Abkehr von der Subordination der Bürgerinnen und Bürger unter gouvernementale Entscheide und somit die Entdeckung der "Akzeptanz" voraussetzte.

Die Planungsdiskussion war die Reaktion auf die politischen Bestrebungen zur Erarbeitung grossangelegter *sektioneller Konzepte* (Gesamtenergiekonzept, Gesamtverkehrskonzeption, Mediengesamtkonzeption) und zur Beeinflussung der räumlichen Entwicklung durch raumplanerische und raumordnungspolitische Massnahmen. Im Rahmen der Planungsdiskussion wurden die damals vorliegenden Erfahrungen kritisch reflektiert, und es wurden zahlreiche Verbesserungsvorschläge unterbreitet. Wie im Ausland, namentlich in den USA ("Great society") und in der damaligen Bundesrepublik Deutschland (sozialliberale Koalition) trafen die Reformen allerdings auf praktische Schwierigkeiten; besonders ihre "Implementation" erwies sich als kritische Variable.

³ Von der regen Evaluationstätigkeit legen die Forschungsnachweise der Schweizerischen Koordinationsstelle für Bildungsforschung in Aarau und zwei Publikationen (Meylan 1981 und 1983) Zeugnis ab.

Die politische Entdeckung von "Vollzugsdefiziten" wurde gefolgt von empirischen Studien über den Vollzug in einzelnen Politikbereichen⁴. Die Nationalen Forschungsprogramme "Regionalprobleme" (NFP 5) und "Entscheidungsprozesse in der schweizerischen Demokratie" (NFP6) wurden mit "volets" zur Raumwirksamkeit staatlicher Massnahmen bzw. zur Implementation in ausgewählten Politikbereichen versehen. In zahlreichen weiteren Nationalen Forschungsprogrammen⁵ sowie in der freien wissenschaftlichen Forschung wurden Vollzugsfragen behandelt.

Die weitere politische Diskussion gab der Implementationsfragestellung dann eine neue Wende. In den Nationalratswahlen 1979 warb die FDP mit dem Slogan "Mehr Freiheit und Selbstverantwortung - weniger Staat". Er war symptomatisch für die namentlich im bürgerlichen Lager vollzogene ideologische Neuausrichtung. Während zu Beginn der siebziger Jahre Vollzugsdefizite noch als Ausnahmeerscheinungen empfunden wurden, verbreiteten sich gegen Ende des Jahrzehnts und in den achtziger Jahren die Zweifel daran, dass der Staat alle ihm übertragenen Verpflichtungen sinnvoll wahrnehmen könne. Diese Zweifel waren auch im sozialdemokratischen und entstehenden grünen Lager vorhanden; das Verbesserungspotential des Staates wurde hier aber wesentlich grösser eingeschätzt.

Den immer weiter verbreiteten Ungewissheiten über die Wirksamkeit staatlichen Handelns nahmen sich die freie wissenschaftliche Forschung (z.B. Knoepfel 1979, Mader 1985) und *Nationale Forschungsprogramme* ab Mitte der achtziger Jahre an. Zu erwähnen ist vor allem das Nationale Forschungsprogramm "Wirksamkeit staatlicher Massnahmen" (NFP 27); zahlreiche weitere Forschungsprogramme⁶ enthielten aber ebenfalls Evaluationsfragestellungen. Die "Wirksamkeit" in allen Facetten wurde zu einem Standardbegriff politischer und wissenschaftlicher Diskussionen.

⁴ Germann et al. 1979 und Bussmann 1980.

⁵ "Probleme der sozialen Integration in der Schweiz" (NFP 3), "Wirtschaftlichkeit und Wirksamkeit im schweizerischen Gesundheitswesen" (NFP 8), "Bildung und das Wirken in Gesellschaft und Beruf" (NFP 10) und "Nutzung des Bodens in der Schweiz" (NFP 22).

⁶ 5. Serie: "Stadt und Verkehr" (NFP 25), "Wandel der Lebensformen und soziale Sicherheit" (NFP 29) und "Die Gesundheit des Menschen in seiner heutigen Umwelt" (NFP 26; Teilbereich Luftverschmutzung-Gesundheit). 6. Serie: "Klimaänderungen und Naturkatastrophen" (NFP 31), "Alter" (NFP 32), "Wirksamkeit unserer Bildungssysteme" (NFP 33), "Komplementärmedizin" (NFP 34) und "Frauen in Recht und Gesellschaft" (NFP 35).

Es verdient hervorgehoben zu werden, dass in der Schweiz wie übrigens in Deutschland die Entdeckung der Wirkungs- und Evaluationsproblematik weitgehend den Phasen des Politikzyklus (Planung, Implementation, Evaluation) folgte, während die Entwicklung in den USA in einer anderen Reihenfolge verlief: bereits in den sechziger Jahren wurden im Sinne von Black-box-Evaluationen die Wirkungen von Sozialprogrammen zu ermitteln versucht. Dass der Misserfolg nicht nur am Programm selbst (=program failure), sondern auch am mangelhaften Vollzug (=implementation failure) liegen könnte, war eine ernüchternde Erkenntnis, die im Anschluss an Pressmans und Wildavskys Studie (1973) vermehrt ins Bewusstsein der (sozial)wissenschaftlichen Gemeinschaft trat und zur Entdeckung der *Implementation* als "missing link" zwischen Planung und Wirkung(slosigkeit) führte. Die Entwicklung in der Schweiz wurde durch die deutsche Vollzugsforschung, insbesondere durch den von Renate Mayntz und Fritz Scharpf geleiteten Forschungsverbund "Implementation politischer Programme" (Mayntz 1980 und 1983) stark beeinflusst, an dem auch Schweizer Forscher beteiligt waren. Durch die erst spätere Erschliessung der Wirkungsproblematik konnten in Europa einige Fehler der frühen amerikanischen Wirkungsforschung (namentlich das allzu starke Abstützen auf Black-box-Evaluationen) vermieden werden.

2.2.3. Evaluationen und andere Instrumente der Erfolgskontrolle (Controlling und NPM)

Bestrebungen zur *Rationalisierung* staatlichen Handelns setzen bei der *Entscheidungsfindung* sowie bei der *Kontrolle des Vollzugs und der Wirkungen* staatlichen Handelns an. Sie versuchen, die beiden Aktivitäten besser aufeinander abzustimmen. Neben den Evaluationen, die primär retrospektiven Charakter haben, aber durchaus auch Ansätze eines integrierten Management-Konzepts im öffentlichen Bereich enthalten⁷, ist eine Vielzahl von weiteren Ansätzen der Erfolgskontrolle (Bussmann 1995d und 1996a) mit meist spezifischen Fokussierungen zu erwähnen (Planung öffentlicher Politiken, Monitoring⁸, Leistungsindikatoren, Controlling, Aufgaben- und Organisationsüberprüfungen, Total Quality Management (z.B. Emery

⁷ Zu erwähnen sind namentlich die von Wholey (1979; 1983) entwickelte Perspektive einer Verbindung von (Management durch) Zielsetzung und Vollzugs- sowie Wirkungskontrolle sowie der Ansatz von Love (1991).

⁸ Näheres dazu unter Kapitel 8.

1996) u.a.m.). Herausgegriffen seien das von der deutschen Betriebswirtschaftslehre entwickelte Konzept des Controlling, dessen Anwendung auch in der öffentlichen Verwaltung versucht wird, und das im angelsächsischen Bereich entwickelte New Public Management.

Das *Controlling* hat mit dem amerikanischen Total Quality Management (TQM) gemeinsam, dass für wichtige Outputfaktoren Vorgaben gesetzt werden und dass anschliessend der Vollzug laufend und systematisch überprüft wird. In der Praxis erweist es sich allerdings als äusserst schwierig, Indikatoren zu finden, welche schlüssig über die Wirksamkeit von Massnahmen Bescheid geben; hierbei müssen Ansätze des Controlling auf die Methodik von Evaluationen zurückgreifen (Bussmann 1995d; Knoepfel 1996a). Evaluation und Controlling erweisen sich als zwei Ansätze, die in der (weiten) Interpretation ihrer Protagonisten deckungsgleich sein können, in ihrer praktischen Anwendung dagegen unterschiedliche Fokussierungen aufweisen. Das Controlling konzentriert sich auf betriebswirtschaftliche, interne finanzielle sowie auf Führungs- und Vollzugsaspekte. Evaluationen konzentrieren sich z. T. auch auf den Vollzug, vor allem aber auf die gesamtgesellschaftlichen Wirkungen und Nebenwirkungen sowie auf externe Nutzen-Kosten-Verhältnisse.

Evaluationen dienen zwar auch als Führungsinstrument innerhalb der Verwaltung, werden daneben aber vor allem als breit angelegtes Instrument der Politikbeobachtung verwendet, welches mehreren Institutionen und Gruppen (Parlament, Verwaltung, Parteien, Verbände, Bürgergruppen u.a.m.) zur Verfügung steht.

Unter dem Titel "*New Public Management*" (NPM) werden seit einiger Zeit in angelsächsischen Ländern (v.a. Neuseeland, Australien) umfassende Reformen des Staates in Angriff genommen. Seine zentralen Elemente sind (Pollitt 1995; Naschold 1995; Schedler 1995; Damkowsky und Precht 1995): Die Qualitätsorientierung gegenüber den "Kunden", die Bemühungen um Kostenersparnisse und insbesondere bessere Transparenz über die Produktkosten, die Ausdifferenzierung von Leistungsfinanzierern, Leistungskäufern und Leistungserbringern, die organisatorische Verselbständigung von Verwaltungseinheiten (oft im Rahmen von Leistungsvereinbarungen und Globalbudgets), die Dezentralisierung von Führungsverantwortung (Trennung in strategische und Operationelle Entscheide), die Einführung von Leistungs- und Wirkungsmessung (z.B. Benchmarking, Leistungsindikatoren), die Einführung von Wettbewerbselementen (z.B. competitive tendering, contracting out, vouchers) und die Reform des öffentlichen Dienstrechts (Abkehr vom Beamtenstatus, Flexibilisie-

rung der Gehälter). In der Schweiz steht die Einführung des NPM noch in der Anfangsphase (Hablützel et al. 1995), und nicht alle in der internationalen Diskussion relevanten Aspekte stehen im Vordergrund. Angestrebt wird vor allem die Lockerung bürokratischer Vorgaben durch Globalbudgets und durch Kontraktmanagement (Leistungsvereinbarungen). Dabei gibt es auch erste Versuche der Output- und Wirkungserfassung.

Die angestrebte, aber noch nicht ausreichend verwirklichte Ergebnis- bzw. Wirkungsorientierung des NPM wird auch Auswirkungen auf die Ausgestaltung der *Kontrollen* haben. Mit der Lockerung bürokratischer Vorgaben und der Zuteilung vermehrter operationeller Verantwortung an dezentrale oder verselbständigte Verwaltungseinheiten wird die an der Rechtmässigkeit der Tätigkeiten orientierte Kontrolle zunehmend überlagert von einer Kontrolle des Outputs und der Outcomes bzw. der Wirksamkeit (Bussmann 1995e). Daraus erwächst ein erhöhter Bedarf nach Evaluationen. Es ist daher kein Zufall, dass gerade die Vorläufer in Sachen NPM (Australien und Neuseeland) seit einiger Zeit ein stark erhöhtes Interesse an Evaluationen bekunden.

NPM und andere Bestrebungen zur Regierungs- und Verwaltungsreform dürften den Bedarf nach Evaluationen aber noch in anderer Hinsicht erhöhen. Während in der Vergangenheit vor allem der Staat die Wirtschaft und Gesellschaft zu reformieren versuchte, was einen entsprechenden Bedarf nach Evaluationen nach sich zog (vgl. Kapitel 1.3.), wird heute angestrebt, den Staat zu reformieren. Dabei wird es sich mehr und mehr als notwendig erweisen, Modelle wie das New Public Management auf ihre Wirksamkeit zu evaluieren (Pollitt 1995). Es ist nämlich noch unsicher, ob und inwieweit sie sich in der Praxis im erhofften Ausmass bewähren werden (Knoepfel 1995a und 1995c; Mastronardi 1995).

2.3. METHODIK: GEGEBENHEITEN UND TENDENZEN

Genese und Kontext haben - über den Weg der wissenschaftlichen Ausbildung der Forscherinnen und Forscher - auch die Methodik der Evaluationsforschung mitgeprägt. Erwähnt wurden die Pionierarbeiten in der Bildungsforschung sowie die frühen Versuche zur Abklärung der Wirkungen in der Regionalpolitik. Das dort entwickelte Instrumentarium leuchtete vor allem Mitnahmeeffekte sowie direkte und indirekte (sekundäre oder gar tertiäre) Wirkungen staatlicher Transferzahlungen oder Dienstleistungen aus. Eingehend behandelt wurde auch die Logik des Handelns

öffentlich-rechtlicher Körperschaften (v.a. spillover bzw. Trittbrettfahrer-Problematik).

Einen weiteren Zugang zur Wirkungsforschung bot vor allem die *Vollzugsforschung* (vgl. u.a. Linder 1988; Kriesi 1993). Diese war und ist in der Regel qualitativ angelegt (Dokumentenanalysen, Interviews). Sie bedient sich meist einfacher oder vergleichender Fallstudien⁹ und versucht, den Ablauf möglichst umfassend und mit dem Ziel zu erfassen, daraus Gesetzmässigkeiten abzuleiten. Die Erfahrung mit qualitativen Methoden hat namentlich in der Westschweiz (C.E.A.T.) zur Weiterentwicklung *interaktiver Evaluationsmethoden* (Forumsansatz, "évaluation par les acteurs", "évaluation pluraliste")¹⁰ geführt, auf denen ein beachtliches Niveau erreicht werden konnte. Weniger leicht ist indessen die Ergänzung der Vollzugsforschung in Richtung einer Erfassung der Wirkungen gefallen, weil hier oft auch *quantitative Methoden* gefragt sind. In letztere Richtung ist die in der Schweiz recht gut ausgebaute ökonometrische Forschung gestossen, die mit ihren ausgefeilten quantitativen Methoden und Modellen aber häufig mit dem Problem ungenügender Datenqualität konfrontiert wird. Erfahrungen mit quantitativen Methoden bestehen aber auch in anderen Disziplinen (Pädagogik, Psychologie, Soziologie, Politikwissenschaft). Insgesamt sind im Bereiche sowohl qualitativer wie auch quantitativer Methoden Verbesserungen erzielt worden.

Eine Besonderheit schweizerischer und generell europäischer Evaluationsforschung gilt es hervorzuheben und gleichzeitig in Frage zu stellen: Die Ermittlung der Wirkungen auf dem Wege einer detaillierten Erfassung des Wirkungspfades. Oder umgekehrt: Die Ablehnung von *Black-box-Evaluationen*. Dafür gibt es namentlich aufgrund der weiter oben referierten US-amerikanischen Erfahrungen durchaus gute Gründe. Die Geschichte medizinischer Evaluation zeigt indessen, dass Wirkungen *entdeckt* und Medikamente angewendet werden können, ohne dass der genaue Wirkungsmechanismus geklärt ist. Dies setzt systematische Prozesse des "trial and error" voraus, die in den USA namentlich durch das "föderative Labor" gegeben sind. In der Schweiz sind aufgrund der Gesetzestradiation und des Gebots der Gleichbehandlung "full coverage pro-

⁹ In der Vollzugs- und generell der Policy-Forschung wird unter "Fallstudie" oft die Untersuchung der Interaktionen bzw. des Vollzugs in einem Politik- bzw. "Interpolicy"-Bereich verstanden. Damit wird der Forschungsansatz der Fallstudie über die in den Sozialwissenschaften sonst üblichen Kategorien (Einzelpersonen, soziale Aggregate) ausgeweitet (vgl. Stake 1994: 236).

¹⁰ Vgl. v.a. INFRAS 1990, Rey et al. 1993, Bierter und Binder 1993 sowie Kapitel 13.3.

grams"¹¹ die Regel; immerhin erlaubt auch der schweizerische Föderalismus das *Erproben* neuer Möglichkeiten, und hat auch auf Bundesebene das experimentelle Paradigma in Form von Evaluationsklauseln, Versuchsregelungen u.a.m. FUSS gefasst.

Die Zukunft wird vermutlich von einer Verbindung von Methoden und deren eklektischen Nutzung gekennzeichnet sein. Disziplinbezogen betrifft dies die Verbindung betriebswirtschaftlicher sowie anderer sozialwissenschaftlicher Ansätze, themenbezogen die Verbindung finanzieller, organisationeller und gesellschaftlicher (Wirkungs-)Gesichtspunkte. Ein Beispiel für solche Verbindungen ist etwa die Evaluation des BUWAL (Knoepfel und Zimmermann 1992)(Verbindung von Ansätzen der Organisations- und der Politikanalyse)¹².

2.4. RECHTLICHE VERANKERUNG VON EVALUATIONEN BZW. ERFOLGSKONTROLLEN

Seit rund 20 Jahren finden die Bemühungen zur Erfassung der Wirksamkeit und der Wirtschaftlichkeit staatlicher Massnahmen auch ihren rechtlichen Niederschlag. *Evaluationen ohne rechtliche Verankerung* ("*spontane Evaluationen*") bildeten indessen den notwendigen Hintergrund für eine rechtliche Verankerung, da sie das Sammeln von Erfahrungen ermöglichen. Die Impulse dazu waren und sind verschiedener Art. Evaluationen können ausgelöst werden durch wissenschaftliche Forschung (Nationale Forschungsprogramme und eigenständige Forschung), Anstösse von Amtsdirektoren oder Vollzugsbehörden, Vorstösse von Parlamentariern und Tätigkeit der Kontrollorgane sowie durch Aufträge von Parteien, Verbänden und Bürgervereinigungen. Die entsprechenden Bestrebungen können punktuellen Charakter haben, wie beispielsweise die einmalige Überprüfung eines Erlasses im Hinblick auf dessen Novellierung; sie können aber auch umfassend und systematisch angelegt sein, ohne direkt in eine rechtliche Verankerung zu münden¹³.

¹¹ Vgl. Rossi und Freeman 1993: 333 ff. Es handelt sich um Programme bzw. Gesetze, welche alle potentiellen Adressaten abdecken und deshalb Kontrollgruppenvergleiche (experimentelles und quasi-experimentelles Evaluationsdesign) erschweren.

¹² Vgl. auch den Bericht der Geschäftsprüfungskommission des Nationalrats an den Bundesrat über die Evaluation des BUWAL (Bundesamt für Umwelt, Wald und Landschaft) vom 18. Mai 1992 (BB1 1992 III 1502).

¹³ Beispiel hierfür sind etwa die Bestrebungen zu einer systematischen Hochschul-Evaluation mit dem Mittel von Befragungen von Studierenden und Hochschulabsolventen.

2.4.1. Evaluationen zum bewussten Gewinnen von Erfahrungen ("experimentelle Evaluationen")

Der selbstreflexive Aspekt, der Evaluationen zugrundeliegt, ist bei der Konzipierung von Evaluationsprogrammen "in eigener Sache" genutzt worden; Evaluationen sind dabei nicht Selbstzweck, sondern dienen der Ermittlung der Stärken und Schwächen dieses Instruments.

Erstes Beispiel ist die *Arbeitsgruppe "Gesetzesevaluation" (AGEVAL)*. Sie wurde 1987 eingesetzt und bestand aus Vertretern des Bundes, der Kantone und der Wissenschaft. Die AGEVAL hatte den Auftrag, "zweckmässige Massnahmen zur Verstärkung der Wirkungsprognose und der Wirkungskontrolle im Bereich der Gesetzgebung vorzuschlagen und deren Realisierungsmöglichkeiten zu prüfen". Die AGEVAL gab fünf Evaluationen in Auftrag, um mit diesem Instrument erste Erfahrungen zu sammeln¹⁴. Gestützt auf die Erfahrungen mit diesen Evaluationen und auf weitere Abklärungen über die aktuelle Praxis der Erlassvorbereitung auf Bundesebene (Klöti und Schneider 1989), auf ausländische Beispiele der Institutionalisierung von Evaluationen und auf eine Umfrage zu Umfang und Verbreitung von Evaluationen auf Bundesebene (Zimmermann, Kolly und Mahon 1989) erarbeitete die Arbeitsgruppe 1991 ihren Schlussbericht (Arbeitsgruppe "Gesetzesevaluation" 1991). Darin forderte sie insbesondere eine Verstärkung des Dispositivs im Bereich prospektiver Evaluation.

Zweites Beispiel ist das *Nationale Forschungsprogramm "Wirksamkeit staatlicher Massnahmen" (NFP 27)*. Das Forschungsprogramm wurde vom Bundesrat ebenfalls 1987 lanciert und mit 5 Millionen Franken dotiert. In der Periode 1990 - 1995 sind 12 Evaluationen, 3 Projekte zum Thema "Lernen in Verwaltungen" sowie eine Metaevaluation durchgeführt sowie 7 Kurzevaluationen mitfinanziert worden, um ein Know-how über Evaluationsmethoden aufzubauen und diese zu erproben. Damit sollten Anhaltspunkte für die Zeitgerechtigkeit, Zuverlässigkeit und Relevanz von Evaluationen gewonnen und überprüft werden, inwieweit durch Evaluationen Lernprozesse in Gang gesetzt werden können¹⁵.

¹⁴ Anwander u.a. 1990; Schaetti 1990; INFRAS 1990; Gerheuser 1991; Linder et al. 1992.

¹⁵ Synthesen der Arbeiten bieten v.a. Bussmann 1994, 1995a, 1995c sowie der vorliegende Band; die Publikationen aus diesem NFP werden im Anhang zu diesem Buch angeführt.

2.4.2. Bestimmungen über die Durchführung von Evaluationen bzw. Erfolgskontrollen

Rechtliche Bestimmungen ermöglichen oder verlangen - entweder vor dem Entscheid über eine Massnahme (selten) oder nach deren Inkraftsetzung und Vollzug - die Durchführung einer Evaluation bzw. einer weniger systematischen und stringenten Erfolgskontrolle. Entsprechende *Verfahrvorschriften* können in die Form einer Evaluationsklausel, eines Versuchserlasses, von Bestimmungen über die Bereitstellung von Budgetmitteln für Evaluationen oder einer Kombination der ersten und dritten Form gekleidet werden.

2.4.2.1. Evaluationsklauseln

Evaluationsklauseln sind rechtliche Verpflichtungen zur Durchführung einer Evaluation bzw. einer Erfolgskontrolle. So einfach die Grundidee auch erscheint, so vielfältig sind die Anwendungsmöglichkeit und die tatsächliche Praxis. Auf Bundesebene allein gibt es ein knappes Dutzend solcher Klauseln. Die meisten davon haben retrospektiven Charakter und beziehen sich auf einzelne Massnahmen¹⁶. Die wichtige Ausnahme davon ist *Artikel 43 Absatz 3 des Geschäftsverkehrsgesetzes* (SR 171.11), welcher die Anforderungen an die Botschaften und Berichte des Bundesrats festhält. Danach hat der Bundesrat ex ante die personellen und finanziellen Auswirkungen der Vorlage auf den Bund, die Folgekosten für die Kantone und Gemeinden, die Auswirkungen auf die Wirtschaft sowie weitere Punkte (soweit möglich das Verhältnis von Kosten und Nutzen der Vorlage sowie bei Finanzhilfe- und Abgeltungsvorlagen die Übereinstimmung mit dem Subventionsgesetz) darzustellen. Einen ähnlich breiten

¹⁶ Artikel 44 des Umweltschutzgesetzes (SR 814.01); Artikel 9 des Bundesgesetzes vom 19. März 1976 über die internationale Entwicklungszusammenarbeit und humanitäre Hilfe (SR 974.0); Artikel 8 und 9 des Bundesgesetzes vom 5. Oktober 1984 über die Leistungen des Bundes für den Straf- und Massnahmenvollzug (SR 341); Artikel 11 der Verordnung vom 18. November 1992 über die Hilfe an Opfer von Straftaten (SR 312.51); Artikel 24 des Energienutzungsbeschlusses vom 14. Dezember 1990 (SR 730.0); Artikel 8 der Verordnung vom 12. April 1995 über den Risikoausgleich in der Krankenversicherung (SR 832.112.1); Artikel 32 der Verordnung vom 27. Juni 1995 über die Krankenversicherung (SR 832.102); Artikel 18 der Verordnung vom 3. Oktober 1994 über die Wohneigentumsförderung mit Mitteln der beruflichen Vorsorge (SR 831.411); Artikel 20 der Verordnung vom 3. Oktober 1994 über die Freizügigkeit in der beruflichen Alters-, Hinterlassenen- und Invalidenvorsorge (SR 831.425).

Geltungsbereich, allerdings nicht prospektiv und bezogen auf Berichte und Botschaften des Bundesrats, sondern retrospektiv und bezogen auf Subventionserlasse entfaltet *Artikel 5 Absatz 1 des Subventionsgesetzes* (SR 616.1). Danach ist der Bundesrat gehalten, periodisch, mindestens aber alle sechs Jahre, zu prüfen, ob die Bestimmungen über Finanzhilfen und Abgeltungen den Grundsätzen des Subventionsgesetzes entsprechen.

Die Begrifflichkeit der Evaluationsklauseln ist noch recht uneinheitlich. In der rechtlichen Praxis findet sich eine Vielzahl von Formulierung, die von einfachen Prüfungsaufträgen¹⁷ oder Überwachungsaufträgen¹⁸ bis hin zu expliziten Evaluierungsaufträgen¹⁹ reichen.

Nebst der Bundesebene finden sich Evaluationsklauseln auch auf der Ebene der Kantone sowie vermutlich auch der Gemeinden²⁰. Besondere Erwähnung als bisher wohl umfassendste Evaluationsklausel verdient Artikel 101 der Verfassung des Kantons Bern vom 6. Juni 1993, welcher dazu verpflichtet, alle Aufgaben periodisch auf ihre Notwendigkeit und Zweckmässigkeit sowie auf ihre finanziellen Auswirkungen und deren Tragbarkeit zu überprüfen.

2.4.2.2. Versuchserlasse

Versuchserlasse (Mader 1988 und 1993; Mastronardi 1991) sind befristet und zudem - im Unterschied beispielsweise zu (befristeten) einfachen Bundesbeschlüssen - zusätzlich mit Vorkehren zur wissenschaftlichen Begleitung/Erfolgskontrolle/Evaluation versehen. Letzteres dient idealerweise als Grundlage für den Entscheid, ob die Massnahme nach der Versuchsphase aufzuheben²¹ oder in eine dauerhaft gedachte Regelung zu überführen ist. Es gibt in der Schweiz rund zwanzig Versuchserlasse und zwar in den Sachgebieten: Datenbanken, Verkehr, Verwaltungsorganisation, Strafrecht, Fernmeldewesen, Gesundheit, Sozialversicherung und

¹⁷ Z.B. der eben erwähnte Artikel 5 Absatz 1 des Subventionsgesetzes sowie Artikel 44 des Umweltschutzgesetzes.

¹⁸ Z.B. Artikel 9 des Bundesgesetzes vom 19. März 1976 über die internationale Entwicklungszusammenarbeit und humanitäre Hilfe.

¹⁹ Artikel 11 der Verordnung vom 18. November 1992 über die Hilfe an Opfer von Straftaten.

²⁰ Die Untersuchung beschränkt sich primär auf die Ebene des Bundes mit einigen Beispielen aus den Kantonen.

²¹ Dieser Fall war in der schweizerischen Praxis indessen bisher (noch) nicht zu beobachten.

Landwirtschaft. In der Öffentlichkeit am bekanntesten ist die versuchsweise Drogenabgabe an stark drogenabhängige Personen²².

2.4.2.3. Budgetmittel für die Evaluation

Mehrere befristete Massnahmen im Bildungsbereich waren in jüngerer Zeit mit Bestimmungen verbunden, welche bestimmte Budgetmittel für die "wissenschaftliche Begleitung und Evaluation" vorsahen²³. Die Formulierungen lassen allerdings offen, ob die Mittel für die Steuerung und Verbesserung des Vollzugs oder für eine Wirkungsanalyse und Bilanz der Massnahmen oder allenfalls beides eingesetzt werden sollen.

2.4.3. Institutionen mit dem Auftrag, Evaluationen bzw. Erfolgskontrollen durchzuführen

Die Schaffung neuer bzw. die Neuumschreibung der Pflichtenhefte bestehender Institutionen bildet seit Ende der achtziger Jahre einen der Kristallisationspunkte im Bereich der Evaluationen. Eine wichtige Rolle spielten dabei - nebst den vorangehend erwähnten Arbeiten der "Arbeitsgruppe Gesetzesevaluation" (AGEVAL) - die Geschäftsprüfungskommissionen (und deren Sekretariat). Die Geschäftsprüfungskommission des Ständerats lud in ihrem Bericht vom 12. November 1987 (BEI 1988 , 748 ff.) den Bundesrat ein, eine gemeinsame Verwaltungskontrollstelle zu schaffen. Der Bundesrat widersetzte sich dieser Idee und beschloss, eine eigene "*Dienststelle für Verwaltungskontrolle*" zu schaffen. Sie unterstützt den Bundesrat bei seiner Aufsicht über die Bundesverwaltung (Art. 1), befasst sich mit den Aufgaben, Tätigkeiten und Mitteln der Verwaltung sowie mit deren Organisation, Arbeitsweisen und Entscheidungsverfahren (Art. 4). Sie

²² Vgl. Verordnung vom 21. Oktober 1992 über die Förderung der wissenschaftlichen Begleitforschung zur Drogenprävention und Verbesserung der Lebensbedingungen Drogenabhängiger (SR 812.121.5).

²³ Art. 6 der Verordnung vom 4. Juli 1990 über die Sondermassnahmen zugunsten der universitären Weiterbildung (SR 414.124.1), Art. 4 Abs. 2 Bst. b des Bundesbeschlusses vom 30. Januar 1992 über Sondermassnahmen zur Förderung des akademischen Nachwuchses in den Jahren 1992-1995 (SR 414.204; 2 Mio Franken für die Begleitung und Evaluation der Massnahmen), Art. 7 der Verordnung vom 9. Juli 1991 über Massnahmen zur Förderung der gegenseitigen Anerkennung von Studienleistungen und der Mobilität in der Schweiz (SR 414.411).

überprüft dabei die Ziel- und Aufgabenkonformität, die Rechtskonformität des Vollzugs, die Wirkungen sowie die Effizienz (Art. 5)²⁴. Die eidgenössischen Räte richteten anschliessend ihre eigene "*Parlamentarische Verwaltungskontrollstelle*" ein. Sie überprüft die Aufgaben der Verwaltung und ihre Erfüllung sowie die Wirkungen des Handelns von Behörden und Verwaltung. Dabei richtet sie sich nach den Kriterien der Rechtmässigkeit, der Zweckmässigkeit, der Leistungsfähigkeit und der Wirksamkeit²⁵. Die daraus mögliche Ausdehnung des Einflussbereichs der Geschäftsprüfungskommissionen zulasten der Finanzdelegation bzw. der Finanzkommissionen veranlassten letztere, den Einbezug von Wirkungsaspekten in den Tätigkeitsbereich der *Eidgenössischen Finanzkontrolle* zu verankern, eine Bestrebung, die zudem in Übereinstimmung mit den Bestrebungen der Internationalen Organisation der Obersten Rechnungskontrollbehörden (INTOSAI) stand²⁶.

Auf Bundesebene können aufgrund der geltenden Rechtslage somit drei Institutionen nachträgliche Wirkungsanalysen vornehmen, nämlich die Verwaltungskontrolle des Bundesrats, die Parlamentarische Verwaltungskontrolle und die Finanzkontrolle. Umfassende Wirkungskontrollen hat bisher vor allem die Parlamentarische Verwaltungskontrollstelle durchgeführt.

Neben diesen drei Institutionen, deren Pflichtenhefte auf Gesetzes- bzw. Verordnungsebene definiert werden, gibt es auf Bundesebene verschiedene weitere Institutionen, welche in ihrem jeweiligen Bereich Evaluationen in Auftrag geben oder durchführen (vgl. auch Fussnote 16 weiter vorne). Dazu zählen etwa das Bundesamt für Energiewirtschaft (im Zusammenhang mit "Energie 2000"), das Bundesamt für Gesundheitswesen (AIDS-Kampagne, kontrollierte Drogenabgabe), das Bundesamt für Justiz (Modellversuche im Strafvollzug, Opferhilfe), das Bundesamt für Umwelt, Wald und Landschaft (Beobachtungsnetze in den wichtigen

²⁴ Vgl. Art. 1-5 der Verordnung vom 11. Dezember 1989 über die Dienststelle für Verwaltungskontrolle (SR 172.210.11).

²⁵ Vgl. Artikel 47sexies des Geschäftsverkehrsgesetzes (SR 171.11), eingefügt durch Ziffer I des BG vom 22. Juni 1990 (BB1 1990 I 1065). Als Beispiel für die Tätigkeit dieses Kontrollorgans sei eine Evaluation der Rechtsetzung auf der Verordnungsstufe im Bereich der Beruflichen Vorsorge erwähnt, die der Geschäftsprüfungskommission des Ständerates im Oktober 1994 zugeleitet worden ist. Im Evaluationsbericht werden Fragen der Gesetzeskonformität des Ordnungsrechts eingehend erörtert.

²⁶ Bundesgesetz vom 28. Juni 1967 über die Eidgenössische Finanzkontrolle (SR 614.0) sowie Botschaft vom 30. März 1994 zur Änderung des Finanzkontrollgesetzes, insbesondere Artikel 5 des Entwurfs.

Umweltbereichen), das Bundesamt für Sozialversicherung (Krankenversicherung, zweite Säule) und die Direktion für Entwicklung und Zusammenarbeit (Evaluation und Monitoring in der Entwicklungshilfe²⁷).

Auch verschiedene Kantone haben Effizienz- und Wirkungsfragen in die Pflichtenhefte ihrer Organisationen aufgenommen. So verfügt beispielsweise der Kanton Bern in der Gesundheits- und Fürsorgedirektion über ein Amt für wissenschaftliche Auswertung, und mehrere grössere Kantone haben im Erziehungs- oder Umweltbereich Forschungs- und Evaluationseinheiten eingerichtet.

In der jüngsten Vergangenheit wurden in Institutionen mit ausgebauten Evaluationsaktivitäten (Parlamentarische Verwaltungskontrollstelle, Direktion für Entwicklung und Zusammenarbeit, Bundesamt für Gesundheitswesen) Reorganisationen durchgeführt, welche (unter anderem) bezweckten, die Evaluation näher an die operativen Tätigkeiten heranzuführen. So wurde namentlich die Parlamentarische Verwaltungskontrollstelle, die bisher ein selbständiges Organ war, administrativ dem Sekretariat der Geschäftsprüfungskommissionen unterstellt. Mit diesen Reformen wird an sich angestrebt, die Chancen für die Nutzung der Evaluationsresultate zu verbessern. Dabei besteht aber auch die Gefahr, dass die Evaluation zu stark in die Tagesgeschäfte eingebunden wird und ihre unabhängige und kritische Funktion verliert. Die Zukunft wird zeigen, ob und inwieweit sich die jüngsten Reformen bewähren.

2.5. ERFAHRUNGEN

Wie dargelegt lässt sich eine Vielzahl von Impulsen und Initiativen zur Politikevaluation in allen Sachgebieten und auf allen staatlichen Ebenen (Bund, Kantone, Gemeinden) beobachten. Rechtliche Verpflichtungen wurden festgeschrieben, und die Pflichtenhefte verschiedener Institutionen verlangen, die Wirkungen staatlichen Handelns zu überprüfen. In der Verwaltung haben die Kenntnisse über die Möglichkeiten und Grenzen von Evaluationen zugenommen. Es gibt sowohl auf Seiten der Universitätsinstitute wie der privaten Beratungsbüros kompetente Personen, welche sich mit dem Instrument "Evaluationen" auskennen. Immer häufiger werden Evaluationen vorgenommen. Nach unserer Schätzungen werden

²⁷ Vgl. Botschaft vom 20. April 1994 über die Weiterführung der technischen Zusammenarbeit und der Finanzhilfe zugunsten von Entwicklungsländern, Ziffer 236 (BB1 1994 1017 f.).

heute jährlich rund 70 - 100 Vollzugs- und Wirkungsanalysen (im weiteren Sinne) durchgeführt. Allein auf Bundesebene wurden gemäss einer Umfrage des Bundesamtes für Justiz in der Legislaturperiode 1991-1995 534 Evaluationen durchgeführt; sie betrafen vor allem Projekte der Entwicklungszusammenarbeit, der Drogenbekämpfung, der AIDS-Prophylaxe, der Energiepolitik und des Strafvollzugs (vgl. Bulletin der Schweizerischen Evaluationsgesellschaft, Nr. 2, November 1996: 2-4). Evaluationen sind aus Politik und Verwaltung kaum mehr wegzudenken.

Allerdings müssen auch verschiedene Mängel erwähnt werden. Erstens: Das *rechtliche Dispositiv* im Bereich von Evaluationen ist zwar mittlerweile recht *eindrucksvoll*, die zuständigen Stellen ringen indessen noch um deren *zweckmässige Umsetzung*. Zweitens ist festzustellen, dass die *bisherige Handhabung* der rechtlichen Vorkehrungen zur Sicherstellung von Evaluationen der damit verbundenen *inhärenten Handlungslogik teilweise nicht entsprochen hat*, indem beispielsweise über die Weiterführung der in der Krankenversicherung eingeführten neuen Versicherungsformen (HMO) beschlossen wurde, noch ehe die Evaluation überhaupt abgeschlossen war. Auch im Bildungsbereich hielt die Realität nicht ganz, was die Experimentierfreude versprochen hatte: Anstatt als Test und Vergleichsbasis zu dienen und den Wettbewerb unter Schulmodellen und die Auslese des besten Modells zu beflügeln, wie ursprünglich gehofft worden war, wurden neue Schulmodelle einfach zur Erweiterung der Angebotspalette im Schulbereich benutzt. In ähnlicher Weise wurden alle bisherigen Versuchsregelungen in definitive Lösungen übergeführt. Drittens sind die Evaluationsaktivitäten bisher noch *dispers* und *unkoordiniert*. Es fehlt eine Stelle, welche Erfahrungen sammelt und Know-how aufbaut. Damit hängt viertens zusammen, dass in zahlreichen Zweigen der Verwaltung, in den politischen Parteien sowie den Verbänden noch *keine genügende kritische Masse* vorhanden ist, damit das Instrument "Evaluationen" massgeschneidert und zielgerecht eingesetzt werden kann. Dazu hat sicher auch die Verteilung der Evaluationsfunktionen auf drei administrative Einheiten des Bundes beigetragen. Auch kennen Departementsvorsteher und Vollzugsinstanzen die Vor- und Nachteile von Evaluationen noch nicht ausreichend, um Möglichkeiten für ihren Einsatz abzuschätzen bzw. realistische Erwartungen zu entwickeln.

2.6. EVALUATIONEN UND KONKORDANZDEMOKRATIE

Das schweizerische politische System ermöglicht direktdemokratische Beteiligung, die quantitativ weit über das hinausgeht, was andere Länder kennen (Linder 1995:138). Als Konsequenzen sind damit Machtteilung (Konkordanz) und Betroffenen einbezug (im Rechtsetzungsverfahren) verbunden. Die Aufteilung der staatlichen Funktionen auf drei Ebenen (Bund, Kantone, Gemeinden) trägt wie die vorerwähnten Faktoren zur politischen Integration der Bürgerinnen und Bürger bei. Die zahlreichen Abstimmungen sowie der Einbezug aller Betroffenen in die Gestaltung neuer Massnahmen kommen den Idealen, wie sie in "stakeholder-based evaluations" (Bryk 1983) oder den "fourth generation evaluations" (Guba und Lincoln 1989) verkörpert werden, recht nahe. Insbesondere im Bereich der Bedürfnisabklärung sowie der prospektiven Wirkungsabschätzung erbringt der politische Prozess Leistungen, die anderenorts durch Evaluationen erfüllt werden (Bussmann 1996b). Besteht überhaupt noch Bedarf für Evaluationen?

Einschränkungen für Evaluationen ergeben sich in dreierlei Hinsicht: Erstens beschränken Volksabstimmungen nach dem Motto "vox populi, vox dei" das Feld des politisch Machbaren stark. Sie setzen Rahmenbedingungen, an welche sich auch Evaluationen mit Vorteil halten. Wenn Evaluationen zu Ergebnissen führen, die mit den vom Volk sanktionierten Entscheiden kollidieren, haben sie eine gute Chance, zu Kontroversen zu führen und/oder in der Schublade zu landen. Zweitens ist der Kompromiss Wesenszug schweizerischer Politik. Kompromisse lassen sich bei den betroffenen Interessengruppen nur gut verkaufen, wenn sie als Sieg dargestellt werden. Evaluationen, welche Kosten und Nutzen von Entscheidungen für die einzelnen Betroffenen transparent machen wollen, stossen deshalb oft auf Widerstand, weil Intransparenz die Kompromissbildung erleichtert. Drittens führt die geringe Grosse der Kantone und Gemeinden zu besserer Überschaubarkeit staatlichen Handelns und zu starker persönlicher Verantwortung der Magistratspersonen und der Vollzugsinstanzen. Diese schätzen intermediäre Institutionen wie Ombudsstellen oder Wissenschaftler, welche sich in ihre Geschäfte einmischen, nicht. Es herrscht politischer Voluntarismus in der Form direkter und spontaner Reaktionen auf politische Forderungen.

Trotz dieser Hindernisse hat sich der Rückgriff auf Evaluationen als unvermeidlich erwiesen. Hierzu lassen sich fünf Gründe anführen: Die gesellschaftlichen Hintergründe und Probleme wurden bereits eingehend

dargestellt. Zweitens betreffen Volksabstimmungen selektiv einzelne Sachgebiete und erlauben zudem als blosser Ja-Nein-Ergebnisse keine differenzierten Rückschlüsse über die Gründe einer Annahme oder Ablehnung einer Vorlage²⁸. Evaluationen ermöglichen viel stärker massgeschneiderte Resultate im Hinblick auf die Beurteilung staatlicher Dienstleistungen durch die Betroffenen. Drittens wird auch bei Volksabstimmung die Information des Stimmbürgers durch wirtschaftlich und politisch mächtige Verbände (z.B. Agrarlobby) stark beeinflusst, während weniger gut organisierte Interessen (z.B. Konsumenten) sich oft nur geringes Gehör verschaffen können. Evaluationen können angesichts blockierter Machtverhältnisse durchaus auch aufklärerische Funktionen wahrnehmen, auch wenn es naiv wäre anzunehmen, dass Wahrheit kurzfristig stärker als Macht sei. Viertens können Evaluationen auch jenen Gruppierungen eine Stimme geben, die im politischen Prozess schlecht vertreten sind, sei es, weil sie als Stimmbürger resigniert haben und gar nicht mehr an Abstimmungen teilnehmen, sei es, weil sie als Ausländer²⁹ gar kein Stimm- und Wahlrecht haben³⁰. Als fünfter und letzter Punkt ist zu erwähnen, dass Politiker und Vollzugsinstanzen in technisch schwierigen Fragen sowie im Fall schleichender und kaum wahrnehmbarer Veränderungen einer wissenschaftlichen Unterstützung bedürfen. Evaluationen können in diesem Sinne den Gehalt politischer Auseinandersetzungen verbessern.

Es hat sich gezeigt, dass das schweizerische Regierungssystem mit seinen Eigenheiten (Volksrechte, Konkordanzregierung) der Entwicklung von Evaluationen nicht unbedingt förderlich war. Es stellt den Bemühungen um bessere Kenntnisse über die Wirksamkeit aber auch keine unüberwindbaren Hindernisse entgegen. Nachdem der institutionelle Aufbau im Bereich von Evaluationen zwar zähflüssig, aber stetig verlaufen ist, dürfte das bisher Erreichte auch in naher Zukunft Bestand haben. Verbesserungen werden vor allem dann zu erzielen sein, wenn die vielfältigen Bemühungen zur Erfassung und Verbesserung der Wirksamkeit staatlichen Handelns in einem Netzwerk von Personen und Organisationen als gemeinsames Anliegen gesehen, wenn Erfahrungen ausgetauscht und

²⁸ Die Vox-Analysen werden deshalb als ergänzendes Instrument der Meinungsforschung verwendet.

²⁹ Der Ausländeranteil macht in der Schweiz 19 Prozent aus; bei den Erwerbstätigen beträgt er gar 29 Prozent.

³⁰ So haben im Laufe der letzten Jahrzehnte Anteile von 5 - 20 Prozent der Gesamtbevölkerung in Abstimmungen jeweils zustimmende Mehrheiten erbringen können; die übrigen 80-95 Prozent bestanden aus Ausländern, nicht Stimmberechtigten (unter 18, früher Frauen), Stimmbastinenten und Neinstimmen (Linder 1995: 92).

die daraus gezogenen Lehren in die entsprechenden Organisationen eingebracht werden können.

TEIL : CHARAKTERISTIKA, OBJEKTE UND NUTZUNGSZUSAMMENHÄNGE

Dieser Teil enthält mehrere zentrale Definitionen und soll Evaluation als Forschungsprozess bzw. Forschungskonzeption, nach möglichen Objekten und nach entsprechenden Evaluationskriterien genauer situieren. Er will die Leserinnen und Leser auf breiter Front einführen in die Evaluations-thematik und in die darin verwendete Begrifflichkeit. Dabei verfolgen wir folgende Zielsetzungen:

- Abgrenzung wissenschaftlicher Politikevaluationen von anderen evaluativen Tätigkeiten im öffentlichen und privaten Sektor durch die Wirkungsorientierung, die Betonung der Wissenschaftlichkeit und einer exakten Untersuchungsanlage bzw. berufsethischer Standards, an denen sich Evaluatorinnen und Evaluatoren zu orientieren haben (Kapitel 3).
- Darstellung der Palette unterschiedlicher Objekte, auf die sich Evaluationen beziehen. Zu diesem Zweck wird der Begriff der "öffentlichen Politik" definiert, deren Zwischen- und Endprodukte Gegenstand der in diesem Buch beschriebenen Politikevaluationen sind, und es wird eine Übersicht über die gängigen Gegenstände von Evaluationen vorgestellt; diese reichen von der Politikformulierung bis zu den Outcomes (Kapitel 4).
- Darstellung der je Politikstufe und Evaluationsgegenstand variierenden Evaluationskriterien. Diese reichen von der Konsistenz der Kausal- und Wirkungshypothesen über die Programmkohärenz, die Eignung der Vollzugsstrukturen und der Politikplanung bis hin zur Effektivität, zur Wirksamkeit und zur Wirtschaftlichkeit bzw. zu Qualitätskriterien der Evaluation selbst (Kapitel 5).
- Präsentation und Diskussion der wichtigsten Nutzungszusammenhänge, in denen Evaluationen in der Schweiz stehen können (Kapitel 6).

Die vier Kapitel von Teil enthalten damit die wesentlichen konzeptionellen Grundlagen, die bei der Durchführung von Politikevaluationen berücksichtigt werden müssen. Die Abgrenzung der Politikevaluationen von

anderen Instrumenten der Politikkontrolle bzw. Politiksteuerung wird erst in Teil III vorgenommen. Die im engeren Sinne methodisch-technischen Fragen der Durchführung von Politikevaluationen tragen wir in Teil IV vor. Diese aus pädagogischen Gründen gewählte Grundstruktur führt an der einen oder anderen Stelle zu geringfügigen Überschneidungen. Diese werden im Text durch entsprechende Verweise gekennzeichnet.

KAPITEL 3: INHALTLICHE UND METHODISCHE ANFORDERUNGEN AN WISSENSCHAFTLICHE POLITIKEVALUATIONEN

Ulrich Klöti

Evaluationen sind ein Informationsinstrument (Bussmann 1995d und 1995e). Sie stellen Informationen über staatliches Handeln bereit. Es gibt mehrere solcher Informationsinstrumente wie etwa die Staatsrechnung, Audits, Monitoring, Controlling, Leistungsindikatoren u.a.m. (vgl. auch Kapitel 2.2.3.). Wo die Grenzen des Begriffs "Evaluation" gegenüber diesen anderen Instrumenten gesetzt werden, ist umstritten. Im angelsächsischen Sprachgebiet wird der Begriff z.T. sehr breit verwendet und schliesst auch das Controlling (das unter diesem Titel im angelsächsischen Sprachraum allerdings nicht verwendet wird) und das Monitoring mit ein. Unbestritten ist jedoch, dass der Kern der Evaluationen in der Untersuchung der Wirksamkeit von Massnahmen und Programmen besteht; hier kommen denn auch die methodische Schärfe und die von der Evaluationsforschung entwickelten Standards voll zum Tragen. Dieser Kern wird von keinem anderen Informationsinstrument abgedeckt; er bildet vielmehr Voraussetzung für andere Managementkonzepte im öffentlichen Bereich (wie "wirkungsorientierte Verwaltung" oder "New Public Management"). Unter Evaluationen im engeren Sinne verstehen wir deshalb Evaluationen der Wirkungen öffentlicher Politiken ("Outcomes"). Im weiteren Sinne darunter zählen wir ferner alle weiteren Informationsinstrumente, welche die Ermittlung der Wirkungen unterstützen können (z.B. vorgelagerte Rechtmässigkeitskontrolle, prospektive Wirkungsabschätzung, Monitoring, Vollzugsuntersuchungen u.a.m.).

Das Kapitel 3 arbeitet die wesentlichen Merkmale heraus, die Evaluationen im Sinne der Autoren dieses Buches auszeichnen. Wir verstehen Evaluation als Beurteilung und Bewertung der Wirkungen staatlicher Programme und Massnahmen mit wissenschaftlichen Methoden. Diese Definition verweist auf *vier Dimensionen*, anhand derer Evaluationen von anderen gesellschaftspolitischen Tätigkeiten abgegrenzt und die Anforderungen an ideale Politikevaluationen umschrieben werden:

- *Staatliches Handeln:* Politikevaluationen befassen sich mit staatlichem Handeln; dabei stehen gesetzgeberische Programme und Massnahmen zur Lösung gesellschaftlicher, wirtschaftlicher und kultureller Probleme im Vordergrund.
- *Systematische Analyse und Bewertung:* Das staatliche Handeln und seine Wirkungen werden einer systematischen Analyse und Bewertung unterzogen, mit dem Ziel, bessere Problemlösungen zu erreichen bzw. schlechtere Lösungen zu eliminieren.
- *Wissenschaftliche Methoden:* Die Beurteilung staatlichen Handelns stützt sich auf Informationen und Erkenntnisse, die mit einem wissenschaftlichen Instrumentarium bereitgestellt werden, die einer Nachprüfbarkeit standhalten.
- *Qualitätsstandards:* Es gibt wissenschaftliche Kriterien und professionelle Standards, an denen die Durchführung von Evaluationen gemessen wird.

Im folgenden gehen wir auf diese vier Dimensionen näher ein. Dabei wollen wir zuerst aufzeigen, wie vielfältig die mit Evaluation verbundenen Fragestellungen sein können. Daraus leiten sich die Schwierigkeiten und Probleme der Evaluationsforschung ab. Schliesslich wird gezeigt, welches die Anforderungen an Evaluationen und Evaluatoreninnen und Evaluatoren sind.

3.1. STAATLICHES HANDELN

Evaluationen gibt es überall. So werden in der Schule Noten erteilt, oder im Berufsleben Mitarbeiterinnen und Mitarbeiter beurteilt. Wirtschaftliche Güter müssen eine Evaluation durch die Akteure am Markt bestehen. Im Sport werden Leistungen bewertet. Selbst in Kunst und Kultur scheuen sich Fachleute und Publikum nicht, Werke und Aufführungen einer oft strengen Kritik zu unterziehen.

Von all diesen Formen von Evaluation ist in diesem Buch nicht die Rede. Denn hier geht es einzig um die Evaluation *staatlicher Tätigkeiten*. Diese reichen allerdings in die meisten Lebensbereiche hinein (Ellwein

1976; Rose 1984; Schmidt 1988; vgl. unten Kapitel 4.2.1. zu den öffentlichen Politiken). Der Staat führt Schulen, regelt sein Verhältnis zu Kirche und Religion und ordnet mit dem Ehe- und Erbrecht die Familie. Am augenfälligsten und am stärksten umstritten sind seine Eingriffe in das Wirtschaftsleben. Der Staat führt dazu nicht nur eine allgemeine Konjunktur-, Wettbewerbs- und Strukturpolitik. Er greift in das Wohnungswesen, das Gesundheitswesen, die Medien, die Energiewirtschaft oder in die Landwirtschaft ein. Zudem ist er bestrebt, mit der Steuerpolitik, der Sozialpolitik und der Regionalpolitik Umverteilungen vorzunehmen und mit Hilfe der Umwelt-, Raumordnungs-, Ver- und Entsorgungspolitik die natürlichen Lebensbedingungen zu schützen und die Ressourcen zu schonen. Schliesslich bietet er selbst Dienstleistungen an, wie etwa bei der Ver- und Entsorgung sowie beim Verkehr.

Schon allein wegen der Vielfalt solcher Tätigkeit ist die Evaluation staatlicher Programme und Massnahmen eine anspruchsvolle Aufgabe; denn sie setzt sie vertiefte Kenntnis des Politikbereichs voraus, in dem evaluiert werden soll. In verschiedenen Politikbereichen werden zwar Spezialisten ausgebildet. Ein besonderes Problem der Evaluation besteht aber darin, dass oft die besten Kenner eines Bereiches die diesbezügliche Politik selbst mitgestalten und damit weder über die nötigen zeitlichen Kapazitäten verfügen noch besonders motiviert sind bzw. über die nötige innere Distanz verfügen, um sich an kritischen Evaluationen zu beteiligen. Umgekehrt fehlt es den Spezialisten der Evaluation oft an detaillierten Kenntnissen der einzelnen Politikbereiche.

Das Problem wird dadurch erschwert, dass der Staat verschiedene Formen der Intervention einsetzt (Lowi 1972). Die klassische Weise des Staates, aktiv zu werden, besteht darin, in konditionalen Gesetzen nach der Formel "Wenn-Dann" Gebote und Verbote aufzustellen. Wer diese nicht befolgt, hat mit Sanktionen zu rechnen. Diese - klassische - Interventionsform ist auch heute noch von grosser Bedeutung. Beispiele sind die zahlreichen polizeirechtlichen Genehmigungsvorbehalte (Umwelt-, Arbeits-, Wirtschaftsverwaltungsrecht) oder das allgemeine oder besondere Strafrecht. So bestimmt etwa das Betäubungsmittelgesetz, dass Personen, die unbefugt mit Drogen handeln, mit einer Strafe zu rechnen haben.

Schon früh begann der Staat, selbst *Leistungen zu erbringen*. Diese Interventionsform umfasst etwa die Bereitstellung von Infrastrukturanlagen, die grundsätzlich allen zur Verfügung steht. Bestes Beispiel dafür ist wohl der Strassenbau. Der Staat beschränkt sich dabei nicht nur auf grosse

Investitionen. Längst ist er dazu übergegangen, auch Güter und Dienstleistungen für den Kleinkonsum anzubieten. In der Schweiz waren während Jahrzehnten die PTT und die SBB die bekanntesten Staatsbetriebe. In anderen westlichen Industrieländern sind zum Teil deutlich mehr Industrien, aber auch in grösserem Masse als bei uns Banken und Versicherungen verstaatlicht worden. In einem weiteren Sinne ist auch der grösste Teil des Schulwesens als staatliches Dienstleistungsangebot zu verstehen. Zudem sind die meisten Spitäler staatlich geführt.

In jüngerer Zeit versucht der Staat in grossem Stil, auch *milfinanziellen Anreizen* zu steuern. Von positiven Anreizen spricht man dann, wenn ein erwünschtes Verhalten eine finanzielle Begünstigung zur Folge **hat**. Klassisches Beispiel dafür ist der Versicherungsbonus, der nach einer längeren schadenfreien Zeit geltend gemacht werden kann und der in einer tieferen Prämie besteht. Von negativen Anreizen spricht man dann, wenn unerwünschtes Verhalten zu einer finanziellen Einbusse führt. In einem weiteren Sinne sind sämtliche Bussen unter diese Interventionsform zu subsumieren. In einem engeren Sinne stehen Preisdifferenzierungen im Vordergrund, die von einem unerwünschten Verhalten abhalten sollen. So wird man etwa die Gebühr für die Benützung einer Strasse oder eines Verkehrsmittels zu bestimmten Zeiten erhöhen, um zu verhindern, dass die Strassen bzw. Verkehrsmittel während der Stosszeiten überlastet sind. Die gleichzeitige Verbilligung von bleifreiem und Verteuerung des bleihaltigen Benzins stellt eine umweltpolitisch motivierte Kombination von positivem und negativem Anreiz dar.

Schon immer hat der Staat auch eine eigentliche *Umverteilungspolitik* betrieben. Im Rahmen einer (progressiven) Besteuerung zieht er Mittel von begüterten Subjekten ab und verteilt diese im Rahmen einer (redistributiven) Sozialpolitik an bedürftige Subjekte. Die Umverteilung ist auch zwischen unterschiedlich entwickelten und unterschiedlich finanzstarken Regionen möglich.

Im weiteren ist der Staat dazu übergegangen, auch *Überzeugungspolitiken* zu entwickeln. In diesem Falle versuchen staatliche Akteure, durch Information und Überzeugung die Politikadressaten zu einem bestimmten Verhalten zu überzeugen oder davon abzuhalten. Die Aids-, Anti-Raucher- oder Luftreinhaltekampagnen sind neuere Beispiele für diese Form staatlichen Handelns. Älter sind Beratungsstellen verschiedener Art (etwa in der Landwirtschaft).

In jüngster Zeit zeichnet sich schliesslich ein weiterer Typ staatlichen Handelns ab, der in der *gesellschaftlichen Strukturbildung* besteht: Die

staatlichen Behörden bauen dabei Netzwerkstrukturen (vgl. dazu unten Kapitel 8.2.) aus staatlichen und gesellschaftlichen Akteuren auf, denen die Produktion bestimmter Politikleistungen übertragen wird. Beispiele dazu finden sich in der Sozial-, der Gesundheits- oder in der Umweltpolitik.

Schliesslich legt der Staat in *Institutionen und Organisationen* die Rahmenbedingungen für das soziale Handeln fest. Direkte Demokratie, Föderalismus, siebenköpfige Kollegialregierung und Aufteilung der Bundesverwaltung in sieben Departemente und über 70 Bundesämter sind die bekanntesten institutionell-organisatorischen Grundvoraussetzungen für staatliche Interventionen auf Bundesebene in der Schweiz.

Diese Vielfalt der staatlichen Interventionsformen stellt die Evaluationsforschung deshalb vor zusätzliche Probleme, weil die Beurteilung der Wirkung der verschiedenen Instrumente unterschiedliche Kompetenzen erfordert. Die klassische Gebots- und Verbotsgesetzgebung gilt als die eigentliche Domäne der Juristen. Betriebswirtschaftler sind am ehesten in der Lage zu beurteilen, ob ein Gut oder eine Dienstleistung effizient produziert wird. Umverteilungsprogramme können am besten von Ökonomen evaluiert werden. Zur Analyse der Wirkung von Anreizen und vor allem von Überzeugungsstrategien sind Sozialwissenschaftler beizuziehen. Schliesslich haben immer mehr Massnahmen auch eine technische und naturwissenschaftliche Dimension. Es genügt somit nicht, über vertiefte Sachkenntnisse in einem Politikbereich zu verfügen; beim vielfältigen Instrumentarium des Staates müssen zudem immer mehr auch Methoden und Theorien aus verschiedenen wissenschaftlichen Disziplinen eingesetzt werden.

Wie sehr die Anforderungen kumulieren können, zeigt sich mit aller Deutlichkeit, wenn man in Rechnung stellt, dass es bei der Evaluation häufig nicht allein um die Beurteilung einzelner isolierter Massnahmen, sondern um die Bewertung ganzer *Programme* geht. Es ist zwar nicht immer einfach, Massnahmen und Programme auseinanderzuhalten; häufig werden die beiden Begriffe synonym verwendet. Grundsätzlich lassen sie sich aber doch gut unterscheiden. Bei einer Massnahme handelt es sich um einen einzelnen Eingriff des Staates. Programme dagegen sind umfassender und beinhalten in der Regel mehrere Einzelmassnahmen. Leicht veranschaulichen lässt sich die Unterscheidung anhand der Luftreinhaltepolitik. Dort sind die Kantone aufgrund der Verordnung des Bundes gehalten, Massnahmenpläne zu erarbeiten. Diese Pläne können als Programme oder als Aktionspläne (Kapitel 5.4.) verstanden werden. Sie um-

fassen eine immense Vielzahl von Einzelmassnahmen (Knoepfel, Imhof und Zimmermann 1995: 12 ff.).

An diesem Beispiel lassen sich nun die *kumulierten Schwierigkeiten* der Evaluation von Programmen leicht verdeutlichen. Wer den Auftrag erhält, den Erfolg der Luftreinhaltepolitik des Bundes zu untersuchen, muss sich erstens mit verschiedenen Politikbereichen auseinandersetzen, enthalten die Programme doch Massnahmen, die den Verkehr, die Energie, die Industrie (Normierungen), die Raumplanung, das Wohnungswesen und allenfalls die Information und sogar die Schule tangieren. Wer Luftreinhalteprogramme evaluiert, ist zweitens mit dem Umstand konfrontiert, dass darunter Massnahmen aus dem ganzen Instrumentarium des Staates vorkommen. Es sind Ge- und Verbote, z.B. Geschwindigkeitsbeschränkungen, vorgesehen. Der Staat investiert in die Infrastruktur des öffentlichen Verkehrs, um eine Umlagerung von der Strasse auf die Schiene zu erreichen, und baut deshalb etwa eine S-Bahn oder die NEAT. Als Anreiz ist beispielsweise die Preisdifferenzierung zwischen bleihaltigem und bleifreiem Benzin zu verstehen. Weiter in dieser Richtung ginge die noch umstrittene CO₂-Abgabe. Diese hätte gleichzeitig auch Umverteilungswirkungen. Schliesslich haben staatliche Stellen grössere Informationskampagnen zum Energiesparen lanciert und damit auch Überzeugungsstrategien eingesetzt.

Zusammenfassend ist festzuhalten, dass die Evaluation staatlicher Massnahmen und, a fortiori, staatlicher Programme einen guten Überblick über die angesprochenen Politikbereiche und eine Vertrautheit mit den Mechanismen der eingesetzten Instrumente staatlicher Interventionen voraussetzt. Dazu ist es unumgänglich, dass Kenntnisse aus verschiedenen wissenschaftlichen Disziplinen für eine Evaluation zur Verfügung stehen. Evaluationen sind deshalb schon fast definitionsgemäss multidisziplinäre Forschungsvorhaben. Einzelne Forscherinnen und Forscher sind immer weniger in der Lage, im Alleingang Evaluationen durchzuführen, selbst wenn sie über eine beträchtliche Erfahrung verfügen. Gefragt ist immer mehr Teamarbeit.

3.2. BEWERTEN VON WIRKUNGEN

3.2.1. Explizite Kriterien

Evaluation hat schon nach dem ursprünglichen Wortsinn mit Werten und Bewerten zu tun. Die frühe Evaluationsforschung schenkte diesem Aspekt allerdings wenig Beachtung, denn sie glaubte, durchaus wertfrei arbeiten zu können (vgl. dazu Kapitel 12.1.). Aus zwei Gründen konnte sich diese Position allerdings nicht halten. Einmal zeigte die Forschungserfahrung, dass in einem hochpolitischen Umfeld immer wieder Entscheidungen der Evaluatoren nötig wurden, die auf bestimmte Wertungen abgestützt werden mussten. Es wurde ausserdem offensichtlich, dass sich die Beurteilung einer Politik auf *Kriterien* berufen muss. Dabei ist es grundsätzlich unerheblich, ob dem "*präskriptiven*" Ansatz (Scriven 1980) gefolgt wird, der zuerst ein normatives Wertesystem entwickelt und daraus anschliessend ein Kriterienraster ableitet, oder ob ein "*deskriptiver*" Ansatz gewählt wird, der sich auf die Werte und Kriterien der an einem staatlichen Programm Beteiligten (Gesetzgeber, Vollzugsverwaltung, Adressaten und Interessengruppen) stützt. Entscheidend für die Evaluation ist, dass die (in Kapitel 5 weiter erläuterten) Bezugskriterien explizit gemacht werden und dass sie wenn möglich öffentlich, zumindest aber zwischen den Evaluatoren und den Auftraggebern diskutiert werden. Beurteilungen von staatlichem Handeln aufgrund von diffusen, nicht einsichtigen und nicht nachvollziehbaren Wertvorstellungen der Forscherinnen und Forscher genügen den Ansprüchen an eine Evaluation nicht.

3.2.2. Ausrichten auf Wirkungen

Wenn Evaluation aufgrund expliziter Kriterien öffentlicher Politiken beurteilt, dann lässt sie sich in die Gruppe der Kontrollinstrumente einordnen, mit denen seit jeher das Handeln der öffentlichen Hand überwacht wurde. Evaluation ist deshalb von der Geschäftsprüfung, der Vollzugs- und Finanzkontrolle, vom Monitoring, Controlling und Auditing¹ abzugrenzen. Dies kann nicht mit trennscharfen Definitionen geschehen. Überschneidungen sind unvermeidlich. Hingegen ist es möglich, die Schwergewichte der Evaluationsforschung hervorzuheben. Wir sind uns

¹ Diese Instrumente werden anderweitig erläutert. Vgl. Kapitel 2.2.3. und 8.

dabei bewusst, dass nicht alle Evaluatorinnen und Evaluatoren die damit verbundenen Orientierungen teilen.

Evaluation hat immer mit der Erfassung von Wirkungen staatlichen Handelns zu tun. Es geht nicht nur darum zu beurteilen, was der Staat tut und wie er es tut. Es geht nicht um das Zustandekommen von öffentlichen Politiken, sondern um deren Wirkungen. Gewiss ist es zum besseren Verständnis eines Politikprogrammes wichtig, auch seine Entstehung und den dazugehörigen Entscheidungsprozess zu untersuchen. Eine solche Notwendigkeit liegt schon deshalb vor, weil aus dem Programm selbst nicht immer klar hervorgeht, welches seine Ziele sind. Immer bleibt aber die Analyse des Politikformulierungsprozesses für die Evaluation nur Mittel; Zweck ist die Beurteilung der Wirkung der Vorkehrungen dieser Politik in der gesellschaftlichen Realität.

Das Maximalprogramm einer Evaluation würde die Beurteilung sämtlicher Wirkungen einer öffentlichen Politiken beinhalten. In der Praxis wird es kaum je möglich sein, eine derart vollständige Evaluation zu verwirklichen. Drei Gründe sprechen in der Regel dagegen. Erstens fehlt es zu meist an Mitteln, um derart umfangreiche empirische Forschungen zu realisieren. Darüber hinaus ist den Auftraggebern kaum je an einer derart umfassenden Evaluation gelegen. Sie sind viel eher an der Abklärung bestimmter Aspekte interessiert. Die freie Forschungsförderung ihrerseits stellt nur beschränkte Ressourcen für die angewandte Forschung zur Verfügung. Unter diesen Umständen müssen sich die meisten Evaluationsteams auf spezifische Fragestellungen beschränken.

Zweitens fehlt es häufig an brauchbaren Daten. Das gilt vor allem für neuere Politiken. Statistiken werden meist erst mit dem Inkrafttreten eines Gesetzes, in dem die Massnahmen verankert sind, erstellt. Will man indessen die Wirkung der Policy beurteilen, dann muss man in der Regel die Situation nach dem Beginn des Wirkens einer Massnahme mit derjenigen vergleichen, die vorher bestand. Rückwirkend werden aber die erforderlichen Daten nur selten erhoben; ein Vorher/Nachher-Vergleich (vgl. Kapitel 5) ist damit oft ausgeschlossen. Aber auch in anderen Fällen kann es an Daten fehlen. Die Schweiz gilt in mancher Beziehung als statistisches Entwicklungsland. Dieser Umstand schränkt die Möglichkeiten der Evaluation drastisch ein.

Drittens schliesslich sind nicht immer alle nötigen Informationen frei zugänglich. Personenbezogene Daten sind oft zu Recht geschützt. Andere Informationen werden vertraulich behandelt, weil sie heikel sind und die Evaluierten in Schwierigkeiten bringen könnten.

Wenn schon ein umfassendes Evaluationsprogramm in der Regel nicht zu verwirklichen ist, dann wird es umso wichtiger, die getroffene Auswahl der evaluierten Aspekte der Massnahme zu begründen. Es gehört zu den Grundanforderungen an eine seriöse Evaluation, dass genau angegeben wird, welche Wirkungen erfasst und beurteilt werden sollen, auf welche möglichen anderen Wirkungen nicht eingegangen wird und warum diese Entscheidungen getroffen werden.

Mit der generellen Wirkungsorientierung ist nicht zwingend ein *summativer Evaluationsansatz* verbunden, der einzig daran interessiert ist, ob mit einem Programm die gesteckten Ziele erreicht worden sind. Solche summative Evaluationen ("Bilanzevaluationen") neigen bei der Feststellung, dass die Ziele verfehlt worden sind, zur Empfehlung, das ganze Programm zu beenden. Auch sog. *formative Evaluationsansätze* ("Verbesserungsevaluationen") sind der Wirkungsorientierung verpflichtet. Diese beziehen die verschiedenen Zwischenschritte vom Programm bis zur Wirkung in die Untersuchung ein und beurteilen diese laufend. Solche formativen (und begleitenden) Analysen erfolgen immer mit der Absicht, das Programm insgesamt zu verbessern. "The most important purpose of evaluation is not to prove but to improve" (Stufflebeam und Shinkfield 1985: 191).

Die Wirkungsorientierung der Evaluation ist nicht unbedingt an einen *zielorientierten Ansatz* gebunden, wie er in den Anfängen der Evaluationsforschung gang und gäbe war. Damals überprüfte man Wirkungen fast ausschliesslich auf ihre Übereinstimmung mit den expliziten Programmzielen. Dieses Vorgehen erwies sich auf die Dauer als nicht gangbar. Denn solche Ziele waren und sind oft nur vage formuliert, umstritten oder latent, weil sie von den verschiedenen Akteuren etwa im Parlament und bei der Regierung oder auf Bundes- und Gliedstaatenbene unterschiedlich interpretiert werden und weil sie sich zudem über die Zeit hinweg verändern können, ohne dass dies in expliziten Programmänderungen sichtbar würde. Darauf gehen wir in Kapitel 5 näher ein (Programmanalyse). Eine ausschliessliche Konzentration auf die (explizit postulierten) Programmziele verschliesst zudem den Blick auf nicht (ausdrücklich oder implizit) intendierte Wirkungen, die bekanntlich wichtiger sein können als die geplanten. Aus diesen Gründen wurde in der Evaluationsforschung der sogenannte *"goal-free"-Ansatz* (Scriven 1973; 1980; 1991: 180) entwickelt. Dieser versucht auch die nicht beabsichtigten Wirkungen erwünschter wie unerwünschter Art in die Analyse einzubeziehen. Auch dieser nicht allein

an den expliziten Programmzielen orientierte Ansatz konzentriert sich freilich auf die Ermittlung von Wirkungen.

Mit der Wirkungsorientierung von Evaluationsforschung ist mithin nicht zwingend eine *Wirksamkeitsanalyse* gemeint. Letztere vergleicht die dem Politikprogramm zugrunde liegenden Ziele mit dem tatsächlich Erreichten. Soll etwa mit einer Geschwindigkeitslimite auf Autobahnen die Luftqualität verbessert werden, so wird in einem Soll-Ist-Vergleich überprüft, ob nach Einführung einer neuen Temporeduktion die Luft qualitativ tatsächlich im erwünschten Ausmass besser geworden ist. Damit wird in der Tat nach der Wirksamkeit der Massnahme gefragt. Aber Evaluationen können auch *Wirtschaftlichkeitsanalysen* beinhalten. Diese vergleichen die Wirkungen eines Programms nicht mit den Zielen, sondern mit den eingesetzten Mitteln und Ressourcen. Die Frage lautet im obigen Beispiel, was die Einführung der Geschwindigkeitslimite, die Überwachung ihrer Einhaltung und die Sanktionierung der Verstösse an Material und Polizeieinsatzstunden, aber auch an Zeitverlusten und allenfalls Ärger bei den Adressaten kosten. Ein solcher Vergleich von Kosten und Nutzen oder von Aufwand und Ertrag dient allein der Bewertung der Effizienz der getroffenen Massnahmen.

Zusammenfassend lässt sich festhalten, dass Evaluationsforschung bezweckt, anhand der in Kapitel 5 näher erläuterten Kriterien und Standards die Wirkungen staatlicher Politiken zu analysieren, zu bewerten und Vorschläge für ihre Verbesserung zu entwickeln. Im Vordergrund des Interesses standen ursprünglich summative, an den Programmzielen orientierte Wirksamkeitsanalysen. In der aktuellen Forschungspraxis sind formative, an Zwischenergebnissen und Nebenwirkungen interessierte Analysen, die sich auch mit der Wirtschaftlichkeit auseinandersetzen, ebenso häufig. Analysen, die allein die Politikformulierung, oder die Entscheidungsprozesse im Vollzug ins Blickfeld nehmen und deren Teil- oder Endergebnisse nicht mitberücksichtigen, gehören nicht zur Evaluationsforschung, so wichtig sie sonst auch sind.

3.3. WISSENSCHAFTLICHE UNTERSUCHUNGSANLAGE

Ein drittes Merkmal der Evaluation besteht darin, dass sie Informationen über staatliches Handeln und über seine Wirkungen mit Hilfe eines wissenschaftlichen Instrumentariums bereitstellt. Die damit geforderte Wissenschaftlichkeit des Vorgehens wird durch zwei Kriterien bestimmt: Er-

stens hat eine Evaluation einer klar umschriebenen, die Empirie anleitenden und steuernden Untersuchungskonzeption mit klar formulierten Fragestellungen zu folgen. Zweitens sind die für die Evaluation erforderlichen Informationen mit einer einsichtigen und nachvollziehbaren Methode zu beschaffen, auszuwerten und zu interpretieren. Zur Frage, welche Verfahren und Techniken dabei zur Anwendung gelangen sollen, gibt Teil IV dieses Buches (Kapitel 11 und 12) Auskunft. In diesem Kapitel befassen wir uns nur mit konzeptionellen Aspekten, die im Rahmen der Erstellung der sog. Untersuchungsdesigns zu beachten sind.

Wie erläutert, befassen sich Evaluationen mit den Wirkungen von staatlichen Massnahmen. Übersetzt in die Forschungslogik heisst dies, dass untersucht wird, ob und in welchem Ausmass räumlich und zeitlich identifizierbare *Veränderungen in der gesellschaftlichen Wirklichkeit* auftreten, die sich möglicherweise auf eine öffentliche Politik zurückführen lassen. Im Sprachgebrauch des empirisch-analytischen Forschungsparadigmas, dem wir hier zunächst folgen, werden diese möglichen Veränderungen als *zu erklärende (abhängige) Variablen* bezeichnet. Es stellt sich somit als erstes die Frage, inwiefern sich in der gesellschaftlichen Wirklichkeit Veränderungen eingestellt haben oder - technisch formuliert - ob bei der abhängigen Variablen Varianz festzustellen ist.

→ Y = Programmwirkungen als abhängige (zu erklärende) Variable

Konkret lauten die entsprechenden Fragen etwa: Ist die Qualität der Luft in einer bestimmten Periode und an bestimmten Orten besser geworden? Sind die Lohndifferenzen zwischen Männern und Frauen beseitigt? Hat die Zahl der eingereichten Volksinitiativen abgenommen?

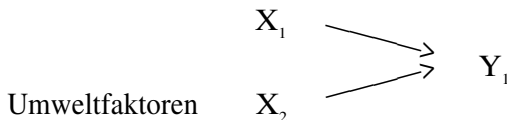
In einem zweiten Schritt ist zu fragen, welches die möglichen *Ursachen für die allfällige Veränderung* in der gesellschaftlichen Wirklichkeit sind. Die erste Vermutung (= Hypothese) besteht in der Regel darin, dass dies etwas mit dem zu evaluierenden staatlichen Programm zu tun hat. Als unabhängige (erklärende) Variable kommen deshalb zunächst einmal das staatliche Programm und darauf abgestützte Umsetzungsmassnahmen in Betracht.

Programm und darauf abgestützte Umsetzungsmassnahmen als unabhängige (erklärende) Variable:

$$X_1 \rightarrow Y$$

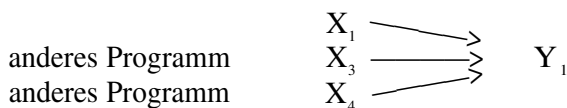
Die entsprechenden konkreten Fragen lauten etwa: Ist aufgrund der in einem Erlass eingeführten Geschwindigkeitslimite die Luft besser geworden? Hat das Gleichstellungsgesetz zu einer Verminderung der Lohndifferenzen zwischen Frauen und Männern geführt? Geht wegen einer Erhöhung der Zahl der erforderlichen Unterschriften die Häufigkeit von Volksinitiativen zurück?

Das grösste Problem der Evaluationsforschung besteht wohl darin, dass nicht immer klar festgestellt werden kann, ob eine Veränderung der gesellschaftlichen Wirklichkeit dem zu evaluierenden Programm und darauf abgestimmten Massnahmen zugeordnet werden kann, oder ob dafür andere Faktoren verantwortlich sind. Als unabhängige Variablen kommen nämlich durchaus auch "externe" *Umweltfaktoren* politischer, sozialer ökonomischer, kultureller oder struktureller Art in Frage. Die zweite Vermutung (= Hypothese) besteht daher darin, dass die gesellschaftlichen Veränderungen zu einem bestimmten Ausmass auf solche "externe" Umweltfaktoren zurückzuführen sind.



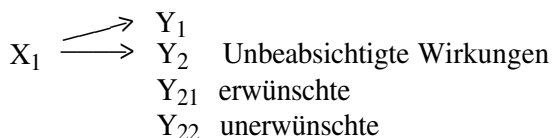
Die konkreten Fragen können in diesem Zusammenhang etwa wie folgt lauten: Ist die Qualität der Luft an bestimmten Orten aus klimatischen Gründen (strukturell) gestiegen? Haben sich Lohndifferenzen zwischen Mann und Frau wegen einer stärkeren Anspannung auf dem Arbeitsmarkt in einer konjunkturellen Boom-Phase (ökonomisch) verringert? Ist die geringere Häufigkeit von Volksinitiativen auf eine abnehmende Konfliktivität der Politik oder auf eine bessere Integration von Oppositionsgruppen (politisch) zurückzuführen?

Das bis jetzt entwickelte einfache Modell entspricht noch in keiner Weise den komplexen Zusammenhängen, die in der Gesellschaft zu berücksichtigen sind. Komplizierend kommt einmal hinzu, dass auf eine gesellschaftliche Wirklichkeit gleichzeitig *mehrere staatliche Programme* und entsprechende Massnahmen einwirken. Als dritte Hypothese ist daher zu überprüfen, ob eine Veränderung in der Zielgrösse (abhängige Variable) auf das zu evaluierende oder auf ein anderes Programm zurückzuführen ist.



Dazu wären etwa die folgenden konkreten Forschungsfragen zu beantworten: Ist die Luftqualität wegen einer Intensivierung der Kontrollen bei festen Anlagen (Heizungen) und nicht etwa wegen der Tempolimits gestiegen? Sind die Lohndifferenzen zwischen den Geschlechtern wegen bildungspolitischer Massnahmen zugunsten der Frauen abgebaut worden? Sind weniger Verfassungsinitiativen eingereicht worden, weil neu das Instrument der Gesetzesinitiative eingeführt wurde?

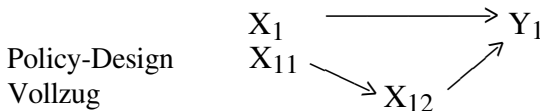
Der nicht ausschliesslich auf die expliziten Programmziele orientierte "goal-free"-Ansatz interessiert sich, wie gezeigt, für verschiedene mögliche Wirkungen staatlicher Aktivitäten. In der Forschungslogik, der bisher gefolgt wurde, heisst dies, dass einer unabhängigen Programm-Variablen mehrere abhängige Folge-Variablen zugeordnet werden. Wirkungen können nicht nur auf der Zieldimension eintreten; es sind unter Umständen auch unbeabsichtigte Wirkungen verschiedener Art festzustellen. Dabei können sich diese *unbeabsichtigten Wirkungen* als erwünscht oder als unerwünscht herausstellen. Als vierte Hypothese ist daher zu überprüfen, inwieweit (erwünschte oder unerwünschte) vom Programm nicht intendierte Wirkungen eingetreten sind.



Die hinter diesen Vermutungen stehenden Forschungsfragen sind etwa wie folgt zu formulieren: Führt die Einführung von Geschwindigkeitsbeschränkungen auf Autobahnen zu einer erwünschten Reduktion der Zahl und Schwere von Unfällen oder zu einer unerwünschten Rückkehr des Verkehrs auf Nebenstrassen und Schleichwege? Hat ein Gleichstellungsprogramm die erwünschte zusätzliche Konsequenz, dass Frauen vermehrt Karrieren in "Männerdomänen" wagen, oder sind auch unerwünschte Wirkungen für Frauen selbst etwa bezüglich ihrer Lebensqualität, ihrer Gesundheit und ihrer subjektiven Zufriedenheit in Rechnung zu stellen? Hat die Erhöhung der Unterschriftenzahl zur erwünschten Folge, dass das Parlament als Entscheidungsinstanz relativ stärker wird, oder ist uner-

wünschterweise damit zu rechnen, dass das Vertrauen in die Regierung weiter sinkt?

Formative Evaluationsansätze können sich nicht damit begnügen, den Zusammenhang zwischen Programm und Zielerreichung zu ermitteln. Sind die Ziele nicht zur Zufriedenheit der Programmgestalter verwirklicht worden, so ist zu prüfen, woran dies liegt. Die unabhängige Programmvariable ist deshalb aufzuschlüsseln. Dabei werden in der Regel zwei Teilvariablen unterschieden. Werden Ziele nicht erreicht, so kann dies einerseits an der *Ausgestaltung der Politik*, d.h. am Policy-Design (vgl. dazu Linder und Peters 1989a) liegen. Andererseits ist denkbar, dass das Scheitern eines Programms auf *ungenügende Umsetzungsmassnahmen* zurückzuführen ist. Dabei sind auch indirekte Wirkungen zu beachten. So ist es denkbar, dass das Design selbst bereits so gestaltet ist, dass ein sinnvoller Vollzug verunmöglicht oder doch erschwert wird (programmierte Vollzugsdefizite)². Als fünfte Hypothese ist daher der vermutete Anteil von Veränderungen des Verwaltungsprogrammes (X_{11}) und jenes spezifischer Umsetzungsmassnahmen (X_{12}) empirisch zu überprüfen.



Die im einzelnen zu überprüfenden Fragestellungen würden etwa wie folgt lauten: Ist die Luftqualität nur marginal verbessert worden, weil die Reduktion der zulässigen Geschwindigkeit zu wenig radikal ausfiel oder weil ihre Einhaltung gar nicht kontrolliert und durchgesetzt wurde? Ist Lohngleichheit nicht erreicht worden, weil keine griffigen Massnahmen (etwa Quoten) eingesetzt wurden oder weil die Hürden bei den Rechtsmittelverfahren zu hoch sind? Ist die Zahl der gültigen Volksinitiativen nicht zurückgegangen, weil die Erhöhung der Zahl der erforderlichen Unterschriften durch eine Intensivierung der Sammeltätigkeit leicht wettgemacht werden konnte oder weil die Vollzugsbehörden die eingegangenen Unterschriften speditiver überprüften und validierten?

In einem weiteren Schritt ist zu bedenken, dass Wirkungen nicht immer einstufig vom Programm zur Zieldimension zu ermitteln sind. In der Regel ist mit *mehrstufigen staatlichen und gesellschaftlichen Wirkungsketten* zu rechnen, die etwa in der Raumplanung über kantonale Richtpläne und Konzepte bzw. kommunale Pläne und Baugenehmigungen (staatliche Wir-

² Vgl. unten Kapitel 5.

kungsketten) oder in der Luftreinhaltung über technologische Innovationen (gesellschaftliche Wirkungsketten) führen. Das Untersuchungsmodell wird damit weiter kompliziert. Zwischen die unabhängige und die abhängige Variable sind Zwischenstufen zu setzen, für die bestimmte Hypothesen zu formulieren sind.

$$\text{Programm } X_1 \rightarrow M_1 \rightarrow U_1 \rightarrow Y_1 \text{ Wirkung}$$

(Stufen der Wirkungskette)

So wird die Luftqualität nur besser, wenn erstens die Autos tatsächlich langsamer fahren und zudem langsamer fahrende Autos auch weniger Schadstoffe emittieren. So tragen etwa Führungskurse für Frauen nur dann zur Gleichstellung bei, wenn die Frauen vom Angebot Gebrauch machen, wenn das im Führungskurs Erlernte zur besseren Ausübung von Führungsfunktionen beiträgt und wenn schliesslich Auswahlgremien für Führungskräfte das Absolvieren von Führungskursen bei Personalentscheiden auch berücksichtigen. So trägt schliesslich eine Erhöhung der Unterschriftenzahl nur dann zu einer Eindämmung der "Initiativenflut" bei, wenn mit dieser Massnahme nicht gleichzeitig die Neigung der Bürgerinnen und Bürger erhöht wird, neue Volksbegehren zu unterschreiben.

Angesichts der teilweise langen Wirkungsketten ist im weiteren daran zu erinnern, dass der *Zeitfaktor* in der Evaluation eine wichtige Rolle spielt. Es ist denkbar, dass kurzfristig keine Wirkungen einer Massnahme zu verzeichnen sind, langfristig aber um so nachhaltigere Konsequenzen eintreten. Ein solcher Wirkungsverlauf ist vielleicht bei einem Gleichstellungsprogramm zu erwarten. Es dürfte längere Zeit dauern, bis sich verfestigte Werthaltungen ändern. Umgekehrt ist die Situation vorstellbar, bei der ein Programm kurzfristig wirksam ist, langfristig die Wirkungen aber durch kompensatorisches Verhalten wieder aufgehoben werden. Einen solchen Ablauf könnte man sich bei der Erhöhung der Unterschriftenzahlen vorstellen. Zunächst wirkt die Massnahme durchaus in der erwarteten Richtung. Nach einigen Jahren stellen sich indessen potentielle Initianten auf die neue Situation ein und entwickeln Strategien, die es ihnen erlauben, im früheren Umfang Initiativen einzureichen. Aus diesen Beispielen ist abzuleiten, dass Evaluationen immer auch dynamische Aspekte zu berücksichtigen haben und dass die *langfristigen Effekte* nicht schon wenige Monate nach Inkrafttreten eines neuen Programms oder einer Massnahme beurteilt werden können.

Die bisherigen Ausführungen zur Forschungskonzeption könnten den Eindruck erwecken, einer Evaluation werde nur dann Wissenschaftlichkeit

attestiert, wenn sie alle bisher aufgeführten Forschungsfragen beantworte, die damit verbundenen Hypothesen erhärte oder verwurfe und somit ein umfassendes Wirkungsmodell überprüfe. Dieser Eindruck ist in doppelter Hinsicht zu korrigieren.

Erstens ist das dem *empirisch-analytischen Forschungsparadigma* entnommene Denken in Hypothesen und ihrer Überprüfung nicht der einzige mögliche evaluative Ansatz. Zwar hat diese wissenschaftstheoretische Richtung die Evaluationsforschung bisher dominiert. Angesichts der zunehmenden sozialen Komplexität und mit Blick auf das wenig gefestigte Wissen der Policy-Forschung sowie den bescheidenen Stand der sozialwissenschaftlichen Theoriebildung ist es aber in vielen Fällen gar nicht möglich, auf Anhieb zu allen denkbaren Wirkungszusammenhängen präzise Fragestellungen und operationalisierbare und überprüfbare Hypothesen zu entwickeln. Hier drängt sich ein alternatives Vorgehen auf (vgl. dazu Kapitel 12.1.), das auf eine *rekonstruktivistische Wissenschaftstheorie* und auf hermeneutische Verfahren zurückgreift. Diese Vorgehensweise besteht darin, dass zunächst einmal Informationen gesammelt werden, die es vielleicht erlauben, auf explorativem Wege überhaupt Vermutungen über Wirkungszusammenhänge zu entwickeln. Allerdings kann erwartet werden, dass sich auf diese Weise rekonstruierte Zusammenhänge sehr wohl in das zu Beginn dieses Abschnitts ausgebreitete Raster einordnen lassen. Deshalb kommen auch nichthypothesentestende Evaluationsansätze letztlich nicht darum herum, deutlich zu machen, welche Wirkungszusammenhänge sie ermitteln und beschreiben können und wollen und welche nicht.

Zum zweiten ist es selbstverständlich nicht die Meinung der Verfasser dieser Übersicht, dass jede Evaluation auf alle postulierten und erwogenen Wirkungszusammenhänge eingehen muss. Vollständigkeit der Abhandlung von möglichen Fragen ist kein Kriterium der Wissenschaftlichkeit. Gefragt ist vielmehr eine rechtzeitige Prüfung der Evaluierbarkeit und eine darauf abgestützte sorgfältige Auswahl der zu analysierenden Zusammenhänge. Entscheidend ist zudem, dass sich Forscherinnen und Forscher ihrer Auswahl und der damit verbundenen Selektivität bewusst sind und deshalb deutlich offenlegen, inwieweit die Aussagekraft ihrer Forschungsergebnisse durch das Untersuchungsdesign eingeschränkt sind.

In jedem Fall, und zwar unabhängig vom befolgten Wissenschaftsverständnis, ist deshalb in Evaluationen klar darüber Rechenschaft abzulegen, welche Forschungsstrategie verfolgt wird und welche Forschungskonzeption der Analyse zugrunde liegt.

3.4. EVALUATIONSSTANDARDS

Professionelle Evaluationen zeichnen sich dadurch aus, dass sie sich an handwerkliche Regeln halten. Zwar werden im vorliegenden Buch die Konzepte, Methoden und Verwaltungszusammenhänge von Evaluationen im einzelnen erörtert. Hier sollen die qualitativen Anforderungen zusammengefasst werden, die an das Metier des Evaluators gestellt werden.

Die Evaluationsforschung wird charakterisiert durch eine Vielfalt von stark unterschiedlichen methodologischen Orientierungen. Sie hat sich als angewandte Forschung zudem aus den verschiedensten wissenschaftlichen Disziplinen entwickelt. Angesichts der Heterogenität der verwendeten Ansätze ist es nicht einfach, geeignete Kriterien zur Beurteilung von Evaluationen zu finden, die allen methodischen und disziplinären Besonderheiten gerecht werden.

Eine Möglichkeit, trotz dieser Schwierigkeiten allgemein anerkannte Kriterien zur Beurteilung von Evaluationen zu formulieren, besteht in der Übernahme der in den Vereinigten Staaten von den Standesorganisationen der professionellen Evaluatorinnen und Evaluatoren entwickelten Standards (ERS-Standards Committee 1982; Joint Committee 1981). Diese breit anerkannten Kataloge von Prinzipien sollen denjenigen, die Evaluationsforschung betreiben, als handlungsanleitende Orientierungsrahmen dienen. Sie können freilich auch dazu verwendet werden, im Nachhinein die Qualität von Evaluationen zu bewerten und eigentliche Metaevaluationen durchzuführen.

Die Standards der *"Evaluation Research Society"* (ERS-Standards) sind entlang des Ablaufs eines Evaluationsprozesses gruppiert, gründen auf einem traditionell positivistischen Wissenschaftsverständnis und präferieren experimentelle und quantitative gegenüber anderen Ansätzen deutlich.

Wir halten uns wegen dieser u.E. zu restriktiven Grundhaltung der ERS-Standards an die *"Standards for Evaluation of Educational Programs, Projects and Materials"* (Joint Committee 1981) und folgen dabei der Zusammenfassung von Widmer (1996). Diese Standards sind zwar ursprünglich für die erziehungswissenschaftliche Evaluationsforschung entwickelt worden, haben aber über ihren Fachbereich hinaus weite Verbreitung gefunden. Sie wurden in jüngster Zeit überarbeitet und auf den neuesten Stand gebracht (Joint Committee 1994). Überdies haben Guba und Lincoln (1989) für jene Standards, die sich für die Beurteilung interpretativer und rekonstruktivistischer Evaluationsstudien nur bedingt

eigen (Validität, Reliabilität, Objektivität), parallele Kriterien entwickelt. Die insgesamt 30 Standards sind in vier Gruppen eingeteilt³:

Die acht *Nützlichkeitsstandards* sollen dafür sorgen, dass Evaluationen die Informationsbedürfnisse der Beteiligten (Entscheidungsträger, Programmverantwortliche, Programmpersonal, Programmteilnehmer, Programmbetroffene) befriedigen. Dabei sollen zunächst (1) die Beteiligten und Betroffenen identifiziert werden, damit deren Interessen berücksichtigt werden können. (2) Dann sollen die Personen, welche die Evaluation durchführen, vertrauenswürdig und kompetent sein, damit die Ergebnisse der Evaluation hohe Glaubwürdigkeit und Akzeptanz erreichen. (3) Die beschafften Informationen sollen vom Umfang und von der Auswahl her die Behandlung sachdienlicher Fragen ermöglichen und die Interessen der Beteiligten und Betroffenen berücksichtigen. (4) Die für die Interpretation der Ergebnisse verwendeten Perspektiven, Verfahren und Begründungen sind sorgfältig zu umschreiben. (5) Der Evaluationsbericht hat zudem das evaluierte Objekt, den Kontext, die Absichten, Verfahren und Ergebnisse der Evaluation für den Leser leicht verständlich zu beschreiben und die gezogenen Schlüsse und Empfehlungen zu begründen. (6) Die Resultate der Evaluation sind allen Beteiligten und Betroffenen zugänglich zu machen. (7) Die Veröffentlichung des Berichts hat rechtzeitig zu erfolgen. (8) Die Beteiligten und Betroffenen sollten durch die Evaluation ermuntert werden, ihren Ergebnissen zu folgen.

Die drei *Anwendbarkeitsstandards* sollen sicherstellen, dass Evaluationen praktikabel und politisch tragbar sind. (1) Dazu sind praktische Verfahren zu wählen, mit denen Störungen gering gehalten werden, damit die benötigten Informationen beschafft werden können. (2) Die politische Überlebensfähigkeit kann gesichert werden, indem auf die unterschiedlichen Positionen der verschiedenen Interessengruppen Rücksicht genommen wird. (3) Kostenwirksamkeit wird dann erreicht, wenn der Wert der in der Evaluation gewonnenen Informationen deren Kosten übersteigt.

Die acht *Korrekttheitsstandards* zielen auf die ethische und rechtliche Verantwortlichkeit in der Evaluationsforschung. (1) Dazu sollen die formellen Parteien einer Evaluation den wechselseitigen Verpflichtungen in schriftlicher Form zustimmen. (2) Die oft unvermeidlichen Interessenkonflikte sind offen und aufrichtig zu behandeln. (3) Evaluationsberichte sollen in der Darlegung der Ergebnisse wie der Grenzen der Evaluation offen, direkt und ehrlich sein. (4) Die formellen Parteien sollen das Recht

³ Vgl. dazu die umfangreiche Anwendung bei Widmer 1996.

der Öffentlichkeit auf Information unter Vorbehalt konkurrierender Prinzipien (wie öffentliche Sicherheit oder Recht auf Privatsphäre) respektieren. (5) Individuelle Persönlichkeitsrechte und (6) die Menschenwürde sind zu respektieren. (7) Die Berichterstattung soll vollständig sein und in fairer Weise Stärken und Schwächen der Evaluation aufzeigen. (8) Die finanzielle Verantwortlichkeit und die damit verbundenen Verfahren sind genau einzuhalten.

Die elf *Genauigkeitsstandards* beziehen sich direkt auf die Einhaltung bestimmter Regeln der sozialwissenschaftlichen Arbeitsweise. (1) Die Evaluationsforschung hat den zu evaluierenden Gegenstand genau zu spezifizieren und (2) seinen Kontext ausreichend zu untersuchen. (3) Die Ziele und das Vorgehen der Evaluation sind genau zu beschreiben. (4) Die Informationsquellen sind präzise und detailliert anzugeben. (5) Die Verfahren zur Erhebung von Informationen sollten deren Gültigkeit (Validität) für den gegebenen Zweck sicherstellen. Beim alternativen Paradigma sollen die Rekonstruktionen durch die Evaluation die Konstruktionen der Akteure korrekt abbilden. (6) Die Verfahren zur Erhebung von Informationen sollen auch die Zuverlässigkeit (Reliabilität) der erhaltenen Information gewährleisten. (7) Die verwendeten Daten sind systematisch zu überprüfen und gegebenenfalls zu korrigieren. (8) Quantitative wie (9) qualitative Informationen sollen angemessen analysiert werden, wobei die entsprechenden methodischen Regeln einzuhalten sind. (10) Die Folgerungen der Evaluation sind explizit zu begründen. (11) Die Berichterstattung sollte durch geeignete Vorkehrungen vor Verzerrungen geschützt werden, die auf persönliche Präferenzen oder Vorurteile der an der Evaluation Interessierten zurückzuführen wären. Beim alternativen Ansatz müssen die gewählten Konstrukte eine klare Grundlage im empirischen Material haben.

Die beschriebenen Standards stellen hohe Anforderungen an die Evaluationspraxis. Sie können selten vollumfänglich eingehalten werden. Häufig setzen gerade in der Schweiz knappe Ressourcen der Erfüllung aller Qualitätsstandards enge Grenzen. Zudem hat die Evaluationsforschung noch bei weitem nicht einen Professionalisierungsgrad erreicht, der mit jenem in den USA vergleichbar wäre. Dennoch zeichnet gerade das Bestreben, den grossen Herausforderungen gerecht zu werden, die professionelle, wissenschaftliche Evaluationsforschung aus und hebt sie ab von Parteigutachten, Polemik und Dilettantismus (Widmer 1996).

KAPITEL 4: DIE ÖFFENTLICHE POLITIK ALS EVALUATIONSOBJEKT

Peter Knoepfel, Werner Bussmann

Gemäss Kapitel 3 beziehen sich Evaluationen der hier diskutierten Art auf staatliches Handeln. Betrachten wir die im Unterkapitel zur Untersuchungsanlage (Kapitel 3.3.) angeführten drei Beispiele, so fällt auf, dass die ersten beiden (Luftreinhaltung und Frauenförderung) unmittelbar die Lösung eines konkreten gesellschaftlichen Problems zum Gegenstand haben, während sich das dritte auf die Veränderung einer staatlichen Institution (Heraufsetzung der Unterschriftenzahl für das Referendum) bezieht. Gemäss einer von Salisbury (1968: 120) bereits in den siebziger Jahren vorgeschlagenen und von Mény und Thönig (1989: 363) Ende der achtziger Jahre wieder aufgegriffenen Unterscheidung sprechen wir in den ersten beiden Fällen von substantiellen und im dritten Falle von institutionellen Politiken. Wir legen im ersten Teil dieses Kapitels zunächst dar, dass sich das vorliegende Buch fast ausschliesslich mit der Evaluation von Aktivitäten im Rahmen substantieller Politiken befasst. Diese Begrenzung ist nicht als Folge einer diesbezüglich ausschliesslichen Definition des Begriffs Evaluation anzusehen. Evaluationen können sich sowohl auf substantielle als auch auf institutionelle Politiken beziehen. Die bewusste Konzentration auf substantielle Politiken entspricht vielmehr dem Schwerpunkt der heutigen Evaluationspraxis in der Schweiz.

Die Schwerpunkte könnten sich freilich in nächster Zeit ändern. Dies hat mit neueren politischen und wissenschaftlichen Diskussionen um die Ausgestaltung der Staatsfunktion zu tun. Das Feld der Vorschläge reicht von solchen des "reinventing government" (Osborne und Gabler 1993) bis zu solchen des "deinventing government", der radikalen Beschneidung der Staatsaufgaben. Während nach dem Paradigma der "experimenting society" (Campbell 1972) der Staat versuchte, die Gesellschaft zu reformieren, wollen Politik und Gesellschaft heute den Staat reformieren. Eine Vielzahl neuer, oft aus der Privatwirtschaft entliehener, Konzepte macht die Runde: New Public Management, Controlling, "schlanker Staat", Total Quality Management, Erlassabbau, grössere Spielräume für unternehmerisches Verhalten, Aufgabenüberprüfung und -kritik u.a.m. Die bereits angelaufenen Reformen des Managements öffentlicher Politiken rufen allerdings nach seriöser Evaluierung (Pollitt 1995). Einzelne Anfänge

wurden dabei in der Schweiz bereits gemacht (etwa Knoepfel und Zimmermann 1992). Es ist zu erwarten, dass sich im Zusammenhang mit Reformvorhaben und Pilotprojekten inskünftig bedeutsame Herausforderungen stellen, die insgesamt eine Verlagerung von der Evaluation substantieller zur Evaluation institutioneller Politik zur Folge haben könnten.

4.1. INSTITUTIONELLE UND SUBSTANTIELLE ÖFFENTLICHE POLITIKEN

Substantielle Politiken zeichnen sich dadurch aus, dass sie unmittelbar die Lösung eines spezifischen gesellschaftlichen Problems zum Gegenstand haben. So bezweckt die Raumplanungspolitik die Sicherstellung einer geordneten Besiedelung des Landes und insbesondere die Trennung von Bauland und Nichtbauland, die Umweltpolitik die Sicherstellung bzw. Wiederherstellung der ökologischen Grundlagen unseres Lebens und die Gesundheitspolitik die Bereitstellung einer hinreichenden Infrastruktur für die medizinische Versorgung der Bevölkerung etc. Wie erwähnt stehen neben diesen substantiellen Politiken nach allgemeiner Auffassung (Mény und Thönig 1989: 363; Salisbury 1968: 120) auch sog. *institutionelle öffentliche Politiken*. Diese bezwecken die Veränderung des Institutionengefüges eines Landes, eines Kantons oder einer Gemeinde im Hinblick auf eine Veränderung der Verteilung der politischen Macht. Der Kampf um die politische Macht seinerseits wird allgemein als "Politics" bezeichnet. Öffentliche institutionelle Politiken (*les politiques publiques institutionnelles*, *institutional public policies*) bezwecken die Sicherstellung bzw. Veränderung der institutionellen Rahmenbedingungen dieses Kampfes um die politische Macht. Sie befassen sich etwa mit Regierungsreformen, mit der generellen Restrukturierung der Verwaltung, mit der Stärkung der Position des Parlaments oder der Gerichte, mit generellen Veränderungen des politischen oder administrativen Föderalismus oder gar mit Reformen im Bereich der Volksrechte bzw. mit der Integration der Schweiz in die Europäische Staatengemeinschaft.

Auch heute noch besteht wenig Klarheit über den Zusammenhang zwischen institutionellen und substantiellen Politiken¹. Die alten Fragen, "do

¹ Zum Verhältnis "Policy"- "Politics" s. die "Klassiker" Lasswell 1951; 1971; Lowi 1972 sowie zusätzlich Heintel 1993.

politics matter?" oder "do institutions matter?", die insbesondere in der international vergleichenden Politikforschung seit den späten sechziger Jahren diskutiert werden², müssen jedenfalls für ihre Auswirkungen auf die konkreten substantiellen Politiken auch in der Schweiz als empirisch immer noch weitgehend ungelöst bezeichnet werden. Es besteht heute auch in der Schweiz eine umfangreiche, ökonomische, politik- und verwaltungswissenschaftliche, aber auch rechtliche Literatur zum Beitrag institutioneller Faktoren zu Teilaspekten öffentlicher Politiken. Diese Debatte wurde und wird jedoch selten unter einem explizit evaluativen Gesichtswinkel geführt. Empirisch analytische Arbeiten sind uns nur wenig bekannt; diese wurden denn auch in erster Linie im Rahmen grösserer international vergleichender Projekte durchgeführt (etwa: Schmidt 1982, Schmidt 1996, Flora und Heidenheimer 1981, Ashford 1978 und Ashford 1990). Wir müssen im Rahmen dieses Buches darauf verzichten, diese Beiträge einer spezifisch evaluativen Analyse zu unterziehen, obwohl dies sicherlich in künftigen Evaluationsforschungen zum Beitrag institutioneller Faktoren zu ausgewählten substantiellen Politiken erforderlich sein wird³. An dieser Stelle seien daher zum wechselseitigen Verhältnis zwischen institutionellen und substantiellen Politiken lediglich folgende, relativ gesicherten Hypothesen angeführt:

- *Auswirkungen institutioneller auf substantielle Politiken:* Forschungsergebnisse aus Frankreich (Mény und Thönig 1989), Deutschland (Scharpf et al. 1976) oder aus der Schweiz (Germann et al. 1979; Klöti, Haldemann und Schenkel 1993) zeigen, dass zwischen der Ausprägungsart des Föderalismus und den substantiellen Politikleistungen einzelner öffentlichen Politiken ein Zusammenhang besteht. So wird darauf hingewiesen, dass mit zunehmender Vertiefung der Dezentralisierung eine zunehmende Varianz der Politikleistungen entsteht. Ähnliches gilt für die Stellung der höchsten Gerichte: Je stärker ihre Machtposition ausgebaut ist, desto homogener sind vermutlich die Leistungen der einzelnen substantiellen Politiken innerhalb eines Landes.

² Lowi 1972, deLeon 1987, Flora und Heidenheimer 1981, Ashford 1978, Ashford 1990, Dierkes, Weiler und Berthoin Anthal 1987, Schmidt 1982, Schmidt 1996, Heidenheimer, Heclou und Teich-Adams 1990, Larson 1980.

³ Ein Beispiel stellt das Projekt Klöti, Haldemann und Schenkel (1993) zum Thema "Stadt im Bundesstaat" im Rahmen der Analyse von Umwelt- und Verkehrspolitik oder das gegenwärtig (1995-97) am IDHEAP laufende, international vergleichende Projekt zum Stellenwert des Subsidiaritätsprinzips in der Umweltpolitik dar (Leitung: P. Knoepfel und K. Horber-Papazian; Durchführung: M. Benninghoff, S. Terribilini, S. Wälti).

Bezüglich der Stellung der Regierung und ihrem Verhältnis zum Parlament lassen sich auf den gegenwärtigen wissenschaftlichen Grundlagen kaum Hypothesen formulieren. So ist weder erwiesen, dass ein gegenüber der Regierung an sich starkes oder aufgrund institutioneller Reformen gestärktes Parlament zu detaillierteren Gesetzen und Verwaltungsprogrammen in den einzelnen substantiellen Politiken führt, noch der Nachweis erbracht, dass eine Stärkung der Position der Bundesämter gegenüber der Regierung und dem Parlament zwangsläufig zu stärker problemorientierten und weniger politisch ausgerichteten Verwaltungsausgaben führen muss (wie das etwa das New Public Management annimmt). Ebenfalls offen ist die Frage des Verhältnisses zwischen dem Stand des Ausbaus der Volksrechte und der Qualität der Leistungen der substantiellen öffentlichen Politiken. Obgleich verschiedene Überlegungen und empirische Untersuchungen (Frey und Kirchgässner 1993; Bohnet und Frey 1994; Eichenberger und Frey 1994; Frey und Bohnet 1994) die These stützen, dass stärker ausgebauten Volksrechte sich vorteilhaft auf die Bürgerorientierung und Effizienz substantieller Politiken auswirken, ist dieser Zusammenhang bisher empirisch noch nicht ausreichend geklärt.

- *Zur Auswirkung von Veränderungen substantieller Politiken auf die institutionellen Politiken* bestehen heute u.E. keine hinlänglich gesicherten empirischen Aussagen. Plausibel, wenn auch nicht im einzelnen empirisch belegt, ist etwa die Vermutung, dass die Bildung von Netzwerken (vgl. Kapitel 7.2.), die aus staatlichen und gesellschaftlichen Akteuren zusammengesetzt sind, in immer mehr öffentlichen substantiellen Politiken negative Auswirkungen auf die Machtposition von Parlamenten haben dürfte. Parlamentarische und gouvernementale Steuerung würden, so die Hypothese, beim Zutreffen dieser Beobachtung zunehmend in die einzelnen substantiellen Politiken hineinverlagert und dort auf pluralistisch zusammengesetzte Gremien "übertragen". Kontrovers ist, ob sich die verstärkte Fokussierung auf die Qualität der Dienstleistungen (bei einer vermehrten Outputsteuerung nach Massgabe des New Public Management) in einer Stärkung (Schedler 1995) oder letztlich ebenfalls in einer Schwächung von Parlament und der Regierung niederschlägt (Knoepfel 1995a: 457, 463 f.). Diese Beispiele zeigen, dass Veränderungen in substantiellen Einzelpolitiken jedenfalls mittelfristig auf die institutionellen öffentlichen Politiken durchschlagen können. Ähnliche Auswirkungen werden etwa

von der Generalisierung des ursprünglich im Bereich der Umweltpolitik angesiedelten Instruments der Mediation in Richtung "verhandelnder Staat" (Weidner 1993 und Kapitel 9), oder von der sich etwa in der Verkehrspolitik abzeichnenden Schaffung neuer - regionaler - Verwaltungseinheiten zwischen Gemeinden und Kantonen in Richtung einer Neustrukturierung des Föderalismus vermutet (Klöti, Haldemann und Schenkel 1993; Knoepfel, Imhof und Zimmermann 1994). Welche Hypothesen zu diesen Auswirkungen von Veränderungen substantieller auf institutionelle Politiken auch immer formuliert werden, so steht doch fest, dass solche Auswirkungen nur in mittel- bis langfristigen Zeiträumen beobachtbar sind.

4.2. EVALUATION SUBSTANTIELLER ÖFFENTLICHER POLITIKEN UND MASSNAHMEN

4.2.1. Die (substantielle) "öffentliche Politik" und ihre Akteure⁴

Unter einer öffentlichen Politik (public policy; politique publique)⁵ verstehen wir ein Ensemble kohärenter und zielgerichteter Handlungen und Entscheidungen unterschiedlicher Rechtsqualität, die dazu berufene staatliche, verbandliche oder private Personen im Hinblick auf die Lösung eines gesellschaftlichen Problems vornehmen bzw. treffen. Das Ensemble besteht regelmässig aus einem mehr oder minder konkreten Policy-Design⁶ (*Verwaltungsprogramm*’, programme administratif), das rechtliche Grundlagen für die Ziele, die Interventionsinstrumente und die institutioneilen, organisatorischen, finanziellen und instrumentellen Merkmale der administrativen und/oder gesellschaftlichen Politikumsetzung bereitstellt. Das Ensemble beinhaltet ausserdem Entscheidungen zur konkreten Organisation und zu den Ressourcen des Politikumsetzungsprozesses (Behörden-

⁴ Unter "öffentlicher" Politik wird im folgenden immer eine substantielle öffentliche Politik verstanden.

⁵ Die hier vorgetragene Definition hat Knoepfel (1995e) in Anlehnung an zahlreiche (leicht abweichende) Definitionen bei Dente (1989), Jobert und Müller (1987), Linder (1988), Mény und Thönig (1989), Müller (1989) oder Windhoff-Heritier (1987) entwickelt.

⁶ Wir verwenden den Begriff konkreter und starker auf formelle Dokumente ausgerichtet als in der angelsächsischen Diskussion (Bobrow und Dryzek 1987; Dryzek und Ripley 1988; Linder und Peters 1988; 1989a; 1990; 1991; Schneider und Ingram 1990).

arrangement; arrangement politico-administratif) sowie andere Zwischenoutputs und Endoutputs. Als Zwischenoutputs gelten namentlich die erwähnten Entscheidungen über das Policy-Design (Verwaltungsprogramm) und das Behördenarrangement; daneben finden sich auch Entscheidungen über Umsetzungspläne. Als Endoutputs sind alle jene Entscheidungen und Handlungen aufzufassen, die ein direktes rechtliches oder faktisches Verhältnis zwischen den Politikadressaten, den weiteren Politikbetroffenen und der zuständigen Vollzugsinstanz schaffen. Die *Politikadressaten* werden in der Regel im Policy-Design definiert. Es ist diejenige Gruppe von Akteuren, deren Verhalten die öffentliche Politik als relevant für die Lösung des angegangenen Problems ansieht. Deshalb nimmt sich die Politik vor, dieses Adressatenverhalten durch geeignete Massnahmen zu verändern oder zu stabilisieren (z.B. Gebote oder finanzielle Beiträge). Die Definition der *Politikbetroffenen* leitet sich oft nur mittelbar aus dem Policy-Design ab; sie findet sich in den oft vage formulierten Politikzielen oder ist mitunter gar nicht explizit formuliert. Als Politikbetroffene ist jene Gruppe von Personen anzusehen, die in direkter oder indirekter Weise positiv oder negativ vom Versuch betroffen ist, das angegangene gesellschaftliche Problem im Rahmen einer öffentlichen Politik in bestimmter Weise zu lösen.

An einer öffentlichen Politik sind immer mehrere, in den meisten Fällen sogar sehr viele *Akteure* beteiligt. Als Akteur bezeichnet man in der Politikanalyse gemeinhin eine Personengruppe, die kraft eigener Ressourcen und spezieller Aufgabenstellung berufen oder in der Lage ist, auf einzelne oder mehrere Entscheidungsprozesse inhaltlich Einfluss zu nehmen. Solche Akteure können sein: Organe (Vorstand, Vorstandsausschuss, Generalversammlung etc.) von Interessengruppen (Adressaten und Politikbetroffene), einzelne oder mehrere Einzeladressaten (Eigentümer der betroffenen Parzelle XX, Direktion der Firma YY), einzelne oder mehrere Personen aus dem Kreis der Politikbetroffenen (Nachbar der Parzelle X, Vorstand der Mietervereinigung Y) sowie Personen bzw. Personengruppen, die in der öffentlichen Verwaltung spezifische Aufgaben wahrnehmen. Als solche treten etwa auf: Die Sektion Y der Abteilung X des Bundesamtes für Polizeiwesen, die Direktion des Bundesamtes für Polizeiwesen, das Generalsekretariat eines Departements, der Departementsvorsteher, die Geschäftsprüfungskommission oder der Gesamtbundesrat. Letztere Beispiele zeigen, dass "die Verwaltung" nicht einfach als ein einheitlicher Akteur aufgefasst werden kann. Vielmehr wirken an einer öffentlichen Politik, für welche eines oder mehrere Ämter haupt-

sächlich zuständig ist, eine Vielzahl von Verwaltungsakteuren mit, die aufgrund ihres Pflichtenhefts oder ihrer hierarchischen Stellung unterschiedliche, sich nicht selten konkurrierende Aspekte dieser Politik zu betreuen haben. So kann es ohne weiteres vorkommen, dass innerhalb ein und desselben Amtes von verschiedenen Akteuren (z.B. Sektionen) unterschiedliche, sich teilweise widersprechende Belange und Anliegen in die Entscheidungsprozesse eingebracht werden.

Gemeinhin wird in der Politikanalyse zwischen *gesellschaftlichen* und *institutionellen Akteuren* unterschieden. Die erste Gruppe umfasst die bereits erwähnten beiden Akteurskategorien der Adressaten bzw. der Politikbetroffenen; sie schliesst aber auch Organisationen ein, die die eine oder andere Gruppe vertreten. Die Mitwirkung dieser gesellschaftlichen Akteure an der Politikumsetzung ist in aller Regel formell nicht obligatorisch; die Gruppen können selbst entscheiden, ob und unter Einsatz welcher Ressourcen sie in den Politikprozess eingreifen wollen⁷. Diese Wahlfreiheit besteht für die *institutionellen* (= behördlichen) *Akteure* meistens nicht. Diese letzteren sind durch ihr Pflichtenheft gehalten, zu einem durch interne Verfahrensregeln bestimmten Zeitpunkt unter Einsatz der ihnen eigenen Ressourcen und nach Massgabe ihrer gesetzlich festgelegten Aufgaben in Politikformulierungs- und -Umsetzungsprozesse zu intervenieren.

Bei den institutionellen Akteuren unterscheiden wir unter den spezifisch schweizerischen Gegebenheiten des Vollzugsföderalismus Akteure, die vornehmlich an der *Politikformulierung* (Bundesinstanzen) und solche, die für dieselbe Politik am *Vollzug* beteiligt sind (kantonale und kommunale Behörden). Eine ähnliche Arbeitsteilung findet sich auch bei den gesellschaftlichen Akteuren: es intervenieren in der Regel kantonale und kommunale Interessenvereinigungen in kantonale Vollzugsprozesse, während gesamtschweizerische Organisationen hauptsächlich an Politikformulierungsprozessen mitwirken. Die Gesamtheit der institutionellen und gesellschaftlichen Akteure einer öffentlichen Politik bilden ein relativ konstantes Ensemble, das man gemeinhin als *Arena* der betreffenden öffentlichen Politik, als *Politiknetzwerk* (vgl. unten Kapitel 7) oder als Akteure eines *Politikraums* (*espace de la politique publique*) bezeichnet. Diese Akteure bleiben für eine bestimmte öffentliche Politik oft über Jahre hinaus konstant. So findet man in sämtlichen landwirtschaftsorientierten

⁷ Allerdings müssen die individuellen Politikadressaten am Prozess der Formulierung des Endoutputs nach dem Verwaltungsverfahrenrecht des Bundes und der Kantone bestimmte Mitwirkungsleistungen erbringen (z.B. Informationsbereitstellung).

Politikprozessen auf der Ebene der institutioneilen Akteure regelmässig verschiedene Sektionen, eine oder zwei Abteilungsleitungen und die Direktion des Bundesamtes für Landwirtschaft, das Generalsekretariat und den Vorsteher des eidgenössischen Volkswirtschaftsdepartements, eine oder mehrere Sektionen sowie die Amtsleitung der kantonalen Landwirtschaftsämter und, auf der Ebene der gesellschaftlichen Akteure, den Schweizerischen Bauernverband, einen oder mehrere spezielle Branchenverbände (Tier- oder Pflanzenwirtschaft), Vertreter der Importeure, der landwirtschaftlichen Hilfsstoffe-Industrie sowie der Verwertungsindustrien und Vertreter der grossen Lebensmittelketten. Die Zusammensetzung dieses Politikraums verändert sich erst, wenn die betroffene Politik politisch in "Schieflage" gelangt und neue Anliegen und Inhalte aufnehmen muss (Beispiel: Ökologisierung der Landwirtschaft, die den Einbezug stärker ökologisch ausgerichteter Akteure notwendig macht; vgl. dazu Knoepfel und Zimmermann 1987, Benninghoff 1995, Sciarini 1994).

Gesellschaftliche und institutioneile Akteure verfügen über bestimmte *Ressourcen* (Dente 1989). Als solche gelten Recht, Information und Wissen, Personal, Organisation, Geld, Zeit und Konsens bzw. die Möglichkeit, Dissens zu mobilisieren (Klok 1995). Die Art und Weise der Verteilung dieser Ressourcen auf die verschiedenen institutioneilen und gesellschaftlichen Akteure ist von grosser Bedeutung für die Qualität der aus den jeweiligen Entscheidungsprozessen resultierenden Zwischen- und Endoutputs. Dies gilt auch dann, wenn die Vorgaben einer Politik in ihrem Verwaltungsprogramm inhaltlich recht detailliert sind und Verhandlungen unter den politikbeteiligten Akteuren relativ enge Grenzen setzen. Die Konzentration der wichtigsten Ressourcen bei einem einzelnen oder einigen wenigen Akteuren verschafft diesen eine Machtstellung, die es ihnen erlaubt, mitunter auch Entscheidungen durchzusetzen, die im Widerspruch zu inhaltlichen Vergaben des Verwaltungsprogramms stehen. Als solche müssen auch Nichtentscheidungen angesehen werden, die etwa darin bestehen, dass trotz des Vorliegens an sich interventionsauslösender Tatbestände keine entsprechende Massnahme getroffen wird.

Eine *öffentliche Politik* besteht nach der vorgetragenen Definition aus Ensembles von Handlungen und Entscheidungen, die sich aus den erwähnten Zwischen- und Endoutputs zusammensetzen. Vollständig ist eine öffentliche Politik erst dann, wenn sie neben den Zwischen- auch Endoutputs aufweist. In der Aufbauphase ist damit jede öffentliche Politik unvollständig. Sie entsteht in verschiedenen Schüben. Ist indessen eine Politikumsetzung weder beabsichtigt noch empirisch beobachtbar, so kann

nicht von einer öffentlichen Politik gesprochen werden. Das gilt etwa für eine blosser Absichtserklärung einer Regierung, in der für eine bestimmte gesellschaftliche Problemlage eine gesetzliche Regulierung in Aussicht gestellt wird oder für eine singuläre Gerichts- oder Genehmigungsentscheidung. In beiden Fällen fehlt das für die öffentliche Politik konstitutive Ensemble. Denn die analytische Innovation des Begriffs der öffentlichen Politik liegt gerade darin, isolierte Einzelentscheidungen in den Kontext all jener anderen Entscheidungen zu stellen, die die Lösung des von der Einzelentscheidung betroffenen Problems ebenfalls mitgestalten. Dieses Ensemble konstituiert ein vielschichtiges Entscheidungskonglomerat, an dem eine Vielzahl von Personen bewusst und zielgerichtet auf unterschiedlichen Ebenen (Bund, Kantone, Gemeinden etc.) in verschiedenen Funktionen (Parlamentsmitglieder, Programm- und Vollzugsverwaltung) und in unterschiedlichen rechtlichen Formen (generell-abstraktes Gesetz, konkrete Budget- oder Planungsentscheidung, individuell-konkrete Verfügung) an der Lösung ein und desselben gesellschaftlichen Problems zusammenwirken.

4.2.2. Gesamt- und Teilevaluationen (Einzelmassnahmen, Aktionsprogramme und Projekte) öffentlicher Politiken

Diese prinzipielle Ausrichtung politik- und verwaltungswissenschaftlicher Evaluationen an den öffentlichen Politiken hat zweifellos den Vorteil, dass Einzelobjekte stets im Rahmen der Gesamtheit der auf die jeweilige Problemlage angesetzten Einzelmassnahmen und ihrer übergreifenden Ziele analysiert werden. In diesem Sinne deckt sich die in diesem Buch vorgeschlagene politikorientierte Evaluation mit der Politikanalyse (Vgl. Kapitel 7.1.). Die Vielfalt und Komplexität der zu berücksichtigenden Zusammenhänge stellt indessen auch schwere und mitunter kaum einlösbare Anforderungen. Auch in Anbetracht der Selektivität der zu beantwortenden Fragestellungen drängt sich oft eine Begrenzung der Evaluation auf spezifische Teilgegenstände oder spezifische Wirkungsaspekte öffentlicher Politiken auf. Sinnvoll ist in vielen Fällen eine Beschränkung der Evaluation auf konkrete staatliche Massnahmen, auf Pläne aller Art oder gar auf Projekte.

Als staatliche *Massnahmen* werden hier analytisch isolierbare Ensembles von Einzeltätigkeiten des Staates im Hinblick auf die Realisierung bestimmter, in den Verwaltungsprogrammen der entsprechenden Politiken

oft nur vage formulierter Ziele verstanden. Die Bündelung mehrerer Massnahmen bezeichnen wir als *Aktionsprogramm*.

In der Schweiz wie in den übrigen kontinentaleuropäischen Ländern gibt das Gesetz dem staatlichen Handeln die Form vor. Evaluationen können sich auch auf Gesetze beziehen (Gesetzesevaluation). Oft sind indessen Gesetze nicht der ideale Ansatzpunkte für Evaluationen. Zum Teil ist ihr Geltungsbereich breiter, als dies für eine Evaluation sinnvoll ist, zum Teil werden sie erst auf Verordnungs- oder Weisungsebene ausreichend als Verwaltungsprogramm konkretisiert. Es erweist sich deshalb oft als sinnvoll, der Evaluation einzelne, analytisch im Hinblick auf bestimmte Ziele isolierbare Teile des Gesetzes zugrunde zu legen. Anstatt das ganze Heimarbeitsgesetz auf seine Wirkungen zu evaluieren, kann sich eine Studie beispielsweise auf den für das Heimarbeitsgesetz zentralen Grundsatz der Lohngleichheit zwischen Heim- und Betriebsarbeit konzentrieren (Gerheuser und Schmid 1994).

Staatliche Einzelaktivitäten können neben ihrer Finalität auch einen eng begrenzten zeitlichen, sachlichen und/oder örtlichen Rahmen aufweisen, weshalb sie oft auch in einer besonderen organisatorischen Struktur durchgeführt werden. Dafür hat sich der Begriff des *Projekts* durchgesetzt. Projekte bilden recht häufig Gegenstand von Evaluationen (z.B. Aids-Prävention, Drogenprävention, Entwicklungshilfe, Strafvollzug).

Sowohl die Evaluierung ganzer öffentlicher Politiken als auch deren Begrenzung auf spezifische Fragestellungen, Massnahmen, Aktionsprogramme oder Projekte ist mit Vor- und Nachteilen verbunden. Bei einer *breiten Fragestellung* können alle bedeutsamen Problemzusammenhänge rekonstruiert und die Interdependenz der Handlungen aufgezeigt werden. Dies erfordert allerdings einen hohen Aufwand, eine lange Zeitdauer und einen Verzicht auf eingehende Analysen der Politik-Wirkungen. Eine *eng eingegrenzte Fragestellung* reduziert den Kosten- und Zeitaufwand und erlaubt zum Teil präzise Aussagen zum Wirkungspfad. Umgekehrt liegt hier die Gefahr in einer nicht ausreichenden Ermittlung des Massnahmenkontexts und in der Nichtberücksichtigung von Querbeziehungen im Rahmen des Wirkungsgefüges. Während in den USA in der Regel mit eng eingegrenzten Fragestellungen gearbeitet wird, ist etwa in Frankreich eine Tradition breit angelegter Evaluationen vorherrschend.

Namentlich im Rahmen konkreter Mandatsarbeit für staatliche Stellen ist aus Zeit- und Budgetgründen oft eine ausreichende Fokussierung der Fragestellung angezeigt. Eine Erfolgsvoraussetzung ist eine ausreichende Kontextanalyse (vgl. die Aussagen zu den Nützlichkeitsstandards in Kapi-

tel 3.4.). Nur wenn das zu evaluierende Objekt und die Beteiligten ausreichend identifiziert und in ihrem Kontext dargestellt werden, sind die Voraussetzungen gegeben, dass auch die richtigen Schlüsse gezogen werden können.

4.3. GEGENSTÄNDE VON POLITIK-EVALUATIONEN

4.3.1. Traditionelle Evaluationstypisierung

In den anfangs der achtziger Jahre entwickelten Standards der amerikanischen Evaluation Research Society werden sechs Evaluationstypen unterschieden (ERS Standards Committee 1982): *front-end analysis*, *evaluability assessment*, *process evaluation*, *impact evaluation*, *program monitoring* und *metaevaluation* (Chelimsky 1985: 9 ff; Rist 1990b: 76 ff; Chelimsky 1987: 90 ff).

Als *front-end analysis* werden Evaluationen mit prospektivem Charakter bezeichnet, die sich vor allem für die Praktikabilität und die möglichen Wirkungen einer Politik interessieren. Beim *evaluability assessment*, das sowohl vor als auch nach der Einführung einer Politik vorgenommen werden kann, geht es primär darum, abzuklären, ob eine Politik überhaupt evaluierbar ist. Dies setzt eine eingehende Auseinandersetzung mit den Zielen einer Politik und mit den geplanten oder getroffenen Maßnahmen voraus. Die *process evaluation* (auch *formative evaluation*) konzentriert sich vor allem auf Fragen der Implementation. Untersucht werden somit in erster Linie die Tätigkeiten und Verhaltensweisen, welche den Vollzug einer Politik prägen bzw. prägten, denn diese Evaluation erfolgt stets aus der Retrospektive. *Impact evaluations* (oder auch *effectiveness evaluations*) untersuchen die tatsächlichen Wirkungen einer Politik und gehen insbesondere auf die Frage ein, ob die Ziele einer Politik realisiert wurden. Beim *program monitoring* handelt es sich um eine politikbegleitende Evaluation, die vor allem als Grundlage für Korrekturen im Vollzug oder allenfalls auch auf der Ebene der Politikentwicklung dient. Unter einer *Metaevaluation* versteht man nach dieser Typologie die Würdigung der Ergebnisse einer oder mehrerer Evaluationen.

Diese in der amerikanischen Evaluationsliteratur entwickelte und verwendete Typologie beruht nicht auf klaren analytischen Kriterien. Sie unterscheidet beispielsweise nicht deutlich zwischen den verschiedenen Evaluationsmomenten (vor, während oder nach einer Politik), den eigent-

lichen Evaluationsfragen (was steht im Zentrum des Interesses: z.B. Tätigkeiten und Verhaltensweisen der Adressaten oder Vollzugsorgane einer Politik, oder deren Wirkungen im Vergleich zu den angestrebten Zielen, d.h. deren "Erfolg") und den Erkenntnisinteressen, die Evaluationen zugrunde liegen. In dieser Hinsicht ist sie nicht unbedingt befriedigend. Auch in terminologischer Hinsicht vermag sie im übrigen nicht völlig zu überzeugen. Es gibt denn zum Teil auch andere Typologisierungsansätze. So hat beispielsweise die 1987 vom Eidgenössischen Justiz- und Polizeidepartement eingesetzte Arbeitsgruppe "Gesetzesevaluation" (AGEVAL) einerseits die verschiedenen zeitlichen Perspektiven (Wirkungsprognose, begleitende Evaluation, Wirkungskontrolle) und andererseits die möglichen Evaluationskriterien (Effektivität, Wirksamkeit, Impact und Effizienz) unterschieden (Arbeitsgruppe "Gesetzesevaluation" 1991; Mader 1985: 46 ff., 55 f.).

4.3.2. Typisierung nach den Stufen des Policy-Cycle⁸

Im Gegensatz zur Typologie der amerikanischen Evaluation Research Society und der Arbeitsgruppe "Gesetzesevaluation" schlagen wir hier eine an der Praxis der Politikgenerierung und -Umsetzung orientierte Aufgliederung von Evaluationsgegenständen vor. Diese hat den Vorteil, dass sie stärker an politikwissenschaftlichen Kategorien orientiert ist und dass sie den Evaluationsgegenstand im Kontext eines - kohärenten - Gesamtprozesses abbildet.

Abbildung 1 Acht Stufen der Politikgenerierung und -Umsetzung

⁸ Vgl. eine ähnliche Konzipierung bei Windhoff-Héritier 1987: 64 ff.

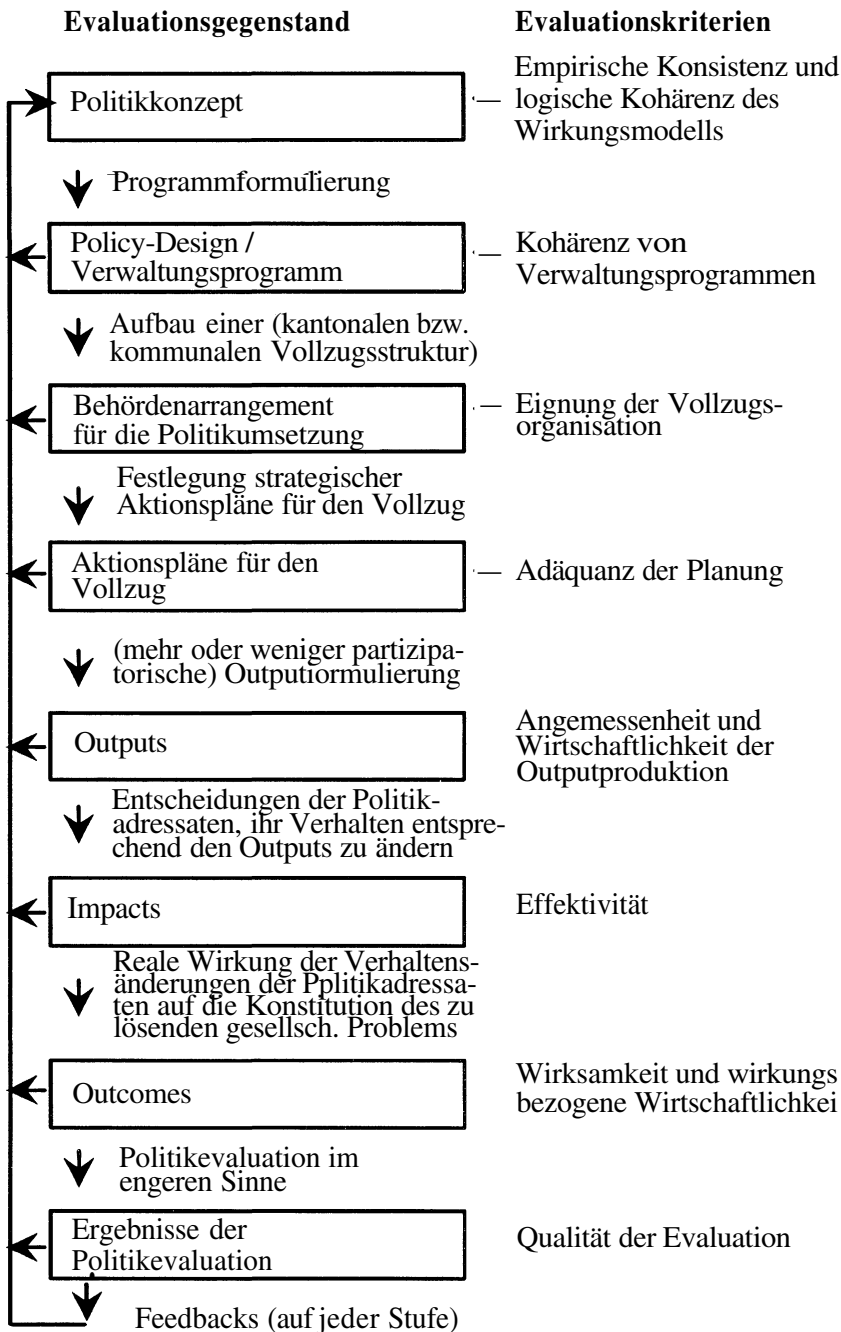


Abbildung 1 unterscheidet acht, auch im schweizerischen Politikalltag nur selten sequentiell hintereinander, meistens eher parallel nebeneinander ablaufenden Politikgenerierungs- und -Umsetzungsstufen (Pfeile), die jeweils zu bestimmten Zwischenergebnissen führen (im Kasten). Diese führen von der administrativen und parlamentarischen Politikkonzeption ("Politikkonzept") bis zur Evaluation der Wirksamkeit der gesamten Politik und den entsprechenden Evaluationsergebnissen. Auf jeder Stufe sind korrigierende Rückkoppelungen zu den vorgelagerten Entscheidungen denkbar und in der Praxis recht häufig.

Jedes Teilergebnis der im folgenden kurz skizzierten und in Kapitel 5 ausführlich dargelegten Stufen kann zum Gegenstand einer Evaluation gemacht werden. Dabei gelangen, wie in Kapitel 5 im einzelnen ausgeführt wird, üblicherweise unterschiedliche Evaluationskriterien zur Anwendung.

Skizzenhaft lassen sich die Stufen wie folgt umschreiben:

- *Politikkonzept*: Ensemble von Entscheidungen der zuständigen politisch-administrativen Akteure zur Definition des als kollektiv anerkannten Politikproblems (Abgrenzung von nichtkollektiven (privaten) Problemen; Abgrenzung von Problemen, die durch andere Politiken reguliert werden bzw. werden sollen; gesellschaftliche und politische Bedeutung des Problems), zum Agendasetting (Erlangen des Status eines entscheidwürdigen und entscheidfähigen Problems), zu den allgemeinen Politikzielen, zum - massgeblichen - Wirkungsmodell bestehend aus den Hypothesen zum kausalen Mechanismus, zur grundsätzlichen Wirkungsweise der gewählten Intervention und deren mutmasslichen konkreten Auswirkungen im Handlungsfeld.
- *Policy-Design / Verwaltungsprogramm*: Ensemble all jener Rechtsätze⁹ sowie diese konkretisierender Weisungen des Bundes und der Kantone, die Regierungen und Parlamente für notwendig erachten, um das Politikkonzept in rechtlich einwandfreier Weise in administrative Aktionspläne und insbesondere in rechtsgültiges Verwaltungshandeln gegenüber den Politikadressaten umzusetzen. Politiken verfügen über mehr oder weniger detaillierte (Regelungsdichte), mehr oder weniger zentralisierte (hohe versus niedrige Regelungsdichte desjenigen Teils

⁹ Im Sinne von Art. 5 Abs. 2 des Geschäftsverkehrsgesetzes vom 23. März 1962 (SR 171.11.)

des Verwaltungsprogramms, den der Bund beisteuert) bzw. über mehr oder weniger kohärente Verwaltungsprogramme (Abgestimmtheit unter den fünf konstitutiven Elementen von Verwaltungsprogrammen). Diese enthalten Zieldefinitionen, evaluative und operative Elemente (Instrumente), Entscheidungen zu den Behördenarrangements, zu den Ressourcen und zu den Verfahren.

- *Behördenarrangement für die Politikumsetzung*: Gesamtheit der das Verwaltungsprogramm konkretisierenden, substituierenden oder gar modifizierenden politisch-administrativen Entscheidungen über die administrative Umsetzungsstruktur einer Politik. Darunter fallen Entscheidungen zur Kompetenzordnung, zum Aufbau neuer bzw. zur Zuordnung der neuen Politik zu bestehenden Verwaltungszweigen, zu den finanziellen, personellen, instrumenteilen etc. Ressourcen der zuständigen Behörden, aber auch zu den managementmässigen Aufbau- und Ablaufstrukturen der implizierten Verwaltungseinheiten.
- *Aktionspläne für den Vollzug*: (Nicht in jeder Policy als verselbständigte Stufe der Politikumsetzung vorhandene) Gesamtheit der für die spätere Output-Produktion notwendigen Planungsentscheidungen, die unter Umständen bereits im Verwaltungsprogramm angeordnet sein können. Diese beinhalten Prioritäten für die Produktion der konkreten Outputs und entsprechende Mittelzuweisungen. Aktionspläne als intermediäre Umsetzungsstufe zwischen Verwaltungsprogramm und oft planlos anfallenden Politikoutputs werden in öffentlichen Politiken in zunehmendem Masse eingesetzt.
- *Outputs*: Gesamtheit der in Umsetzung einer bestimmten öffentlichen Politik erstellten Endprodukte des politisch-administrativen Entscheidungsprozesses einer öffentlichen Politik. Beispiele dafür sind Verfügungen aller Art¹⁰, aber auch Geldleistungen, direkte Dienstleistungen staatlicher Behörden oder staatliche Beratungsaktivitäten. Neben förmlichen treten zunehmend informelle Outputs auf. Outputs charakterisieren sich dadurch, dass sie zwischen den Politikadressaten und der zuständigen staatlichen Vollzugsinstanz eine direkte Beziehung (oft sogar ein Rechtsverhältnis) schaffen. Outputs haben aber in vielen Fällen

¹⁰ Im Sinne von Art. 5 des Verwaltungsverfahrensgesetzes vom 20. Dezember 1968 (SR 172.021).

auch Auswirkungen auf Dritte (sog. Politikbetroffene). Das Verwaltungsverfahrensgesetz sieht daher bestimmte Mitwirkungspflichten bzw. -rechte für Politikadressaten und (zunehmend auch) Politikbetroffene vor. Nur selten gelingt es Verwaltungen, alle Politikadressaten (rechtzeitig und gleichzeitig) mit Outputs zu bedienen bzw. zu bestrafen. Man spricht in diesen Fällen von zeitlich oder örtlich auftretenden Vollzugsdefiziten.

- *Impacts*: Gesamtheit der bei den Politikadressaten eintretenden intendierten oder nichtintendierten Verhaltensänderungen, die unmittelbar auf das Inkrafttreten von Verwaltungsprogrammen, in den meisten Fällen jedoch auf diese konkretisierende Politikoutputs zurückführbar sind. Nach dem Rechtsstaatsprinzip müssen die Politikadressaten grundsätzlich bereits im Verwaltungsprogramm (operative Elemente) aufgeführt werden. Nur in wenigen Fällen entsprechen Zahl und Ausmass der feststellbaren Verhaltensänderungen den Erwartungen bzw. den formulierten Sollwerten. Man spricht in diesen Fällen von einer mangelhaften Effektivität öffentlicher Politiken.
- *Outcomes*: Gesamtheit der intendierten oder nichtintendierten Auswirkungen der Verhaltensänderungen der Politikadressaten auf die Konstitution des Problems, das sich die Politik zu lösen vorgenommen hat. Wo diese Outcomes in einer positiven und dem Politikziel entsprechenden Auswirkung auf das Problem bestehen, spricht man von einer wirksamen Politik; diese ist "effizient", wenn der Verwaltungsaufwand relativ niedrig ist. Unwirksam sind jene Politiken, die gegebenenfalls eine Menge Outputs, eventuell sogar beachtliche Impacts, aber unerwartet wenig Outcomes zeitigen. Für die Messung der Outcomes ist möglichst auf die Zieldefinitionen in den Verwaltungsprogrammen zurückzugreifen.
- *Politikevaluation im engeren Sinne*: Diese (bei weitem nicht in allen öffentlichen Politiken vorgesehene und praktizierte) Stufe des Policy-Cycle umfasst die Gesamtheit der methodisch und wissenschaftlich mehr oder weniger gut abgesicherten Entscheidungen behördlicher oder wissenschaftlicher Instanzen über die Wirksamkeit und über die Wirtschaftlichkeit einer öffentlichen Politik. Solche Politikevaluation sollte in der Regel die in diesem Buch enthaltenen Anforderungen (Kapitel 3)

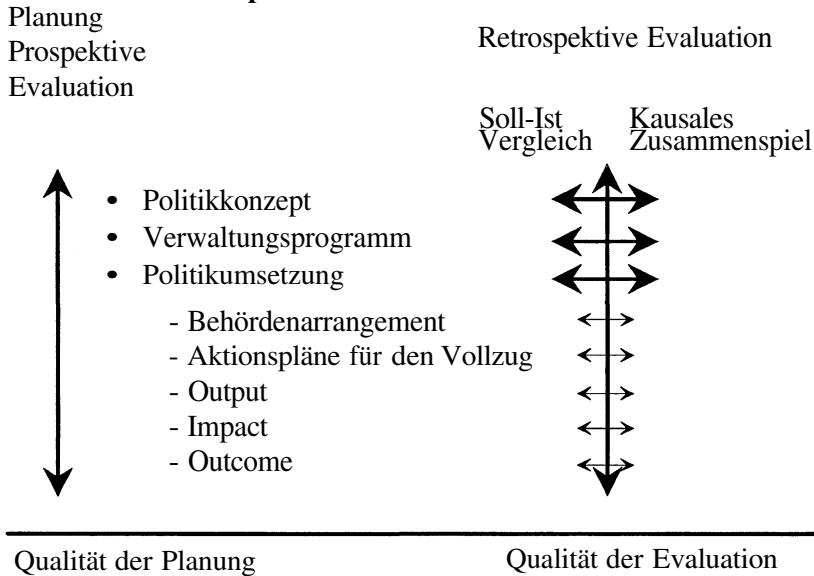
erfüllen. In der Praxis ist dies aber bei weitem nicht immer der Fall. Die Entscheidung über die Wirksamkeit einer Politik ist und bleibt eine bewertende Entscheidung; faktisch kann sie auf Glaubenssätzen, auf politischen Überzeugungen etc. basieren. Es ist u.a. Aufgabe der politikorientierten Evaluationsforschung, solche Politikevaluationen im engeren Sinne zu erstellen und zu evaluieren (Metaevaluation).

Wie oben dargelegt, umfasst politikorientierte Evaluationsforschung keineswegs nur diese Politikevaluation im engeren Sinne. Wie im Kapitel 5 im einzelnen dargelegt wird, kann das Ergebnis jeder dieser acht Entscheidungsstufen zum Gegenstand politikorientierter Evaluation gemacht werden. Wie aus der rechten Kolonne der Abbildung 1 hervorgeht, sind dabei die einschlägigen Evaluationskriterien unterschiedlich.

4.4. PROSPEKTIVE UND RETROSPEKTIVE EVALUATION

In Abbildung 2 werden die verschiedenen Stufen des Politikprozesses dargestellt. Darin zeigen wir überdies die Zusammenhänge der Politikprogrammierung und -Umsetzung mit der prospektiven, auf Planung ausgerichteten Evaluation sowie mit der retrospektiven Evaluation auf.

Abbildung 2 Prospektive und retrospektive Evaluation der Komponenten öffentlicher Politiken



Bereits die Formulierung eines Verwaltungsprogramms erfordert in der Regel eine implizite oder explizite Planung der Umsetzung: Das Programm selbst muss die Grundzüge des Behördenarrangements, seiner Finanzierung und die wesentlichen Elemente der Umsetzungsverfahren festlegen. In (meist) späteren Entscheidungen legen die zuständigen Behörden die für den Vollzug nötigen Budgetmittel, Pflichtenhefte, Vollzugsprioritäten (oft im Rahmen eigentlicher Aktionspläne) etc. fest. Bei dieser Planung wird oft eine *prospektive Evaluation* in Form einer Vorwärtsplanung (forward mapping) durchgeführt. Dabei können bereits ausgehend vom Entwurf eines Verwaltungsprogramms die mutmasslichen Kosten etc. für die Outputs, die Impacts und die Outcomes abgeschätzt werden. Sinnvoll ist indessen auch eine Rückwärtsplanung (backward mapping). Diese geht von möglichen (simulierten) Reaktionen der Adressaten des Programms (Impacts) und den davon ausgehenden Outcomes aus und versucht daraus die notwendigen Outputs, Verfahren, Mittel oder Behördenarrangements zu bestimmen. Der Vorteil eines Einbezugs der Rückwärtsplanung liegt darin, dass Probleme der Umsetzung, besonders im Hinblick auf mögliche Ausweichreaktionen der Adressaten, deutlicher als bei der Vorwärtsplanung hervortreten.

Die reale Umsetzung ist aus dieser Sicht der eigentliche Test für die Planung und die prospektive Evaluation. Sie bedarf allerdings der Ergänzung durch die Evaluation: "Implementation and evaluation are the opposite sides of the same coin, implementation providing the experience that evaluation interrogates and evaluation providing the intelligence to make sense out of what is happening." (Pressman und Wildavsky 1984: xvi).

Evaluationen können sich auf *normative* und auf *kausalorientierte Fragen* beziehen (Chen 1990: 52 ff.; s. auch Bussmann 1995a: 60-62). Erstes beinhaltet den *Soll-Ist-Vergleich* zwischen Politikkonzept, Verwaltungsprogramm und Politikumsetzung (in Abbildung 2 die horizontale Dimension). So lässt sich mit einem solchen Vergleich etwa ermitteln, ob auf kantonaler Ebene die als nötig erachteten Mittel tatsächlich eingesetzt und ob die im Programm vorgesehenen Verfahren tatsächlich benutzt werden. Für diese Form des Soll-Ist-Vergleichs bis und mit Stufe Impacts wird in den USA der Begriff Evaluation (neuerdings auch performance review), im deutschsprachigen Raum häufig auch der Begriff Controlling verwendet.

Die zweite Perspektive ist kausalorientiert: Dabei wird versucht, das *Zusammenspiel der die Outcomes beeinflussenden Faktoren* - in ihrer Ganzheit oder bezogen auf einzelne Aspekte - zu erhellen (in Abbildung 2 die vertikale Perspektive). Darin liegt u.E. das Hauptschwergewicht der Evaluation. Solche Evaluationen können ein Verständnis für die hauptsächlichen Komponenten vermitteln, welche den Output, die Impacts und die Outcomes staatlicher Politik prägen.

In der Politikanalyse (vgl. Kapitel 7) werden üblicherweise die normative und die kausalorientierte Perspektive in der Weise kombiniert, als das normative Programm als ein den Outcome erklärender Faktor unter vielen konzipiert wird.

Es ist vor dem Missverständnis zu warnen, dass prospektive Evaluation immer vor und retrospektive Evaluation immer nach der Umsetzung stattfindet. Prospektive und retrospektive Evaluation sind in der Regel sehr viel enger mit dem Programmformulierungs- und Politikumsetzungsprozess verknüpft. So sind in manchen Politikumsetzungsstufen selbst Planungselemente enthalten; das gilt etwa für die strategische und operationelle Ressourcenplanung im Rahmen der Aktionspläne. Schon in frühen Stadien des Politikumsetzungsprozesses sind begleitende oder retrospektive evaluative Elemente eingebaut, namentlich in Form der bereits erwähnten Soll-Ist-Vergleiche. Planung und Evaluation sind mit der Politikumsetzung deshalb eine sehr enge Verbindung eingegangen, z.B. in Form

rollender Planung und sog. begleitender Evaluation. Solche Evaluationen erfüllen auch Bedürfnisse nach Frühwarnsystemen für öffentliche Politiken. Freilich ist gleichzeitig darauf hinzuweisen, dass für eine Evaluation der Outcomes (insb. Wirksamkeit) einer Politik eine gewisse Zeit seit Inkrafttreten des Programms vergangen sein sollte, damit nicht Fehlschlüsse im Sinne einer anfälligen Unwirksamkeit der Politik getroffen werden. Bei sog. latenten Wirkungen (erst längerfristig auftretende Outcomes; z.B. bei Forschungsergebnissen) sind sogar sehr viel längere Beobachtungszeiträume angebracht.

KAPITEL 5: EVALUATIONSGEGENSTÄNDE UND EVALUATIONSKRITERIEN

*Peter Knoepfel, Frédéric Varone, Werner Bussmann,
Luzius Mader*

In diesem Kapitel werden die im Kapitel 4 lediglich grob skizzierten acht Stufen der Politikgenerierung und -Umsetzung - soweit in unserem Kontext erforderlich (vgl. detaillierter etwa Faganini 1991; Germann 1986; Germann 1996; Kriesi 1980; Kriesi 1995; Linder 1988; Linder 1995; Moor 1992; Nüssli 1985; Wälti 1997) - etwas ausführlicher beschrieben. Wie angekündigt, sollen dabei insbesondere die für die einzelnen Politikgenerierungs- und -Umsetzungsstufen üblicherweise verwendeten Evaluationskriterien erläutert werden. Soweit möglich sollen dabei auch die Dimensionen prospektiver und retrospektiver bzw. normativer und kausalerorientierter Evaluationen berücksichtigt werden.

5.1. POLITIKKONZEPT - EMPIRISCHE KONSISTENZ UND LOGISCHE KOHÄRENZ DES WIRKUNGSMODELLS

In Policy-Cycles lässt sich ganz allgemein ein mehr oder weniger *sequentielles Ausfiltern* einzelner Aspekte des anfangs von den Politikakteuren als regulierungsbedürftig betrachteten Problems feststellen (Windhoff-Heritier 1987: 64 ff.; Knoepfel 1995e: 159 ff.). Diese Ausfilterung ist besonders markant bei der Politikgenerierung, sie lässt sich jedoch auch bei der Politikumsetzung ("Augenschliessen" vor interventionsbedürftigen Tatbeständen) und bei der Politikevaluation im engeren Sinne (bewusstes Wegschauen von wirkungslosen Politiken) feststellen.

In der ersten Stufe der Politikgenerierung, die ihren Abschluss im Politikkonzept findet, bezieht sich dieses Ausfiltern namentlich auf die in vielen Fällen zunehmend stärker eingeeengte Definition desjenigen Anteils des konstatierten Problems, der als *kollektives Problem* einer staatlichen Regulierung zugeführt werden soll. Aus ideologisch-parteilpolitischen, aber auch aus parteiübergreifenden Zweifeln an der Problemlösungsfähigkeit des heutigen Staates (vgl. dazu Kapitel 6.5.) werden in dieser Phase oft ursprünglich als kollektive, staatlich zu regulierende Probleme anerkannte Sachverhalte zu gesellschaftlichen und damit allein durch die gesellschaftlichen Akteure zu lösenden Probleme umdefiniert. Diese

Tendenz wird oft auch dadurch verstärkt, dass anlässlich der Wahl des Interventionstyps (Kapitel 3.1.) augenfällig wird, dass keiner der gängigen Typen das Problem angemessen lösen könnte. Es kann sich nämlich herausstellen, dass der rechtliche und/oder politische Preis für die bei den Politikadressaten notwendige Intervention zur Erreichung der (oftmals euphemistisch formulierten) Ziele zu hoch wäre, und weniger eingreifende Massnahmen die Ziele wiederum in weite Ferne rücken liessen. Schliesslich erfährt auch die Definition der Adressaten insofern eine Ausfilterung, als sich diese Gruppe in aller Regel im Laufe des Politikgenerierungsprozesses verkleinert; potentiell wichtige Adressaten werden aus der Policy entlassen, wenn sie über ein genügend hohes politisches Machtpotential verfügen und lautstark dagegen opponieren. Oft gebärt der Berg nach langen Gesetzgebungsverfahren dann eben eine Maus (definitives Politikkonzept und - ceteris paribus - definitives Verwaltungsprogramm).

Die für die Wirksamkeit der generierten Politik wohl entscheidende Variable in diesem Prozess ist das je nach politischem Standort und verfügbaren sozialwissenschaftlichen Informationen unterschiedlich formulierte *Wirkungsmodell*. Dieses enthält Mutmassungen über das problemverursachende Wirkungsgefüge (Kausalhypothese) und die möglichen Formen kollektiven Handelns, durch welche diesem Zustand abgeholfen werden kann (Interventions- und Handlungshypothese). Prägenden Einfluss auf die Art und Weise der Problempерzeption hat die *Kausalhypothese* ("causal hypothesis", vgl. Rossi und Freeman 1993: 120 f.). Sie gibt Antwort auf die Frage, wer oder was am politisch als untragbar bezeichneten Zustand in einem bestimmten gesellschaftlichen Bereich "schuld" bzw. dafür (auch ohne subjektives Verschulden) "objektiv" verantwortlich ist. Aus der Antwort auf diese Frage definieren sich die Politikadressaten. Werden diese "falsch" ausgewählt (vgl. Beispiel unten Kapitel 5.5.3. - Anwohnerparkkarte), so ist die Politik von Anfang an falsch programmiert und zur Unwirksamkeit verdammt. Bekanntlich ist die Geschichte voll von Schuldzuweisungen, die sich später als falsch erwiesen haben. Mächtige gesellschaftliche Gruppen können ihre Verantwortung oft auf schwächere Glieder der Gesellschaft abwälzen, weshalb denn auch die Formulierung dieser grundlegenden Kausalhypothese immer auch mit politischen Wertungen und mit der Art der Problempерzeption verbunden bleibt. Häufig kann aber auch die Wissenschaft nicht mit ausreichender Sicherheit Aufschluss über die problemverursachenden Mechanismen geben. So ist auch nach jahrelanger Forschung noch nicht ausreichend ge-

klärt, wie Luftschadstoffe und andere Einflussfaktoren (v.a. Klima) im Hinblick auf den Zustand von Wäldern zusammenwirken. Häufig kann erst ergiebige empirische Forschung zeigen, ob die einem Problem zugrundegelegten Faktoren wirklich zutreffen. Eine solche durchaus unsichere Kausalhypothese liegt z.B. dem Massnahmendispositiv zur Bekämpfung des Drogenhandels zugrunde, die stark auf dem Konzept des organisierten Verbrechens (mit hierarchischen Strukturen und einer Befehlsspitze) basiert. Es ist aber durchaus möglich, dass sich ein solches Konzept als Fiktion entpuppt und dass der Drogenhandel mehr einem sich selbst organisierenden Netzwerk entspricht. Aus letzterer Sicht müsste das Schwergewicht vermehrt auf andere, das Drogenproblem verursachende Faktoren in den Drogenkonsum- und -produktionsländern gelegt werden.

Das Politikkonzept bzw. das ihm zugrundeliegende Wirkungsmodell beinhaltet ferner eine *Interventionshypothese* ("intervention hypothesis", vgl. Rossi und Freeman 1993: 121 f.) darüber, wie das zu lösende Problem mittels staatlicher Politik gelindert oder gar zum Verschwinden gebracht werden kann. Die Interventionshypothese umschreibt, über welchen Ansatzpunkt staatliches Handeln auf den problemverursachenden Kausalmechanismus einwirken soll. Je nach vorgängiger Analyse der gesellschaftlichen Wirkungsmechanismen wird beispielsweise eine staatliche Intervention zur Bewältigung des Arbeitslosenproblems bei staatlichen Konjunkturmassnahmen, bei einer Veränderung des regulatorischen Umfelds oder bei einer Umgestaltung der Arbeitslosenversicherung ansetzen. In ähnlicher Weise wird in der Umweltpolitik abzuklären sein, ob Informationskampagnen bei den Adressaten eine höhere Bereitschaft auslösen, das Verhalten zu ändern, als Gebote und Verbote. Es müssen realistische Überlegungen entwickelt werden, auf welchen Pfaden eine in Aussicht genommene Politik ein gesellschaftliches Problem lindert oder behebt. Auch hier bestehen wissenschaftlich oft unzureichend erhärtete Erkenntnisse. Dies hängt mit der Dynamik sozialer Prozesse zusammen. Es können neue Konstellationen entstehen, welche die Wirksamkeit der gewählten Intervention in Frage stellen. Ein Beispiel dafür ist die Kampagne gegen das Zigarettenrauchen in den USA, welche durch die Anprangerung der gesundheitlichen Folgen und die starke Ächtung des Rauchens Gegenreaktionen auszulösen scheint; zumindest bei der jüngeren Generation ist neuerdings ein Anstieg des Zigarettenkonsums beobachtbar.

Wenn über die Wirkungen von Pilotprojekten (z.B. regionale Arbeitsvermittlungstellen) oder gegebenenfalls neuer, in Aussicht genommener

Instrumente (z.B. Informationskampagnen, prozedurale Vorkehrungen) bereits empirische Erfahrungen bestehen, können und müssen Überlegungen darüber angestellt werden, ob und inwieweit diese Erfahrungen generalisiert werden können. Diese Überlegungen werden als *Handlungshypothese* ("action hypothesis": Rossi und Freeman 1993: 122 f.) bezeichnet. Denn es besteht die Möglichkeit, dass im Rahmen einer "Freisetzung" Faktoren eine Rolle spielen (z.B. Gewöhnungseffekte, administrative Routinisierung), die im Rahmen des Pilotprojekts nicht in dieser Form aufgetreten sind. So stellt sich etwa bei einem Schulversuch oder einem Verwaltungsreformprojekt die Frage, ob sich die positiven Effekte, die während des Pilotversuchs zu beobachten waren, auch bei einer breiten Verwirklichung des entsprechenden Modells einstellen werden. Dabei sind beispielsweise Überlegungen darüber anzustellen, ob sich bei einer breiteren medizinischen Heroinabgabe an stark drogenabhängige Personen nicht Effekte einstellen könnten (z.B. internationale Implikationen, Entwendung von Drogen, Drogentourismus, Erhöhung der gesellschaftlichen Akzeptanz des Drogenkonsums), welche im Rahmen eines begrenzten Versuchs noch keine Rolle spielten.

Die *prospektive Evaluation* des Politikkonzepts orientiert sich an den *Kriterien* der empirischen Konsistenz und logischen Kohärenz. Die Überprüfung des Wirkungsmodells im Hinblick auf dessen *empirische Konsistenz* fragt danach, ob die wichtigsten Bestandteile des Wirkungsmodells empirisch ausreichend abgestützt sind oder zumindest nicht in Widerspruch zu empirischen Beobachtungen stehen. Aufgrund der bisherigen Erfahrungen empirisch nicht ausreichend abgestützt wäre beispielsweise die Interventionshypothese, wonach sich eine Konjunkturspritze ohne grössere Verzögerung wirtschaftlich bemerkbar machen und der Volkswirtschaft Innovationsimpulse geben würde. Ebenfalls auf zweifelhaften Grundlagen stünde eine Interventionshypothese, welche davon ausgeht, dass das Androhen erhöhter Strafen (ohne entsprechende Verstärkung der Kontrollen) einen unmittelbaren Einfluss auf das Verhalten der Adressaten habe. Bei der Konzipierung eines empirisch ausreichend abgestützten Wirkungsmodells ist es empfehlenswert, die Resultate vorliegender Evaluationsstudien zur Kenntnis zu nehmen. So wurde aufgrund der Ende der achtziger Jahre vorliegenden Erfahrungen mit Sackgebühren in einem Evaluationsbericht (INFRAS 1990: 22, 30, 39) die klare Empfehlung ausgesprochen, dass deren Einführung sinnvollerweise von erhöhten Informationsanstrengungen und wenn nötig auch von Kontrollen begleitet sein sollte. Diese Erfahrungen wurden beispielsweise bei der Einführung der

Kehrichtsackgebühr in der Stadt Bern nicht in ausreichendem Masse beachtet. Dies hatte Folgen (illegale Entsorgung über öffentliche Abfalleimer), die sich in dieser Grössenordnung hätten vermeiden lassen.

Bei der prospektiven Evaluation ist ferner die *logische Kohärenz* des Wirkungsmodells zu überprüfen. Dabei ist einerseits abzuklären, ob die einzelnen Bestandteile des Wirkungsmodells - auch in den Grössenordnungen der Wirkungsbeziehungen - zueinander passen. Zu fragen ist dabei insbesondere, ob mit dem vorgesehenen Instrument überhaupt ein namhafter Effekt auf die Problemdimensionen zu erwarten ist. Es geht hier um die innere Stimmigkeit oder Kohärenz des Politikkonzepts (intrapolicy Dimension). Andererseits ist abzuklären, ob das Politikkonzept nicht im Widerspruch zu bestehenden oder geplanten anderen öffentlichen Politiken steht, was gegebenenfalls zur Folge haben könnte, dass seine Wirksamkeit stark beeinträchtigt wird oder dass anderswo unerwünschte Nebenwirkungen eintreten (interpolicy Dimension). Letzteres kann etwa dann eintreten, wenn die Umweltpolitik von den Landwirten ein ökologisches Verhalten (z.B. Reduktion des Düngemiteleinsatzes) fordert, während die ökonomischen Anreizmechanismen der Landwirtschaftspolitik diesen Aspekten kein Gewicht zumessen.

Ob ein Wirkungsmodell und die ihm zugrundeliegenden Hypothesen tatsächlich (auch nach Inkrafttreten der Politik) zutreffen (*retrospektive Evaluation*), lässt sich meist erst nach einer Ermittlung ihrer Outcomes (s. Kapitel 5.6.1.) ermitteln (s. Kapitel 5.6.2.). Eine staatliche Politik kann zwar zu den erwünschten Outputs und gleichzeitig auch zu den erwarteten Verhaltensänderungen der Adressaten (Impacts) führen, wenn daraus kein namhafter Beitrag zur Linderung des Problems resultiert, dann ist dies ein untrüglicher Hinweis darauf, dass das zugrundeliegende Wirkungsmodell falsch konzipiert war oder überholt ist. So lag beispielsweise der amerikanischen Armutsbekämpfung der sechziger Jahre ein eindimensionales Bild einer "Kultur der Armut" zugrunde, welches sich in der Folge nicht mehr halten Hess. Das schweizerische Wirkungsmodell der Landwirtschaftspolitik war während langer Zeit ausschliesslich auf die Einkommenseffekte der Subventionierung ausgerichtet, ignorierte aber deren Wirkungen auf die Produktionsmengen und auf die Ökologie. Auch das ursprüngliche Wirkungsmodell der öffentlichen Verkehrspolitik, welches davon ausging, mit einer Angebotserweiterung und Verbilligung des öffentlichen Verkehrs allein Hesse sich ein namhafter Umsteigeeffekt erzielen, ist durch die Entwicklung falsifiziert worden. Mitunter geben auch unerwartete Reaktionen der Adressaten (Impacts) Hinweise auf ein falsch

konzipiertes Wirkungsmodell. Als Beispiel sei etwa auf das Weissweimportregime des Jahres 1996 hingewiesen. Weisswein konnte solange frei eingeführt werden, bis das entsprechende Kontingent ausgeschöpft war. Die darüber hinausgehenden Importe mussten verzollt werden. Die Behörden waren ihren eigenen Verlautbarungen zufolge offenbar überrascht, dass das Kontingent nach zwei Tagen bereits ausgeschöpft war. Bereits der Name dieses Vergabesystems ("Windhundverfahren") hätte aber eigentlich auf eine solche Reaktionsweise schliessen lassen müssen.

5.2. VERWALTUNGSPROGRAMM - KOHÄRENZ

5.2.1. Einführung

Die Ziele und die wesentlichen Umsetzungselemente öffentlicher Politiken finden sich in den staatlichen Handlungsprogrammen ("*Verwaltungsprogramme*" = "*Policy-Designs*"). Diese formulieren die vom Gesetzgeber artikulierten politischen Aufträge für die zu erarbeitende Problemlösung in Form von Rechtssätzen. Diese letzteren sind die primäre Legitimationsquelle einer öffentlichen Politik (Moor 1994: 31 ff. und 309. ff.). In formeller Hinsicht bestehen sie aus mehreren schriftlichen Dokumenten (v. a. Gesetze, Durchführungsverordnungen, Verwaltungsvorschriften etc.). Materiell beinhalten sie normative Sollwerte für die Art der angestrebten Problemlösung, die dazu bereitgestellten Mittel, die administrative Organisation der Politikumsetzung sowie für die Definition der Politikadressaten und deren Rolle bei der Politikrealisierung.

Wir vertreten die Auffassung, dass sich die Evaluation einer öffentlichen Politik primär an den Inhalten dieser Verwaltungsprogramme orientieren soll. Denn die Entscheidung darüber, welche Massnahmen "wirksam" sein sollen, ist nach diesem Ansatz primär als eine Entscheidung des Gesetzgebers anzusehen. Allerdings kann sich der Wille des parlamentarischen Gesetzgebers von demjenigen des Ordnungsgebers (Bundesrat) unterscheiden¹, oder er kann weitgehend verborgen bleiben. In diesen Fällen ist es Sache des Evaluationsteams, die genauen Inhalte des einschlägigen Verwaltungsprogramms zu definieren. Die Evaluation kann aber zeigen, dass diese gesetzgeberische Entscheidung "falsch" war, weil ihr unzureichende Annahmen über gesellschaftliche oder administrative

¹ Vgl. als Beispiel dafür: Organe parlementaire de contrôle de l'administration 1994.

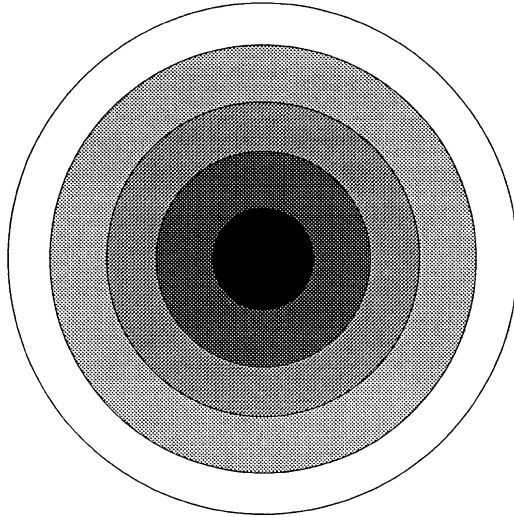
Wirkungszusammenhänge zugrunde lagen. Sie bedarf für solche Aussagen einer Analyse der Problemsituation und namentlich der Bedürfnisse der mit dem Verwaltungsprogramm anvisierten Gruppen.

5.2.2. Typische Elemente von Verwaltungsprogrammen

Jede Evaluation muss daher eine detaillierte Analyse des einschlägigen Verwaltungsprogramms vornehmen. Für diese Programmanalyse hat sich in der Praxis das folgende Schema bewährt, das nach dem Kern-Schale-Modell fünf aufeinander bezogene Elemente von Verwaltungsprogrammen unterscheidet (Abb. 3). Es wurde 1982 von Knoepfel und Weidner (1982: 93) als Instrument zur vergleichenden Analyse umweltpolitischer Verwaltungsprogramme entwickelt, seither aber für die Analyse vieler anderer Politiken verwendet (vgl. Knoepfel 1995e: 173 ff.). In der Evaluationsforschung dient es dem (prospektiv planenden oder retrospektiven) Test der Kohärenz bzw. der Vollständigkeit von Verwaltungsprogrammen.

**Abbildung 3 Die fünf konstitutiven Elemente eines
Verwaltungsprogrammes (Policy-Design)**

(nach Knoepfel und Weidner 1982: 93)



Legende:

Ebene	Funktion
Äussere Schalen	Instrumentelle und prozedurale Elemente
	Behördenarrangement, finanzielle Mittel und weitere Ressourcen
Innere Schalen	Operative Elemente (Instrumente)
	Evaluative Elemente
Kern	Konkrete Zielsetzungen

Verwaltungsprogramme weisen in der Regel folgende fünf unterschiedlichen Elemente auf:

5.2.2.1. Ziel(e)

Ziele umschreiben den mit der Problemlösung anzustrebenden gesellschaftlichen Zustand im Problemfeld. Auf Gesetzesebene sind solche Ziele oft recht abstrakt formuliert ("keine gesundheitsgefährdenden Luftverschmutzungen", eine "angemessene Wohnung" etc.). Auf der Ebene der einschlägigen Verordnungen und Verwaltungsvorschriften finden sich demgegenüber vielfach sehr konkrete quantifizierte und messbare Zielwerte (z.B. $30 \text{ pg SO}_2/\text{m}^3$ im Jahresmittelwert; 15 m^2 Wohnfläche pro Person etc.). Je konkreter solche Zielwerte formuliert sind, desto einfacher lässt sich ermitteln, ob diese erreicht wurden. Dementsprechend steigt die Chance der Politikbetroffenen, verbesserte Problemlösungen politisch oder gerichtlich einzufordern. Gerichtliche Schritte sind im Prinzip nur möglich, wenn entsprechende Werte in einer Rechtsverordnung niedergelegt sind. Um eine solche Einklagbarkeit zu verhindern, werden konkrete Ziele deshalb oft im Rahmen nicht einklagbarer Verwaltungsvorschriften festgelegt. Die Sozialwissenschaft nennt solche Handlungsprogramme, die über ausgebaute Ziele verfügen, sog. finale Programme (Luhmann 1973; Morand 1993).

Konkrete Ziele setzen die *Festlegung von Messzahlen oder Indikatoren* voraus, die sich auf die gesellschaftliche Wirkung der Programme beziehen (Wirksamkeitsindikatoren). Letztere sind nicht zu verwechseln mit Indikatoren, die Massnahmen oder Verwaltungsaktivitäten beschreiben. Ziel einer öffentlichen Politik ist nicht die Produktion von Massnahmen, sondern die Veränderung gesellschaftlicher Wirklichkeiten im Sinne der im Programm enthaltenen impliziten oder expliziten Vorstellungen. Ziel des Raumplanungsgesetzes ist es nicht, Bauten ohne Baugenehmigung zu verhindern, sondern eine "auf die erwünschte Entwicklung des Landes ausgerichtete Ordnung der Besiedlung" zu verwirklichen². Ziel des Investitionshilfegesetzes ist es nicht, Leistungen an Bergregionen zu entrichten, sondern die "Existenzbedingungen im Berggebiet zu verbessern"³.

² Artikel 1 des Bundesgesetzes vom 22. Juni 1979 über die Raumplanung (SR 700)

³ Artikel 1 des Bundesgesetzes vom 28. Juni 1974 über Investitionshilfe für Berggebiete (SR 901.1)

5.2.2.2. *Evaluative Elemente*

Namentlich jene Politiken, die mit vergleichsweise konkreten Zielbestimmungen operieren und die darüber hinaus einen stark technischen Charakter aufweisen, enthalten in ihren Verwaltungsprogrammen oft auch Angaben darüber, wann und welche Daten mit welchen naturwissenschaftlichen, ökonometrischen oder sozialwissenschaftlichen Methoden wo zur genauen Bestimmung des Zielerreichungsgrades erhoben werden müssen. Sie enthalten ausserdem Hinweise dazu, wie diese Daten zu interpretieren sind. Beispiele dafür finden sich in der Umweltpolitik (Definition der Messflächen und der chemischen Analysemethoden etc.)⁴, in der Wohnbaupolitik (Definition der anrechenbaren Wohnungsfläche) oder in der Einkommenspolitik (Definition der Zusammensetzung des berühmten Warenkorb⁵ für die Festsetzung der Preisindizes).

Der politische Stellenwert solcher evaluativer Elemente für die inhaltliche Politiksteuerung kann unter Umständen sehr bedeutsam sein, weshalb sich die politischen Auseinandersetzungen vor Ort nicht selten auf die Angemessenheit und die wissenschaftliche Vertretbarkeit vorgelegter Messdaten konzentrieren. Für Natur- und Sozialwissenschaftler ist es bekanntlich eine Binsenwahrheit, dass man je nach der gewählten Methode unterschiedliche Ergebnisse "herbeimessen" kann. Dies gilt auch für die sozialwissenschaftliche Politikevaluation (vgl. dazu Kapitel 12). Veränderungen der Messanlagen können u.U. vollständig wirkungslose Politiken im nachhinein "wirksam machen", ohne dass an der Substanz der getroffenen Massnahmen auch nur ein Jota verändert würde.

5.2.2.3. *Operative Elemente (Instrumente)*

Diese Programmbestandteile enthalten die für die Zielerreichung einer bestimmten Politik vorgesehenen Massnahmen. Sie sind die für den Charakter einer öffentlichen Politik vermutlich entscheidenden Programmelemente, weil sie ihre Interventionsintensität bestimmen, unmittelbare Rechte und Pflichten der Adressaten festlegen und damit Ausmass und Qualität der Politikleistungen determinieren.

⁴ Als Beispiel vgl. Ziffern 6 und 7 der Verordnung vom 15. Juni 1977 über die Beurteilung der Abbaubarkeit von grenzflächenaktiven Waschmitteln (SR 814.226.227)

⁵ Der Warenkorb wird jeweils im Rahmen von Bundesratsbeschlüssen definiert.

Das instrumentelle Setting, das sich in den operativen Elementen eines Verwaltungsprogrammes findet, hängt stark von der gewählten Interventionsform ab (unmittelbare Leistungserstellung, Anreize, Umverteilung, Überzeugung, gesellschaftliche Strukturbildung, Institutionen und Organisationen, vgl. Kapitel 3.1.)- Wegen ihrer (mehr oder weniger weitgehenden) Auswirkungen auf die Rechtsposition der Adressaten bedarf sie einer ausdrücklichen gesetzlichen Grundlage. Sowohl rechtlich wie verwaltungswissenschaftlich bedeutsam ist ferner, dass im Rahmen der operativen Elemente auch Angaben dazu gemacht werden, unter welchen Bedingungen welche Massnahmen zum Einsatz gelangen können oder müssen. Die Rechtswissenschaft spricht in diesem Zusammenhang von konditionalen Gesetzesbestimmungen, die in der Regel nach einem Wenn-dann-Schema aufgebaut sind: Wenn jemand ein Haus bauen will, dann braucht er eine Baubewilligung. Wenn eine übermässige Luftbelastung mit Schadstoffen vorliegt, dann muss der verursachende Betrieb eine Sanierung durchführen. Wenn jemand ohne sein Verschulden seinen Arbeitsplatz verliert, dann hat er Anspruch auf Arbeitslosenentschädigung etc. Im Rahmen der neueren Debatte zum New Public Management wird oft beklagt, dass öffentliche Politiken konditional über- und final untersteuert seien (Hablützel 1995: 499). In der Tat engen konditionale Programme den Spielraum der Verwaltung ein; sie sichern demgegenüber aber Vorhersehbarkeit und Rechtsstaatlichkeit (Knoepfel 1996a: 158 f.).

5.2.2.4. Behördenarrangement, finanzielle Mittel und weitere Ressourcen

Diese Programmelemente bezeichnen die für verschiedene Umsetzungsstufen zuständigen *Behörden* und *Amtsstellen*, und sie legen ausserdem fest, welche anderen Verwaltungsinstanzen gegebenenfalls je Umsetzungsstufe in den Politikvollzug einzubeziehen bzw. anzuhören sind. Man spricht in diesem Zusammenhang von der Kompetenzordnung einer Politik, die sich sowohl in der einschlägigen Gesetzgebung, als auch in den Satzungen und Pflichtenheften der betroffenen Dienststellen findet. Namentlich in der Agrar- oder Sozialpolitik gibt es im Rahmen dieser Programmelemente auch Beleihungen verbandlicher Träger. Infolge der primären Vollzugszuständigkeit der Kantone findet sich die Mehrzahl dieser verwaltungsstrukturellen Elemente helvetischer öffentlicher Politiken in der Regel in den von den Kantonen beigesteuerten Bestandteilen des

Verwaltungsprogramms ("Einführungs"- bzw. "Vollzugsgesetze"). Deshalb kommt es zu unterschiedlichen Kompetenzordnungen und damit auch zu variierenden Verwaltungsstrukturen innerhalb ein und derselben öffentlichen Politik. Allerdings kann der Bundesgesetzgeber im Interesse der Vereinheitlichung die kantonale Organisationshoheit teilweise durchbrechen und den Kantonen die Einrichtung spezialisierter Fachstellen verbindlich vorschreiben (z.B. in der Raumordnungs-, Gesundheits-, Umwelt- und - neuerdings (1996) - in der Arbeitsmarktpolitik).

Die Ausstattung der öffentlichen Politiken mit bestehenden oder neugeschaffenen Amtsstellen und Behörden verschiedener staatlicher Ebenen (im folgenden: "*Behördenarrangement*") ist nicht ohne Bedeutung für ihre Effektivität (vgl. Kapitel 5.4.3.). Ungeeignete Behördenarrangements können zu erheblichen Effektivitätseinbussen führen und dadurch die Tragweite der substantiellen Programmelemente erheblich abschwächen. Umgekehrt können besonders geeignete Arrangements Beschleunigungseffekte zeitigen, die zu Vorsprüngen in der Problemlösung einzelner Kantone gegenüber den anderen führen (vgl. dazu v. a. Kissling-Näf 1996: 69 ff. und 282 ff.). Die Entscheidung über diese verwaltungsstrukturellen Elemente stellt damit eine wichtige politische Weichenstellung dar. Dementsprechend kommt es denn auch nicht selten zu ausführlichen politischen Debatten über solche Kompetenzzuweisungen.

Politikumsetzung kostet Geld. Das gilt nicht nur für jene öffentlichen Politiken, die ohnehin mit *finanziellen Leistungen* operieren (z.B. Hochschulförderung, Strassenbau), sondern schlichtweg für jede öffentliche Politik. Denn es müssen hierzu Personal, minimale Sachmittel oder ganz einfach Lokalitäten für neue Verwaltungszweige zur Verfügung gestellt werden. Soweit dies nicht ausdrücklich vorgesehen ist, zahlt der Bund den Kantonen keine Beiträge an die Finanzierung des Vollzugs seiner öffentlichen Politiken; die Kantone haben dafür im Rahmen ihres allgemeinen Haushalts aufzukommen. Wenn dafür gesetzliche Grundlagen bestehen, können sie sich einzelne Vollzugsaktivitäten teilweise durch Gebühren seitens der Adressaten finanzieren lassen (z.B. Finanzierung teurer Emissionsmessungen im Rahmen der Luftreinhaltepolitik durch die Betreiber von Feuerungsanlagen).

5.2.2.5. *Instrumentelle undprocedurale Elemente*

Die im Rahmen öffentlicher Politiken zusammenwirkenden Amtsstellen und Behörden sollen jedenfalls nach der Vorstellung der normativen Verwaltungsprogramme untereinander und gegenüber den Adressaten Interaktionsformen pflegen, die es erlauben, den Austausch von Informationen, Sachmitteln, Geldleistungen etc. nachvollziehbar zu machen. Intransparenz, über den Tisch geschobene Geldscheine oder Geheimnistueri sollen verhindert werden, und Kommunikation soll in geordneten Bahnen erfolgen. So wollen es das Rechtsstaats- und das Demokratieprinzip. Zu diesem Zweck stellen die Verwaltungsprogramme öffentlicher Politiken den Akteuren eine beschränkte Zahl von *Handlungsformen* zur Verfügung, die sie bei ihrer Kommunikation gegenüber innen und aussen benützen sollen. Es sind dies etwa die Verfügung, der verwaltungsrechtliche Vertrag oder der Verwaltungsbescheid.

Diese Förmlichkeit administrativen Handelns dient der Objektivierbarkeit, der späteren Nachvollziehbarkeit und der definitiven Abnabelung und Verselbständigung der Politikleistung von den staatlichen Organen. Die Verfügung wird zum zirkulationsfähigen Wertpapier (Rechtstitel) und stellt für dessen Inhaber eine gegenüber anderen Gesellschaftsmitgliedern einsetzbare Handlungsressource dar (z.B.: Baubewilligung, Subventionsbescheid etc.). Die beschränkte Zahl (numerus clausus) anerkannter administrativer Handlungsformen, die überdies prinzipiell politikübergreifend und für Bundespolitiken meist bundeseinheitlich geregelt sind, dient nicht zuletzt der einfachen Zirkulationsfähigkeit und gesellschaftlichen Anerkennung solcher Politikoutputs.

Gerade bezüglich dieser instrumentellen Elemente von Verwaltungsprogrammen ist nachdrücklich daran zu erinnern, dass Verwaltungsprogramme nicht Verwaltungswirklichkeiten beschreiben, sondern anzustrebende Sollwerte festlegen. Gleich wie die tatsächliche Struktur der Behördenarrangements infolge faktischer Kontraktions- oder Erweiterungsprozesse von den normativen Vorgaben abweichen kann, oder Verwaltungsressourcen durch geschickten Stellenabtausch unabhängig vom Verwaltungsprogramm veränderbar werden, so treten auch und gerade bezüglich der Handlungsinstrumente deutliche Unterschiede zwischen den Sollwerten und der Realität auf. Vielfach werden aus den verschiedensten Gründen (Verwaltungsökonomie, Verhinderung von gerichtlichen Kontrollverfahren etc.) informelle Verwaltungsoutputs förmlichen Vorgehen vorgezogen. Solche informelle Outputs reichen in vielen Fällen auch aus, um

bei den Adressaten das erwünschte Verhalten zu bewirken. Sie sind daher oftmals effizienter als förmliche Handlungsweisen.

Wer Struktur sagt, sagt auch Prozedur. Die Struktur des Behördenarrangements einer öffentlichen Politik soll nicht aus einem chaotischen Gemenge unterschiedlicher Behörden und Amtsstellen bestehen, die nach aussen in dauernd ändernden Konstellationen auftreten, sondern in geordneten Interaktionen zwischen diesen einzelnen Entscheidungsträgern. Im Hinblick auf die Schaffung einer funktionsfähigen, vorhersehbaren und beliebig wiederholbaren Interaktion sehen Verwaltungsprogramme eine Vielzahl von *Verfahrensregeln* vor. Diese betreffen zum einen die innere Struktur der Behördenarrangements und regeln etwa das Vernehmlassungsverfahren, den Weg einzelner Verwaltungsvorgänge durch verschiedene Amtsstellen, interne Genehmigungs- oder Mitberichtsverfahren etc.

Wohl noch wichtiger als diese internen Verfahrensregeln sind jene prozeduralen Elemente, die Austauschprozesse zwischen den Behörden und den Adressaten oder den Politikbetroffenen "draussen" in der Gesellschaft regulieren. Die Erstellung von Politikleistungen erfordert in der Regel nicht nur aus Gründen der Sinn- und Konsensstiftung, sondern bereits aufgrund des rechtsstaatlichen Gebots der vorgängigen Anhörung der Betroffenen die Beobachtung minimaler *Verfahrensstandards*. Diese sind für Bundespolitiken im Verwaltungsverfahrensgesetz vom 20. Dezember 1968 (SR 172.021) niedergelegt. Sie dienen nach klassischem Verständnis dem Schutz der Bürgerinnen und Bürger vor Willkür und vor Grundrechtsverletzungen. Ausserdem sollen sie die Bereitstellung der für die Politikumsetzung erforderlichen Informationen seitens der Adressaten und der Politikbetroffenen sicherstellen. Darüber hinaus sollen sie auch die Behörden vor einer unkontrollierten Invasion durch stille Lobbyisten oder lauthals schreiende Interessenvertreter abschirmen.

5.2.3. Kohärenz und Rechtmässigkeit von Verwaltungsprogrammen

Zentrales Kriterium der Evaluation von Verwaltungsprogrammen ist die *Kohärenz* der verschiedenen Programmelemente. Dies ist in der föderalistischen Schweiz umso bedeutsamer, als die meisten Elemente von Verwaltungsprogrammen aus Teilentscheidungen des Bundes und diese konkretisierenden Teilentscheiden des betroffenen Kantons zusammengesetzt sind. Die wichtigsten Elemente eines Verwaltungsprogramms müssen

aufeinander abgestimmt sein, müssen sich ergänzen, dürfen sich nicht widersprechen. Bei der Analyse ist dabei nicht nur die innere Logik und Stimmigkeit der einzelnen Teile des Verwaltungsprogramms einzubeziehen (innere Kohärenz). Die Analyse muss auch die Übereinstimmung mit anderen (gegebenenfalls widersprechenden) staatlichen Handlungsprogrammen (äussere Kohärenz) umfassen. Damit kann sich die Kohärenzprüfung entweder auf die spezifische betroffene Politik beziehen, in welcher das Verwaltungsprogramm angesiedelt ist (Intrapolicy-Bereich) oder aber auch auf das gesamte Problemfeld, welches durch das betreffende Verwaltungsprogramm beeinflusst wird (Interpolicy-Bereich, vgl. Knoepfel 1995d: 209 ff.). Es liegt in der Natur der Sache, dass die Abschätzung der Kohärenz bei der Planung und prospektiven Evaluation oft weniger verlässlich ist als im Rahmen einer retrospektiven Evaluation.

Die Analyse von Verwaltungsprogrammen entlang den fünf soeben skizzierten Elementen stellt die Grundlage für eine Prüfung der Kohärenz von Verwaltungsprogrammen dar. Sie erlaubt, Widersprüche zwischen einzelnen Programmelementen (z.B. zwischen Zielfestlegungen und operativen Elementen oder zwischen Zielen und Ressourcen) aufzudecken, die sich unmittelbar auf die Qualität und Quantität des Politikoutputs auswirken können. Man spricht in diesem Zusammenhang von *"vorprogrammierten Vollzugsdefiziten"* (Knoepfel und Weidner 1982: 92). Diese sind nicht den Politikumsetzungs-, sondern den Politikprogrammierungsinstanzen zuzuschreiben. So können etwa ausbleibende Politikoutputs darauf zurückzuführen sein, dass das Verwaltungsprogramm für die Politikumsetzung "ungeeignete" Behördenarrangements festgelegt hat, die (vorhersehbar) nicht in der Lage sind, die postulierten Politikziele oder die vorgesehenen Politikleistungen zu realisieren. Eine andere Möglichkeit für vorprogrammierte Leistungsdefizite sind "falschen" Definitionen der Politikadressaten, die ihrerseits auf "falschen" Kausalanalysen (Kapitel 5.1.) beruhen können. Solche Programmierungsfehler sind ihrerseits darin begründet, dass die Ursachen des zu bekämpfenden gesellschaftlichen Problems in unzutreffender Weise den Verhaltensweisen einer Gruppe von Personen zugerechnet wird, die dafür nicht oder nur teilweise verantwortlich sind. Ein Beispiel für eine solche Fehlattribution stellen etwa die (im Kasten in Kapitel 5.4.3. dargestellten) Politiken gegen den privaten Pendlerverkehr in städtischen Agglomerationen dar, die (fälschlicherweise) an einer Reduktion der öffentlichen Parkierungsfläche ansetzen, obwohl der weitaus grösste Teil dieser Flächen in privater Hand ist.

Für Politiker ist es in der Regel schwierig abzuschätzen, inwieweit sich die von ihnen formulierten Ziele durch die vorgeschlagenen Massnahmen tatsächlich realisieren lassen. Meistens wird denn auch die Debatte über die vorzusehenden Massnahmen nicht mit Blick auf die Ziele, sondern aus der Sicht der "Härte des Eingriffs" etc. und damit aus der Perspektive der konkurrierenden Parteiideologien oder der finanziellen Aufwendungen für den Massnahmenvollzug etc. geführt. Ähnliche Entkoppelungen zwischen Zielgrössen und Massnahmen kommen aber auch im Rahmen hochtechnischer Programmformulierungsprozesse im Bereich der Standardfestsetzung etwa im Planungs- oder Umweltrecht, aber auch im Bereich der Sozialversicherungen vor. So werden beispielsweise aufgrund politischen Drucks seitens der betroffenen Branchenverbände im Umweltrecht anlagenspezifische Schadstoffmaximalwerte (sog. Emissions-Grenzwerte) so hoch angesetzt, dass bereits einfache physikalische Berechnungen genügen, um nachzuweisen, dass in Ballungsgebieten der Zielwert (sog. Immissions-Grenzwert) zwangsläufig überschritten wird. Damit wird die lobenswerte Zielgrösse bereits auf der Ebene des Verwaltungsprogramms schachmatt gesetzt.

Zu ähnlichen Programminkonsistenzen kann es auch zwischen Ressourcen für das Behördenarrangement und Politikzielen oder Politikinstrumenten kommen. Mit der zunehmenden Professionalisierung und den steigenden Anforderungen an die technisch-instrumentelle Ausstattung der Behörden verteuert sich der Vollzug öffentlicher Politiken. Im Rahmen der Stellenbewirtschaftung und der allgemeinen Haushaltsplanung sind diesbezüglich Prioritäten zu setzen. Oftmals fernab von den zentralen Akteuren der betroffenen Politiken werden auch im Rahmen solcher Budgetbeschlüsse wesentliche Entscheidungen über die Effektivität öffentlicher Politiken getroffen. Und vielfach werden sich die Parlamente der Tragweite solcher Beschlüsse erst im nachhinein bewusst. Bekanntlich sind die für die Umsetzung öffentlicher Politiken erforderlichen Ressourcen (Gehälter, Investitionen etc.) ein wichtiges Instrument der öffentlichen Wirtschafts- und Konjunkturpolitik. Die Wirksamkeit dieser Politik folgt nicht den gleichen Gesetzen wie jene der von solchen Budget-Entscheidungen ressourcenmässig betroffenen Politiken. Es ist deshalb für Aussenstehende nicht immer leicht, haushaltspolitische und sachpolitische Motive von Budgetbeschlüssen auseinanderzuhalten. Ausserdem stellen angeblich haushaltspolitisch motivierte Budgetkürzungen für bestimmte Ressorts in Tat und Wahrheit ein probates Mittel zur Begleichung einer

Rechnung zwischen Mitgliedern von Haushaltsausschüssen und Einzelpolitiken dar.

Neben dem Kriterium der Kohärenz spielt bei der Evaluation von Verwaltungsprogrammen in praktischer Hinsicht auch das Kriterium der *Rechtmässigkeit* eine bedeutsame Rolle. In politischen Systemen wie der Schweiz, deren rechtliche Ordnung vom Legalitätsprinzip geprägt ist, beruhen praktisch alle Politiken auf einer rechtlichen Grundlage, und ihre Ausgestaltung, insbesondere das konkrete Instrumentarium, ist rechtlich geregelt. Analysen öffentlicher Politiken können diesen rechtlich-normativen Aspekt staatlicher Politiken nicht ignorieren oder ausklammern. In der Praxis ist die rechtliche Regelung häufig ein zentrales Element oder zumindest ein wichtiger Ausgangspunkt.

Die Überprüfung der rechtlichen und materiellen Kohärenz der normativen Elemente einer Politik kann sowohl im Vorfeld des politischen Entscheids als auch im Nachgang zu diesem erfolgen. Sie kann daher proaktiv (und zugleich präventiv) oder retrospektiv sein. Die präventive Beurteilung der Rechtmässigkeit rechtlicher Regelungen wird im wesentlichen durch verwaltungsinterne Prozeduren und Organe sichergestellt. Was die Bundesebene anbelangt, sei in diesem Zusammenhang namentlich erwähnt, dass die Rechtmässigkeit aller Erlassentwürfe durch das Bundesamt für Justiz überprüft wird⁶. Im weiteren werden vor allem auch im Rahmen der Ausübung von Aufsichtsbefugnissen Rechtmässigkeitsüberprüfungen vorgenommen. Besondere Bedeutung kommt in dieser Hinsicht auf Bundesebene zwei erst vor wenigen Jahren geschaffenen Organen zu, nämlich der Dienststelle für Verwaltungskontrolle und der parlamentarischen Verwaltungskontrollstelle (vgl. Kapitel 2.4.3).

Ebenfalls zu erwähnen sind hier die Rechtmässigkeitsüberprüfungen, die anlässlich der Genehmigung kantonaler Erlasse durch den Bund vorgenommen werden. Wie eingangs dargelegt, umfassen Verwaltungsprogramme in der Schweiz sehr häufig Elemente, die auf unterschiedlichen staatlichen Ebenen angesiedelt sind. Der Vollzug der Programme obliegt in der Regel den Kantonen und Gemeinden, wobei der Begriff des Vollzugs nicht nur reine Normanwendung bedeutet, sondern meistens auch

⁶ Siehe Art. 7 der Verordnung über die Aufgaben der Departemente, Gruppen und Ämter (SR 172.010.15). Die Mitwirkung bei der Vorbereitung von Erlassen ist nicht ausschliesslich auf Fragen der Rechtmässigkeit beschränkt. Wichtige Querschnittsaufgaben kommen in diesem verwaltungsinternen Verfahren auch der Bundeskanzlei (Rechts- und Sprachdienst) zu. Siehe dazu auch die Richtlinien für die Vorbereitung und Erledigung der Bundesratsgeschäfte, vom 15. Oktober 1985.

den Erlass von inhaltlichen, organisatorischen und prozeduralen Ausführungsbestimmungen einschliesst, die zum Verwaltungsprogramm gehören. Diese Ausführungsbestimmungen unterliegen der Genehmigung des Bundes, soweit das Gesetz es vorsieht⁷. Die Genehmigung hat dabei konstitutiven Charakter, d.h. sie ist Gültigkeitsvoraussetzung, so dass die in diesem Zusammenhang erfolgende Überprüfung der Bundesrechtskonformität an sich präventiver Natur ist.

5.3. BEHÖRDENARRANGEMENT - EIGNUNG UND RECHTMÄSSIGKEIT DER VOLLZUGSSTRUKTUREN

Obwohl (namentlich auf der Ebene der kantonalen Einführungsgesetzgebung) wesentliche Vorentscheidungen über das einschlägige Behördenarrangement bereits im Verwaltungsprogramm enthalten sind (z.B. Zuständigkeitsordnung, interne Koordinationsmechanismen, Kommunikation mit anderen Dienststellen), muss der konkrete Aufbau eines neuen Behördenarrangements bzw. der Einbau einer neuen Politik in ein bestehendes Arrangement als eine von der Programmformulierung (teilweise) unabhängigen Politikumsetzungsstufe angesehen werden (1. Stufe der sog. Politikimplementierung; vgl. dazu schon Mayntz 1980: 10 ff.). Auch in der schweizerischen Politikforschung wird zunehmend evident, dass die Struktur des *Behördenarrangements* namentlich im Hinblick auf die Behörde, welche die Hauptlast der Politikumsetzung trägt, den Produktionsprozess der öffentlichen Politik stark prägt (Kissling-Näf 1996: 36 ff.; Knoepfel 1996b: 5, 39; 1995: 81 ff.; Ruch 1993; Kapitel 7.1.). Dabei interessiert in erster Linie, wie die Vollzugsstruktur die Qualität des Verwaltungsprodukte beeinflusst. Letztere fächert sich u.a. auf in Produktionsdauer, Widerspruchsfreiheit der Produkte, Klientennähe des Produktionsprozesses, angemessene Berücksichtigung der Interessen von Politikbetroffenengruppen, Produktionsaufwand (personelle und finanzielle Kosten pro Produktgruppe), Anzahl Schnittstellen zwischen den Akteuren für die Produktion einer Produktgruppe sowie Verfügbarkeit hinreichender Problemmonitoringskapazitäten (vgl. etwa Kanton St.

⁷ Siehe dazu Art. 7a des Verwaltungsorganisationsgesetzes (SR 172.010). Nach der 1989 beschlossenen Neuregelung ist die Genehmigung nur noch notwendig, wenn sie ausdrücklich in einem Bundesgesetz oder in einem allgemeinverbindlichen Bundesbeschluss vorgesehen ist. Dadurch und durch die Aufhebung von Genehmigungspflichten ist die Anzahl der Genehmigungen erheblich reduziert worden.

Gallen: Knoepfel, Baitsch und Eberle 1995). Überprüft wird damit sowohl die (outputbezogene) Wirtschaftlichkeit als auch die Zweckmässigkeit des administrativen Vollzugsgeschehens. Als Vergleichsmassstab dienen Werte aus Simulationen (Abschätzung der verschiedenen Grossen unter verschiedenen Ablauf- oder Aufbauorganisationen desselben Amtes) oder aus anderen Kantonen. Hinzu kommen (subjektive) Bewertungen durch die verschiedenen Zielgruppen und durch Betroffenenorganisationen (customer complaints index) sowie weitere Indikatoren (z.B. Zahl von Gerichtsverfahren). Solche Evaluationen bilden in der Regel den Ausgangspunkt für Verwaltungsreformen, in denen u.a. auch Elemente des New Public Management (vgl. dazu Schedler 1995 sowie Hablützel et al. 1995) zum Tragen kommen können. Im Zentrum stehen oft Bestrebungen nach Verfahrensvereinheitlichung oder Verfahrensverkürzungen (vgl. dazu Verwaltungskontrolle des Bundesrates 1994).

Bundesgesetze werden in der Schweiz in der Regel durch die Kantone vollzogen (oder zumindest haben die Kantone in den meisten Fällen einen grossen Anteil am Vollzug). Die kantonalen Vollzugsstrukturen sind somit für die praktische Umsetzung politischer Ziele und für den Erfolg einer Politik sehr bedeutsam (vgl. bereits Knoepfel 1979: 53 f.; Bussmann 1986: 23 f. und 39 ff.). Politikevaluation muss aus diesem Grund notwendigerweise auch der Frage nachgehen, ob die kantonalen Vollzugsstrukturen überhaupt eine korrekte und effektive Umsetzung einer Bundespolitik ermöglichen. Diese Frage ist mit dem Kriterium der *Eignung der Vollzugsstrukturen* angesprochen. Die entsprechende Prüfung erfolgt mitunter erst im Rahmen der Aufsicht des Bundes über den Vollzug durch die Kantone. Bereits im Rahmen der Politikentwicklung, d.h. aus prospektiver Warte, ist jedoch diesem Aspekt Rechnung zu tragen. In gewissen Fällen kann auch die vorgesehene Genehmigung kantonaler Ausführungsbestimmungen durch den Bund Anlass für eine Überprüfung der kantonalen Vollzugsstrukturen sein.

Bei dieser Überprüfung ist davon auszugehen, dass die Kantone in Fragen der kantonalen Organisation und Verfahren grundsätzlich autonom sind; Eingriffe des Bundes in die kantonale Organisations- und Verfahrensautonomie bedürfen zumindest einer besonderen Begründung. Andererseits dürfen die Kantone den Vollzug des Bundesrechts oder - in einem etwas umfassenderen Sinne - von Bundespolitiken auch nicht vereiteln. Der Bund muss daher abwägen zwischen der Organisations- und Verfahrensautonomie der Kantone einerseits und der adäquaten Aufgabenerfüllung andererseits; er kann in den kantonalen Vollzug nur soweit eingrei-

fen, als dies für die Verwirklichung eines politischen Programms erforderlich erscheint⁸.

Ein aktuelles Beispiel aus der helvetischen Praxis ist die Gesetzgebung über die Hilfe an Opfer von Gewaltverbrechen⁹. Die Opferhilfegesetzgebung überlässt den Vollzug entsprechend den allgemeinen Grundsätzen unserer föderalistischen Ordnung grundsätzlich den Kantonen. Sie verpflichtet jedoch die Kantone, Beratungsstellen einzurichten, die in der Lage sind, den Opfern von Straftaten die erforderliche medizinische, psychologische, soziale, materielle und juristische Hilfe zu bieten. Die Kosten für die Schaffung der notwendigen Strukturen für den Vollzug des Bundesrechts werden normalerweise von den Kantonen getragen. Um einen raschen und wirksamen Vollzug der Opferhilfegesetzgebung sicherzustellen, hat der Gesetzgeber in diesem Fall jedoch beschlossen, dass den Kantonen - zusätzlich zur Gewährung von Subventionen für die Förderung der Ausbildung des Personals der Beratungsstellen und der anderen mit der Hilfe an Opfer betrauten Personen - eine auf sechs Jahre befristete Finanzhilfe für den Aufbau der notwendigen Vollzugsstrukturen ausgerichtet werden soll. Allerdings müssen die Kantone dem Bundesrat alle zwei Jahre über die Verwendung dieser Aufbauhilfe Bericht erstatten (Art. 18 Abs. 2 des Gesetzes). Diese Berichte bilden die Grundlage für eine Evaluation der Wirksamkeit der Opferhilfe durch das zuständige Bundesamt (Art. 11 Abs. 4 der Verordnung)¹⁰. Ähnlich weitgehende Eingriffe in die kantonale Organisationshoheit hat der Bund kürzlich (1995) auch für den Bereich der Arbeitsmarktpolitik beschlossen (Schaffung regionaler Arbeitsvermittlungsämter gemäss dem 1995 revidierten Arbeitslosenversicherungsgesetz vom 23. Juni 1995¹¹; vgl. dazu Piazza 1997).

Nicht zuletzt geben die Vollzugsstrukturen auch wichtige Hinweise darauf, ob einer Politik tatsächlich eine überwiegend instrumentale Funktion

⁸ Zur Erörterung dieser Problematik aus rechtlicher Sicht siehe Saladin in Kommentar BV, Art. 3, Rz. 104.

⁹ Siehe Bundesgesetz und Verordnung über die Hilfe an Opfer von Straftaten (SR 312.5 und 312.51).

¹⁰ Nachdem die Opferhilfegesetzgebung auf den 1. Januar 1993 in Kraft getreten ist, haben die Kantone im Frühjahr 1995 erstmals ihre Berichte abgeliefert. Darauf aufbauend wurde ein Evaluationsbericht zuhanden des Bundesrates erarbeitet, der sich namentlich auf diese Rückmeldungen der Kantone, auf Erfahrungen der verschiedenen Vollzugsinstanzen und auf eine Studie abstützt, die sich speziell mit der Sichtweise der Opfer befasst (Boggio et al. 1995).

¹¹ Bundesgesetz über die organisatorische Arbeitslosenversicherung und die Insolvenzentschädigung (SR 837.0).

zukommt bzw. zukommen soll oder ob es sich primär um eine symbolische Politik handelt. Vollzugsstrukturen, die nicht geeignet sind, eine rasche und wirksame Konkretisierung und Durchsetzung eines Programms sicherzustellen, können bereits nach dem Kriterium der Kohärenz (Kapitel 5.1.) zumindest den Schluss nahelegen, dass diese Ziele gar nicht angestrebt werden.

5.4. AKTIONSPROGRAMME FÜR DEN VOLLZUG - ADÄQUANZ DER PLANUNG

Namentlich im Zusammenhang mit der Gewährung von Bundessubventionen verlangen verschiedene Bundesgesetze von den Kantonen die Vorlage von Ausbauprogrammen (Nationalstrassenbau, Hauptstrassen, Niveauübergänge; Hochschulbauten), von Sanierungsprogrammen (Luftreinhaltungspläne, Lärmschutz, Gewässerschutz, Altlasten etc.), von Richtplänen (Raumplanung) oder die Ausarbeitung von räum- oder projektbezogenen Plänen, die den eigentlichen Ausführungsplanungen vorgelagert sind (Nationalstrassen: generelle Projektierung; Raumplanung: Nutzungsplan). All diesen Planungswerken ist gemeinsam, dass sie der Produktion von Endoutputs vorgelagert sind. Sie bezwecken eine zeitliche Staffelung und inhaltliche Priorisierung der späteren Endoutputs im Hinblick auf die etappenweise Realisierung der strategischen Ziele. Sie geben Antwort auf die Frage, welche Adressaten und Politikbetroffenen mit den entsprechenden Politikleistungen früher als andere bedient werden sollen, welcher Aufwand dafür notwendig ist und inwieweit sich das zu lösende gesellschaftliche Problem durch diese Prioritätensetzung vermutlich lösen lässt. Die Priorisierung kann dabei in funktioneller (Art der betroffenen Aktivitäten), zeitlicher und räumlicher Hinsicht geschehen. Solche Prioritätensetzungen gehen bei regulativen Politiken oft mit der Entscheidung einher, partielle Vollzugslücken bei bestimmten Adressatengruppen für eine bestimmte Zeit in Kauf zu nehmen. Das daraus resultierende Privilegierungs-/Diskriminierungsmuster kann auf politischen Druck, auf erhöhten Problemdruck (in bestimmten Branchen oder Regionen) oder auf Bestimmungen im Verwaltungsprogramm zurückgehen (vgl. Terribilini 1995).

Zentrales Evaluationskriterium zur Beurteilung solcher Aktionsprogramme ist die *Adäquanz der Planung*. Die Planung muss ausreichende Grundlagen für eine zeitliche Staffelung der Umsetzungsaktivitäten bilden

und dabei vor allem eine effiziente Ressourcennutzung sicherstellen. Dies ist dann der Fall, wenn mit den gewählten Massnahmen und den bereitgestellten Ressourcen bei gegebener Verhaltensänderungsbereitschaft der Adressaten ein höherer Problemlösungsbeitrag erzielt werden kann als mit anderen Massnahmen. Bei diesem Kriterium spielt die ökonomische Grenznutzenüberlegung eine zentrale Rolle: Wie lässt sich unter Einsatz gegebener Mittel ein möglichst hoher Grenznutzen im Hinblick auf die Problemlösung erzielen? Die Evaluation hat dabei zunächst jene Adressatengruppe zu identifizieren, deren Verhaltensänderung einen möglichst grossen mutmasslichen Problemlösungsbeitrag ergibt. In einem zweiten Schritt ist zu prüfen, welche Adressatengruppe die höchste Verhaltensänderungsbereitschaft aufweist (z.B. jene Branche, in der ohnehin ein neuer Investitionszyklus bevorsteht, oder jene Gruppen, die wirtschaftlich eine Verhaltensänderung verkraften können). In einem dritten Schritt wird schliesslich geprüft, bei welcher dieser potentiell bedeutsamen Adressaten der Verwaltungsaufwand relativ am geringsten ist. Die Planung von Aktionsprogrammen beinhaltet damit prospektive Überlegungen zur Effektivität, Wirksamkeit und Effizienz.

Derartige Überlegungen fliessen auch in die Genehmigung solcher Pläne und in die Formulierung einer positiven oder negativen Stellungnahme durch die Bundesinstanzen oder gegebenenfalls durch kantonale Behörden ein, wobei den nachgeordneten Behörden auch ein ausreichender Ermessensspielraum einzuräumen ist.

Überprüfungen der Adäquanz der Planung enthalten oft mehr oder weniger spekulative Momente, deren empirische Evidenz lediglich aus Vergleichen mit ähnlichen Sachverhalten in anderen Kantonen bzw. in anderen Politikbereichen oder Zeitabschnitten besteht. Um retrospektive Evaluationen handelt es sich nur, wenn es etwa darum geht, die entsprechenden Pläne zu revidieren und dabei die zentralen Steuerungsdimensionen bzw. die Modalitäten der Prioritätensetzung im Lichte der Erfahrungen anders festzulegen. Von entscheidender Bedeutung ist dabei erfahrungsgemäss nicht die Definition der privilegierten, sondern diejenige der (vorläufig noch) diskriminierten Gruppen. Wie aus der Beobachtung des Vollzugs von Bundespolitiken bekannt ist, werden solche Politiken in der Regel schubweise umgesetzt, weshalb in den ersten Jahren bewusst mit Umsetzungsdefiziten gelebt werden muss. In der Regel wird man zu Beginn bei jenen Adressaten ansetzen, von deren Verhaltensänderung man unter den Voraussetzungen möglichst geringen Verwaltungsaufwands den grössten Beitrag für die Lösung des anstehenden Problems erwarten kann.

In der Umweltpolitik beispielsweise werden grosse Industriebetriebe als erste saniert, während die vielen kleineren Gewerbebetriebe unter Inkaufnahme eines Vollzugsdefizits vorläufig zurückgestellt werden. Ähnliche Priorisierungen müssen im Rahmen des Lärmschutzes (Grosse der exponierten Bevölkerungsgruppe) oder der Politik des öffentlichen Verkehrs (Umfang der neu angeschlossenen Populationen) vorgenommen werden.

5.5. OUTPUT, IMPACT UND EFFEKTIVITÄT

5.5.1. Output

Im Sinne der Ausführungen in Kapitel 4.2. handelt es sich bei der Output-evaluation nicht um Zwischen- sondern um Endoutputs. Diese umfassen die von der Verwaltung und von weiteren, mit hoheitlichen Aufgaben beladenen (privaten und parastaatlichen) Organen erbrachten Direktleistungen gegenüber den Politikadressaten. Solche bestehen in Verfügungen aller Art (Bewilligungen mit Auflagen, individuelle Verbote, Genehmigungen etc.), in Verfügungen und anschliessenden Geldzahlungen (Subventionen), Geldforderungen (Steuern) oder Gewalteinwirkungen, in einer direkten Leistungserbringung (Kontrollen, Eichungen, Ausbildungs-, Pflege- etc. -leistungen), in Beratungs- oder Umsetzungsleistungen oder in organisatorischen Vorkehrungen.

Der Output ist eine analytische Kategorie, welche auf unterschiedlichen Aggregationsniveaus ansetzen kann. Die Perspektive kann sich auf einzelne der oben angeführten Tätigkeiten beziehen, bei welchen vor allem überprüft wird, ob und inwieweit Vorschriften befolgt und Pläne umgesetzt wurden. Sie kann sich aber auch auf grössere Leistungseinheiten beziehen, die sich aus einzelnen Tätigkeiten zusammensetzen; für diese letzteren wird seit einiger Zeit der Begriff "Produkte" verwendet. Die Erfahrungen zeigen dabei, dass die Produkte staatlichen Handelns nicht ohne weiteres ermittelt, sondern in einem oft schwierigen Prozess der Selbstdefinition, bei welchem unterschiedliche Typologisierungsmuster verwendet werden können, konstruiert werden müssen. Eine solche bewusste Definition der Outputs bzw. Produkte ist aber in der Regel trotz des damit verbundenen Aufwands insbesondere für die Reorganisation von Verwaltungen sinnvoll, weil sich auf diese Weise die Aufbau- und Ablauforganisation stärker an den Erfordernissen der Produkte und ihrer Herstellung

orientieren lässt (produkt- anstatt gesetzesorientierter Verwaltungsaufbau). Darüberhinaus kann die gemeinsame Definition der zu erbringenden Produkte die "corporate identity" der betreffenden Dienststellen und parastaatlichen Organisationen fördern. Schliesslich erlauben solche Produktdefinitionen es auch, die Outputs gegenüber politischen Instanzen besser als bisher kenntlich zu machen, zu belegen und zu messen. Damit erlaubt sie eine stärkere Outputorientierung der Steuerung öffentlicher Verwaltungen, die etwa vom NPM gefordert wird (Schedler 1995:13, 47, 127 ff.).

Die Ermittlung des Outputs einer Verwaltungseinheit erfolgte bisher *selten prospektiv*. Zum Teil ist aufgrund der rechtlichen Erlasse nämlich nur schwierig abschätzbar, welche konkreten Tätigkeiten zu unternehmen und welche Produkte herzustellen sind. Oft resultiert daraus eine Unterschätzung des Umsetzungsaufwands. In dieser Beziehung besteht im öffentlichen Sektor durchaus noch ein Planungsdefizit.

Im Rahmen von Evaluationen erfolgt die Ermittlung der Outputs daher meistens *retrospektiv*. Dabei stehen Fragen wie die folgenden im Vordergrund: Wie setzen die vorgesehenen Vollzugsorgane ein Verwaltungsprogramm in die Praxis um? Welche Tätigkeiten üben sie aus (räumlich, zeitlich, personell etc.)? Welche Strategien entwickeln sie dabei? Diese Fragen sind besonders bedeutsam bei Politiken, die sich auf Instrumente abstützen, welche einer Konkretisierung und Umsetzung durch die Verwaltung bedürfen. Die Verwaltung - die, in einem umfassenden Sinn, auch private oder parastaatliche Organisationen einschliesst, die allenfalls mit dem Vollzug betraut sind - ist in diesen Fällen sozusagen eine unerlässliche Relais-Station, ohne deren Tätigkeit politische Entscheide nicht zum Tragen kommen.

Bei der Formulierung der Endoutputs kommt den Vollzugsinstanzen eine zentrale Stellung zu. Von diesen hängt es zu einem guten Teil ab, ob eine Politik zu den erhofften Ergebnissen führt oder nicht. Insbesondere können die Vollzugsorgane durch die Art und Weise, wie sie vom ihnen eingeräumten Ermessen Gebrauch machen, die praktische Umsetzung einer Politik wesentlich beeinflussen. Es erstaunt deshalb nicht, dass die Outputs gerade auch für Aufsichtsinstanzen von grossem Interesse sind und dass zahlreiche Gesetze *Berichtspflichten* enthalten, mittels derer die Vollzugstätigkeiten qualitativ oder quantitativ erfasst werden. Diese Berichtspflichten ermöglichen oder erleichtern zumindest die Aufsicht über den (in der Regel kantonalen) Vollzug. Sie lassen sich als Grundlage für ein eigentliches Monitoring der konkreten Vollzugstätigkeiten verwenden

(vgl. Kapitel 8). Als Beispiel sei hier etwa die Berichtspflicht im Bundesgesetz über den Erwerb von Grundstücken durch Ausländer erwähnt¹². Ein geradezu klassisches Beispiel sind die Berichte der Kantone über die Verwendung des Alkohol-Zehntels: Gemäss Artikel 44 Absatz 2 des Alkoholgesetzes geht der Reinertrag der Eidgenössischen Alkoholverwaltung zu 90 Prozent an den Bund und zu 10 Prozent an die Kantone. Der Anteil der Kantone ist nach Art. 45 Abs. 3 zur Bekämpfung des Alkoholismus sowie des Suchtmittel-, Betäubungsmittel- und Medikamentenmissbrauchs zu verwenden, wobei die Kantone dem Bund jährlich Bericht zu erstatten haben über die Verwendung ihres Anteils¹³.

Auch bei Regelungen, die primär an Private gerichtet sind und nur subsidiär eine Intervention staatlicher Organe erfordern (z.B. Kontrolle der Einhaltung einer Norm und Ergreifen von Sanktionen bei Missachtung), spielen die Tätigkeiten der Vollzugsinstanzen eine grosse Rolle. Der Grad der Befolgung einer Norm hängt nämlich wesentlich davon ab, ob Kontrollen vorgenommen und die für die praktische Durchsetzung einer Norm allenfalls notwendigen Sanktionen auch tatsächlich ergriffen werden.

Bei einer Analyse des Outputs müssen in der Regel je nach Fragestellung eine Vielzahl von *Daten* erhoben werden, beispielsweise über die Quantität, die zeitliche, räumliche und adressatenspezifische Verteilung sowie die inhaltliche Ausrichtung sämtlicher behördlicher Outputs, die von den Vollzugsbehörden im Zusammenhang mit der untersuchten Politik erbracht wurden. Im Hinblick auf die Impact- und Outcomedaten (Kapitel 5.4.1. und 5.5.2.) sind ausserdem auch Outputs anderer, nicht im Zentrum des Interesses der untersuchten öffentlicher Politiken zu dokumentieren, die vermutlich einen Einfluss auf das Verhalten der Adressaten bzw. der Politikbetroffenen ausgeübt haben. Bei grosser Häufigkeit der Outputs empfiehlt es sich, diese graphisch zu erfassen (Outputprofile).

Zur vollständigen Erfassung der Outputdaten ist eine klare Definition der Adressaten erforderlich, an die sich die Outputs richten. Die Adressaten (natürliche und juristische Personen) werden im schweizerischen Kontext nach rechtsstaatlichen Grundsätzen normalerweise bereits in den Verwaltungsprogrammen definiert als jene Personen, denen die Behörden zur Erreichung der Ziele der jeweiligen Politiken Rechte und Pflichten

¹² Art. 24 Abs. 3 des Bundesgesetzes (SR 211.412.41); siehe auch Art. 20 der Verordnung über den Erwerb von Grundstücken durch Personen im Ausland, welcher die Details der Erstellung einer Statistik in diesem Bereich präzisiert (SR 211.412.411).

¹³ Art. 45 Abs. 2 Alkoholgesetz (SR 680).

übertragen können oder deren Verhalten durch anderweitige behördliche Aktivitäten beeinflusst werden soll. Gerade bei der Outputevaluation ist immer wieder daran zu erinnern, dass Politiken niemals "nur" Outputs, sondern Wirkungen erzielen sollen. Evaluationen, die sich nur auf die Outputs, nicht aber auf die Impacts (Kapitel 5.5.2.) und auf die Outcomes (Kapitel 5.5.3.) konzentrieren, sind unvollständig, weil sie Gefahr laufen, Outputs mit Politikwirkungen zu verwechseln.

5.5.2. Impact

Als Impacts bezeichnen wir die realen Effekte öffentlicher Politiken bei den Politikadressaten: In welcher Weise und in welchem Ausmass fuhr eine Politik zu einer Verhaltensänderung bzw. zu einer Nicht-Änderung des Verhaltens (Stabilisierung von Verhaltensweisen, die ohne Intervention geändert würden)? Die Beantwortung dieser Frage gibt Aufschluss darüber, ob ein Instrument, dass für die Realisierung einer Politik eingesetzt wird, überhaupt geeignet ist, die erhoffte oder erwartete Verhaltensänderung in Wirklichkeit zu bewirken. Es geht mithin um die Frage der unmittelbaren Wirkung der Outputs bei diesen Adressaten. Zu erinnern ist allerdings daran, dass dabei nur die angestrebte Verhaltensänderung, nicht aber die reale Problemlösung erfasst wird, die durch diese Verhaltensänderung eintreten sollte.

Die Evaluation der Impacts begnügt sich nicht damit, die tatsächlichen Verhaltensweisen zu erfassen, sondern sie interessiert sich für die Wirkungsbezüge zwischen einer Politik und ihren Outputs und den wahrnehmbaren Verhaltensweisen ihrer Adressaten. Sie beschränkt sich also nicht darauf zu prüfen, ob das tatsächliche Verhalten dem normativen Modell entspricht, sondern schliesst eine Analyse der Kausalbeziehungen ein: vom Impact einer Norm oder eines normanwendenden Outputs kann nur dann gesprochen werden, wenn das tatsächlich beobachtbare Verhalten dem normativen Modell entspricht und wenn dieser Umstand auch wirklich auf die Norm und/oder den darauf aufbauenden Output zurückzuführen ist. Nicht selten trifft diese Gleichung entgegen der Meinung von Vertretern der politikumsetzenden Verwaltung nicht zu. Die in der realen Welt auftretenden Verhaltensänderungen sind öfter als von diesen Verwaltungen angenommen auf andere Faktoren (vgl. Kapitel 3.3.) zurückzuführen; in Frage kommen dabei Outputs anderer Politiken (z.B. für umweltrelevante Verhaltensänderungen nicht Outputs der Umwelt-, sondern solche der Energie-, der Agrar- oder gar der Fiskalpolitik), Ver-

änderungen der Marktbedingungen (z.B. Entwicklung der relativen Energiepreise) oder gesellschaftlicher Druck seitens der Mitbewerber, Konsumenten, Nachbarn etc., die sich direkt (und damit ausserhalb irgendeiner öffentlichen Politik) auf die Politikadressaten auswirken.

Bereits klassische Beispiele für die Evaluation von Impacts sind die zahlreichen Studien zu verschiedenen Normen im Bereich des Strassenverkehrs: Alkohol am Steuer (drinking and driving), Gurtentragpflicht und Geschwindigkeitsbeschränkungen (EJPD 1975; Arbeitsgruppe Verkehrssicherheit 1983; Universität Zürich 1977; Volvo Car Corporation 1980). Zu diesen Themen werden seit bald 30 Jahren in zahlreichen Ländern Evaluationen durchgeführt. Dies ist sicher nicht nur dem politischen Interesse an diesen Themen zuzuschreiben, sondern auch dem Umstand, dass es sich dabei um vergleichsweise einfache Fragestellungen handelt, dass die Bezüge zwischen staatlichen Normen und individuellem Verhalten leicht quantifizierbar und mit relativ einfachen Evaluationsdesigns (Vorher/Nachher-Vergleiche, Kontrollgruppen) erfassbar sind.

Bei der Ermittlung der Impacts werden etwa Angaben zum Realverhalten der Adressaten vor und nach der erfolgten Massnahme erhoben. Klassischerweise werden dabei etwa Daten zu Verstössen gegen behördliche Anordnungen, zu Kontrollbefunden, Sanktionen und zu Dunkelziffern¹⁴ beigezogen. Zu den Impactdaten gehören, wie gezeigt, auch Angaben zu jenen Verhaltensänderungen, die gegebenenfalls auf Outputs anderer (möglicherweise entgegenstehender) öffentlicher Politiken zurückzuführen sind, oder Angaben zu möglichen Verhaltensänderungen, die (vermutlich) nicht aus öffentlichen Politiken erklärbar sind (z.B. sich wandelnde gesellschaftliche Wertvorstellungen oder ökonomische Einflussfaktoren).

5.5.3. Effektivität

Bei der Evaluation der Impacts kommt das Kriterium der Effektivität (effectivité, effectivity/compliance) zum Tragen. Dieses misst den Grad der Übereinstimmung von rechtlich geforderten Politikanordnungen und realem Adressatenverhalten. Die Überprüfung der Effektivität einer öffentlichen Politik bedingt einen systematischen Vergleich zwischen dem "Soll-Vollzug" gemäss den Bestimmungen des Verwaltungsprogramms,

¹⁴ Diese sind bekanntlich sehr schwierig zu ermitteln, weil der Grad der tatsächlichen gesellschaftlichen Ächtung von Straftatbeständen oft stark variiert.

gegebenenfalls vorhandener Aktionspläne und - vor allem - den Outputs und den ausgelösten Verhaltensänderungen bei den Adressaten (Impacts). Je nach Art der untersuchten Massnahme eignen sich für die Messung und die Beurteilung dieser Effektivität unterschiedliche Indikatoren: Es sind dies etwa der *Vollzugsgrad* bei öffentlichrechtlichen Normen, die der Umsetzung durch Verwaltungsinstanzen bedürfen; der *Befolgungsgrad* bei Geboten und Verboten; der *Grad der Inanspruchnahme* bei Normen mit Anreizcharakter oder der *Beachtungsgrad* bei persuasiven Massnahmen (Arbeitsgruppe "Gesetzesevaluation" 1991: 15).

Die Evaluation der Effektivität bezieht sich somit auf den Soll-Ist-Vergleich im Sinne eines Output-Impact-Vergleichs. Sie kann sich aber auch auf einen Soll-Ist-Vergleich zwischen Verwaltungsprogrammen und Outputs bzw. Impacts beziehen. Obgleich bei Outputs und Impacts verschiedene Akteure angesprochen sind, sind beide Kategorien eng miteinander verknüpft. Um reale Verhaltensänderungen bei den Adressaten zu bewirken (Impacts), sind fast immer behördliche Aktivitäten (Outputs) nötig: Die Einhaltung von Geboten und Verboten muss überprüft, Leistungen müssen erbracht, finanzielle Beiträge ausgerichtet und Bussen eingezogen oder Informationen übermittelt werden. Umgekehrt werden viele behördliche Outputs erst erbracht, nachdem private Personen oder Gruppen tätig geworden sind: Untersuchungen werden auf eine Klage hin eingeleitet, Bewilligungen aufgrund von Baueingaben erteilt, Beiträge aufgrund von Gesuchen zugesprochen, Bussen nur bei entsprechendem Fehlverhalten ausgestellt und Informationen erst auf Anfrage erteilt. Outputs können nicht nur aufgrund des Verhaltens staatlicher Organe, und Impacts nicht nur aufgrund des Verhaltens der Adressaten ausbleiben; in manchen Fällen sind tatsächlich beide Gruppen beteiligt. In solchen Fällen sollen unter dem Kriterium der Effektivität Outputs *und* Impacts in die Politikevaluation einbezogen werden. Namentlich bei der Erbringung staatlicher oder öffentlich unterstützter Dienstleistungen (z.B. Bildung, Sozialarbeit, Arbeitslosenberatung, Drogentherapie, psychiatrische Kliniken) wirken Leistungserbringer und Kunden dermassen eng zusammen, dass die Qualität der Ergebnisse oft von der Qualität der gegenseitigen Kooperation abhängt.

Die Effektivitätsüberprüfung konzentriert sich auf quantitative und qualitative Aspekte. Die quantitative Überprüfung soll darauf Antwort geben, ob überall dort, wo sich politikrelevante Adressaten befinden, entsprechende Outputs produziert worden sind (klassische Frage nach den sog. Vollzugsdefiziten) bzw. ob und inwiefern diese Outputs auch entspre-

chende Verhaltensänderungen bewirkt haben. Die Definition der Adressaten befindet sich in den operativen Elementen des Verwaltungsprogramms (Definition der Massnahmen). Die qualitative Evaluation bezieht sich auf die inhaltliche Ausgestaltung und auf die gewählten Verfahrensweisen. Die Sollwerte finden sich in den operativen, in den instrumentellen und in den prozeduralen Elementen des Verwaltungsprogramms, sowie in den kasuistischen Entscheidungen der Vollzugspolitik vor Ort.

Das Kriterium der Effektivität hat stark normativen Charakter. Es hat deshalb vor allem in stark verrechtlichten Politiken (etwa Frankreich, Deutschland, z.T. auch Schweiz) eine grosse Bedeutung. Hier sind die Vollzugsaktivitäten auf Gesetzes- und Verordnungsstufe oft detailliert festgeschrieben. Eine Evaluation von Effektivität ist dort schwieriger, wo die Umsetzung rechtlich weniger stark fixiert ist und wo den Vollzugsbehörden ein grösserer Handlungsspielraum eingeräumt wird (z.B. USA und andere angelsächsische Länder). In solchen Kontexten steht eher eine kausalerorientierte Evaluation der Effektivität im Sinne einer realen Normbefolgung ("compliance") oder im Zusammenhang mit strafrechtlich relevanten Tatbeständen ("abuse", "fraud") im Vordergrund. Dass die Umsetzung nicht immer in den geplanten (normativen) Pfaden verläuft, wird nicht als dermassen bedeutsam erachtet; seit Pressman und Wildavskys (1973) klassischer Implementationsstudie wird dieser Umstand kausal-analytisch der "complexity of joint action" zugeschrieben.

Die Evaluation der Effektivität kann zum einen die *Ist*-Outputs und -Impacts mit den vorgeschriebenen oder geplanten (=Soll-)Outputs und -Impacts vergleichen und die entsprechenden Übereinstimmungen oder Abweichungen dokumentieren. Solche Evaluationen stellen oft nichts anderes als eine systematische Überprüfung der Rechtmässigkeit des Vollzugs dar. Sie verfolgen eine Kontroll- oder (neudeutsch) Controlling-Absicht. Eine solche normative Perspektive kann durchaus wertvolle Ergebnisse zeitigen und beispielsweise darüber Aufschluss geben, inwieweit Normen überhaupt befolgt, Beiträge abgerufen, Informationen beachtet werden u.a.m. Daraus können namentlich in prospektiver Hinsicht Schlüsse über die Adäquanz des Einsatzes der entsprechenden staatlichen Instrumente gezogen werden.

Zum zweiten lassen sich dabei in einer kausalerorientierten Perspektive (vgl. Abb. 2) die *Wirkungszusammenhänge* öffentlicher Politiken rekonstruieren. Eine solche systematische Untersuchung des Zusammenspiels der Aktivitäten der öffentlichen Verwaltung (Outputs) und der Verhaltensweisen der durch die staatlichen Massnahmen anvisierten Adressaten

(Impacts) fragt zwar (noch) nicht nach den beabsichtigten Wirkungen (Outcomes) einer öffentlichen Politik. Gleichwohl ist dieser - wirkungstheoretisch beschränkte - Output-Impact-Vergleich bedeutsam, da die Effektivität einer öffentlichen Politik oftmals eine notwendige Voraussetzung für deren Wirksamkeit ist; defizitäre Ergebnisse einer öffentlichen Politik können oft bereits auf fehlerhafte oder ungenügende Outputs oder Impacts zurückzuführen sein. So stellt man nicht selten fest, dass öffentliche Politiken z.B. infolge ungeeigneter Vollzugsstrukturen und -prozeduren gar keine oder unzureichende administrative Outputs produzieren. Für solche ineffektiven (outputlosen) Politiken besteht kaum eine Chance, die Adressaten jemals zu erreichen.

Ineffektive (outputlose) Politik: Energiepolitische Massnahme in kommunalem Kontext

Mit der Verabschiedung von Art. 5 des Energienutzungsbeschlusses vom 14. Dezember 1990 (SR 730.0) wurde die Installation von neuen ortsfesten Elektroheizungen bewilligungspflichtig. Danach sollten die nach kantonalem Recht zuständigen Behörden die Bewilligung für eine Elektroheizung nur erteilen, wenn gewisse Bedingungen kumulativ erfüllt sind (z.B. keine Anschlussmöglichkeit an Gas oder Fernwärme, Unverhältnismässigkeit einer Wärmepumpe). Verschiedene Kantone haben die Gemeinden mit dem Vollzug dieser Bundesvorschrift beauftragt.

Eine Evaluation des Vollzugs und der energetischen Auswirkungen der Bewilligungspflicht für Elektroheizungen (Aebischer et al. 1993) zeigt, dass die Delegation der Bewilligungszuständigkeit an die Gemeinden die Effizienz dieser Politik beeinträchtigen kann. Wegen mangelnder fachlicher Kompetenz und Routine sind die Milizbeamten auf kommunaler Ebene bei der Operationalisierung und Beurteilung der technisch-ökonomischen Bewilligungskriterien häufig überfordert. In kleineren und mittleren Gemeinden spielen auch Interessenverflechtungen (z.B. direkte Beteiligung der Gemeinde am lokalen Elektrizitätswerk) eine entscheidende Rolle für die mangelhafte Qualität des Vollzugs dieser energiepolitischen Massnahme. Diese und weitere Faktoren (z.B. soziale Kontrolle) führen oft dazu, dass die Bewilligungspflicht gar nicht vollzogen wird (d.h. keine Outputs). Ausserdem bewirken Konjunktur, Wertewandel, Selbsthilfe der Betroffenen, soziales Lernen usw., dass sich die geforderte restriktive Bewilligungspraxis nur mit grosser Zeitverzögerung durchsetzt (d.h. ungenügender Vollzugsgrad).

Es finden sich aber auch öffentliche Politiken, deren Behördenarrangements durchaus beachtliche Verwaltungsaktivitäten hervorbringen, deren Outputs aber bei den Adressaten keine der beabsichtigten Verhaltensänderungen auslösen. Dermassen ineffektiven (impactlose) Politiken sind **z.B.** auf die fehlende Akzeptanz des gewählten Instruments bei den Adressaten

oder auf eine falsche Einschätzung der Organisations- und Partizipationsfähigkeit dieser Gruppen zurückzuführen¹⁵.

Ineffektive (impactlose) Politik: Städtische Anwohnerparkkarte, Privatparkplätze und Pendlerverkehrsaufkommen

Die Bewirtschaftung öffentlicher Parkplätze gilt heute in der städtischen Verkehrsplanung als die Schlüsselgrösse zur Kanalisierung, Plafonierung und Reduzierung des Motorisierten Individualverkehrs (MIV), insbesondere des Pendlerverkehrsaufkommens. Die Anwohnerparkkarte als neu eingeführte verkehrspolitische Massnahme sieht eine Parkzeitbeschränkung mit blauer Zone vor, die den Pendlerverkehr aus den Wohnquartieren verdrängen und das Parkplatzangebot und die Lebensqualität für die Anwohner und das ansässige Gewerbe verbessern soll.

Evaluationsstudien zum Beitrag des Parkkartenmodells zur Reduzierung des Pendlerverkehrsaufkommens in den Städten Zürich, Basel und Bern (Schneider et al. 1990; Schneider, Häberling und Keiser 1995) kommen zu dem Schluss, dass eine solche Massnahme zum grössten Teil ineffektiv (impactlos) bleibt. Zwischen 70 und 85% der Pendler mit MIV verfügten schon vor der Einführung der Anwohnerparkkarte über einen Privatparkplatz. Die eindeutige Mehrheit der anvisierten Zielpopulation (d.h. die Pendler) sind im Besitz eines (firmen-)eigenen Privatparkplatzes, so dass sie ihr Mobilitätsverhalten gar nicht ändern müssen (somit kein Umstieg auf öffentlichen Verkehr).

Weitere Beispiele solcher impactlosen, sich mitunter aber durch hohe Kosten und hohe Outputproduktion auszeichnenden Politiken finden sich etwa im Strassenverkehr (unbefriedigende Reduktion der Durchschnittsgeschwindigkeit trotz aufwendiger Kontrollen und hinaufgesetzter Bussenstrafen), im Steuerwesen (fortwährende Steuerhinterziehung) oder im Bereich anderer "Bagatell-" oder "Kavaliersdelikte". Diese und andere Beispiele aus dem Alltag zeigen einmal mehr, dass die Finalität öffentlicher Politiken eben nicht in der Produktion von Outputs, sondern in der über diese Outputs zu erreichenden Verhaltensänderungen der Politikadressaten in der realen Welt und der auf diese Weise zu bewerkstellenden Problemlösung liegt. Output-Produktion als solche mag Personal beschäftigen und Ressourcen konsumieren und auf diese Weise gegebenenfalls die Konjunktur beleben; sie kann und darf indessen nicht als Selbstzweck angesehen werden. Obwohl diese Feststellung recht einleuchtend ist, droht sie im bürokratischen Alltag mitunter vergessen zu gehen. Und selbst neuere Konzepte des öffentlichen Managements, wie etwa das NPM

¹⁵ Zum Thema Charakteristika der Zielpopulation und Wirksamkeit von "policy tools" siehe z.B. Scharpf (1983) und Windhoff-Héritier (1989).

in seiner "outputorientierten" Variante ist dagegen nicht gefeilt (Knoepfel 1996a: 158 ff.).

5.6. OUTCOMES, WIRKSAMKEIT UND WIRTSCHAFTLICHKEIT

5.6.1. Outcomes

Namentlich aus politischer Sicht gilt das Hauptinteresse von Evaluationen der Frage, ob die Ziele oder Zwecke einer Politik mit den eingesetzten Massnahmen und Ressourcen tatsächlich erreicht werden können. Die Verhaltensänderungen, die dafür als notwendige Voraussetzungen angesehen werden, interessieren oft höchstens mittelbar. Wichtiger als die Outputs und Impacts sind aus dieser - politischen und wissenschaftlichen - Sicht die Outcomes einer Politik. Unter Outcomes verstehen wir alle ursächlich auf öffentliche Politiken zurückföhrbaren Änderungen in der Gesellschaft oder der natürlichen Umwelt. Outcomes sind - entsprechend dem Sinn des Wortes - das, was aus staatlichem Handeln "herauskommt", d.h. was daraus resultiert. Dabei werden die intendierten und un intendierten, erwünschten und unerwünschten, direkten und indirekten Auswirkungen einbezogen (vgl. Kapitel 3.2.2.).

Die Evaluation der Outcomes kann deutlich machen, dass optimale Outputs oder Impacts lediglich eine notwendige, nicht aber eine ausreichende Bedingung für optimale Outcomes sind. Die veränderten Verhaltensweisen vermögen nur dann zur Zielrealisierung beizutragen, wenn die der Instrumentenauswahl zugrundeliegenden Annahmen über das Verhalten der Adressaten (Kausalhypothese; vgl. Kapitel 5.1.) richtig sind und keine kontraproduktiven Nebenwirkungen auftreten.

Ein einfaches Beispiel dafür liefern etwa die Energiesparmassnahmen: Eine allgemeine Absenkung der Raumtemperatur durch eine entsprechende Einstellung der zentralen Heizanlagen föhrt nur dann zum Ziel, wenn die Massnahme nicht durch den Einsatz individueller, aus energetischer Sicht wesentlich weniger effizienter (z.B. Nutzung von hochwertiger Elektrizität anstelle anderer Energiequellen) Heizgeräte unterlaufen wird.

Wesentlich komplizierter sind die Wirkungszusammenhänge - und entsprechend schwieriger ist auch die Ermittlung und Beurteilung der Outcomes - etwa im Falle der Emissions- und Immissionsvorschriften der

Luftreinhaltepolitik¹⁶ oder von Massnahmen zur Bekämpfung der Arbeitslosigkeit (Freiburghaus 1987).

Es ist wichtig hervorzuheben, dass Outcomes eine analytische Kategorie sind. Unmittelbar beobachtbar ist nur die gesellschaftliche Wirklichkeit. Outcome-relevante Daten sind etwa Angaben zur Entwicklung des von der öffentlichen Politik anvisierten gesellschaftlichen Problems. Darunter fallen namentlich Angaben zu den sich gegebenenfalls verändernden Lebensumständen der Politikbetroffenen. Politikbetroffene (vgl. Kapitel 4.2.1.) können gesellschaftliche Gruppen wie etwa Nachbarn, Mieter, Einwohner und Besucher (im Falle der Raumplanungspolitik), Geschäftskunden (im Falle der Konsumentenschutzpolitik), Patienten (im Falle der Gesundheitspolitik) sein. Weil es sich bei diesen Gruppen meist nicht um individuell genau identifizierbare Personen handelt, ist es angezeigt, dafür auch die gängigen Sozialindikatoren zu verwenden. Denkbar sind aber auch direkte Befragungen der Politikbetroffenen (z.B. PTT-Kunden, Patienten etc.) Bei der Erfassung Outcome-relevanter Daten muss die Problembeobachtung auf jeden Fall auch offen sein für nichtintendierte, wenn möglich sogar kontraproduktive Effekte der beobachteten Politik.

Alle diese Daten, die im Zusammenhang mit Zielen öffentlicher Politik stehen, geben aber bloss Hinweise auf mögliche Outcomes. Ein blosser Vergleich der Zielwerte und der tatsächlichen Werte (Zielindikatoren) sagt nämlich noch nichts darüber aus, ob die Zielwerte wegen der betreffenden staatlichen Politik oder aufgrund anderer Einflüsse (nicht) erreicht worden sind. Die Outcomes als eigentliche Resultate öffentlicher Politik unter Abzug aller anderen möglichen Einflüsse (Konjunktur, Wertewandel, Selbsthilfe der Betroffenen, soziales Lernen usw.) können letztlich nur forschungsmässig erschlossen werden. Hierfür ist ein impliziter oder expliziter Vergleich mit der Situation nötig, die ohne die untersuchte Politik oder Massnahme bestehen würde. Hierzu stellt die Evaluationsforschung spezifische Methoden bereit (vgl. Kapitel 10-12).

¹⁶ Siehe dazu Luftreinhalte-Verordnung (SR 814.318.142.1); Knoepfel und Weidner 1984; die Auseinandersetzungen um Tempo 80 auf einzelnen Autobahnabschnitten liefern dazu aufschlussreiche Illustrationen. Erwähnt sei in diesem Zusammenhang auch, dass Bund und Kantone nach Art. 44 des Umweltschutzgesetzes Erhebungen über die Umweltbelastung durchführen und den Erfolg der getroffenen Massnahmen prüfen sollen (SR 814.01).

5.6.2. Wirksamkeit

Das Kriterium der Wirksamkeit schliesst unmittelbar an die Kategorie der Outcomes an. Es bezieht sich auf das Verhältnis zwischen den beabsichtigten und den tatsächlich eingetretenen gesellschaftlichen Wirkungen einer öffentlichen Politik. Die Feststellung und Beurteilung der Wirksamkeit einer öffentlichen Politik erfolgt anhand eines Vergleichs zwischen den im Verwaltungsprogramm formulierten Zielwerten (Kapitel 5.2.2.1.) und den tatsächlich ausgelösten Wirkungen bei den Politikbetroffenen (zielbezogene Wirkungen/Outcomes).

In seiner logischen Struktur beinhaltet das Kriterium einerseits (aufbauend auf den Darlegungen zu den Outcomes) eine *kausalanalytische* Rekonstruktion des Verhältnisses zwischen öffentlicher Politik und/oder Endoutputs und der gesellschaftlichen Wirklichkeit. Andererseits bezieht es sich *normativ* auf die daraus ermittelte Differenz zwischen den politischen Zielen und der realen Problemveränderung.

Die im politischen Prozess aufgestellten Ziele werden dieser komplexen logischen Struktur des Wirksamkeitskriteriums allerdings selten gerecht. In der Regel werden im politischen Raum absolute Ziele formuliert, wobei offengelassen wird, welchen *Beitrag* die öffentliche Politik an deren Erreichung leisten soll. Es ist dann durchaus möglich, dass diese Ziele erreicht werden, ohne dass die öffentliche Politik dazu einen wesentlichen Beitrag leistet. Möglich ist umgekehrt auch, dass diese Ziele verfehlt werden, dass ohne öffentliche Politik sich das Problem aber wesentlich schlimmer darstellte. In diesen Fällen geben in absoluten Werten formulierte Ziele keine Handhabe zur Beantwortung der Frage, inwieweit die Politik wirksam war. Hier wäre es nötig, dass in der Zielbestimmung angegeben würde, welche relative Verbesserung die Politik erreichen soll¹⁷. Diesem Erfordernis kommen die in Verwaltungsprogrammen formulierten Ziele indessen selten nach. Sie sind ohnehin oft wenig explizit, eindeutig, quantifiziert und über die Zeit hinweg stabil (Hellstern und Wollmann 1983: 11-22).

Evaluationen können zeigen, dass im Rahmen von öffentlichen Politiken Outputs erbracht werden und sich auch das Adressatenverhalten (Impacts) im erwünschten Masse verändert, dass daraus aber nicht die be-

¹⁷ Beispiele: Mit den vorgesehenen Arbeitsloseneingliederungsprogrammen wird eine Reduktion der Arbeitslosigkeit um einen Prozentpunkt angestrebt; mit der neuen Stipendienregelung wird eine Erhöhung des Studentenanteils der unteren Schichten um einen Drittel angestrebt.

absichtigten Outcomes resultieren. Der Grund für diesen politisch hoch-explosiven Umstand liegt oft in unzutreffenden Annahmen über den Beitrag der anvisierten Adressaten zur Entstehung des zu lösenden kollektiven Problems ("falsche" Kausalhypothese; vgl. Kapitel 5.1.), in extern bedingten Problemverschärfungen oder in naturwissenschaftlichen oder technischen Fehlannahmen zum Wirkungsgefüge der angegangenen Problematik.

Wirkungslose (outcomelose) Politik: Staatliche Hotellerieförderung und langfristige Wettbewerbsfähigkeit der Schweizer Ferienhotels

Die betriebswirtschaftliche Lage mancher Hotels in den Schweizer Ferienregionen präsentiert sich wenig günstig. Tiefe Renditen oder sogar Verluste erschweren die Finanzierung von Investitionen, die zur Verbesserung der Angebotsqualität notwendig wären. Der Bund und verschiedene Kantone bieten Hoteliers mit erfolgversprechenden Projekten finanzielle Unterstützungleistungen an (z.B. zinslose Darlehen, Zinskostenbeiträge). Dadurch soll u.a. die internationale Wettbewerbsfähigkeit der Schweizer Ferienhotellerie verbessert werden.

Eine Evaluation der staatlichen Hotellerieförderung (Hanser et al. 1994) zeigt, dass diese Investitionsförderung intensiv genutzt wird. Drei Viertel der realisierten Projekte hätten ohne staatliche Unterstützung nicht im geplanten Umfang oder Zeitpunkt getätigt werden können. Die geförderten Hotels weisen nach Realisierung der Investitionsprojekte zwar eine günstigere Umsatzentwicklung als nicht geförderte Hotels auf. Mittelfristig lässt sich aber statistisch keine günstigere Entwicklung ihrer Renditen nachweisen. Bauliche Investitionen allein reichen also nicht aus, um die Wettbewerbsfähigkeit eines Hotelbetriebs zu verbessern. Auf Branchenebene drohen deshalb unerwünschte strukturerhaltende Nebeneffekte. Diese staatliche Förderungsstrategie zeigt eindeutig positive Outputs (grosses Volumen der unterstützten Investitionen) und Impacts (Anzahl der realisierten Projekte) und damit eine Erfüllung des Kriteriums der Effektivität, bleibt aber auf der Ebene der Problemlösung unwirksam (d.h. keine langfristige Förderung der Wettbewerbsfähigkeit).

Derartige Evaluationsbefunde sind vermutlich alarmierender als jene impactloser Politiken. Denn sie führen zu Frustrationen und Gefühlen von Ungerechtigkeit und bürokratischem Leerlauf, die mitunter auch bei den Politikadressaten, sicherlich aber bei den Politikbetroffenen eine eigentliche Staatsverdrossenheit auslösen können. Solche Anzeichen zeigen sich etwa, wenn Automobilistinnen und Automobilisten, die eine Smogwarnung ernstnehmen und vorschriftsgemäss ihr Fahrtempo reduzieren, anderentags in der Zeitung lesen, dass die Ozonkonzentrationen Rekordwerte

erreicht hätten und dass Geschwindigkeitsreduktionen darauf keinen Einfluss hätten.

Schliesslich gibt es aber auch öffentliche Politiken, deren Outcomes den intendierten Zielen entsprechen, weil sie Outputs und Impacts produzieren, die tatsächlich in der Lage sind, die Problemstruktur in der beabsichtigten Weise zu verbessern.

Wirksame Politik: Staatliche Wohneigentumsförderung

Im Jahr 1970 lebten in der Schweiz im Durchschnitt nur 28.1% der Haushalte im selbstgenutzten Eigentum. Diese Wohneigentumsquote war im internationalen Vergleich niedrig. Der Bund hat deshalb Ende 1974 ein Wohnbau- und Eigentumsförderungsgesetz (WEG; SR 843) in Kraft gesetzt, das zur Senkung der Anfangsbelastung der Erwerber von Wohneigentum folgende Massnahmen vorsieht: eine Bundesbürgschaft, Grundverbilligungen und nicht rückzahlbare Zusatzverbilligungen. Hauptziel dieser Politik ist die Verbesserung der Wohnselbstversorgung.

Laut einer Evaluation der Massnahmen des WEG (Schulz et al. 1993) wirkte diese Wohneigentumsförderung des Bundes tatsächlich im erwünschten Sinne. Im Zeitraum zwischen Mitte der siebziger Jahre und 1991 wurden durch den Staat insgesamt 15'747 Eigentumsobjekte finanziell unterstützt (Outputs). Wohneigentum mit Bundeshilfe wird insbesondere von jüngeren Familienhaushalten mit beschränkten finanziellen Mitteln erworben, die auf dem freien Markt ansonsten kaum eine Chance hätten (Impacts). Dank dieser Massnahme ist die Wohneigentumsquote im untersuchten Zeitraum von ca. 15 Jahren auf 31,3% gestiegen (= zielkonforme Outcomes). Weiter löste das WEG auch positive Nebenwirkungen aus. So ist z.B. die Wohneigentumsförderung in der Rezession eine willkommene Konjunkturförderung, wie die Förderungsquote von 20% beim Einfamilienhausbau im baukonjunkturschwachen Jahr 1991 zeigt.

5.6.3. Wirtschaftlichkeit (Effizienz)

Das Kriterium der Wirtschaftlichkeit (=Effizienz) bezieht sich auf das Verhältnis zwischen den für eine öffentliche Politik bereitgestellten Ressourcen und den erreichten Outcomes (Wirkungen). Es umschreibt damit das Verhältnis zwischen Kosten und Nutzen einer öffentlichen Politik. Der Vergleich konzentriert sich in der Regel auf die beabsichtigten Hauptwirkungen eines Verwaltungsprogramms. In diesem Zusammenhang spricht man von ziel- oder wirkungsbezogener Wirtschaftlichkeit (Effizienz). So wird etwa gefragt, ob zur Erzielung der gleichen Wirkun-

gen weniger Ressourcen eingesetzt werden könnten bzw. ob mit den gleichen Ressourcen ein höherer Zielerreichungsgrad möglich wäre. Zur Beurteilung der Wirtschaftlichkeit von Politiken oder einzelner ihrer Massnahmen werden in der Evaluationsforschung vor allem zwei Methoden angewandt (vgl. Rossi und Freemann 1993: 363-401): Bei *Kosten-Nutzen-Analysen* werden die Kosten (materiell und immateriell) und die Wirkungen (=Nutzen) in Geldwerten quantifiziert und einander gegenübergestellt. So wurden etwa die Schneeräumungskosten früherer Passöffnung dem daraus resultierenden Nutzen (v.a. für den Tourismus) gegenübergestellt (Ecoplan 1993). Namentlich die Monetarisierung des Nutzens stellt die Forscherinnen und Forscher oft vor heikle Probleme. Diesbezüglich stellt die *Kosten-Effektivitäts-Analyse* niedrigere Anforderungen. Dabei werden einander unterschiedliche Massnahmen (z.B. Reduktion der Geschwindigkeit, Ausbau von Strassen, Gurtentragepflicht auf den Rücksitzen) zur Erzielung einer bestimmten Wirkung (z.B. Reduktion der Strassentoten um eine vorgegebene Anzahl Personen) gegenübergestellt. Daraus kann ermittelt werden, welche der Massnahmen im Hinblick auf die Erreichung des Ziels am wirtschaftlichsten ist. Bei dieser vergleichenden Vorgehensweise stehen nicht die oft nur schwer ermittelbaren absoluten Werte von Massnahmenkosten und -nutzen, sondern die leichter ermittelbaren relativen Unterschiede (teurer/billiger) der entsprechenden Werte für verschiedene Varianten im Zentrum des Interesses.

Solche Überlegungen zur Wirtschaftlichkeit einer öffentlichen Politik erweisen sich erst dann als sinnvoll, wenn der Grad der Wirksamkeit empirisch feststeht. Das in der Evaluationsforschung verwendete Wirtschaftlichkeitskriterium darf daher nicht dem betriebswirtschaftlichen Rationalitätsbegriff gleichgesetzt werden. Letzterer bezieht sich in der Regel auf das Verhältnis zwischen dem Output und den eingesetzten Ressourcen (=outputbezogene Wirtschaftlichkeit). Die in der Gesellschaft oder in der natürlichen Umwelt durch den - gegebenenfalls reduzierten (Sparmassnahmen) - Einsatz von Ressourcen im Rahmen einer öffentlichen Politik erzielten Outcomes werden bei solchen, nur outputbezogenen Wirtschaftlichkeitsanalysen zu unrecht ausgeblendet. Ausserdem werden kostenseitig meist nur die direkten und materiellen, nicht aber die immateriellen Kosten bzw. Ressourcen (wie z.B. Konsens - vgl. Kapitel 8.1.) berücksichtigt. Es ist an der Zeit, die seit den achtziger Jahren geforderten Rationalisierungs- und Effizienzsteigerungsprogramme unter diesem erweiterten Gesichtswinkel der Problemlösungsfähigkeit öffentlicher Politiken kritisch zu evaluieren. In der Tat wird auch eine intern hochgradig

rationalisierte und damit kostengünstige Staatstätigkeit zum sinnentleerten Verwaltungsaktivismus, wenn sie extern keine oder nicht die erstrebten Wirkungen zeitigt; der Nachweis effizienter Erledigung von Verwaltungsvorgängen allein genügt nicht.

Als effizient gelten für die politikorientierte Verwaltungswissenschaft nur jene öffentliche Politiken, die erstens wirksam sind und die zweitens die materiellen und immateriellen Ressourcen, die für den Vollzug erforderlich sind, optimal einsetzen.

Effiziente Politik: Kombination von Kehricksackgebühr, Separatsammlungen und Informationskampagne

Seit Mitte der achtziger Jahre sind immer mehr Gemeinden von pauschalen Abfallgebühren (pro Haushalt oder pro Raumeinheit) zu einem System der mengenproportionalen Kehricksackgebühr übergegangen. Neben diesem Lenkungsinstrument werden meistens auch flankierende Massnahmen eingesetzt, wie beispielsweise Separatsammlungen (von Papier, Glas, Metall, etc.) oder staatliche Kampagnen zur Information und Motivation der Bevölkerung.

In einer Evaluationsstudie (INFRAS 1990) wurden der Vollzug, die Haupt- und Nebenwirkungen sowie die Kostenwirksamkeit dieser drei, als Einzelmassnahme oder in Kombination angewandten Instrumente untersucht. Abschliessend wurde die Wirksamkeit und die Wirtschaftlichkeit solcher (damals alternativer) kommunaler Abfallpolitiken wie folgt beurteilt: Die Kehricksackgebühr wirkt am stärksten, wenn sie mit einem System von Separatsammlungen und Informationsmassnahmen kombiniert wird. Die Kehricksackgebühr reduziert die Menge gemischten Kehrichts kostenwirksamer als eine starke Verdichtung des Sammelstellen- und Informationsangebots. Die optimale Kombination der Einzelmassnahmen Kehricksackgebühr, Separatsammlungen und Information minimiert die Entsorgungskosten pro Tonne gesammelter Werkstoffe. Die bisherigen Erfahrungen ergeben, dass eine Sammelstelle für ca. 2000-3000 Einwohner ein optimaler Bereich ist. Seither sind zu diesem Thema zwei weitere Studien (Hausheer 1991; Schwarz und Brozovic 1992) veröffentlicht worden, die (mit erheblich besserem Datenmaterial) zu ähnlichen Ergebnissen führten.

5.7. META-EVALUATIONEN, EVALUATIONSSYNTHESEN UND METAANALYSEN - QUALITÄT DER EVALUATION

Evaluationen können (wie in der Abbildung 1 dargestellt) den letzten Schritt des Politikumsetzungsprozesses bilden oder aber darin verwoben sein. Im ersteren Falle sprechen wir von einer Politikevaluation im engeren Sinne, die die gesamten Haupt- und Nebenwirkungen der über die

sieben oben identifizierten Stufen vom Politikkonzept bis hin zu den Outcomes ablaufenden öffentlichen Politik zu einem gegebenen Zeitpunkt in getrennten Wirkungsräumen zum Gegenstand hat. Eine solche als Politikevaluation im engeren Sinne bezeichnete "Evaluation" wird heute auch vom Gesetzgeber zunehmend als abschliessende Stufe einer Politikumsetzung gefordert (Evaluationsklauseln; vgl. Kapitel 2.4.). Im zweiten Falle sprechen wir von Evaluationen der Ergebnisse einzelner Stufen, wie sie in den bisherigen Abschnitten dieses Kapitels dargestellt wurden. Liegt gemäss einem gesetzlichen Auftrag oder aufgrund eines politischen oder wissenschaftlichen Interesses an einer bestimmten Politik eine Politikevaluation im engeren Sinne vor, so bezieht sich die Evaluation auf die Qualität dieser letzten Politikumsetzungsstufe. Wir sprechen von einer Evaluation der Evaluation. Dieselbe Situation kann dann eintreten, wenn etwa eine Evaluation einzelner Umsetzungsstufen (z.B. die Evaluation des Verwaltungsprogrammes oder der Outputs) politisch hinterfragt wird, weshalb in einer Expertise die Qualität der angestellten Evaluation überprüft werden soll. Diese letzteren Evaluationen liegen logisch auf einer anderen Ebene als die soeben dargestellten. Während früher (etwa von der Evaluation Research Society) für all diese evaluierenden Abklärungen der Begriff Metaevaluation verwendet wurde, hat sich in der Zwischenzeit eine differenziertere Terminologie durchgesetzt, welche insbesondere zwischen Metaevaluation, Evaluationssynthese und Metaanalyse unterscheidet.

Metaevaluationen bezwecken die Bewertung von Evaluationsprozessen und der Qualität ihrer Ergebnisse. Sie sind ein wichtiger Bestandteil des Selbstreflexionsprozesses der Evaluationsforschung. Metaevaluationen geben Informationen über die *methodische Qualität von Evaluationen* und können als Instrument der Qualitätssicherung gesehen werden. Sie können Aufschluss darüber geben, ob die Ergebnisse einer bestimmten Evaluation wissenschaftlich ausreichend abgesichert sind. Als Grundlage von Metaevaluationen können die in Kapitel 3.4. vorgestellten Evaluationsstandards herangezogen werden. Eine bestimmte Evaluation kann somit entlang sämtlicher Kriterien von Evaluationsstandards bewertet werden. So hat Widmer (1996) in einer breitangelegten Studie diese Standards auf zehn in der Schweiz im Laufe der letzten zehn Jahre durchgeführte Evaluationsstudien angewandt. Diese vergleichende Metaevaluation hat Aufschluss über typische Fehler bzw. Stärken schweizerischer Evaluationen und über allfällige Ausbildungsdefizite gegeben. In einer zusätzlichen Studie haben Widmer, Rothmayr und Serdült (1996) 8 Kurzevaluationen

auf ihre methodologische Qualität überprüft und einen Quervergleich der Resultate mit denjenigen grösserer Evaluationen vorgenommen.

Metaevaluationen beziehen in der Regel auch die Verwendung der Evaluationsergebnisse mit ein. Zu dieser Frage hat sich ein eigenständiger Forschungsweig, die *Verwendungsforschung* ("utilization research"), herausgebildet¹⁸. Evaluationen werden bei dieser Betrachtungsweise als Teil des Meinungsbildungs- und Entscheidungsprozesses, ja sogar als Instrument einer Politik verstanden. Verwendungsstudien können Fragen wie die folgenden ansprechen: Welchen Einfluss haben Evaluationsstudien auf politische Entscheidungsprozesse? Wie kann die Umsetzung der Ergebnisse von Evaluationen möglichst erfolgreich vorbereitet werden? In der Schweiz haben sich namentlich Freiburghaus und Zimmermann (1985) mit der Nutzung der Ergebnisse Nationaler Forschungsprogramme sowie Klöti und Schneider (1989) mit der (Nicht-)Nutzung von Evaluationen bei der Gesetzgebung befasst.

Mit der Zunahme von Evaluationsstudien hat auch die Möglichkeit zugenommen, Studien durchzuführen, die Ergebnisse mehrerer Evaluationen auswerten. Im Rahmen von *Metaanalysen* (Cook et al. 1992; Hunter und Schmidt 1989; Rosenthal 1987) werden die Ergebnisse bestehender Wirkungsstudien quantitativ zusammengefasst. Zum Teil lassen sich dadurch methodische Schwächen der einzelnen Studien eliminieren und Ergebnisse von grösserer Glaubwürdigkeit erzielen. Insbesondere lassen sich damit die Fehlerquellen, die aufgrund einer mangelnden Stichprobengrösse der einzelnen Studien entstehen, zum Teil vermindern. Als Beispiel einer in der Schweiz erarbeiteten Metaanalyse, die sich allerdings zu einem grossen Teil auf ausländische Studien bezieht, sei auf Grawe, Donati und Bernauer (1994) verwiesen.

Evaluationssynthesen (vgl. GAO 1992c; Balthasar et al. 1995) beruhen wie Metaanalysen auf Sekundärdaten. Sie stützen sich ebenfalls auf bereits vorhandene Evaluationsstudien. Anders als Metaanalysen beruhen sie indessen nicht ausschliesslich auf einer quantitativen Aggregation der Daten, sondern beziehen deren qualitative Würdigung mit ein.

Evaluationssynthesen setzen ein Screening vorhandener Dokumente im Hinblick auf die zu untersuchende politikrelevante Frage, die retrospektiver oder prospektiver Natur sein kann, voraus. Ziel dieses ersten Schritt ist, die relevanten Evaluationsstudien zu ermitteln. Diese werden in einem

¹⁸ Vgl. v.a. die beiden Zeitschriften "Knowledge: Creation-Diffusion-Utilization" und "Knowledge and Policy" sowie Kapitel 6 und 13.

zweiten Schritt kritisch auf ihre inhaltliche Relevanz und methodische Qualität überprüft; dieser Teil hat inhaltlich zu einem grossen Teil den Charakter einer Metaevaluation. In einem dritten Schritt schliesslich werden auf der Grundlage der vorangegangenen methodischen Analyse die Resultate zusammengefasst; diese Synthese kann quantitativer und/oder qualitativer Art sein.

Sowohl Metaanalysen wie auch Evaluationssynthesen erlauben es, das Wissen und die Kenntnisse bestehender Evaluationsstudien zusammenzufassen und in einem grösseren Kontext zu verwenden. Beide setzen ausreichende methodische Kenntnisse voraus. Metaanalysen sind zeitlich recht aufwendig und werden vor allem bei retrospektiven Fragen eingesetzt. Evaluationssynthesen stellen vergleichsweise geringere Anforderungen an das empirische Material und sind zeitlich weniger aufwendig. Sie werden in den USA vor allem prospektiv zur Beurteilung bestimmter politikrelevanter Vorschläge eingesetzt.

KAPITEL 6: TYPISCHE NUTZUNGSZUSAMMENHÄNGE VON EVALUATIONEN

Werner Bussmann, Peter Knoepfel

6.1. EINLEITUNG: EVALUATIONEN ALS INSTRUMENT ÖFFENTLICHER POLITIK

Wie in Kapitel 3 ausgeführt wird, sind Evaluationen ein *Informationsinstrument*. Sie stellen Wissen über die Effektivität, die Wirksamkeit oder die Wirtschaftlichkeit staatlichen Handelns bereit. Solches Wissen kann Selbstzweck sein bzw. der Lehre und Forschung dienen, namentlich wenn es im universitären Bereich erarbeitet wird. In der Regel werden Evaluationen aber mit einem konkreten Staats- oder verwaltungspolitischen Ziel durchgeführt: Novellierung einer Massnahme, Legitimierung eines Entscheids, Verbesserung des Vollzugs, Kontrolle, Kürzung von Subventionen u.a.m. Evaluationen sollen Grundlagen für anstehende Entscheide bereitstellen oder getroffene Massnahmen legitimieren.

Wie staatliche Einzelmassnahmen bzw. öffentliche Politiken haben auch Evaluationen intendierte und unintendierte, direkte und indirekte, hauptzielbezogene und andere politische Wirkungen. Es ist unmöglich, all diese mittels Evaluationen angestrebten Ziele und ihre möglichen Wirkungen eingehend zu beschreiben. Wir beschränken uns deshalb auf typische Konstellationen, die uns aus der Erfahrung bekannt sind. Denn auch Evaluationen selbst gehören heute zum Repertoire staatlichen Handelns. Wie alle anderen Instrumente staatlicher oder gesellschaftlicher Akteure sind sie Artefakte. Ihr Einsatz ist deshalb nicht naturgegeben, sondern von Menschen in spezifischen Situationen für spezifische Zwecke gemacht. Typologien möglicher Ziele und Wirkungen von Evaluationen sind daher unvollständig, weil die Instrumente in der Praxis ständig neu- und weiter-erfunden werden.

6.2. MASSNAHMEN, AKTEURE UND POLITISCHE KONSTELLATIONEN ALS ANSATZPUNKTE

Für eine Beschreibung von Zielen und Wirkungen von Evaluationen sind viele Ansatzpunkte möglich (z.B. Politikbereiche, historische Konstellationen).

tionen, wirtschaftliches Kalkül). Wir legen hier eine möglichst neutrale (geschichtlich und politikbereichsspezifisch unabhängige) Systematisierung vor, die sich auf die zu evaluierenden Massnahmen und Politiken (Kapitel 6.3.), die evaluationsbetroffenen Akteure und deren Strategien (Kapitel 6.4.) und umfassendere, politisch-ideologische Positionen (Kapitel 6.5.) konzentriert, zu denen Evaluationen Bezüge eröffnen.

6.3. ZU EVALUIERENDE MASSNAHMEN ODER POLITIKEN

Evaluationen können zum Ziel oder zur Folge haben, Massnahmen oder Politiken neu zu gestalten oder abzuändern, deren Vollzug zu verbessern oder Rechenschaft abzulegen. Über diese zentralen Formen der Nutzung von Evaluationen besteht in der Literatur ein breiter Konsens (Chelimsky 1987; Rist 1989; Bussmann 1995a: 36-45). In der Praxis bestehen allerdings oft Meinungsunterschiede über die Zeitdimension von Evaluationen: Denn recht oft werden bestimmte Evaluationsarten einseitig bestimmten Phasen des Politikzyklus (Massnahmenformulierung, Vollzugsverbesserung, Rechenschaftsablegung) zugerechnet. So wird beispielsweise davon ausgegangen, dass im Hinblick auf die Neuformulierung einer Massnahme ausschliesslich prospektive Evaluationen sinnvoll seien. Demgegenüber vertreten wir die Meinung, dass verschiedene Evaluationsarten im Hinblick auf die aufgeführten drei Hauptzwecke miteinander kombiniert werden können. Wir zeigen dies an folgenden Beispielen:

1. Für die *Neugestaltung oder Änderung einer Massnahme* (prospektive Dimension) kann eine Evaluation der Wirksamkeit oder der Wirtschaftlichkeit bisheriger Massnahmen beispielsweise Hinweise über Wirkungsverläufe geben, und Effektivitätsanalysen können zeigen, wie sich der Vollzug entwickelt hat. Schliesslich können Evaluationssynthesen bestehende Evaluationsresultate und Erfahrungen aufarbeiten, die in verwandten Gebieten oder anderen Ländern mit dem in Aussicht genommenen Instrumentarium gemacht wurden.
2. In ähnlichem Sinne gibt es verschiedene Evaluationsarten, welche der *Begleitung und Verbesserung einer Massnahme* dienen. Zu erwähnen sind dabei in erster Linie Monitoring-Daten, welche über die Problemlage (z.B. Entwicklung der Umwelt) informieren oder Effektivitätsdaten zum Stand des Vollzugs (was wurde getan?). Diese geben einen

vertieften Einblick in die (kausalanalytischen) Zusammenhänge des Vollzugs. Alle diese Informationen können zur Begleitung der Massnahme und für Vollzugsverbesserungen eingesetzt werden.

3. Hauptinstrument zur *Rechenschaftsablegung* sind (retrospektive) Evaluationen der Wirksamkeit und allenfalls auch der Wirtschaftlichkeit. Häufig begnügt man sich auch mit Evaluationen des Outputs und der Impacts oder mit entsprechenden Monitoring-Daten. Gelegentlich kommt es auch vor, dass Überlegungen angestellt werden, welche Auswirkungen alternative Massnahmen gezeitigt hätten (hypothetische Gedankenkonstruktionen, ökometrische Modellsimulationen u.a.m.). In solchen Fällen werden Methoden miteinander kombiniert, die in der retrospektiven (empirische Wirkungsfeststellung) und in der prospektiven (hypothetische Gedankenexperimente) gängig sind.

6.4. EVALUATIONEN UND IHRE AKTEURE

Für die Akteure stellen Evaluationen eine spezifische Form (sozial)-wissenschaftlicher Beratung dar (vgl. Bussmann 1989 und 1993; Kissling-Näf und Wildi-Ballabio 1993 sowie generell Linder 1989). Sie können damit verschiedene Absichten verfolgen. Wir unterscheiden dabei die folgenden Möglichkeiten der Nutzung: (1) Evaluationen können in einem exklusiven und engen Beratungszusammenhang stehen; die Informationen und Handlungsempfehlungen stehen ausschliesslich dem Auftraggeber zur Verfügung. (2) In der überwiegenden Zahl der Fälle sind die Informationen indessen der Öffentlichkeit zugänglich. Ihre Bereitstellung kann mit der Absicht verbunden sein, die politisch-strategische Position eines Akteurs zu stärken oder zu schwächen. (3) Eine Evaluation kann jedoch auch "neutrale" Informationen über bestimmte Sachverhalte bereitstellen. (4) Ausserdem kann der Aufbau eines kontinuierlichen Beratungsverhältnisses zur Bildung einer administrativ-wissenschaftlichen Koalition beitragen, deren Zweck in der Absicherung des Aufgabenbereichs vor allfälliger Kritik liegt. (5) Als letztes kann Politikberatung von einem Akteur als symbolische Handlung zum blossen Zeitgewinn eingesetzt werden.

6.4.1. Sich einen direkten Informationsvorteil verschaffen

Wenn Evaluationen in Form von direkter, exklusiver und von der Öffentlichkeit abgeschirmter (wenn auch nicht unbedingt geheimer) Beratung erfolgen, entziehen sie sich weitgehend sozialwissenschaftlicher Beobachtung. Darüber finden sich normalerweise weder in Publikationen noch in Zeitungsmeldungen Informationen. Am ehesten gewinnt man dazu Zugang über intime Verwaltungskennntnisse oder über Gespräche mit sozialwissenschaftlichen Beratungspraktikern. Solche Beratung erfolgt eher durch *private Büros* als durch Universitätsinstitute, da für letztere der Verzicht auf Verwertung der Resultate in Form allgemein zugänglicher Publikationen wenig attraktiv ist.

Der Grossteil solcher Beratung besteht aus *Organisations- und Führungsberatung*. Evaluationen im Sinne von Vollzugs- und Wirkungsstudien sind noch eher selten. Beratung verbessert die Voraussetzungen, um Schwachstellen zu erkennen und Verbesserungen einzuleiten, ohne Fehler eingestehen zu müssen. Obgleich sich die Verwaltungsstellen das Recht auf ausschliessliche Verfügung über die Evaluationsdaten sichern, werden sie diese verbreiten, wenn ihnen dies einen Nutzen verspricht (Imagegewinn, Argumente zur Stärkung der eigenen Position). Derartiges Evaluationswissen verschafft dessen alleinigem Besitzer einen enormen Machtvorsprung gegenüber anderen staatlichen oder gesellschaftlichen Akteuren. Nebst dieser exklusiven Beratung gibt es auch an sich öffentlich zugängliche, meist *technisch orientierte Evaluationen* (z.B. Rentabilität einzelner Busstrecken, Evaluationen spezifischer Unterrichtsmittel), deren Resultate nicht an die Öffentlichkeit gelangen, weil die Medien davon kaum Notiz nehmen. Beispiele dafür sind etwa Umweltverträglichkeitsberichte von wenig umstrittenen Anlagen (Projektelevaluationen) oder bautechnische Variantenevaluationen, wie sie etwa beim Amt für Bundesbauten routinemässig durchgeführt werden.

6.4.2. Die eigene strategische Position stärken

Viele Evaluationen werden in Auftrag gegeben, weil sich die betreffende Organisation daraus einen Vorteil verspricht. Die Evaluation soll ihre Position unterstützen oder zumindest jene oppositioneller Akteure neutralisieren. Aus der Praxis greifen wir hierzu folgende typischen Situationen heraus:

- *Unterstützung für ein bestimmtes Projektleine bestimmte Massnahme mobilisieren:* Im Jahre 1988 gab das Bundesamt für Konjunkturfragen Dieter Freiburghaus eine Evaluation der "Kommission zur Förderung der wissenschaftlichen Forschung (KWF)" in Auftrag. Das Bundesamt wollte im Hinblick auf einen neuen Kredit eine unabhängige Beurteilung der Wirkung der geleisteten Förderung erreichen. In der zeitlichen Abwicklung war die Durchführung der Studie mit der Erarbeitung der Botschaft koordiniert, und die Ergebnisse (Freiburghaus, Zimmermann und Balthasar 1990) wurden als Argument für die Weiterführung des Kredites in der entsprechenden Botschaft des Bundesrates benützt. Zudem wurden Optimierungen in Aussicht gestellt, welche im Rahmen der Evaluation vorgeschlagen worden waren.
- *Widerstand gegen ein bestimmtes Projekt mobilisieren:* Evaluationen werden auch eingesetzt, um bestimmte Projekte/Massnahmen, die für einzelne Akteure nachteilig sind, abzublocken. So erwirkte das Solothurner Finanzdepartement, das sich mit einem zunehmenden Finanzbedarf für die Denkmalpflege konfrontiert sah, eine Evaluation der Wirksamkeit dieser Aufwendungen. Mit einem erfolglosen Versuch, eine Politikinitiative zu durchkreuzen, sahen sich die Autoren dieses Beitrags kurz vor der Einführung des Phosphatverbots in Waschmitteln Mitte der achtziger Jahre konfrontiert. Die grossen Waschmittelkonzerne befürchteten Signalwirkungen in anderen Ländern und suchten deshalb nach Argumenten, um ein Phosphatverbot in der Schweiz zu verhindern. Sie versuchten, uns zur Mitwirkung an einer solchen Studie zu gewinnen; zeitliche Gründe und der brisante Kontext verunmöglichten dies indessen. Die Studie wurde schliesslich von einem privaten Beratungsteam durchgeführt. Auch diese Personen liessen sich indessen nicht für die allzu evidenten Ziele der Waschmittelindustrie vereinnahmen. Das sozialwissenschaftliche Störmanöver gegen das Phosphatverbot musste schliesslich abgeblasen werden.
- *Unterstützung für ein bestimmtes Politikanliegen mobilisieren:* Mit der Ausdehnung des Wohlfahrts- und Leistungsstaates seit den späten fünfziger Jahren sind ehemals getrennte Politikbereiche zusammengewachsen. Politik erfolgt heute nicht in fein säuberlich abgetrennten Arenen, sondern mehr und mehr im Bereich der "interpolicies", d.h. dort, wo sich die Kontinentalplatten sektoraler Politikbereiche (Verkehr, Gesundheit, Landwirtschaft, Gewässerschutz usw.) aneinan-

der reiben und Verwerfungen produzieren. Viele relevanten Auseinandersetzungen finden heute an diesen Rändern der Politikbereiche statt: Soll die Finanzpolitik den sozialpolitischen "acquis" vollumfänglich schonen? Sollen dem Strassenverkehr auch die von ihm verursachten Umwelt- und Gesundheitskosten angerechnet werden? Soll die Landwirtschaft stärker umweltverträglich ausgestaltet werden? Diese und viele weitere Fragen siedeln sich entlang den Grenzlinien zwischen den einzelnen Politikbereichen an, die je nach Standort und Perspektive anders gezogen werden können. Evaluationen können und sollen oft Argumente liefern, wie diese Abgrenzung zu lösen ist.

Ein Beispiel dafür ist die Evaluationsstudie zum Gewässerschutz in der Landwirtschaft, mit der die landwirtschaftliche Forschungsanstalt für Agrikulturchemie und Umwelthygiene, Bern/Liebelfeld (FAC), das Bundesamt für Umwelt, Wald und Landschaft (BUWAL) sowie die zuständigen Departemente für Umweltschutz der drei Kantone Bern, Luzern und Appenzell-Ausser rhoden 1987 das IDHEAP beauftragten (Knoepfel und Zimmermann 1992). Diese Studie zeigte deutlich die verhältnismässig schwache Position der Umweltpolitik gegenüber der im ländlichen Gewässerschutz dominierenden Landwirtschaftspolitik auf. Sie machte Vorschläge für eine Verstärkung der Umweltseite, die auch von den landwirtschaftspolitischen Akteuren zur Kenntnis genommen und später teilweise aufgegriffen wurden.

Ein weiteres Beispiel bilden Studien zur Evaluation des Strassenverkehrsrechts unter dem Gesichtswinkel des Umweltschutzes (Baroni, Brunner, Knoepfel und Moor 1991) oder Untersuchungen zur Kostendeckung des Strassenverkehrs, die auch dessen soziale Kosten (v. a. Unfälle, Lärmbelästigungen, Gebäudeschäden) einbeziehen und monetär bewerten (Generalsekretariat EVED 1993).

- *Eine "ordnungspolitische Standortbestimmung" vornehmen:* Auch hier geht es um Standorte und Perspektiven, aber diesmal nicht zwischen einzelnen öffentlichen Politiken, sondern zwischen gesellschaftlichen Subsystemen (Wirtschaft, Politik, Kultur, Religion, Erziehung) und deren vorherrschenden Rationalitäten. Dominant sind namentlich ökonomische Analysen, die im deutschen Sprachgebiet mit dem Terminus "ordnungspolitische Standortbestimmung" bezeichnet werden. Solche Analysen werden zu zahlreichen Themen durchgeführt: Mietrecht, Arbeitsmarkt, Käsemarkt, Anlagefonds, Umweltschutz, Telekommunikation u.a.m. In der Regel handelt es sich um den Vergleich zwischen theoretischen Axiomen und vorgefundenen staatlichen Regelungen, was an sich noch keine Politikevaluation im Sinne dieses Buches ist. Zum

Teil werden die theoretischen Überlegungen aber durch empirische Untersuchungen, namentlich durch Wirksamkeitsstudien, ergänzt, um die Aussagen zu belegen (z.B. international vergleichende Untersuchungen über den Zusammenhang zwischen arbeitsmarktrechtlichen Regelungen und Arbeitslosenquoten). Insofern haben solche Studien durchaus auch Evaluationscharakter. Sie sind oft den Interessen derjenigen Akteure verpflichtet, die auf "Deregulierung" setzen.

6.4.3. Die Wahrheit herausfinden

Es gibt auch Evaluationen, die ohne die Absicht durchgeführt werden, einer Organisation oder Institution Vorteile zu verschaffen. Sie folgen routinierten Pfaden der Wahrheitsfindung oder Bewertung. Der Wahrheitsfindung dient primär das Wissenschaftssystem, welches sich bemüht, in die anderen Subsysteme "objektive Informationen" einzuspeisen. Möglichst objektive (wenn auch inhaltlich keineswegs unumstrittene) Bewertungen im Hinblick auf bestimmte Kriterien werden zunehmend auch von supranationalen Organisationen (z.B. OECD, Weltbank) durchgeführt.

Wissenschaftliche Institutionen (Universitätsinstitute, Forschungsanstalten des Bundes, private Forschungsbüros u.a.m.) produzieren eine Vielzahl von solchen Forschungsergebnissen, von denen nicht wenige "policy-relevant" sind (z.B. Formen der Übertragung des Rinderwahnsinns, Nährstoffversorgung von Fichtennadeln, Treibhausgase und Atmosphäre), weil sie indirekt Potentiale oder Risiken staatlichen Handelns aufzeigen. Akademisch orientierte Forschung kann namentlich in der Phase der Problemidentifizierung wertvolle Dienste leisten (z.B. Armutsstudien in einzelnen Kantonen), indem sie Informationsgrundlagen bereitstellt, auf die sich staatliche Massnahmen stützen können. Sind die entsprechenden Massnahmen einmal in Kraft, so können Studien Aussagen über deren Wirkungen machen. Dazu gehört auch die Inzidenzforschung, welche Aufschluss über Kosten und Nutzen einzelner Massnahmen für Segmente der Bevölkerung (Einkommensschichten, Regionen usw.) gibt (z.B. Steuerinzidenz, Kosten/Nutzen Nationalstrassenunterhalt, Inzidenz der AHV). Ein Teil dieser Wirkungsforschung ist primär akademisch motiviert; dabei ist nicht auszuschliessen, dass Forscherinnen und Forscher implizite politische Präferenzen haben und/oder Forschungsergebnisse faktisch auch Einfluss auf die strategische Position einzelner Akteure haben. Die Forschung verläuft indessen im Kontext von Lehre, Forschung, Pu-

blikationen und etablierter Forschungsförderung (Schweizerischer Nationalfonds), woraus man gemeinhin auf ihre primär akademische Orientierung schliesst.

Der Wahrheitsfindung dient neben der akademischen Forschung nach ihrem eigenen Selbstverständnis auch die Evaluationstätigkeit *staatlicher Kontrollorgane* (Dienststelle für Verwaltungskontrolle des Bundesrats, Parlamentarische Verwaltungskontrolle, Eidgenössische Finanzkontrolle; vgl. Kapitel 2.4.3.). Zu erwähnen ist in diesem Zusammenhang schliesslich auch die (oben angesprochene) international vergleichende Länderevaluation in Bezug auf bestimmte Anforderungen. Besonders bedeutsam sind für die Schweiz namentlich die Länderstudien der *OECD* in verschiedenen Politikbereichen (Finanz- und Wirtschaftspolitik, Bildungspolitik, Landwirtschaftspolitik, Regionalpolitik, Umweltpolitik u.a.m.), die periodisch oder sporadisch durchgeführt werden. Ein Teil dieser Studien hat durchaus den Charakter der oben angesprochenen "ordnungspolitischen Standortbestimmungen", indem sie staatliche Massnahmen mit einem Kanon gesammelter Kenntnisse vergleichen. Supranationale Reviews dieser Art sind indessen zumindest teilweise auch offen für eine Aufdatierung ihres Erfahrungsschatzes durch "success stories", die mit unkonventionellen Massnahmen erzielt wurden (z.B. Jänicke und Weidner 1995).

6.4.4. Durch Studienaufträge Koalitionen bauen

Bisher haben wir den Blick auf die Ziele und Wirkungen von Einzelstudien gerichtet. Ein Grossteil von Beratung, Forschung und Evaluationen vollzieht sich indessen in strukturierten Kontexten (z.B. Spezialisierung von Evaluatoren in bestimmten Politik- und Forschungsbereichen). Dies führt zur Herausbildung konstanter Erwartungen (von Forschern und Forschungsadressaten) oder bestimmter Karrieremuster (z.B. von universitären Forschungseinrichtungen in staatlichen oder parastaatlichen Organisationen). Forschung und Evaluation bilden dann Teil des politisch-administrativen Behördenarrangements (vgl. dazu auch Kapitel 8.2. - Netzwerke - und 9 - Mediation). Sie versorgen staatliche Akteure kontinuierlich mit einigermaßen berechenbaren Informationen und Argumentationsmustern und haben auf diese Weise ein konstantes Auftragsvolumen. Durch Evaluationen lassen sich längerfristig durchaus auch die Interessenpositionen der Forschergemeinschaften absichern. Beispiele für solche kontinuierli-

chen Beziehungen zwischen staatlichen Akteuren und Forschungs-/ Evaluationseinrichtungen finden sich auf Bundesebene beispielsweise in der Umweltforschung des BUWAL (z.B. Natur- und Landschaftsschutz), in der Wohnbauforschung des Bundesamts für Wohnungswesen, im Rahmen der Modellversuche Strafvollzug (einschliesslich Auswertung) des Bundesamts für Justiz oder im Programm "Energie 2000" des Bundesamts für Energiewirtschaft.

6.4.5. Beschwichtigen, indem Studien in Auftrag gegeben werden

Wie jede staatliche Massnahme kann auch die Ankündigung und Durchführung einer Evaluationsstudie als symbolischer Akt eingesetzt werden (Edelman 1976; Kinderman 1988). Die Evaluation hat dabei keine primär instrumentelle, aufklärerische oder legitimierende Funktion. Sie dient vielmehr dazu, auf wissenschaftliche Abklärungen zu vertrusten, damit politischem Forderungsdruck die Spitze gebrochen und Zeit gewonnen werden kann. Die Ankündigung der Durchführung einer Evaluation zeigt, dass das Problem einer seriösen Abklärung bedarf und dass hierfür das Notwendige unternommen wird. Die symbolische Funktion der Evaluation schliesst dabei nicht aus, dass sich Wirkungen zeigen: Die Ankündigung einer Evaluation kann beispielsweise eine vertrauensbildende Massnahme (die allerdings auch Erwartungen wecken kann) oder eine Warnung vor legislatorischen Schnellschüssen sein. Angesichts der Respektabilität wissenschaftlicher Arbeit ist die Forderung nach wissenschaftlicher Analyse bzw. das Abwiegen von Forderungen unter Verweis auf die Notwendigkeit solcher Analysen im politischen Alltag recht üblich geworden. Der Nachweis einer rein symbolischen Funktion von Evaluation ist dabei freilich im konkreten Fall schwierig zu erbringen. Denn man müsste zeigen, dass durch die Anlage der Evaluationsuntersuchung jegliche, also auch legitimatorische Nutzung ausgeschlossen wurde. In Analogie zur symbolischen Gesetzgebung (Kindermann 1988: 229) dürfte auch die symbolische Funktion von Evaluationen selten in reiner Form vorkommen, sondern ein Kontinuum bis hin zu wirksamer Nutzung bilden.

6.5. EVALUATIONEN UND ALLGEMEINPOLITISCHER KONTEXT

Evaluationen können bestenfalls Aussagen über Zusammenhänge zwischen staatlichen Massnahmen und daraus in der Gesellschaft resultierende Wirkungen machen. Ihre Interpretation erfolgt indessen in der Regel im Rahmen eines umfassenden politischen Kontexts. Politik ("politics") ist das Ringen um politische Macht. In diesem Rahmen sind politische Konzepte oder Ideologien für die konkurrierenden Parteien Versuche, Sinn zu stiften. Ähnlich wie wissenschaftliche Aussagen weisen solche Konzepte oft einen "Wenn-dann"-Charakter auf (z.B.: Wenn die Staatsquote beschnitten wird, wird die Wirtschaft leistungsfähiger). Sie enthalten Leitsätze in Form von makro-gesellschaftlichen Kausalbeziehungen (vgl. auch Kapitel 5.1.). Anders als wissenschaftliche Ergebnisse basieren sie aber auf Erfahrungen, auf Vermutungen oder auf Glaubenssätzen. Man kann deshalb sagen, dass dort, wo die Aussagekraft wissenschaftlich-methodisch orientierter Evaluationen aufhört, legitimerweise das Feld der politischen Ideologien beginnt.

Wissenschaftlich abgesicherte empirische Evaluationsresultate lassen sich in der Regel nur bei eng umgrenzten Untersuchungsobjekten erzielen. Die Interpretation und (vor allem) die Übertragung (Generalisierung, externe Validierung) von Evaluationsresultaten erfolgt daher in der Regel unter Beizug solcher politischer Konzepte. Wenn beispielsweise eine Evaluation zum Schluss kommt, dass die eine Massnahme A um X Prozent erwünschte Zunahme des Indikators Z nicht erreicht habe (was in dieser Klarheit selten ist), dann lässt sich daraus, je nach Ideologie, sowohl der Schluss ziehen, die Massnahme sei aufzuheben, als auch jener, die Massnahme sei zu intensivieren. Die Nutzung von Evaluationsresultaten setzt somit in den meisten Fällen eine *politische Wertung* voraus.

Evaluationen beeinflussen durch den empirischen Gehalt, den sie haben, auch politische Ideologien. Sie haben Einfluss auf die Frage, (1) welcher Staat geeignet ist, kollektive Probleme anzugehen, und (2) wieviel Staat sinnvoll ist. Obwohl beide Fragen eng miteinander verwoben sind, wollen wir sie im folgenden einzeln betrachten.

6.5.1. Welcher Staat - welche Politikinstrumente?

Ordnungs-, Wohlfahrts-, Sozial-, Leistungs-, Anreiz-, kontextsteuernder, vermittelnder, verhandelnder, kooperativer, reparierender, vorausschau-

ender etc. Staat: An Eigenschaftswörtern fehlt es wahrhaftig nicht, um den heutigen oder den gewünschten "Staat" zu charakterisieren. Empirisch fächert sich der heutige Staat in vielerlei Facetten auf, und es gibt vielerlei Zugänge, um ihn und seine Aktivitäten zu analysieren bzw. gar zu erklären. Gleichzeitig wird auch die Palette der Instrumente staatlichen Handelns ("Policy tools") immer reichhaltiger, und die Auseinandersetzungen um den Einsatz der "richtigen" Instrumente werden immer heftiger (Beispiel: Umweltpolitik).

Evaluationen haben eine "nachhinkende", historische Funktion, indem sie die Reichhaltigkeit der Instrumentenpalette gewissermassen retrospektiv dokumentieren und die Erfahrungen mit einzelnen Instrumenten in spezifischen Politikbereichen aufbereiten. Sie legen Zeugnis ab vom Erfindungsreichtum der politisch-administrativen Praxis. Dadurch ermöglichen sie es, solche Erfahrungen - zeitlich, gebietsmässig oder politikbereichsbezogen - übermittelbar zu machen (eingehender dazu Bussmann 1993).

Allerdings ist es trotz der seit rund 30 Jahren intensiv betriebenen Evaluationsforschung nicht gelungen, eine *Instrumentenlehre* zu erarbeiten, welche Politikkontext und Instrumente staatlichen Handelns in eine eindeutige Zuordnung zueinander brächte. Es fehlt auch weiterhin an einer "General Theory of Public Intervention", welche rezeptartig für spezifische Kontexte Angaben über das oder die passenden Instrumente machen könnte (Bussmann 1993: 266). Dies hängt einerseits damit zusammen, dass die in den sechziger Jahren gepflegten kontingenztheoretische Ansätze im staatlichen Bereich nur eine beschränkte Anwendbarkeit haben (vgl. dazu Mayntz 1983: 17; Linder und Peters 1989b: 53-55). Andererseits haben auch rund 20 Jahre Politikanalyse die immer noch stark *disziplinbezogene Ausrichtung* von Politikstudien nicht ungeschehen machen können. Dies hat bekanntlich zur Folge, dass sich wissenschaftliche Instrumentenempfehlungen immer noch stark an der disziplinären Herkunft ihrer Autoren (aus Ökonomie, Rechtswissenschaft, Medienwissenschaften, Betriebswirtschaftslehre u.a.m.) orientieren.

Freilich wäre es falsch, auf einen Fehlschlag der wissenschaftlichen Befassung mit Instrumenten staatlichen Handelns zu schliessen. Die Möglichkeiten und Grenzen und die allgemeinen Einsatzbedingungen beispielsweise regulativer und finanzieller Instrumente sind im Laufe der vergangenen Jahre sehr viel präziser herausgearbeitet worden, als dies früher der Fall war. Weitere Arbeit ist indessen nötig. Dies gilt beispielsweise für die präzisere Identifikation der Voraussetzungen für den Einsatz

prozeduraler Regelungen oder von persuasiven Instrumenten (Informationskampagnen).

Damit gelangt jede instrumentenbezogene Evaluation staatlichen Handelns potentiell irgendwann in den politischen Raum. Die Forscherinnen und Forscher müssen damit rechnen, dass ihre Ergebnisse politisch "verwertet" werden. Erweisen sich etwa weiche Politikinstrumente wie Persuasion oder wirtschaftliche Anreize als unzureichend, so werden sich immer Politiker finden, die das Kind mit dem Bade ausschütten und die betreffende öffentliche Politik mit ihren Zielsetzungen insgesamt in Frage stellen werden. Oft sind dies dieselben Politiker, die seinerzeit für die Wahl weicher Instrumente eingestanden waren, weil sie damals politisch zwar die Zielsetzungen, nicht aber harte Vorgehensweisen befürwortet hatten. Evaluation im öffentlichen Bereich hat mit Politik zu tun, weshalb eine derartige politische Nutzung der Evaluationsergebnisse potentiell immer möglich ist und sein muss. Die Forscherinnen und Forscher müssen sich einer solchen Debatte stellen und sich bei der Formulierung ihrer Empfehlungen der gesellschaftlichen Relevanz ihres Tuns bewusst sein. Dies ist die gesellschaftliche Verantwortung, die die Sozialwissenschaftler für ihr Tun und Lassen tragen.

6.5.2. Wieviel Staat?

Evaluationen sollen beurteilen, ob staatliche Massnahmen positive Wirkungen bzw. keine übermässigen negativen Nebenwirkungen zeitigen. Sie sind deshalb geeignet, Bausteine für den Diskurs darüber zu liefern, welche *Rolle* und welchen *Umfang* der *staatliche Bereich* einnehmen soll.

Historisch gesehen sind Evaluationen an sich die Frucht einer skeptischen Haltung gegenüber den Möglichkeiten des Staates, komplexe Probleme zu lösen. Eine solche Haltung war bereits in den sechziger Jahren in den USA bei der Einführung der Sozialprogramme für unterprivilegierte Bevölkerungsschichten zu beobachten; sie hat seither fast alle Länder und fast die ganze Breite des politischen Spektrums erfasst. Evaluationen haben auf diese Weise sicherlich mit zur "*Entzauberung des Staat*" (Willke 1983) beigetragen. So wurde etwa in den USA für viele Sozialprogramme nur ein massiger Erfolg nachgewiesen, und namentlich im Strafvollzug wurden die meisten Versuche, die zu einer verbesserten Resozialisierung von Straftätern beitragen sollten, letztlich als Fehlschläge beurteilt ("nothing works"). Recht ernüchternde Ergebnisse staatlicher Massnahmen wurden auch in der bundesdeutschen, in der englischen und

- in jüngster Zeit - in der französischen Evaluationsforschung (Duran 1990; C.E.O.P.S. 1990; Commissariat général du Plan 1989; Conseil scientifique de l'évaluation 1993, 1994 und 1995; Baslé 1994) festgestellt. In der Schweiz erbrachten empirische Studien über Effektivität und Wirksamkeit öffentlicher Politiken des Bundes *gemischte Resultate*: Es wurden Mängel, aber auch Erfolge festgestellt. Zu wesentlich ungünstigeren Resultaten kamen und kommen auf ökonomischen Axiomen basierende Analysen öffentlicher Politiken ("ordnungspolitische Standortbestimmungen"), beispielsweise in der Landwirtschaftspolitik, der Ausländerpolitik, dem Grundstückerwerb durch Personen aus dem Ausland, dem Mietrecht, der Telekommunikation und bezüglich des im Umweltschutz verwendeten Instrumentariums.

Der Beitrag von Evaluationen zur übergreifenden Beurteilung staatlicher Eingriffe folgt allerdings nicht einem linearen Schema, wonach beispielsweise schlechte Evaluationsresultate automatisch zu einer schlechten Einschätzung der Opportunität staatlichen Handelns führte. Gelegentlich sind sogar gegenläufige Tendenzen zu beobachten wie gegenwärtig in den USA, wo die nicht allzu negative, z.T. sogar positive Beurteilung sozialpolitischer Massnahmen durch die Evaluatorinnen und Evaluatoren deutlich mit einer sehr pessimistischen Einschätzung der breiten Öffentlichkeit kontrastiert. Evaluationsresultate wirken mit anderen Faktoren zusammen, namentlich mit der *Alltagsbeobachtung* der Bürgerinnen und Bürger, mit "*wissenschaftlichen Konstrukten* (wie der ökonomischen Theorie des Staatsversagens, dem Konzept der "Kontextsteuerung" oder dem New Public Management) und mit ideologischen Versatzstücken (wie "mehr Freiheit - weniger Staat"). Eine wichtige Rolle spielt zudem der *internationale Vergleich*, innerhalb dessen "extreme" Fälle besondere Aufmerksamkeit finden. Solche Länder sind beispielsweise Schweden (hohe Staatsquote, Wohlfahrtsstaat), Neuseeland (Cutback Management, Privatisierung) oder die Niederlande (Kommunalreform mit dem bekanntesten Fall Tilburg). Internationale Organisationen wie die OECD sorgen durch ihre Berichterstattung, die sich ihrerseits an für Evaluationen massgebenden Kriterien (wie Nützlichkeit, Sorgfalt, Objektivität und Genauigkeit) orientieren soll, für die Verbreitung und Nutzbarmachung solcher länderspezifischen Erfahrungen.

Ähnlich wie bei der Frage nach der Angemessenheit der Instrumente wird die Evaluationsforschung auch bei der Frage nach dem Ausmass staatlichen Handelns unmittelbar in ihre gesellschaftliche Verantwortung genommen. Trotz des nur relativen, vermutlich niemals genau identifi-

zierbaren Einflusses von Evaluationsforschung auf politische Entscheidungen über das Ausmass staatlicher Tätigkeit, muss sich die Evaluationsforschung ihrer diesbezüglichen politischen Indienstnahme immer bewusst bleiben. Entsprechend vorsichtig sollten daher in Evaluationsberichten auch diesbezügliche Aussagen formuliert werden.

TEIL II : EVALUATION UND VERWANDTE INSTRUMENTE - NEUERE TENDENZEN

Nachdem die verschiedenen Gegenstände und Kriterien der Evaluation ausführlich dargestellt worden sind, soll im folgenden Teil auf neuere und sich zur Zeit entwickelnde Verwendungsformen der Evaluation und verwandter Instrumente hingewiesen werden. Anspruch, Profil und Funktion der Evaluation werden dabei aus einem jeweils anderen Blickwinkel beleuchtet. Evaluationen stehen oft im Kontext von - umfassenderen - Politikanalysen, die sich für die Mechanismen der Erbringung von Politikleistungen in Staat und Verwaltung interessieren, weshalb wir zunächst auf das Verhältnis von Evaluation und klassischer Politikanalyse sowie neueren Netzwerkanalysen eingehen (Kapitel 7). Als permanente Kontrolle im Vollzug hat in den letzten Jahren das Monitoring an Bedeutung gewonnen. Entsprechende Bestrebungen im Rahmen der Informationsverarbeitung und -gewinnung werden durch die Ideen des New Public Managements verstärkt und gefördert (Kapitel 8). Insbesondere zur Beilegung von schwierigen Konflikten haben sich Mediationsverfahren zunehmend als sehr wertvoll erwiesen; darin werden ebenfalls Zustandsbewertungen aller Art vorgenommen (Kapitel 9).

Viele der in diesem Kapitel angeführten Beispiele stammen aus dem engeren und weiteren Bereich der Umweltpolitik. Dies ist zum einen auf die Forschungserfahrungen der Autorin und der Autoren zurückzuführen. Allerdings lässt sich die Auswahl auch damit begründen, dass sich in diesem Politikfeld seit den frühen siebziger Jahren beispielhafte Entwicklungen anbahnten, die später auch in anderen Politikbereichen übernommen wurden. Dies gilt sowohl für die Politikanalyse ganz allgemein, wie auch für die Anwendungen der Netzwerktheorie, für das Politik-Monitoring und für die Mediation.

KAPITEL 7: EVALUATION UND POLITIKANALYSE

Peter Knoepfel, Ingrid Kissling-Näf, Werner Bussmann

7.1. EVALUATION UND KLASSISCHE POLITIKANALYSE

In der sozialwissenschaftlichen Praxis werden Evaluationen vielfach im Rahmen von einfachen oder vergleichenden Politikanalysen ("comparative public policy analyses"; "analyses de politiques publiques comparées") durchgeführt. Die Politikanalyse¹ konzentriert sich vor allem auf die Erklärung von Politikoutputs (vgl. dazu Mayntz 1981), d.h. sie betrachtet sie nicht als zufällig, sondern versucht, sie aus bestimmenden Merkmalen des Politikerzeugungsprozesses zu erklären. Ihre Fragestellung galt und gilt der Erklärung von Vollzugsvarianzen öffentlicher Politik. Dabei hat sie namentlich fünf Erklärungsvariablen herausgearbeitet, die weiter unten im einzelnen dargestellt werden.

Die Evaluationsforschung ist - historisch gesehen - älter als die Politikanalyse. Die frühere Evaluationsforschung konzentrierte sich ausschliesslich auf die Wirkungsfeststellung. Durch Messungen (z.B. Schultests) wurde beispielsweise ermittelt, ob die erhofften Wirkungen (z.B. Verbesserung der schulischen Leistungen) bei der Zielgruppe (z.B. Schüler aus unterprivilegierten Bevölkerungsschichten) eintraten. Die konkreten Umstände der Umsetzung der Massnahmen (z.B. Ausgestaltung des vorschulischen Unterrichts) interessierten dabei wenig ("black-box evaluations"). Erst als zutage trat, dass ungenügende Wirkungen möglicherweise nicht auf ein falsch konzipiertes Wirkungsmodell im Programm ("program failure"), sondern schlicht und einfach auf fehlende Umsetzung ("implementation failure") zurückzuführen waren, begann die Einsicht zu wachsen, dass die Implementation möglicherweise das "fehlende Glied" darstellte. Die Vollzugs- und später die Politikanalyse nahm sich dieser Umsetzungsfragen eingehend an und identifizierte die zentralen Faktoren, die für die Umsetzung von Bedeutung waren.

¹ Die nachfolgend dargestellte Konzeption ist nicht entscheidungstheoretisch ausgerichtet wie ein Teil der amerikanischen Literatur zur "policy analysis". Die Konzipierung und Umsetzung öffentlicher Politiken wird dort z.T. auch unter den Begriffen "policy sciences" und "policy studies" (vgl. Zeitschriften mit den entsprechenden Titeln) subsumiert. Entsprechende Standardwerke sind Brewer und de Leon 1983; Dunn 1994; Dunn und Kelly 1992; Howlett und Ramesh 1995; Lindblom und Woodhouse 1993; Nagel 1994 und Parsons 1995.

Nach und nach drangen die Ergebnisse der Politikanalyse auch in die Evaluationsforschung ein. Den Weg dazu hatte Michael Scriven (1976) bereits in den siebziger Jahren gewiesen, indem er auf die Bedeutung des "modus operandi" (d.h. des konkreten Wirkungspfades) hinwies. Scriven hatte anhand der Komponentenanalyse, d.h. der Aufgliederung des "Produkts" in voneinander unabhängige Einzelteile auch die Voraussetzungen dafür aufgezeigt, dass umfassende Analysen des Wirkungspfades in konkrete Verbesserungen umgesetzt werden können. Seit Ende der achtziger Jahre sind Ergebnisse der Politikanalyse in zunehmendem Masse auch in das konzeptuelle Gerüst der Evaluationsforschung übernommen worden (Palumbo und Oliviero 1989; Chen 1990; Chen und Rossi 1992; Palumbo und Calista 1990), so dass sich beide Forschungsrichtungen in gewissem Masse ineinander verschmolzen haben. Dies erlaubt der Evaluationsforschung heute besser als früher, die einzelnen Komponenten, welche die Erreichung der gesetzten Ziele (bezüglich Output, Impact und Outcome) begünstigen bzw. hindern, zu identifizieren.

Während in der frühen Evaluationsforschung die bereits klassischen "threats to validity" (history, maturation, testing usw., s. Campbell und Stanley 1963) als Erklärungsfaktoren für Variationen der Testergebnisse im Vordergrund standen, sind es heute in der politikanalytisch inspirierten Evaluationsforschung fünf Erklärungsdimensionen, deren Varianz für unterschiedliche Outputqualitäten verantwortlich gemacht wird (Mayntz 1980; Mayntz 1983; Knoepfel 1996a).

- *Varianzen bezüglich Struktur und Inhalt der Verwaltungsprogramme ("Policy-Design")*: Vermutet wird, dass sich zeitlich bzw. örtlich variierende Politikleistungen zumindest zum Teil aus entsprechenden Varianzen der Struktur der Verwaltungsprogramme (z.B. "inhaltlich hoch detailliert" versus "vage"; "Konzentration auf Inhalte" versus "Konzentration auf prozedurale Elemente") und deren Inhalte (vgl. dazu die fünf konstitutiven Elemente von Verwaltungsprogrammen in Kapitel 5.1.2.) erklären lassen. Dieser Hypothese liegt die klassisch-juristische Behauptung zugrunde, Politikumsetzung sei nichts anderes als eine mehr oder weniger "technische" Gesetzesanwendung, die in jedem Falle - unabhängig von Wertungen der beteiligten Verwaltungsakteure - zum gleichen, nachvollziehbaren Ergebnis führen müsse. Bekanntlich liegt einer der Ursprünge der Politikanalyse gerade darin, dass diese Behauptung in Frage gestellt wurde (Knoepfel 1979:36). Ihre tiefere Begründung erfährt diese Hypothese dann, wenn davon ausgegangen

wird, dass zumindest für kontinentaleuropäische Verhältnisse Verfassungsaufträge oder Verwaltungsprogramme immer noch eine wichtige Handlungsressource der beteiligten staatlichen Akteure darstellen. Verwaltungsprogramme haben normativen Charakter, aus dem die Behörden ihre Handlungslegitimität herleiten. Schon die Implementationsforschung der späten siebziger Jahre hat aber gezeigt, dass der Erklärungsgehalt von Verwaltungsprogrammen für den Politikoutput mitunter recht gering sein kann (Knoepfel 1995e: 198 f.). Dies liegt daran, dass die Verwaltungsprogramme nicht die einzige Legitimationsressource der beteiligten Behörden sind und dass dementsprechend empirisch die Politikergebnisse auch mit der Varianz der vier folgenden, nicht normativen Faktoren erklärt werden müssen.

- *Varianzen bezüglich der Konstellation behördlicher Akteure und ihrer Ressourcen ("Behördenarrangement")*: Die meisten öffentlichen Politiken werden nicht von einer einzigen staatlichen Instanz generiert und umgesetzt. Meistens sind daran eine Vielzahl administrativer, in jüngster Zeit auch gesellschaftlicher Akteure beteiligt. Heute wissen wir aufgrund empirischer Untersuchungen, dass Qualität und Quantität der Outputs öffentlicher Politiken stark von der Art und Weise der Kooperation bzw. Nicht-Kooperation behördlicher Akteure verschiedenster politisch-administrativer Institutionen (Bundesämter, Parlamente, öffentlich rechtliche Körperschaften etc.) verschiedener Ebenen unseres Staatswesens (Bund, Kantone, Gemeinden) untereinander und mit den Politikadressaten bzw. den Politikbetroffenen abhängen. Insbesondere Studien zur Politikumsetzung aus föderalistischen Ländern mit relativ hoher organisatorischer Autonomie der Gliedstaaten (Bundesrepublik Deutschland, Schweiz, USA), zunehmend aber auch Studien aus zentralistischen Staaten wie Frankreich oder England haben gezeigt, dass die Politikleistungen öffentlicher Politiken mit gemeinsamem (zentral-staatlichem) Verwaltungsprogramm unter den vollzugszuständigen Verwaltungsräumen stark variieren können.

Als Erklärung für diese Varianzen werden zunächst interne Merkmale der umsetzungsbeteiligten *Behördenarrangements* angeführt (Zahl der beteiligten Behörden, Integrationsgrad des Behördenarrangements, Zentralitätsgrad, Politisierungsgrad etc.). Mitunter lassen sich die Outputvarianzen aber auch erklären durch unterschiedliche *Kommunikations- und Interaktionsstrukturen* zwischen Behördenarrangement und "ausstehenden" gesellschaftlichen Akteuren wie Politikadressaten oder

Politikbetroffenen. Diese können kooperativ oder konfliktiv, locker oder intensiv, organisatorisch vernetzt oder punktuell individualistisch etc. konzipiert sein. Als entscheidend haben sich ausserdem die Ausstattungen der behördlichen und gesellschaftlichen Akteure mit unterschiedlichen *Handlungsressourcen* erwiesen (Recht, Gewalt, Geld, Information/Wissen, Personal, Organisation, Konsens oder Zeit; vgl. Kapitel 4.2.1.). Insbesondere bei stark technischen öffentlichen Politiken (z.B. Raumplanung, Umweltpolitik, Gesundheitspolitik, Forstpolitik, aber zunehmend auch Wirtschaftspolitik oder Gesundheitspolitik) spielt das Recht (= Verwaltungsprogramm) eine vergleichsweise geringere, die übrigen Ressourcen wie Geld, Wissen, Personal oder auch Zeit eine gewichtigere Rolle. Bei solchen Politiken ist der Erklärungsgehalt des Verwaltungsprogramms für die Politikergebnisse gegenüber dem Behördenarrangement und den verwendeten Ressourcen vergleichsweise kleiner (= abnehmende rechtliche Inputsteuerung).

- *Die Verhaltensweise der Adressaten und deren Verhältnis zu den Politikbetroffenen:* Politikleistungen variieren ceteris paribus auch in Abhängigkeit von der Fähigkeit der Adressaten und der Politikbetroffenen (vgl. zu den Begriffen oben 4.2.1.), an der konkreten Politikausgestaltung mitzuwirken. Mitwirkungsrechte bzw. -pflichten finden in vielen Fällen ihre Begründung bereits in den Verwaltungsprogrammen; ihre Wahrnehmung setzt indessen zusätzlich die Ressourcen Information, Zeit und Geld seitens der mitwirkungsberechtigten gesellschaftlichen Gruppen voraus. Diese Ressourcen sind oftmals ungleich verteilt (stärkere Position der Adressaten), weshalb symmetrische Dreiecksverhältnisse zwischen Behörden, Zielgruppen und Politikbetroffenen eher selten anzutreffen sind. Je stärker die Position der Adressaten, desto stärker werden ihre Interessen im Rahmen der Politikleistungen berücksichtigt. Allerdings bilden auch die Adressaten oft keine homogene Gruppe (z.B. Konkurrenzverhältnisse unter verschiedenen Unternehmungen), weshalb ihre Interessen nicht immer leichter organisierbar sind als diejenigen der Politikbetroffenen (z.B. Umweltschutz-, Konsumenten- oder Arbeitnehmerinteressenorganisationen).

Zu beachten ist, dass auch Evaluationen als Politikleistungen anzusehen sind. Sie stellen gewissermassen den letzten Schritt der Realisierung öffentlicher Politiken durch die Behörden dar. Gerade bei solchen Evaluationen ist die Möglichkeit der Mitwirkung der Adressaten und der Politikbetroffenen von zentraler Bedeutung (vgl. auch Kapitel 8.4.

zum Verhältnis zwischen Evaluation und Mediation sowie Kapitel 12.2.1. und 13.3.).

- Politikleistungen können aber auch deshalb variieren, weil sie durch Ereignisse mit spezifischer zeitlicher oder (seltener) räumlicher Inzidenz beeinflusst werden, die ausserhalb des Einflusses der politikbeteiligten Akteure liegen ("*situative Variablen*"). Derartige Veränderungen der politischen oder der politikspezifischen Grosswetterlage können das Kräfteverhältnis innerhalb des Behördenarrangements oder dasjenige zwischen den Behörden und den Adressatengruppen bzw. den Politikbetroffenen schlagartig verändern, ohne dass entsprechende Veränderungen der Verwaltungsprogramme, der Behördenarrangements oder des Verhaltens dieser Gruppen feststellbar wären. Meldungen über Katastrophen, neue wissenschaftliche Erkenntnisse, politische Skandale etc. können auf diese Weise zu wichtigen Einflussfaktoren und damit zu Erklärungsgrössen für plötzlich ansteigende oder sich inhaltlich verändernde Politikleistungen werden (Beispiel: Seveso-Syndrom, Tschernobyl oder Schweizerhalle, Berner Finanzskandal, Fichen-Affäre, New Public Management etc.).
- Namentlich im synchronen Vergleich von Politikleistungen in unterschiedlichen Verwaltungsräumen, aber auch in sehr langfristigen diachronen Vergleichen lassen sich variierende Politikleistungen mitunter auch aus *gesellschaftsstrukturellen Determinanten* erklären, die über mittlere Zeiträume kaum veränderbar sind. Das können etwa Eigentumsverhältnisse (in der Raumplanung), stabile Werthaltungen (etwa in der Gesundheits-, der Familien- oder der Umweltpolitik) oder kulturelle Präferenzen (etwa in der Schulpolitik) sein, die für die Herrschaftsverhältnisse und die Machtverteilung in bestimmten Verwaltungsräumen (Kantonen oder Gemeinden) charakteristisch sind. Diese Variablen können auch aufgefasst werden als eine strukturell verbürgte Zuteilung von politikrelevanten Ressourcen an behördliche und insbesondere an gesellschaftliche Akteure. Ähnlich wie die situativen Variablen beeinflussen sie die Position sämtlicher Akteure einer öffentlichen Politik im Kräftespiel der Produktion der Politikleistungen.

7.2. EVALUATION UND NEUERE POLITIKANALYSE (UNTER EINBEZUG DER VERÄNDERUNG VON NETZWERKEN ÖFFENTLICHER POLITIKEN)

Bereits im Kapitel 5 haben wir auf die Bedeutung des "Behördenarrangements" für die Erbringung von Leistungen öffentlicher Politiken hingewiesen; in Kapitel 6 wurde gezeigt, welchen Stellenwert Evaluationen für einzelne Akteure öffentlicher Politiken haben können. Im Hinblick auf jüngste Entwicklungen in der Politikwissenschaft soll in diesem Kapitel die gleiche Problematik nochmal in einem neueren, netzwerkorientierten Kontext dargestellt werden. Danach werden die Behördenarrangements als Teil umfassender, von staatlichen *und* gesellschaftlichen Akteuren besetzten Politiknetzwerken aufgefasst, für deren Konstitution, innere Machtverhältnisse und Politikleistungen auch Evaluationen einen wesentlichen Beitrag leisten können und die ihrerseits auf die Durchführung von Evaluationen Einfluss nehmen können².

7.2.1. Zum Begriff des Netzwerks

Die wachsenden Ansprüche an die staatliche Steuerungsfähigkeit stehen zunehmend in Konflikt mit dem tatsächlich existierenden obrigkeitlichen Problemlösungspotential. Hierarchische Interventionsformen werden darum zunehmend durch partizipative, prozedurale und diskursive Elemente ergänzt oder sogar ersetzt. Die politischen Entscheidungsprozesse haben sich also insofern gewandelt, als staatliche Akteure heute zunehmend innerhalb von Netzwerken zu korporativen Akteuren in eine gleichberechtigte und damit symmetrische Beziehung treten. Die institutionalisierten Kontakte z.B. in der Form einer interorganisationellen Politikimplementation (vgl. O'Toole 1993: 53) deuten an, dass Steuerung vermehrt über netzwerkartige Strukturen erfolgt und dass das Netzwerk neben dem Markt und der Hierarchie zu einer wichtigen Kooperationsform geworden ist.

In der theoretischen Diskussion lassen sich verschiedene Definitionen und Verwendungen des Begriffs Netzwerk finden. Klar unterschieden werden müssen (1) Struktur- und Entscheidungstheorien, die das Netzwerk als Koordinationstyp und Governanceform thematisieren, und (2) die

² Zum folgenden vgl. Kissling-Näf 1996: 18 ff.

eigentliche Netzwerkanalyse, die eine Methode der Strukturbeschreibung darstellt. Die ältere Unterscheidung zwischen der quantitativen und qualitativen Netzwerkanalyse hat an Bedeutung verloren.

- *(1) Netzwerke als (neuer) Koordinationstyp*

Für Mayntz (1993a) ist die Veränderung der Entscheidungsstrukturen eine Begleiterscheinung des Strukturwandels moderner Gesellschaften überhaupt. Die wachsende Bedeutung des Steuerungsmodus Netzwerk folgt aus der funktionalen Ausdifferenzierung gesellschaftlicher Subsysteme und der Herausbildung von formalen Organisationen, die mit einer zunehmenden Fragmentierung der staatlichen Macht zusammengehen. Netzwerke sind der Hierarchie und dem Markt insofern überlegen, als sie einen "qualitativ anderen Typus von Sozialstruktur repräsentieren, der durch eine Kombination von Elementen der beiden anderen grundlegenden Ordnungsformen charakterisiert wird, nämlich auf der einen Seite das für Märkte typische Vorhandensein einer Vielzahl von autonom Handelnden (oder Subjekten) und auf der anderen Seite die für Hierarchien typische Fähigkeit, gewählte Ziele durch koordiniertes Handeln zu verfolgen." (Mayntz 1993a: 44.) Staatliche Akteure sehen sich zudem zunehmend einer gut organisierten und informierten Klientel gegenüber, mit der verhandelt werden muss. Formale Entscheidungen werden deshalb in partnerschaftlichen Beziehungen zwischen staatlichen und korporativen Akteuren vorbereitet. Scharpf hat dafür den Begriff der "Selbstkoordination in Netzwerken" (vgl. Scharpf 1993: 125 ff. und 147 f.) geprägt. Netzwerke versinnbildlichen demzufolge einen relativ schwachen Staat, der eine grosse Sensibilität für die komplexer werdende politische Steuerung aufweist³.

Mit dieser Konzeption verbindet sich die Vorstellung einer besonders hohen und angemessenen Problemlösungsfähigkeit von Netzwerken. Schneider (1994) geht sogar soweit zu behaupten, dass es sich beim Policy-Netzwerk um ein Umbrella-Konzept handelt, das die Elemente des Pluralismus und Korporatismus aufzunehmen und zu integrieren vermag. In dieser Sichtweise ist das Netzwerk ein reales Phä-

³ Mayntz 1993a: 44; Mayntz 1993b: 13 ff.; Kenis und Schneider 1991 zeigen anhand der gesellschaftlichen Entwicklungen, warum das Netzwerkkonzept an Bedeutung gewonnen hat.

nomen, das für die Erklärung von Policy-Sachverhalten und Evaluationsbefunden beigezogen wird, weil es eine neue Realität beschreibt⁴.

- (2) *Netzwerk als Methode zur Beschreibung gesellschaftlicher Interaktionen*

Im Gegensatz dazu steht in der soziologischen Tradition die generelle Beschreibung der Struktur als Handlungsrahmen für das Beziehungsgeflecht im Vordergrund. Dabei bilden relationale Daten die Ausgangsbasis für die Untersuchung der Struktur gesellschaftlicher Interaktion⁵. Das Netzwerk wird im Sinne einer "generischen Bezeichnung" für unterschiedliche Kooperationsformen zwischen Staat und Interessengruppen aufgefasst (vgl. Van Waarden 1992: 31; Jordan und Schubert 1992: 12). Mehr oder weniger komplexe Kategorienraster dienen der Beschreibung und Charakterisierung von Beziehungen zwischen privaten und öffentlichen Akteuren. Die Beziehungen werden nicht als Merkmale der einzelnen Akteure, sondern der Akteursysteme verstanden⁶.

Vergleichend lassen sich die beiden Konzeptionen wie folgt zusammenfassen:

	(1) Politiknetzwerk	(2) Netzwerkanalyse
Variablentyp	erklärende Variable (in Politikanalyse)	zu erklärende Variable (in soziologischer Analyse)
Ontologischer Status	reales Phänomen	neutrales Konzept, generische Bezeichnung
Theoretischer Rahmen	Struktur- und Entscheidungstheorie (Steuerungsmodus)	Methode zur Strukturbeschreibung

⁴ Die Politikproduktion in politischen Netzwerken wird im Sammelband von Jansen und Schubert 1995 erstmals aus der soziologischen und politikwissenschaftlichen Forschungstradition beleuchtet. Zugleich wird der Versuch unternommen, das Verhältnis von Politikfeldanalyse und Netzwerkforschung zu klären. Zum Verhältnis zwischen Netzwerken und neokorporatistischen Strukturen vgl. Kissling-Näf 1996: 1.

⁵ Typische Vertreter sind Scott, Knoke und Kuklinski. So schreibt Scott (1991: 4) zur strukturellen Analyse: "Relational data are central to the principal concerns of the sociological tradition, with its emphasis upon the investigation of the *structure* of social situation. Structures are built from relations, and the structural concerns of sociology can be pursued through the collection and analysis of relational data."

⁶ Mit Wasserman und Faust 1994 liegt ein Grundlagenwerk und Kompendium zur Netzwerkanalyse vor, das deren Konzepte, Methoden und Anwendungen zusammenfasst und für Anfänger wie auch Spezialisten aufarbeitet.

- *Mischformen* zwischen den beiden dargestellten Konzeptionen sind durchaus denkbar. In unserem Zusammenhang wird der realen Variante der Netzwerkanalyse der Vorzug gegeben. Eine solche Mischform lässt sich etwa folgendermassen konzeptualisieren: "Politiknetzwerke können definiert werden als Konfigurationen von Akteuren, die in einer interdependenten Beziehung zueinander stehen und nicht an formale System- oder Organisationsgrenzen gebunden sind." (Döhler 1990: 45) Die überwiegend informellen Interaktionen sind ein Ergebnis der unterschiedlichen und wechselseitig abhängigen Interessen, die zur dezentralen Bearbeitung des Problems führen (vgl. Héritier 1993b: 432). Dabei kann kein Akteur die Tauschbeziehungen oder Entscheidungsprozesse allein kontrollieren, sondern jeder ist auf die Kooperation mit anderen Netzwerkteilnehmern angewiesen. Im Rahmen von Netzwerken gelingt es, Ressourcen zu mobilisieren und zu poolen, d.h. sie gemeinsam zu nutzen (vgl. Kissling-Näf, Marek und Gentile 1994: 20 f.). Der Verbund besteht als eine lose Koppelung, in der konfligierende Interessen vertreten sind. Manche Autoren sprechen auch von einer Proto-Organisation (vgl. z.B. Van Waarden 1993: 31) oder Ad-hoc-Entscheidungsstrukturen (vgl. Jordan und Schubert 1992: 13). Zur Konstitution eines Netzwerkes ist ein gemeinsamer Problembezug der Akteure erforderlich. Bei Policy-Netzwerken stimmen die Grenzen sehr oft mit dem inhaltlichen Politikfeld überein, weil im Netzwerk alle Akteure des Politikfeldes vertreten sind.

Neben dem Problembezug wirken als strukturelle Faktoren auch nationale Eigenheiten auf die Netzwerke ein, da sie das Reaktions- und Handlungspotential der Akteure eingrenzen und die Netzwerkkonfigurationen mitbestimmen. Für kleine Staaten wie die Schweiz sind etwa internationale Flexibilität und interne Stabilität wichtig. Typisch für unser Land ist des weiteren, dass der Staat und die Sozialpartner mit starken Interessengruppen verhandeln. Ein ausgeprägtes Repräsentationsmodell verbindet sich mit einem schwachen Staat und einem stabilen und dezentralen Kooperationsmodell (vgl. Kriesi 1980: 688 f.; Busch und Merkel 1992: 193-219; Katzenstein 1985: 12).

Entscheidend für die Konstitution des Netzwerks bleibt der *Problembezug*; dieser darf aber nicht mit der Übereinstimmung der Interessen oder der inhaltlichen Problemlösungsvorschläge der beteiligten Akteure verwechselt werden. Der Handlungszusammenhang darf nicht zu abstrakt sein, weil über den Problembezug trotz (oder wegen) inhaltlicher Diver-

genzen minimale kognitive Gemeinsamkeiten geschaffen werden sollen. Eine gewisse Stabilität wird im Netzwerkverbund zwar über Austauschbeziehungen erreicht, aber kollektives Handeln ist so voraussetzungsreich, dass es der Unterstützung durch kognitive Gemeinsamkeiten bedarf. Dabei kann es sich um gemeinsame Problemwahrnehmungen oder um anerkannte Ordnungsprinzipien handeln (vgl. Döhler 1994: 45 f.; Kissling-Näf und Knoepfel 1993: 35 f.)

Die Basis für die Interaktionen im Netzwerk stellen die *gegenseitige Abhängigkeit* und der *Austausch von Ressourcen* dar. Hier werden die verschiedenen von den einzelnen Akteuren eingebrachten Ressourcen zusammengeführt, neu bewertet und gepoolt. In Politiknetzwerken spielen die Ressourcen Konsens, Recht, Informationen, Geld und Zeit eine Rolle. Die Macht einzelner Akteure hängt von der Verfügungsgewalt über eine breite Palette dieser Ressourcen ab. Das Ausmass bestimmt sich aber auch aus der Angewiesenheit eines Akteurs auf die Mittel eines anderen und der Möglichkeit, sich die Mittel anderweitig zu beschaffen (vgl. Kissling-Näf, Marek und Gentile 1994: 20 ff.).

7.2.2. Evaluation als Ressource im Netzwerk⁷

Evaluationen stellen eine besondere Form der Ressource Information innerhalb eines Netzwerkverbundes dar. Vielfach sind Netzwerkteilnehmer erfahrene Spezialisten und aktive Politikbeteiligte, die unabhängig von ihrer Ausbildung Kenntnisse im entsprechenden Bereich aufweisen. Die Aufgabe der öffentlichen Verwaltung besteht heute weniger darin, den Vollzug im einzelnen zu organisieren, als vielmehr weitreichende gesellschaftliche Entscheidungen vorzubereiten und zu fällen. Wie bereits gezeigt, ergeben sich mit Netzwerken und informalen Strukturen neue Wege für die Bearbeitung von interdependenten Problemen und Konflikten, die zur Entlastung der formalen Strukturen beitragen⁸. Im Normalfall muss das notwendige Wissen zusammen mit den Interessengruppen aufgearbei-

⁷ Vgl. auch Kapitel 13.

⁸ Netzwerkartige Beziehungsmuster lassen sich danach unterscheiden, ob sie sich erstens vor allem aus politischen Generalisten zusammensetzen oder zweitens aus Fachbrüdern bestehen. Fachlich spezialisierte Netzwerke dürften stabiler sein als der lose Verbund zwischen politischen Vertretern (z.B. Bürgermeister), da Ausbildung und Professionalisierung sowie deren Zuständigkeit eine bessere Basis für gemeinsames Handeln darstellen. Eine dritte Form ist mit Verbünden zwischen privaten und öffentlichen Organisationen gegeben. Vgl. Benz 1995: 197 ff.

tet oder bei Spezialisten eingeholt werden. Es müssen Fachleute und Beratungsbüros konsultiert werden, da heute niemand mehr genau weiss, was sicher und richtig ist. Hinzu kommt, dass in vielen Politikbereichen aufgrund der zunehmenden Komplexität eine weitreichende *Technisierung und Professionalisierung* eingesetzt hat. Die Produktion von politischen und wissenschaftlichen Gewissheiten wird heute darum sehr oft an Spezialisten delegiert. In konfliktuellen Situationen wird der Sachverstand vermehrt bei neutralen Experten abgerufen (vgl. Heclo 1978: 102 ff.; Knoepfel 1996a: 38 f.).

Insbesondere Evaluationen von naturwissenschaftlichen Zusammenhängen und der Auswirkungen (möglicher) gesetzgeberischer Entscheide tragen zur kollektiven Meinungsbildung bei und sind eine wichtige Hilfestellung bei dieser Abarbeitung von Unsicherheiten. Es ist jedoch nie genau vorhersehbar, wie eine Expertise den Verfahrensablauf beeinflussen wird, denn jeder Akteur wird auf seine Art versuchen, das Expertenwissen für seine Interessen zu nutzen und in die eigene Strategie einzubauen. Dabei ist es durchaus denkbar, dass das neue Wissen keinen Einfluss auf die Problemlösung und die Akteurkonstellation hat. Dieser Fall dürfte dann eintreten, wenn die Studie vorwiegend *legitimatonschen Wert* hat und von Beginn an nur als Gefälligkeitsgutachten konzipiert war.

Im gegenteiligen Fall kann eine Evaluation zur *Stärkung von einzelnen Positionen* im Netzwerk oder gar zur Modifikation der Netzwerkzusammensetzung (Erweiterung um neue Akteure; Ausschluss "unnötiger" Akteure) führen. Der Experte kann durch seine Untersuchung die Meinung einer Koalition unterstützen und so deren Gewicht innerhalb des Netzes verstärken. Die Netzwerkteilnehmer werden darum in strittigen Fällen schon von Beginn an bemüht sein, einen Experten ihrer "Couleur" zu finden, oder aber zu einem späteren Zeitpunkt ein Gegengutachten erstellen zu lassen. Erfahrungsgemäss lassen Evaluationen die Netzwerkkonfigurationen und Machtkonstellationen nur selten völlig unberührt. Denn in Evaluationen fliessen bereits die Beschreibung, das Abbild oder die Beurteilung der Ausgangssituation ein, die in der Folge hinterfragt oder bestätigt werden. Bekanntlich stellt bereits die Beschreibung einen Eingriff in die Wirklichkeit dar.

Je nach Art des Expertenwissens und des Evaluationstyps (partizipativ, technokratisch - vgl. unten Kapitel 11) sind die Auswirkungen von Evaluationsstudien auf den Problemlösungsprozess und das Akteurarrangement unterschiedlich gross. Am grössten sind sie naturgemäss dann, wenn eine Reorganisation der Akteurstruktur einer Politik oder eines Amtes

angestrebt wird⁹ und die Entwicklung der Expertenvorschläge partizipativ im Sinne der Organisationsentwicklung erfolgt¹⁰. Veränderungen der Machtkonfigurationen innerhalb des Netzwerks sind aber auch dann denkbar, wenn die Evaluation nur sogenanntes technisches Wissen zum Gegenstand hat, z.B. die Ergebnisse einer geologischen Bohrung für die Abklärung eventuell möglicher Ablagerungen nuklearen Materials. Dadurch können einzelne Positionen im Netzwerk unterstützt und andere geschwächt oder ausgehöhlt werden. Neues Wissen verändert oder gestaltet den Prozess aber nur dann, wenn die Evaluationsergebnisse von mehreren Akteuren aufgenommen und in den Meinungsbildungsprozess integriert werden, so dass sich einzelne Gruppenmitglieder argumentativ wieder auf die Studie beziehen können. Die relationale Struktur, d.h. die gegenseitige Bezugnahme auf ein Dokument ist Voraussetzung dafür, dass die Evaluation Auswirkungen auf den Ablauf des Geschehens oder gar auf die Netzwerkkonfiguration hat. Auch ältere Gutachten, die in einem anderen Kontext entstanden, können als wichtiges Argument dienen¹¹.

Funktion und Aufgabe der neutralen Experten können sich im Zeitablauf ebenfalls ändern. Als neutraler Experte konsultiert, werden der Fachmann oder die Fachfrau - wie bereits unter Kapitel 6.4.4. ausgeführt - immer stärker in die Gestaltung des Politikfeldes einbezogen. Ihr Know-how öffnet ihnen den Zugang zum Netzwerk. Werden sie mit einem Dauerauftrag betraut, entwickeln sie sich zu Netzwerkteilnehmern wie alle anderen. Es erstaunt darum nicht, dass der Übergang vom Experten zum Politikfeldbeteiligten fließend ist (vgl. dazu unten Kapitel 8.4.: Administrativ-wissenschaftlicher Komplex).

7.2.3. Zusammenfassung

Die neuartigen Problemlagen, die damit verbundene Unsicherheit, die sich vervielfachenden und gleichzeitig überlappenden Zuständigkeitsbereiche, die unsicheren, oft ausbleibenden oder mit Nebenwirkungen ver-

⁹ Vgl. zur Reorganisation des Bundesamt für Umwelt, Wald und Landschaft (BUWAL): Knoepfel und Zimmermann 1991.

¹⁰ Beispiel: Reorganisation des Bundesamtes für geistiges Eigentum (BAGE) oder des sanktgallischen Amtes für Umweltschutz (Ruch 1993 bzw. Knoepfel, Baitsch und Eberle 1995).

¹¹ Majone 1993 spricht in diesem Zusammenhang auch von Post-Entscheidungs-Argumenten, die Handlungszusammenhänge erklären und eine Politik legitimieren.

sehenen Folgen staatlichen Handelns sind schwierig zu meisternde Herausforderungen. Der Anspruch moderner Industriegesellschaften, mittels rationaler Planung auf die eigene Entwicklung einzuwirken, hat einer eigentlichen "Entzauberung" des Staates (Willke 1983) Platz gemacht. Gleichzeitig wurde aber auch der Blick freigemacht für funktionale Äquivalente, die die Entscheidungen in Netzwerke hineinverlegen und stärker partizipative Verfahren begünstigen. Diese Ad-hoc-Gruppen erhöhen die Innovationsfähigkeit und stellen eine Plattform für den Meinungsbildungsprozess dar. Sie setzen sich vorwiegend aus Politikfeldspezialisten zusammen (vgl. Offe 1987; Von Prittwitz 1993). Für die Lösung von komplexen Problemen werden die Mobilisierung bestehenden und die Produktion ausreichenden neuen problembezogenen Wissens zu einem zentralen Faktor. Gelingt es den Politikfeldspezialisten nicht mehr, das entsprechende Know-how selber einzubringen, so wird es extern angefordert. Als zentrale Ressource können Evaluationen die Strategien der Akteure, die vorherrschenden Machtkonfiguration und Entscheide im Politikfeld beeinflussen und kanalisieren. Die Experten können so langsam Bestandteil des Politiknetzwerks werden.

KAPITEL 8: EVALUATION UND MONITORING

Ingrid Kissling-Näf, Peter Knoepfel

Die Evaluation kann auch als ein Kontrollinstrument verwendet werden: Diesbezüglich weist sie Ähnlichkeiten auf mit dem Controlling und dem Monitoring, die heute ebenfalls zum Alltag jeder Verwaltung gehören. Obwohl keine der Evaluierungsformen in reiner Form existiert und sogar je nach Fachgebiet eine Unterordnung des einen unter den anderen Fachterminus erfolgt¹, lassen sich unter Kontrollgesichtspunkten für die drei Typen spezifische Charakteristika ausmachen. In Abgrenzung zum Controlling und zur Evaluation soll im folgenden gezeigt werden, wann das Monitoring zur Überwachung und Kontrolle von staatlichen Aktivitäten geeignet ist, wie es eingesetzt und weiterentwickelt werden kann (vgl. Kissling-Näf und Wildi-Ballabio 1993).

8.1. ZUM BEGRIFF MONITORING

Die Staats- und Verwaltungslehre hat in Anlehnung an die betriebswirtschaftlichen Konzepte zur Überwachung der Unternehmenstätigkeit den Monitoring-Begriff entwickelt und versteht heute darunter "die Aktivitäten der Informationsgewinnung und -Verarbeitung, die für den laufenden Vollzug eines Programmes erforderlich sind" (Freiburghaus 1988: 7). Monitoring meint im wesentlichen die Sammlung der für die praktischen Zwecke notwendigen, laufenden Informationen über Umfang und Richtung von Veränderungen im interessierenden Handlungs- und Politikfeld. Anvisiert sind sowohl Politikleistungen als auch gesellschaftliche Veränderungen. Es handelt sich dabei um eine beschreibende Tätigkeit, bei der eine möglichst zuverlässige Darstellung der Objekte erreicht werden soll. Die Daten beruhen auf periodischen, mehr oder weniger häufigen Erhebungen und haben oft den Charakter von Zeitreihen. Die kausalen Zusammenhänge der untersuchten Gegenstände bzw. ihrer Merkmale werden dabei nicht systematisch erschlossen. Die Sammlung von Daten ist jedoch insofern nicht theoriefrei, als ihr erahnte Zusammenhänge oder

¹ So ist **z.B.** bei Rossi und Freemann (1993) das Monitoring integraler Bestandteil der Evaluation. Im Projektmanagement (Methoden der Projektüberwachung) stellt das Monitoring einen Bestandteil des Controlling dar. Vgl. Madauss 1991: 208 ff.

Hypothesen zugrundeliegen (Spöhring 1989: 42 f.; Hellstern und Wollmann 1984: 37 f.). Der Grad der Kenntnisse über Wirkungsverläufe bestimmt den Charakter des Datenmaterials. Es handelt sich dabei um selektive Informationen über einzelne Elemente des Vollzugs, die zu dessen Feinsteuerung weiterverwendet werden können. Das Monitoring stellt daher eine grobe, begleitende Evaluation einer öffentlichen Politik dar, die sich in der Regel auf deren Effektivität (vgl. Kapitel 5.4.3.) beschränkt. Die Auswertungen werden von Verwaltungen meist auf eine mechanisierte, einfach handhabbare Weise vorgenommen. Diese routinisierte Art der Kontrolle ist im Vergleich zu eigentlichen Wirksamkeitskontrollen relativ kostengünstig.

Monitoring kann auf der Ebene des Gesamtsystems, des Politikfeldes, der Einzelmassnahmen oder der Instrumente vorgenommen werden. Die gewählte Ebene hat einen Einfluss auf die Komplexität des Beobachtungssystems. Politikanalytisch gesprochen kann jeder Monitoring-Prozess bei den Outputs, den Impacts und den Outcomes ansetzen. Bezüglich der Akteure lassen sich Informationen über die Adressaten, die Politikbetroffenen oder über die umsetzenden Verwaltungsstellen einholen². Das Monitoring liefert Informationen für das Management von Programmen. Diese Art der Kontrolle befasst sich z.B. mit:

- der Prüfung der Outputs bzw. der Outputprofile in Raum und Zeit der entsprechenden Politik (im Sinne von Kapitel 5.4.1.);
- der Prüfung der Programmreichweite, d.h. der Erreichung der Zielpopulation (Ebene der Impacts im Sinne von Kapitel 5.4.2.), oder
- dem Versorgungsnachweis und somit der Frage, ob die Adressatengruppen in der gewünschten Art und Weise auf die staatlichen Impulse reagieren und demzufolge die Programmleistungen im richtigen Umfang erbracht werden (Beziehung zwischen Output und Impact).

Monitoring-Daten lassen sich in fast allen Politikbereichen finden; allerdings können Menge und Verarbeitung der Informationen je nach Politikbereich stark variieren. Bis anhin haben Vollzugsüberwachungssysteme vorwiegend die Quantität und (seltener) Qualität der staatlichen Leistungen erfasst. Die Informationen liegen darum mehrheitlich auf der Ebene

² Vgl. Arbeitsgruppe "Gesetzesevaluation" (AGEVAL) 1991: 20, 47.

der Outputs (Zahl der bewilligten Gesuche, Höhe der Subventionen, Anzahl der unternommenen Sanierungen).

Staatsausgaben und Arbeitslosenquote als klassische Monitoringdaten

Eine der am häufigsten verwendeten Beobachtungsgrössen sind die *Staatsausgaben*. Die Verwendung dieses Indikators ist mit zwei grossen Vorteilen verbunden: es handelt sich dabei um eine aggregierte Grosse, die über den Umfang der Staatstätigkeit informiert und für internationale Vergleiche beigezogen werden kann. Nachteilig ist jedoch, dass die Staatsausgaben Input-Daten sind, die uns zwar sagen, wieviel an Mitteln in die Produktion staatlicher Güter fliesst, aber keine Informationen über die tatsächlich erzielten staatlichen Leistungen und Aktivitäten liefern. In einigen Politikbereichen kann es durchaus angezeigt sein, outputbezogene Indikatoren wie z.B. die gesetzgeberischen Tätigkeiten für die Beobachtung staatlichen Handelns zu verwenden. Der Beitrag der jeweiligen Staatsaktivität zur Zielerreichung kann anhand der Staatsausgaben dabei allenfalls geschätzt, die Zielerreichung aber nicht beurteilt werden. Ebenfalls zu beachten ist, dass der gewählte nationale Instrumentenmix in einen bestimmten Politikfeld von kulturellen und historischen Gegebenheiten abhängig ist. Da meist nur öffentliche Ausgaben für die Beurteilung der Leistungserstellung verwendet werden und in manchen Ländern ein Teil der Sozialtransfers privat (über Abgaben der Arbeitgeber) erfolgt, kann sich im internationalen Vergleich beim ausschliesslichen Abstützen auf die öffentlichen Ausgaben das Bild zusätzlich verzerren (vgl. Castie 1993; Simeon 1993).

Im Gegensatz dazu ist das Informationssystem im Bereich des *Arbeitsmarktes* aussagekräftiger, weil dort die Leistungsindikatoren physisch besser messbar sind. Es liegen Zielindikatoren vor, die die öffentliche Tätigkeit und die Handlungsziele kurzschliessen, so dass der Beitrag zur Zielerreichung in etwa eruiert werden kann. Die Höhe der Arbeitslosenrate kann Auskunft über den Erfolg der Arbeitsmarktpolitik geben. Aber auch diese Zahlen sind mit Vorsicht zu gemessen, denn es ist durchaus denkbar, dass aus definitorischen Gründen statistisch weniger Arbeitslose ausgewiesen werden, als tatsächlich Personen eine Erwerbsarbeit (unter besseren konjunkturellen Bedingungen) aufnehmen möchten. So verzeichnet das BIGA seit Februar 1994 ein Stagnieren der Arbeitslosigkeit. Von verschiedenen Seiten wird jedoch darauf hingewiesen, dass der Rückgang zu einem erheblichen Teil auf das Ausscheiden von Erwerbslosen aus der BIGA-Statistik zurückzuführen ist. Deshalb stimmen die Angaben der Arbeitslosenstatistik und der Schweizerischen Arbeitskräfteerhebung (SAKE) nicht vollumfänglich überein³.

³ Dies hängt damit zusammen, dass die Definitionen und der Erhebungsmodus nicht dieselben sind. Bei der Arbeitslosenstatistik des BIGA gilt als erwerbslos, wer bei den Arbeitsämtern eingeschrieben ist und somit den Anforderungen des Bundesgesetzes über die Arbeitslosenversicherung und Insolvenzenschädigung entspricht. Bei der Arbeitskräfteerhebung (SAKE) werden demgegenüber jene Personen als erwerbslos erfasst, die

Im Gegensatz zum *Monitoring* dient die *Evaluation* der vertieften Analyse von Wirkungszusammenhängen. Hohe Kosten und ein entsprechender Zeitaufwand machen aber eine permanente und flächendeckende Anwendung dieses Instruments unmöglich.⁴ Meist handelt es sich dabei um ex post-Studien, die einen Politikbereich auf die Effektivität, Wirksamkeit oder auf die Effizienz hin untersuchen. Im Zentrum steht dabei der Wirkungszusammenhang, wobei die Reichweite der erfassten Wirkungen je nach Studie variiert. Beim Monitoring ist die kausale Zuordnung der untersuchten Veränderungen zweitrangig. Ähnlich wie das Controlling bezweckt das Monitoring eine permanente, auf wenige Schlüsselindikatoren beschränkte Überprüfung der Verwaltungsaktivitäten, die eine Feinsteuerung und Stabilisierung der internen Abläufe ermöglicht.

Monitoring-Daten sind daher in jedem Falle für Politikevaluationen wertvoll. Ihr besonderer Wert liegt darin, dass sie in aller Regel periodisch erhoben werden und damit Zeitreihen enthalten, die im Rahmen einer gewöhnlichen Evaluation oft nur mühsam bereitgestellt werden könnten. Ein weiterer Vorteil kann darin liegen, dass sie flächendeckend angelegt sind und damit Vergleiche zwischen verschiedenen Verwaltungseinheiten ermöglichen. Umgekehrt muss aber auch darauf hingewiesen werden, dass solche Daten selten genügend aussagekräftig sind, um als alleinige Basis für Politikevaluationen ausreichen zu können.

8.2. INTEGRIERTE POLITIKBEOBACHTUNG

Am Beispiel der Umweltbeobachtung oder der Entwicklungszusammenarbeit lässt sich zeigen, dass die Kontrollinstrumente kombiniert und weiterentwickelt werden. Waterman und Wood (1993) schlagen vor, das Policy Monitoring als neue analytische Technik im Sinne der Principal-Agent-

⁴ in der Referenzwoche nicht einer regelmässigen Arbeit oder einem Gelegenheitsjob nachgehen oder in ein Beschäftigungsprogramm integriert sind und in den nachstehenden vier Wochen für eine Beschäftigung zur Verfügung stehen und auf Arbeitssuche sind. Für die SAKE wurde die empfohlene Definition der Erwerbstätigkeit vom Internationalen Arbeitsamt übernommen. Erhebungstechnisch handelt es sich bei den Daten des BIGA um eine Vollerhebung, die SAKE-Daten hingegen beruhen auf einer Stichprobenumfrage. Die Interpretation der statistischen Daten stellt somit hohe Anforderungen an den Beobachter. Vgl. Statistisches Jahrbuch der Schweiz 1993: 106 ff.

Gemäss Bussmann (1995a: 56 ff.) und Widmer, Rothmayer und Serdült (1996: 204 ff.) liegen die Kosten für eine Evaluationsstudie in der Schweiz zwischen SO'000 und SOO'000 Franken.

Theorie zu nutzen. In diesem Denkansatz ist der Auftraggeber (Principal) als Folge der Informationsasymmetrie nicht in der Lage, die Aktivitäten seiner Untergebenen zu überwachen. Da viele Aktivitäten ausschliesslich der Verwaltung bekannt sind, kann sie Regierung und Politiker über die tatsächlich relevanten Outputs im Dunkeln tappen lassen. Policy Monitoring stellt darum eine wirksame Möglichkeit zur Überwachung dar. Die periodische Erhebung von Policy-Daten gibt Auskunft darüber, ob die Bürokratie auf politische Impulse reagiert. Zugleich impliziert sie, dass auch die Auftraggeber vermehrt zur Rechenschaft gezogen werden können (vgl. Waterman und Wood 1993: 686 ff.).

Policy Monitoring wird als iterative Technik beschrieben, bei der nach einem ersten Selektionsprozess die Programmziele und Leistungsindikatoren festgelegt werden, dann eine Datei angelegt und die entsprechenden Informationen gesammelt werden. In einem nächsten Schritt wird das Material mit entsprechenden Hilfsmitteln analysiert, so dass zum Schluss die Resultate anhand von zusätzlichen Quellen überprüft und für die zukünftige Gestaltung der Programme weiterverwendet werden können (vgl. Waterman und Wood 1993 und Henry und Dickey 1993). Diese Beschreibung zeigt, dass das Monitoring nicht mehr nur aus einfachen periodischen Erhebungen bestehen muss, sondern dass es sich - zumindest der Intention nach - der Evaluation annähert. Denn dadurch, dass die Monitoring-Daten über Management-Informationssysteme verarbeitet und ausgewertet werden, erlauben sie umfangreichere und komplexere Schlüsse. Dergestalt liefern sie einen Grossteil der Informationen für die Wirtschaftlichkeitsanalysen (vgl. Rossi und Freeman 1993: 362). Gefördert wurde diese Entwicklung zusätzlich durch das New Public Management, das für den öffentlichen Bereich Zielvorgaben und Qualitätsstandards verlangt und Leistungsindikatoren einführen will, um so die Verwaltungseinheiten stärker zur Verantwortung zu ziehen (Ergebnisverantwortung) (zum New Public Management vgl. Osborne und Gabler 1993; Naschold 1993/1995; Schedler 1995; Hablützel et al. 1995).

Besonders ausgereift sind die Monitoringsysteme im Bereich der Umweltbeobachtung (Grolimund und Peter 1994; SKUB 1993). Das *Umweltmonitoring* erfüllt in der Regel drei elementare Funktionen: Es versorgt die Umweltforschung mit langfristig und kontinuierlich erhobenen Messdaten und Beobachtungen; es beliefert die Umweltpolitik mit Informationen über den Zustand der Umwelt, die wiederum als wichtige Grundlagen in die Planungs- und Gesetzgebungsprozesse einfließen; und drittens informiert es den Bürger und die Bürgerin über den Zustand der

Umwelt, was einem Grundprinzip der demokratischen Gesellschaft entspricht. Das Umweltinformationssystem soll zur Gestaltung der Rahmenbedingungen für eine wirksamere Umweltpolitik beitragen (vgl. Zieschank 1992: 8 ff.; Zu den verschiedenen Erfassungssystemen für die Umweltqualität auf unterschiedlichen Ebenen vgl. Lichtenecker 1994).

Wie in anderen Ländern sind auch in der Schweiz drei verschiedene Ebenen der Umweltberichterstattung zu erkennen (Knoepfel 1992). Die *mediale* oder *sektorielle Umweltbeobachtung* erhebt Daten in einem umweltrelevanten Sektor oder Medium. Als Beispiele dafür können die verschiedenen Messnetze wie NABO (Boden), NABEL (Luft) oder Sanasilva (Wald) angeführt werden, die Indikatoren oder Messdaten über Leitschadstoffe erheben. Als Beispiel für den kombinierten oder additiven Ansatz in der Umweltbeobachtung können nationale (vgl. BUWAL 1991 und die neue Version des nationalen Umweltberichts BUWAL 1992), kantonale oder kommunale (vgl. Stadt St. Gallen 1992; Stadt Zürich 1990; Waadt 1995 etc.). Berichte genannt werden, die Ergebnisse aus verschiedenen Umweltbereichen nebeneinanderstellen und es dem Leser überlassen, intersektorale Bezüge herzustellen. Auf der Ebene der Datengewinnung bestehen zwischen den Messnetzdaten keine Bezüge.

Den *integrierten Ansätzen*, die zur Zeit in Entwicklung sind und an die Managementsysteme der siebziger Jahre erinnern, liegt demgegenüber ein ökosystembezogenes Modell⁵ der Umwelt zugrunde. Die neueren Ansätze in diesem Bereich nehmen als weitere Elemente die Verursacher und die subjektive Bewertung der Umweltbelastung durch die Betroffenen auf. Zudem sollen diese Informationssysteme prognostische und bewertende Aussagen erlauben (Zieschank 1992: 13 ff.). Das Neue an diesen Modellen besteht darin, dass sie Umweltinformationen auf der Basis eines ökosystembezogenen Modells erfassen und weiterverarbeiten. Integriert ist dieses Vorgehen deshalb, weil es sowohl eine möglichst vollständige Erhebung der Daten wie auch ihr Inbezugsetzen in Wirkungs- und Systemzusammenhänge beinhaltet. Die technologischen Errungenschaften der Informatik ermöglichen eine Verbindung der Umweltdatenbasis mit einem Geographischen Informationssystem (GIS) und Programmen, mit denen die Umweltinformationen bearbeitet werden können (vgl. nachfolgend Kapitel 8.3.). Der permanenten und langfristigen Umweltbeobachtung soll ein komplexes systembezogenes Modell zur Verfügung stehen. Bezüglich der Komplexität des Modells und des Erhebungsrhythmus wird das Umweltmonitoring auf integrierter Stufe die klassischen Instrumente übertreffen. Zugleich soll es entscheidungsrelevante Informationen für und über die Verwaltungstätigkeit (Outputs; Impacts; Outcomes) liefern, und zwar wiederum auf verschiedenen Ebenen. Integriertes Monitoring stellt darum den Versuch dar, Datenbedürfnisse für unterschiedliche Auswertungsansätze (Controlling, Monitoring, Evaluation) auf verschiedenen Ebenen (lokal, regional, global) und für variierende Zeithorizonte (kurz-, mittel- und

⁵ Eine ökosystembezogene Betrachtung meint die "Betrachtung eines Systems aus der Sicht des Zusammenwirkens der Lebensgemeinschaften mit der abiotischen Umwelt". (SKUB 1993)

langfristig) zu befriedigen (Grolimund und Peter 1994). Ob es in dieser Komplexität allerdings auch taugliche Entscheidungsunterlagen bereitstellen kann, wird sich in Zukunft noch weisen müssen.

Die Arbeiten der Schweizerischen Kommission für Umweltbeobachtung (SKUB) für die langfristige, integrierte und ökosystembezogenen Umweltbeobachtung haben gezeigt, dass kurz- und mittelfristig in den einzelnen Bereichen integrationsfähige Datensätze ausgearbeitet und Basisprogramme erstellt werden müssen, die als Ergebnisse für die eigentlichen Monitoringtätigkeiten (noch) nicht verwendbar sind. Das fachübergreifend integrierte Langzeitbeobachtungsprogramm hätte nach dem Willen der Schweizerischen Akademie der Naturwissenschaften (SANW) ab 1996 in einem Beobachtungsgebiet angewendet und getestet werden sollen. Dabei hätte es sich noch nicht um ein kausales Monitoring gehandelt, denn die in den einzelnen Bereichen zu erhebenden Wirkungsdaten sollten die Formulierung von Wechselwirkungen und Interaktionshypothesen überhaupt erst erlauben und so die Entwicklung eines integrierten Beobachtungsmodells ermöglichen. Die Verallgemeinerung und Operationalisierung wäre auch damit ein langfristiges Ziel. Der Vorschlag wurde 1995 vom Schweizerischen Nationalfonds abgelehnt⁶.

8.3. GEOGRAPHISCHES POLICY-MONITORING

Wie in der Umweltbeobachtung, werden auch in der Politikevaluation sehr viele Daten erhoben, die wie die, allzu sektoriellen und medienspezifischen Umweltdaten, nicht vernetzt, in Wirkungszusammenhänge gesetzt und integriert werden. Es lässt sich darum fragen, ob ein geographisches Informationssystem (GIS) oder zumindest ein GIS-gestütztes-Politiken-Monitoring wichtige Dienste leisten könnte. Die Vervielfachung der staatlichen Zuständigkeitsbereiche, die Vermehrung der Instanzen und der Eingriffe haben zu räumlich und zeitlich unübersichtlichen "Interventionsniederschlägen" im Raum geführt (Knoepfel und Kissling-Näf 1993). Die Koordination und Kooperation von Politikbereichen könnte durch ein geographisches Policy-Monitoring erleichtert werden. Die Idee der SKUB, die Outputs und Adressaten jener öffentlichen Politiken zu erfassen und auf Output- oder "Politiklandkarten" zu übertragen, die in irgendeiner Weise für Veränderungen im Beobachtungsgebiet verantwortlich sind, könnte auch für ein permanentes Policy-Monitoring verwendet werden⁷.

⁶ Antrag SKUB 1995 (nicht publiziert).

⁷ Antrag SKUB 1995, Modul 3.

Bei den verwaltungswissenschaftlichen Beobachtungsparametern handelt es sich um Output-Informationen, die periodisch erhoben und deren Aussagekraft nach Massgabe eines langfristigen Zielfragenkatalog gesteuert/gesichert würden.

Der Wert dieser Informationen für die Politikevaluation wäre folgender:

- Diese Art des Monitoring soll eine *Kontinuität der Beobachtung* garantieren. Die empirischen Studien und Evaluationen sind bis heute dadurch gekennzeichnet, dass nur zum Zeitpunkt der Evaluation genauere und tiefere Informationen zu Schlüsselvariablen der Untersuchung vorliegen oder dafür speziell erhoben werden. Da Evaluationen meist einmalige Projekte sind, verlassen die Forscherinnen oder Forscher nach getaner Arbeit das empirische Feld wieder. Eine Kontinuität bezüglich der Erfassung von Daten fehlt, weshalb spätere Vergleiche meist nicht mehr möglich sind.
- Dokumentation und Datengrundlage in verschiedenen Politikbereichen sind sehr selektiv und partiell. Meistens fehlen namentlich systematische Informationen zu Instrumenten, Outputs, Impacts und Outcomes. Diachrone sowie mittel- bis langfristige Dokumentationen sind inexistent. Politiken haben oft kein "Gedächtnis". Die *Langzeitbeobachtung* kann die Anlage eines Datensatzes zur Verwaltungsaktivität und somit eine systematischere Dokumentation der staatlichen Aktivitäten ermöglichen. Damit würde es beispielsweise möglich, (in Vergessenheit geratene) Subventionsauflagen auch längerfristig zu kontrollieren.
- Ebenso fehlen Informationen zu erbrachten Politikleistungen im *Raum*. Die Erfassung des räumlichen Bezugs der Outputs ist nicht für jede staatliche Aktivität von Bedeutung. Für jene Massnahmen, die die physische, aber auch kulturelle räumliche Umwelt in irgendeiner Art und Weise mitgestalten, wäre eine Speicherung und Erfassung des Ausgangssachverhaltes entscheidend⁸. Diese Art der Information gäbe die Möglichkeit, die Dichte und Häufigkeit staatlicher Interventionen im Raum zu erfassen ("Politikteppiche", Knoepfel und Kissling-Näf

⁸ Ansätze dazu finden sich im Raumplanungsbericht des Bundesamtes für Raumplanung (BRP) von 1994.

1993: 276). Zudem würde ersichtlich, welche Akteure in einem bestimmten Territorium direkt auf den Raum einwirken.

- Die Erfassung des *räumlichen* Bezugs von *Politikoutputs* könnte zu einem wichtigen Koordinationsinstrument für Verwaltungen überhaupt werden. Aufgrund dieser Informationen wüssten die Behörden, wer wo und wann auf den Raum einwirkt. Die Kartographierung (Atlas) der Outputs wäre eine Voraussetzung für die Koordination und Abstimmung zwischen einzelnen Verwaltungseinheiten. Damit würden auf jeden Fall die bislang fehlende Gesamtschau und eine horizontale und vertikale Integration von Eingriffen ermöglicht.
- Da die Datensätze kontinuierlich erhoben würden, wären die Beobachtungsgebiete ideale Untersuchungseinheiten für die Forschung. Die verwaltungswissenschaftlichen Datensätze könnten zu Daten zur Entwicklung der realen gesellschaftlichen Probleme in Beziehung gesetzt werden. Auf der Grundlage dieses Datenmaterials könnten Wirksamkeits- und Effizienzanalysen vorgenommen werden.

Bei dieser Art von Monitoring erfolgt die Erhebung und Gruppierung der Daten auf der Basis vermuteter Wirkungszusammenhänge und einer spezifischen Fragestellung (Veränderung des Problemzustandes). Die aufgeführten Punkte zeigen, dass die Entwicklung der integrierten, langfristigen Beobachtung für die Verbesserung der Wirksamkeit von staatlichen Massnahmen auch politikwissenschaftlich von Bedeutung sein könnte. Der Karte der Outputs im Raum käme für die Verwaltungspraxis und damit für die Koordination und Integration von Politiksverhalten eine entscheidende Bedeutung zu. Zugleich wird deutlich, dass sich das integrierte Monitoring zunehmend der Evaluation annähert und darüber hinaus dank seiner Langfristigkeit eine unabdingbare Grundlage für vertiefte Politikevaluationen liefert.

Erste Erfahrungen in Zusammenhang mit Umweltinformationssystemen haben im übrigen gezeigt, dass letztere quer zu den medialen und sektoralisierten Verwaltungsstrukturen stehen und damit zu einer medienübergreifenden Vernetzung beitragen können, was wiederum gleichbedeutend ist mit einer Kollektivierung der Umweltinformationen. Zunehmend werden auch GIS in die Umweltinformationssysteme eingebaut, um die Umweltbelastungen und Verursacher räumlich differenziert zu erheben (vgl. Fürst u.a. 1994: 33).

KAPITEL 9: EVALUATION UND MEDIATION

Helmut Weidner, Peter Knoepfel

9.1. UMFELD

Policy-Beratung kann auf sehr unterschiedliche Art und Weise betrieben werden. In einem dezisionistischen Verständnis glaubt der Berater eine objektive Entscheidungsgrundlage für die Politik zu erarbeiten. Diese instrumentelle Sicht sowie das zugrundeliegende Wissenschaftsverständnis werden aus demokratietheoretischer und postpositivistischer Perspektive zunehmend in Frage gestellt. Als Reaktion darauf sprechen etwa Fischer und Forester (1993) von einer argumentativen Wende in der Politikanalyse und Planungsprozessen. Guba und Lincoln (1987: 203-234) erkennen eine vierte Generation von Evaluationen. Konstitutives Element der ersten Generation war der Versuch, Resultate von Massnahmen zu messen und zu beziffern ("measurement"). Die Evaluationen der zweiten Phase waren vor allem deskriptiver Natur ("description"). Die dritte Generation wiederum versuchte das Element der Beurteilung ("judgement") zu integrieren. Die vierte Generation ist geprägt vom interaktiven Vorgehen und von Aushandlungsprozessen ("negociation")¹. Dabei wird Prinzipien wie dem Wertpluralismus² oder dem Stakeholder-Konzept³ Rechnung getragen, und der Evaluationsprozess wird um Elemente der Mediation erweitert. Damit haben sich die Position des Evaluators (Honeyman 1993) und der Evaluationsbegriff (Fischer 1994) in der folgenden Art und Weise geändert:

- Evaluation wird verstanden als ein gesellschaftlicher und politischer Prozess.

¹ Man spricht in diesem Zusammenhang auch von "responsive evaluation", "utilization-focused evaluation" oder "illuminative evaluation", vgl. Kapitel 13.3.

Bis anhin ging man in den meisten Evaluationen davon aus, dass ein Konsens über Programmziele besteht. Die neueren Ansätze (4. Generation) haben mit diesem Paradigma gebrochen und beziehen verschiedene Standpunkte in den Evaluationsprozess ein. Evaluation wird zu einem Konsensbildungsprozess. Vgl. Guba und Lincoln 1987: 209 f. Dies wiederum bedeutet den Einbezug von Akteuren wie den Auftraggebern, Umsetzenden, Adressaten etc. in den Evaluationsprozess.

- Evaluation wird zum Lernprozess, bei dem es darum geht, die Wirklichkeits-Konstruktionen der einzelnen Gruppen zu erkennen, zu thematisieren und gegebenenfalls interaktiv weiterzuentwickeln.
- Der Prozess der Evaluation schafft seine eigene Wirklichkeit.
- In diesem Sinne wird Evaluation zu einem emergenten, rekursiven und divergenten Prozess.
- Der Evaluator wird zum Mediator, der die Aushandlungsprozesse/Lernprozesse unterstützt und animiert.

Die partizipatorische und diskursive Policy-Analyse orientiert sich an den Vorstellungen der relationalen *Konstruktivismus*⁴. Die Individuen greifen für die Bewältigung des Alltags auf Interpretationsmuster und Typisierungen zurück, die sie in früheren Erfahrungen gewonnen haben. "[D]as gesamte Wissen, sowohl wissenschaftliches wie alltägliches, [besteht] aus "Konstruktionen, das heisst einem Verband von Abstraktionen, Generalisierungen, Formalisierungen und Idealisierungen" [...], die die gesellschaftliche Wahrnehmung strukturieren." (Nullmeier 1994: 6) Sowohl die Beteiligten einer gesellschaftlichen Interaktion wie auch der/die aussenstehende wissenschaftliche Betrachter oder Betrachterin werden zu Interpretierenden. Daraus ergibt sich, dass für Laien und Experten der Zugang zur Wirklichkeit derselbe ist. Interpretative Ansätze betonen die sich mit den Deutungen eröffnenden Spielräume und den prozessualen Charakter von gesellschaftlicher Interpretation. Gesellschaftliche Wirklichkeit wird dabei verstanden als Resultat von Verhandlungsprozessen. Als weiteres wichtiges Kennzeichen nennt Nullmeier (1994: 6 ff.) den intersubjektiven Charakter des Alltagswissens. Bezogen auf den politischen Prozess (politics) stellt Politik eine Konfliktarena dar, in der über Konzepte zur Beurteilung von gesellschaftlichen Problemen gestritten wird. "In the case of democracies, this conflict is managed by a public debate on and a negotiated definition of shared meanings." (Hoppe 1994: 55) Wie können nun die in der Policy vernachlässigten Werte in den Ent-

⁴ Beim Konstruktivismus geht man davon aus, dass es objektive Wirklichkeit nicht gibt, sondern dass jedes wissende Subjekt sich seine eigene Wirklichkeit konstruiert (subjektive Realitätskonstrukte). Vgl. dazu z.B. Berger und Luckmann 1996; Dachler 1993; Hosking und Morley 1991. Eine entgegengesetzte Wirklichkeitskonzeption finden wir beim methodologischen Individualismus.

scheidungsprozess einbezogen werden? Eine Möglichkeit stellt die partizipatorische und dem demokratischen Prinzip verpflichtete Politikanalyse (Fischer 1992) dar, die zu einer Verbreitung des Beteiligtenkreises führt. Daher ist es angezeigt, im Zusammenhang mit der Politikevaluation auf Mediationsverfahren einzugehen.

9.2. DEFINITION VON MEDIATION

Unter Mediationsverfahren wird üblicherweise ein Verfahren zur Regelung von verhandlungsfähigen Konflikten verstanden, an dem zwei oder mehrere Streitparteien freiwillig teilnehmen mit dem Ziel, in einem gesiteten und direkten ("face-to-face")-Kommunikationsprozess ihre Differenzen gemeinsam zu erkunden, Handlungsspielräume auszuloten und zu einer von allen Teilnehmern entwickelten und getragenen Lösung in Form einer Vereinbarung zu kommen. Hierbei werden sie von einer neutralen Person (Mediator) unterstützt, deren Hauptaufgabe in der Gestaltung und Durchführung eines fairen Verfahrens liegt. Alle für den Konfliktfall relevanten Gruppen sollen dabei durch Personen mit Verhandlungskompetenz repräsentiert sein (Susskind und Cruikshank 1987; Goldberg, Sander und Rogers 1992; Fietkau und Weidner 1996).

Mediationsverfahren gehören zur Klasse der "alternativen Streitbeilegungsverfahren". Sie werden so genannt, weil sie im Unterschied zu konventionellen Verfahren auf freiwilliger Teilnahme beruhen, weil die Teilnehmer sich ihre Verfahrensregeln und -ziele selbst geben und weil die Ergebnisse nicht von vornherein förmlich bindend sind. Allerdings sind diese Prozeduren kein Ersatz für förmliche Verfahren, sondern ergänzen diese, indem sie vor, nach oder parallel zu ihnen stattfinden (Fietkau und Weidner 1992). Es gibt unterschiedliche Formen alternativer Verfahren, denen allen eines gemeinsam ist: Durch sie sollen Streitigkeiten auf dem *Verhandlungswege* beigelegt werden. Verhandeln (negotiating) und Aushandeln (bargaining) gelten als so zentrale Elemente, dass Mediationsverfahren auch kurzweg als durch einen "unabhängigen Vermittler unterstützte Verhandlungsprozesse" (Cormick und Knaster 1986: 8) definiert werden. Damit können nur verhandlungsfähige *Interessenkonflikte*, nicht aber fundamentale *Wierikonflikte* Gegenstand solcher Verfahren sein. Die Grenze ist allerdings nicht immer eindeutig zu ziehen, und manch ein Wertkonflikt wandelt sich im Zeitverlauf zu einem verhandelbaren Interessenkonflikt (Goldberg, Sander und Rogers 1992). Wesentlich ist überdies zumindest ein stummes Einverständnis, dass es eine "pluralistische

Sicht der Dinge" geben kann. Darin besteht ein wesentlicher Unterschied zu einem positivistischen Wissenschafts- und Evaluationsverständnis. Mediation anerkennt unterschiedliche Rationalitäten, Wertsysteme etc. prinzipiell als legitim; ihre Überzeugungskraft muss sich in einem geregelten Wettstreit der Argumente erweisen (Fisher, Ury und Patton 1991). Besonders aus diesem Grunde ist es schwierig, dem Diskurs selbst und den Ergebnissen solcher Verfahren allein die üblichen (zumeist für konventionelle Verfahren entwickelten) Evaluationskriterien zugrunde zu legen. Das hat zweierlei Konsequenzen: Zum einen werden durch die Teilnehmer im Verfahren Evaluationskriterien *für* das Verfahren **ent**wickelt, zum anderen können externe Evaluatoren den Verfahrensergebnissen nur dann vollständig gerecht werden, wenn ihre Evaluationskriterien die Spezifika auch von Mediationsverfahren reflektieren. Es handelt sich bei diesen Verfahren nicht bloss um sozial erweiterte "Sachverstandsdebatten"; vielmehr ergeben technisch-naturwissenschaftliche, biologische, ökonomische etc. Rationalitäten hier im Wechselspiel mit den Gruppenrationalitäten eine eigene Verfahrensrationalität, die generell so gewollt, wenn auch im Einzelfall nicht immer so gewünscht ist.

Gegenüber konventionellen Verfahren werden Mediationsverfahren auch mit Bezug auf traditionelle Evaluationen erhebliche Vorteile zugeschrieben. So werden sie oft als effektiver, flexibler, fairer, schneller und "preiswerter" bezeichnet. Zumindest für die beiden letztgenannten Kriterien gibt es kaum überzeugende Belege (Amy 1987, Bingham 1986). Vielmehr ist die Mediation häufig mit einem beträchtlichen Kosten- und Zeitaufwand verbunden, den man jedoch gegen eine Entwicklung ohne Mediation abwägen sollte. Das haben die beteiligten Gruppen in aller Regel schon vorab gemacht, indem sie sich bei ihrer Entscheidung für oder gegen ein Mediationsverfahren fragen, was ihre "BATNA" ist, nämlich ihre Best Alternative To a Negotiated Agreement (Carnevale 1986). Für ein Unternehmen oder eine Verwaltung kann es beispielsweise durchaus attraktiv sein, höhere Kosten und Zeitverzögerungen in Kauf zu nehmen, wenn dies zu einer grösseren gesellschaftlichen Akzeptanz ihres Vorhabens oder langfristig zur Akkumulation von "Vertrauenskapital" für zukünftige Entscheidungen beitragen könnte. Schliesslich gehört zu den Standardargumenten für die Mediation, hierdurch würden die Normalverfahren kennzeichnenden Gewinner-Verlierer-Situationen vermieden und statt dessen Gewinne für alle Beteiligten ermöglicht ("win-win-Situationen") (Fisher, Ury und Patton 1991; Susskind und Cruikshank 1987 und Goldberg, Sander und Rogers 1992).

Insgesamt kann man aufgrund bisheriger Praxiserfahrungen insbesondere aus dem Umweltbereich folgende *Vorteile* der Mediation hervorheben (Hoffmann-Riem und Schmidt-Assmann 1990; Holznagel 1990; Claus und Wiedemann 1994; Zillesen und Barbian 1992; Zillesen, Dienel und Strubelt 1993; Daily, Weidner und Fietkau 1993; Knoepfel 1995b):

- Einbezug von ansonsten verdrängten Themen und Aspekten in die Verhandlungen, also eine Ausweitung des politischen Themen- und Optionsrahmens;
- die Schaffung gesellschaftsadäquaterer Entscheidungsprozesse, die besonders den in Industrieländern stattgefundenen Wertewandel besser berücksichtigen. In diese Prozesse können auch rechtlich noch nicht normierte, aber sozial- und konfliktrelevante Aspekte bis hin zu diffusen Ängsten und Unsicherheiten einfließen;
- die Schaffung eines problemangemesseneren Kommunikationsprozesses: Die Gespräche finden nicht direkt in der politischen Arena statt, die üblicherweise zu Dramatisierungen einlädt, und sind auch nicht in bewusst-strategischer Weise auf eine Letztentscheidung durch Gerichte ausgerichtet;
- die gegenseitige Stimulierung von Kreativität zum Überspringen oder Nivellieren von gegensätzlichen Positionen und Interessen hin zu einer für alle Beteiligten relativ vorteilhaften Lösung;
- Gewährleistung grösserer Stabilität der im Verfahren erzielten Entscheidungen im nachfolgenden Implementationsprozess, da sie einvernehmlich und auf der Basis eines breit angelegten, interessenausgleichenden und transparenten Verfahrens entwickelt wurden sowie
- Verbreiterung und Vertiefung gesellschaftlicher Partizipation durch aktive Mobilisierung von üblicherweise oftmals ausgeklammerten Interessen- und Betroffenengruppen.

9.3. ANWENDUNGSFELDER

Mediation ist im allgemeinen Sinne eine gesellschaftliche Konfliktregelungstechnik. Im speziellen Fall von (öffentlich-rechtlichen) Politikentscheidungen ist sie ein "Policy-tool" mit nahezu unbegrenztem Einsatzbereich: Die Konfliktbeteiligten müssen es "nur" wollen. Darüber hinaus gibt es in einigen Bereichen rechtliche Grenzen. Aber auch hier zeigt die internationale Praxis, dass Einsatzspielräume geschaffen werden können, wenn Mediationsverfahren beispielsweise als Ergänzungen zu rechtsförmig geregelten Verfahren konstruiert werden und die Entscheidungsprärogative demokratisch-rechtlich legitimierter Institutionen zumindest symbolisch erhalten bleibt (Susskind und Ozawa 1985). Hierbei ergeben sich nun wieder - besonders in "hoch verrechtlichten" Ländern - Abgrenzungsprobleme, die zu einer lebhaften und kontroversen juristischen und demokratietheoretischen Diskussion geführt haben (Hoffmann-Riem und Schmidt-Assmann 1990). Diese kreist vorzugsweise um die Frage, ob mit solchen Verfahren eine Mitwirkung an Entscheidungsvorbereitungen durch eine definitive Entscheidungsprogrammierung ersetzt wurde, womit rechtsstaatlich gesetzte Grenzen überschritten würden. Es ist hier nicht der Ort, um auf die damit verbundenen demokratietheoretischen Bedenken (begrenzte Öffentlichkeit der Verfahren etc.) und auf politikstrategische Kritiken (Mediationsverfahren als Manipulations- und Konfliktunterdrückungsinstrumente) einzugehen (vgl. Amy 1983: 1987).

Die internationale Praxis zeigt jedenfalls einen weiten und sich in letzten Jahren ausdehnenden Anwendungsbereich von Mediationsverfahren. So gibt es v. a. Scheidungsmediation, Mediation im Täter-Opfer-Ausgleich, bei Tarifkonflikten, in organisationsinternen Streitfällen, bei Rasen- und Nachbarschaftskonflikten sowie bei sogenannten Campus-Streitereien oder bei Umweltkonflikten. Mediation wird ausserdem auf allen Ebenen und Sachgebieten des öffentlich-politischen Bereichs eingesetzt: auf lokaler, regionaler und staatlicher Ebene, in der Landwirtschaftspolitik ebenso wie etwa in der Energie-, Technologie-, Verkehrs- und Umweltpolitik oder gar in der "grossen" internationalen Politik (Knoepfel 1995b; Daily, Weidner und Fietkau 1994; Zillessen, Dienel und Strubelt 1993; Goldberg, Sander und Rogers 1992). Sie finden statt als Leitlinien definierende "Politikdialoge", normprogrammierende Policy-Verhandlungen (regulatory negotiation), intra- und interadministrative Abstimmungs- und Ausgleichsverfahren oder als allgemeine Konventionen anstrebende "Branchendiskurse". Am häufigsten kommen sie im öffentlich-

politischen Bereich wohl im Vor- und Umfeld von umweltrelevanten Planungs- und Genehmigungsverfahren zum Einsatz. Ihre Anwendung in der Umweltpolitik wird im folgenden Kasten dargestellt:

Beispiel Umweltmediation: Ein Überblick

Kein anderer Politikbereich hat sich in den letzten zwei Jahrzehnten in fast allen OECD-Ländern als so konfliktträchtig erwiesen wie die Umweltpolitik, vor allem wenn auch die damit sachlich eng verknüpften (und ebenfalls "umweltpolitisch debattierten") anderen Politikbereiche wie die Verkehrs-, Energie-, Technologie- und Landwirtschaftspolitik mitberücksichtigt werden. Aufgrund zahlreicher, über die Jahre zunehmender Politik- und Investitionsblockaden, die mit dem traditionellen Regelungsinstrumentarium nicht oder nur für eine Konfliktpartei zufriedenstellend aufgelöst werden konnten, ist in diesen Bereichen die Bereitschaft gewachsen, mit neuen Politikinstrumenten zu experimentieren, in denen ein kooperativer Interessenausgleich angestrebt wird (Weidner 1993).

Mediationsverfahren zur Regelung von Umweltkonflikten sind vermutlich zuerst in *Japan* in systematischer Weise eingesetzt worden, wo die konsensorientierte politische Kultur generell "versöhnliche Streitbeilegungsverfahren" unterstützt (Tsuru und Weidner 1989). Es gibt in diesem Land kaum ein grösseres umweltrelevantes Vorhaben, in dem nicht Verhandlungslösungen angestrebt werden. Ihre Bedeutung für die relativ erfolgreiche japanische Umweltpolitik zeigt sich auch darin, dass es gegenwärtig rund 45'000 Umweltschutzvereinbarungen zwischen Firmen und Kommunen sowie Bürgergruppen gibt, in denen für den Einzelfall "massgeschneiderte" Umweltschutzanforderungen festgelegt werden. Sie beruhen auf einem wesentlich breiteren und differenzierteren Set ökologischer, ökonomischer und gesellschaftlicher Bewertungskriterien, als dies die konventionellen Verfahren anwenden.

Ausgangspunkt der Karriere des Mediationsverfahrens als Politikinstrument sind die *USA*⁵. Hier wurde es erstmals 1973 eingesetzt - mit so durchschlagendem Erfolg, dass relativ rasch zahlreiche weitere folgten. Inzwischen wurde in den USA gar von einem "Mediationsboom" gesprochen. Damit verbunden ging eine Professionalisierung der Mediator Tätigkeit einher: in nahezu allen Bundesstaaten wurden private oder halbstaatliche Mediationsinstitute gegründet, die Mediatoren ausbilden und ihre Dienste bei Umweltkonflikten anbieten. Trotz einiger kritischer Stimmen überwiegen die Befürworter von Mediationsverfahren heute ganz eindeutig.

Auch in *Kanada* sind alternative Konfliktregelungsverfahren relativ weit verbreitet (mit steigender Tendenz). Hier ermöglicht das Gesetz zur Durchführung von Umweltverträglichkeitsprüfungen systematisch Mediationsverfahren. In europäischen Ländern wächst seit einigen Jahren das Interesse an Mediation, einige solcher Verfahren sind bereits erfolgreich abgeschlossen worden. In *Österreich*, der *Schweiz*, also Ländern mit einer partizipativen bzw. korporatistischen politischen Kultur, wer-

⁵ Vgl. für einen Überblick zur Mediationssituation in den hier angeführten Ländern die Beiträge von Elliott, Sigurdson, Ocenasek, Knoepfel und Claus, in: Daily, Weidner und Fietkau 1994 sowie Fietkau und Weidner 1996; für die Schweiz: Knoepfel 1995b.

den besonders bei Konflikten über Standorte und Techniken zur Abfallbeseitigung konsensorientierte Konfliktregelungsverfahren eingesetzt (Rey 1995; Renn und Webler 1995). In Ländern mit zentralistischem Staatsaufbau sind die Verfahren eher selten. Immerhin wurde in *Grossbritannien* 1990 mit Unterstützung des britischen Unternehmerverbandes CBI das Centre for Dispute Resolution gegründet. Es bietet Trainingsprogramme und Mediatordienste an. In den *Niederlanden* finden sich ebenfalls vereinzelt Mediationsverfahren, so beispielsweise im Zusammenhang mit dem Planungsverfahren für eine grosse Hafenanlage. Die geringe Zahl solcher Verfahren resultiert vermutlich daraus, dass in der ohnehin hochpartizipativen politischen Kultur der Niederlande Kooperations- und Verhandlungslösungen mit weitreichenden Mitwirkungsmöglichkeiten von Umweltorganisationen (sowohl im Planungs- als auch im Entscheidungsprozess) gang und gäbe sind.

Selbst in der vergleichsweise staatsfixierten *Bundesrepublik Deutschland* mit starker formalrechtlicher Tradition sind Mediationsverfahren im Kommen. In zwei grossen Pionierverfahren, dem "Münchehagener Verfahren" zur Sanierung einer undichten Sonderabfalldeponie und dem "Neusser Verfahren" zur Modifikation und Umsetzung eines Abfallwirtschaftskonzeptes (Weidner und Fietkau 1995) wurden zahlreiche Mediations- und mediationsähnliche Verfahren initiiert und teilweise bereits erfolgreich abgeschlossen. Um nur einige Beispiele zu nennen: Standortsuchen für eine Müllentsorgungsanlage in Berlin, für Abfalldeponien in den Bundesländern Bremen, Sachsen, Niedersachsen; Einrichtung eines Verkehrsforums zu Erarbeitung eines langfristigen Verkehrsleitbildes für Heidelberg; verschiedene Sanierungs- und Kompensationsfälle im Zusammenhang mit Altlasten; Dialoge zu Fragen der Energie- und Gentechnologiepolitik sowie ein mediatisierter Bürgerdialog zum äusserst komplexen Verfahren der Standortsuche für einen Grossflughafen im Raum Berlin-Brandenburg.

9.4. BEZIEHUNGEN ZWISCHEN EVALUATION UND MEDIATION

Wie eingangs unter dem Hinweis auf die Literatur und die Gesellschaftlichkeit von Wahrheitsproduktion ausgeführt, könnte nahezu alles, was als sozialwissenschaftliche Evaluation betrieben wird, auch in Form eines Mediationsverfahrens durchgeführt werden. In fast allen Mediationsverfahren finden Evaluationen statt. Mediationsverfahren können deshalb generell als mehr oder minder begrenzte Evaluationsverfahren interpretiert werden; man kann sie als eine prozedurale Spezialform der Evaluation auffassen. Das aber hat zugleich erhebliche inhaltliche Auswirkungen auf die (konventionelle) Evaluation.

Evaluation in konventioneller (d.h. nicht partizipativer) Form hat als Auftraggeberin üblicherweise eine einzelne Institution, die den Evaluationsrahmen und die -kriterien festlegt sowie die Evaluatoren und Evalua-

torinnen bestimmt. Im Bereich öffentlicher Politiken ist dies zumeist eine Institution des politisch-administrativen Systems. Die Evaluation selbst wird in aller Regel von unabhängigen Experten nach Massgabe des Auftrags in Form einer überwiegend nicht-partizipativen "Einwegkommunikation" (Datenerhebung und Informationsabfrage) durchgeführt; nur bei der abschliessenden Bewertung sind die Auftraggeber oftmals intensiver involviert. Gegenstand der Evaluation sind häufig, jedenfalls im Bereich öffentlicher Politiken, (Teil-)Ergebnisse laufender oder abgeschlossener Prozesse. Hiervon unterscheiden sich Mediationsverfahren erheblich. Es sind prinzipiell kommunikative, partizipationserweiternde Verfahren (Crowfoot und Wondolleck 1991). Bezüglich aller wichtigen strittigen und "streitfähigen" Punkte sollen gemeinsam und einvernehmlich Vorgehensweise und Bewertungskriterien entwickelt werden; auch die Ergebnisbeurteilung ist Sache aller Beteiligten. Unter der Voraussetzung, dass die Verfahren zu Konsensen führen, wird ebenfalls deren Umsetzung von vornherein vereinbart. Die Art und Weise des Umgangs mit den Ergebnissen ist jedenfalls entschieden weniger beliebig und einseitig bestimmbar wie bei konventionellen Evaluationen.

Mediationsverfahren, so zeigt etwa die Praxis in den Pionierländern, kommen auf allen Ebenen des Politikzyklus zur Anwendung (Bingham 1986). Sie treten indessen viel seltener als bei der konventionellen Evaluation erst am Ende eines Implementationsprozesses auf (im Sinne einer ex post-Evaluation eines realisierten Vorhabens). Im Prinzip besteht jedoch kein unüberwindbares Hindernis, Evaluationen in allen Hauptphasen der Mediation systematisch einzubauen. Indirekt finden sie zumindest partiell ohnehin fast immer statt. Die *typischen Phasen eines Mediationsverfahrens* werden im folgenden Kasten skizziert (Susskind und Cruikshank 1987; Fietkau 1994):

1. Vorbereitungsphase

- Identifikation der konfliktrelevanten Gruppen und Informationsgespräche mit ihnen;
- Entscheidung, ob der Streitfall verhandlungsfähig ist;
- Auswahl der Verfahrensbeteiligten;
- Definition der Verhandlungs- und Streitthemen;
- Festlegung von Verfahrensgrundregeln und eines Zeitplans;
- Klärung von Kosten- und Ressourcenfragen;
- Vereinbarung von Verhandlungszielen;
- Auswahl und Benennung eines Mediators/einer Mediatorin.

2. Verhandlungs- und Konsensbildungsphase

- Informationsaustausch, Aufarbeitung von Informationsdefiziten und Abklärung von Informationsproblemen;
- Erarbeitung einer gemeinsamen Wissensbasis und Abklärung von Verständnis- und Interpretationsproblemen;
- Analyse der gruppenspezifischen Interessen;
- Ausloten von Verhandlungsspielräumen;
- Verhandlungen (im Gesamtplenum und in speziellen Arbeitsgruppen) und Entwicklung von "Verhandlungspaketen";
- Entscheidung über einen Vereinbarungsvorschlag;
- Rückkopplung des Vereinbarungsvorschlags zu den repräsentierten Gruppen;
- Schlussvereinbarung, inklusive Festlegung von Verfahrensregeln zur Implementation der Vereinbarung;
- Implementationsplan.

3. Umsetzungsphase

- Verknüpfung der Vereinbarungsinhalte mit dem förmlichen Verfahren;
- Monitoring der Umsetzung (in einer im Mediationsverfahren festgelegten Art und Weise);
- eventuell Nachverhandlungen und Modifikationen;
- Evaluation der Vereinbarungsumsetzung;
- eventuell Nachverhandlungen und Ergänzungsmassnahmen.

Evaluative Handlungsformen und Evaluationen im engeren Sinne spielen in der zweiten und dritten Phase zeitweilig eine zentrale Rolle; sie können aber auch schon in der Vorbereitungsphase stattfinden (etwa indem systematisch eine Betroffenenanalyse in Verbindung mit einer Auswirkungsanalyse des Vorhabens durchgeführt wird). In der zweiten, der "Verhandlungs- und Konsensbildungsphase", entsteht in aller Regel ein erheblicher Zeitbedarf für die gemeinsame Evaluation des strittigen Vor-

habens⁶. Die Gesichtspunkte und Kriterien hierfür werden gemeinsam erarbeitet. Davon hängt u.a. ab, welchen Stellenwert das Projekt im Politikprozess erhält, der wiederum den Evaluationsrahmen, den -aufwand und potentielle Evaluationsfolgen vorbestimmt. So kann beispielsweise ein strittiges Müllverbrennungsanlagen-Projekt anhand der Kriterien evaluiert werden, ob der vorgesehene Standort adäquat ist, ob eine biologisch-mechanische Abfallentsorgungsanlage eine geeignetere Technik sein könnte, ob eine grundsätzliche Änderung der Abfallpolitik durch Fokussierung auf Müllvermeidung, Recycling, Abfallberatung etc. oder gar eine generelle Wende in der Umwelt- und Industriepolitik hin zur ökologischen Modernisierung der Gesellschaft problematische oder ungeliebte Abfallentsorgungsanlagen generell überflüssig machen könnten.

Häufig besteht bei Mediationsverfahren, in denen "heisse" Themen verhandelt werden, eine starke Tendenz der beteiligten Organisationen und Bürgerinitiativen, das Streitthema und damit den Evaluationsrahmen zu "globalisieren" bzw. zu fundamentalisieren, während wirtschaftliche und politisch-administrative Gruppenvertreter den Beurteilungskontext eher eng und technokratisch-pragmatisch, in etwa den förmlichen Verfahrensanforderungen entsprechend, fassen wollen. Dementsprechend ist die Abgrenzung des Verhandlungsrahmens und des -gegenstandes typischerweise ein zentraler Verhandlungspunkt in der ersten und zweiten Phase eines Mediationsverfahrens. Ähnliches gilt für die konkreten Evaluationskriterien, die im Anschluss an die Kontextvereinbarungen festgelegt werden. In beiden Fällen, so zeigt die Praxis, findet im Vergleich zu üblichen Verfahren immer eine Erweiterung der Evaluation aufgrund des pluralistisch partizipativen Prozesses statt (Knoepfel 1995b; Fietkau und Weidner 1996). Die Kraft der Argumente und die Kompromissfähigkeit der beteiligten Gruppen sind ausschlaggebender als der Verweis auf förmliche Erfordernisse und konventionelle Praxis. Die Evaluation verschiebt sich dadurch von einer wissenschaftlich-technokratischen Technik hin zu einem gesellschaftlichen Analyse- und Beurteilungsprozess, in dem multilaterale Konsense den gleichen Rang wie wissenschaftlich-sachliche Argumente einnehmen können.

Im Unterschied zur konventionellen Evaluation (etwa im Umweltbereich), wo inzwischen ein relativ feststehender *Kriterienkatalog* entwickelt wurde (Pictet 1996), gibt es für die *evaluative Mediation* noch keinen

⁶ Vgl. beispielhaft das Mediationsverfahren zu einem Abfallwirtschaftskonzept im Kreis Neuss. Hierzu die Beiträge von Fülgraff, Fietkau und Weidner, in: Daily, Weidner und Fietkau 1994.

solchen Katalog (van der Daele 1993). Er wird auch schwierig zu entwickeln sein, ist er doch jeweils in starkem Masse auch ein Einzelfallergebnis, das von der Teilnehmerzusammensetzung, ihren spezifischen Interessen und der Interaktionsdynamik des Verfahrens abhängig ist. Es ist aber verfahrensimmanent (und aufgrund bisheriger Praxis evident), dass Kriterien der "Sozialverträglichkeit" bezogen auf das Entscheidungsergebnis *und* das -verfahren immer eine wichtige Rolle spielen⁷. Inwieweit Sozialverträglichkeit tatsächlich gegeben ist, zeigt sich oftmals erst in der Umsetzungsphase anhand der Reaktionen von nicht beteiligten Gruppen oder Betroffenen. In einem "schlecht" organisierten Mediationsverfahren kann sie zusätzlich dadurch eingeschränkt sein, dass die Mitglieder der "Mutterinstitutionen" den Konsens ihrer Verfahrensvertreter später nicht mehr mittragen wollen.

Inwieweit das Mediationsverfahren effektiv oder erfolgreich war, ist ebenfalls eine Evaluationsfrage. Über die Frage, auf Basis welcher Kriterien eine derartige Evaluation erfolgen sollte, besteht in Wissenschaft und Praxis noch kein umfänglicher Konsens, selbst in den USA nicht, mit ihrer nunmehr gut zwanzigjährigen Praxiserfahrung im Bereich Umweltmediation. Es lassen sich in etwas holzschnittartiger Vereinfachung zwei Positionen hierzu ausmachen. Nach der einen sind Mediationsergebnisse prinzipiell nach ihrer Umwelt- *und* Sozialverträglichkeit zu beurteilen; die andere Position hält ausschliesslich die Zufriedenheit der Teilnehmer für relevant.

In Übereinstimmung mit dem Projektteam "Mediationsverfahren im Umweltschutz" des Wissenschaftszentrums Berlin für Sozialforschung (WZB) (Fietkau und Weidner 1992) geben wir der erstgenannten Position den Vorzug. Sie macht die Evaluation von Mediationsverfahren allerdings zu einem schwierigen Unterfangen. Dies liegt nicht nur an der damit verbundenen grossen Bandbreite der Evaluationskriterien, sondern auch daran, dass beide Basiskriterien nicht per se kongruent sind: Eine im gesellschaftlichen Konsens getroffene Entscheidung kann sachlich (z.B. ökologisch) durchaus problematisch sein; umgekehrt kann ein sachlich "umweltfreundliches" Ergebnis in der Folge (insbesondere bei der Umsetzung) gesellschaftlich destruktive Wirkungen zeitigen. Deshalb sollten für die Beurteilung der ökologischen Qualität dieselben Evaluationskriterien gelten, wie sie auch bei allen anderen Evaluationen verwendet wer-

⁷ Sozialverträglichkeit wurde hier in Anführungszeichen gesetzt, da ihr Ausmass nach Art der Gruppenpluralität (und der spezifischen Interessen der einzelnen Gruppen) des Verfahrens variiert.

den (im Umweltbereich: ökologische Effektivität, Nachhaltigkeit, Ausschluss von Problemverschiebungen etc.); zusätzlich einzubeziehen sind noch diejenigen ökologischen Kriterien, die von den Beteiligten festgelegt werden. Die Minimalbedingung für einen "gesellschaftlichen Erfolg" des Verfahrens ist ein unmanipulierter Kompromiss der Verfahrensbeteiligten, wobei die "Kompromissfolgekosten" für Nichtbeteiligte nicht erheblich grösser sein dürfen, als sie für Verfahrensbeteiligte sind. Die Einhaltung dieses Kriteriums wird oftmals schwierig zu klären sein. Umso wichtiger wird deshalb die Qualität des Verfahrens an sich: Es muss transparent, fair und offen sein.

In welcher Weise dann für eine abschliessende, zusammenfassende Erfolgsbeurteilung die Kriterien "Sozial- und Umweltverträglichkeit" zu gewichten sind, ist derzeit wissenschaftlich noch ungeklärt⁸. Sie zu beantworten wird in viel grösserem Masse als etwa bei der Entwicklung bisheriger umweltpolitischer Evaluationskriterien eine interdisziplinäre Forschungsk Kooperation erfordern. Man kann zudem davon ausgehen, dass liebgeordnete Idealvorstellungen und sehr rigide Bewertungsprinzipien dann unter Begründungs- und Veränderungsdruck kommen werden, wenn Kompromisslösungen im Umweltbereich zur Regel werden. Es wird wohl notgedrungen weniger das "Entweder-oder" als vielmehr der dazwischen liegende Trennungsstrich berücksichtigt werden müssen, auf dem die Mehrzahl aller Mediationsergebnisse angesiedelt sein wird. Wenn es die erwünschte Funktion von Mediationsverfahren ist, aus diesem Trennungsstrich einen gesellschaftlichen Verbindungsstrich (Mediation und Evaluation als Integrationsinstrumente) zu machen, dann wird auch die Wissenschaft, sofern sie empirische Sozialwissenschaft sein will, darauf reagieren müssen. Auch in dieser Hinsicht ist das Thema "Evaluation" trotz seiner jahrzehntelangen Konjunktur aus sozialwissenschaftlicher Sicht noch nicht ausgereizt.

⁸ Lawrence Susskind, einer der führenden Experten im Bereich Umweltmediation, nennt vier Charakteristika eines guten Verhandlungsergebnisses: Fairness, Effizienz, Weisheit (im Sinne von faktischer Bestätigung zentraler Annahmen und Prognosen im Rückblick) und Stabilität. Vgl. Susskind und Cruikshank 1987: 21 ff. Vgl. auch die multidimensionale Bewertung des Mediationsverfahrens im Kreis Neuss (in: Pfingsten und Fietkau 1995) sowie Buckle und Thomas-Buckle (1986) zu den positiven Lerneffekten fehlgeschlagener Mediationsverfahren.

9.5. ZUSAMMENFASSUNG UND AUSBLICK

Die traditionellen Politikinstrumente stossen in allen demokratischen Industrieländern zunehmend an Grenzen der Problem- und Konfliktregelung (vgl. für viele Kriesi 1993; Willke 1992; Jänicke 1986). Informale, konsensorientierte Verfahren mit Verhandlungscharakter, insbesondere Mediationsverfahren, sind bei einigen hochstrittigen Fällen offensichtlich besser zur Konfliktregelung geeignet. Gleichwohl haben auch Mediationsverfahren ihre Grenzen: Sie sind einerseits nicht zur Regelung von *fundamentalen Wertkonflikten* geeignet; zum anderen ist zu beachten, dass die *Grenzen der Belastbarkeit* von nicht-staatlichen Organisationen und Bürgerinitiativen durch solche sehr zeitaufwendigen Verfahren recht schnell erreicht werden können. Es handelt sich hierbei nicht nur um zeitliche und finanzielle, sondern ganz besonders auch um psychische Grenzen, sind doch die Gruppenvertreter hohen Belastungen ausgesetzt, wenn sie die kompromissorientierten Verfahren(s)ergebnisse ihren Organisationen "vermitteln" und sie zur entsprechenden Selbstverpflichtung bewegen sollen - ganz zu schweigen von dem gesellschaftlichen Druck, dem sie als "Kompromissler" dann meist ausgesetzt sind.

Überdies ist eine Reihe rechtlicher und demokratietheoretischer Fragen (noch) nicht zureichend geklärt. Hierzu gehören insbesondere Fragen nach der Auswahl der Beteiligten am Mediationsverfahren und nach ihrer *demokratischen Legitimation*, stellvertretend für eine diffuse Vielzahl unorganisierter Interessen, ausserhalb der förmlichen Verfahren Verhandlungslösungen zu betreiben. Denn diese können für die gleichwohl erforderlichen förmlichen Verfahren eine Signalwirkung haben. Auch die u.E. hierzu überzeugendste Argumentationslinie, Mediationsverfahren seien in Systemen des organisierten Interessenpluralismus (mit asymmetrischen Einflusschancen) eine partizipationsdemokratische Erweiterung neokorporatistischer Politiknetzwerke, vermag einen Rest an Unbehagen nicht aufzulösen. Der Grund für solche Zweifel liegt u.a. in den sehr hohen Anforderungen an Neutralität, Demokratieverständnis und Geschick des Mediators oder der Mediatorin, schwer organisierbare Interessen zu ermitteln und ihre Partizipation aktiv zu betreiben (Amy 1987; Pruitt und Kressel 1985; Honeyman 1993). Eine Professionalisierung der Mediatorrolle (einschliesslich der Entwicklung eines Berufskodex) sowie eine generelle Normierung des Verfahrens (im Sinne einer Muster-Geschäftsordnung für alle Mediationsverfahren) könnte den Grad der Beliebigkeit, d.h. den Ermessensspielraum von Mediatoren zwar eingrenzen. Es müss-

ten damit voraussichtlich aber auch die aus anderen Politikbereichen hinlänglich bekannten Nachteile einer Professionalisierung und Kommerzialisierung von ausgelagerten staatlichen Aufgaben in Kauf genommen werden (z.B. Bildung von "Fachbruderschaften", "Partizipationseliten", "Zielverschiebung")⁹.

Im Fall einer Verfahrensnormierung bei gleichzeitiger Professionalisierung der Mediatorenaufgabe müsste zudem mit einer "schleichender Verrechtlichung" eines gesellschaftlichen Politikinstrumentes gerechnet werden, das seinerzeit gerade wegen der Probleme einer zu grossen Verrechtlichung konventioneller Problemregelungsverfahren "erfunden" worden ist. In Ländern wie der Bundesrepublik Deutschland, wo Politikfelder und Lebenswelten zunehmend von juristischen Problemregelungsmodi kolonialisiert werden ("Verjuristifizierung")¹⁰, würde das wohl unweigerlich die Verrechtlichung vorantreiben, deren Folge Gesetze, Verordnungen und Verwaltungsvorschriften (möglicherweise eine "Technische Anleitung Mediatorverhalten"), sicherlich auch entsprechende EG-Richtlinien zu Mediationsverfahren wären.

Trotz einiger noch offener Fragen zu diesem für europäische Länder jungen Instrument zur Regelung von Umweltkonflikten kann u.E. auf der Basis bisheriger praktischer Erfahrungen gesagt werden: Mediationsverfahren - wie allgemein alternative Konfliktregelungsverfahren - haben erwiesenermassen das Potential, das bestehende Instrumentarium zur Regelung von Umweltproblemen sinnvoll zu bereichern. Es wäre aber ein überzogener Anspruch, von ihnen zu erwarten, dass sie einen "Königsweg" ebnen könnten. Einen solchen allein selig machenden Weg wird es in unseren hochkomplexen Gesellschaften wohl nicht geben. Auf der Basis von Analysen von Vollzugsdefiziten wie auch erfolgreichen Massnahmen aus der Umweltpolitik ist es dagegen plausibel, für ein möglichst grosses "Instrumentenorchester" (Jänicke) zu plädieren¹¹, in dem regulative Instrumente und staatliche Rahmensteuerung wichtige Parts spielen.

⁹ Vgl. hierzu die neuen Erkenntnisse aus den USA, dokumentiert in RESOLVE 1994 sowie Busterud 1990.

¹⁰ Vgl. hierzu die Zeitschrift International Science Review, Vol. 15, No. 2, 1994 mit dem Schwerpunktthema "judicialization of politics".

¹¹ Das zeigt auch eine Analyse von (relativ) erfolgreichen Umweltschutzmassnahmen im internationalen Vergleich (vgl. Jänicke und Weidner 1995): Relative oder gar absolute Umweltentlastungen wurden in keinem Fall bloss mit einem einzigen Instrument, sondern immer durch eine Kombination von Instrumenten erzielt, wobei sich ihre jeweils

Hieraus lässt sich folgendes Fazit ziehen: Vom politisch-administrativen System initiierte und finanzierte Mediationsverfahren sollten hochstrittigen und -komplexen Fällen vorbehalten bleiben. In jedem Fall sollte geprüft werden, inwieweit die konventionellen Instrumente im Rahmen der üblichen Ermessensspielräume nicht doch so modifiziert werden können, dass sie als sinnvolles Konfliktregelungsinstrument wirken könnten, das auch vor den bei Mediationsverfahren herangezogenen Evaluationskriterien Bestand hat. Es scheint jedenfalls sinnvoller zu sein, die förmlichen Verfahren zu verbessern, anstatt alles irgendwie Problematische an Neutrale und Konfliktregler abzugeben - und dabei, wie in der Bundesrepublik Deutschland teilweise der Fall, die Gesetzesmaschinerie zur Produktion von üblicherweise konfliktschaffenden, zumindest nicht konfliktbewältigenden Regelungen weiterhin auf Hochtouren laufen zu lassen. Mit Bezug auf das Kriterium der "gesamtgemeinschaftlichen Effizienz" scheint es auch nicht sinnvoll zu sein, quasi jeden strittigen Komposthaufen mit einem Mediationsverfahren zu "garnieren". Denn dies würde nur die ohnehin bestehende Tendenz verstärken, jede örtliche Opposition als legitime "Notwehraktion" zu betrachten (vgl. etwa die Literatur zum sog. NIMBY-Problem: Inhaber 1992, Wheeler 1994).

Gleichwohl ist zu vermuten, dass Mediationsverfahren aus guten Gründen zunehmen werden: Sie verschaffen den staatlichen Behörden neue Flexibilitätsspielräume und entwickeln sich zunehmend zu adäquateren Instrumenten sich verändernder Politiknetzwerke, in die Zielgruppen und Politikbetroffene immer stärker integriert werden. Schliesslich sind solche Verfahren prinzipiell adäquate Instrumente für die Aufgaben- und Problembereiche neuer öffentlicher Politiken, die eine aktive und freiwillige Beteiligung gesellschaftlicher Gruppen voraussetzen.

Konventionelle Evaluationsverfahren und Mediations müssen dabei nicht immer als pure Alternativen betrachtet werden (Subirats 1995).

spezifische Verknüpfung und Gewichtung als eine wichtige Erfolgsbedingung erwies. Strukturelle und situationale Rahmenbedingungen, Akteurskonfigurationen, Grade des Umweltbewusstseins (und der öffentlichen Aufmerksamkeit) waren hierbei wichtige, zu berücksichtigende Faktoren. Deutlich wurde anhand vieler Fallstudien, dass staatliche (auch: hierarchische) Steuerung eine massgebliche Einflussgrösse ist und wohl bleibt, wobei "Meta-Instrumente", die die Handlungsbedingungen ökologischer und anderer (z.B. industriepolitischer) Innovateure generell verbessern, voraussichtlich eine zunehmend wichtige Rolle spielen werden. Zu den sogenannten Meta-Instrumenten zählen etwa partizipations- und transparenzfördernde Instrumente, Umwelthaftungsregelungen, funktionale und räumliche Dezentralisierung in Kombination mit neuen kooperativen Integrationsmechanismen sowie Arrangements für Verhandlungslösungen.

Vermutlich ist es sinnvoller, beide Verfahrenstypen als *zergliederungsfähige Systeme* zu konzeptualisieren und ihre Einzelelemente fallbezogen miteinander zu *kombinieren*. So könnte es beispielsweise bei einem komplexen förmlichen Verfahren sinnvoll sein, die durchzuführende Evaluation von üblicherweise strittigen Issues (Stichwort "Expertenstreit") in ein partielles Mediationsverfahren zu überführen. Vielleicht sogar erweist sich das Mediationsverfahren als besser geeignet zur Konsensbildung über komplizierte technisch-naturwissenschaftliche Streitfragen als zu einer abschliessenden Entscheidung (Vereinbarung) zum Gesamtvorhaben. Denn diese stellt in der Praxis insbesondere den gesellschaftlichen Organisationen der Politikbetroffenen grosse Probleme, indem es sie oftmals in "politische Zerreisssituationen" mit ihrer Klientel bringt. Ähnliches gilt auch für öffentliche Verwaltungen, die durch überkomplexe Mediationsverfahren häufig zu schwierigen rechtlichen Gratwanderungen gezwungen sind. Es erscheint daher als ein praktikabler Weg, spezifische Streitelemente aus konventionellen Verfahren in Mediationsverfahren auszulagern, um dann hier erzielte Teilkonsense wieder in konventionelle Verfahren einzuspeisen. Der Streit um "die Gesamtentscheidung" wäre dann Sache des politischen Normalprozesses, der unter diesen Voraussetzungen an umweltpolitischer Innovationsfähigkeit (und Kompromissfähigkeit) gewinnen könnte.

TEIL IV: DURCHFÜHRUNG, KONZEPTE, METHODEN

Die Durchführung von Evaluationen stellt grosse *Anforderungen* an die Forschung. Evaluatorinnen und Evaluatoren müssen eine Reihe von Voraussetzungen erfüllen und die verschiedensten Regeln beachten, wenn sie eine Studie erarbeiten. Im Kapitel 3 werden die Anforderungen an wissenschaftliche Evaluationen vorgestellt. Diese richten sich sowohl an Stellen in Politik und Verwaltung, die Evaluationen in Auftrag geben, wie auch an die Forschung, welche sie durchführt. Im folgenden Teil IV dieses Buchs werden die Kriterien behandelt, denen Evaluatorinnen und Evaluatoren genügen müssen. Im Vordergrund stehen deshalb wichtige Konzepte, Methoden und Verfahren¹, die für die Evaluationsforschung relevant sind.

Die in der internationalen Evaluationsforschung entwickelten Evaluationsstandards (vgl. Abschnitt 3.4) legen grossen Wert auf die Qualität der Methoden bei der Durchführung einer Evaluation. Sie bestimmen aber nur die Kriterien, nach denen einzelne Evaluationen bewertet werden können. Eigentliche *Empfehlungen*, wie Evaluationen anzufertigen sind, enthalten sie nur bedingt. Diese Lücke soll hier nach Möglichkeit für die schweizerische Praxis geschlossen werden. Die folgenden Kapitel (10-13) versuchen deshalb im einzelnen, verschiedene Fragen der Durchführung von Evaluationen zu beantworten.

Kapitel 10 stellt die verschiedenen *Arbeitsschritte* vor, die bei der Erarbeitung von Evaluationsstudien einzuhalten sind. Im wesentlichen werden die Probleme erörtert, die sich bei der Benennung der zu untersuchenden Fragestellungen, bei der Entwicklung eines für die zu analysierende politische Massnahme geltenden Wirkungsmodells, bei der Ausarbeitung des Untersuchungsdesigns, bei der Wahl der Forschungsmethoden und bei der Festlegung einer Umsetzungsstrategie ergeben.

Im Kapitel 11 wird zunächst die grosse Bedeutung der Wahl einer bestimmten *Untersuchungsanlage* hervorgehoben. Die verschiedenen Dimensionen von Evaluationsdesigns, aus denen die Forschung auswählen kann, sind Gegenstand des zweiten Abschnitts. In einem dritten Unterkapitel werden anhand von Beispielen einige gängige Untersuchungsde-

¹ Eine pointiert auf die Bedürfnisse der Verwaltung ausgerichtete Sicht gibt Bussmann (1995a).

signs wie die deskriptive Fallstudie, der Quervergleich, die Zeitreihenanalyse sowie das Experiment vorgestellt. Am Schluss wird betont, dass sich in vielen Fällen ein Methodenmix oder gar eine eigentliche Triangulation von Methoden aufdrängen kann.

Kapitel 12 befasst sich mit den eigentlichen *Forschungsmethoden*. Zunächst wird das dominante rational-analytische Methodenparadigma von alternativen Ansätzen abgegrenzt. In den folgenden Abschnitten werden sowohl für quantitative wie für qualitative Verfahren Fragen der Datenerhebung, der Datenauswertung und der Präsentation der Ergebnisse behandelt.

KAPITEL 10: ARBEITSSCHRITTE

Andreas Balthasar

Am Anfang einer jeden Evaluation stehen Fragen. Die Politik interessiert sich etwa für die Wirkungen von Subventionen in der Energiepolitik oder für die Erfolge der Gleichstellungspolitik. Die Verwaltung zweifelt an den Vorteilen von Privatisierungen. Häufig genügt es, wenn diese Fragen auf der Basis der vorhandenen Kenntnisse beantwortet werden. Oft sind jedoch ausgedehnte, wissenschaftlich fundierte Evaluationen notwendig. Wie werden solche Untersuchungen durchgeführt?

Vereinfacht lässt sich der Weg der Evaluation von der Frage zur Antwort in die fünf Etappen "Problembenennung", "Wirkungsmodell", "Untersuchungsdesign", "Forschungstechnik" und "Umsetzung" unterteilen. Allerdings ist das Bild von der sukzessiven Abfolge einzelner Phasen trügerisch. Die Abgrenzung der einzelnen Stufen erfolgt weniger chronologisch als funktionell. In der Regel muss man sich die Evaluationsforschung als Prozess vorstellen, bei welchem die verschiedenen Schritte überlappend bearbeitet werden.

10.1. PROBLEMBENENNUNG

Die von Politik und Verwaltung gestellten Fragen sind in den meisten Fällen nicht ausreichend präzise. Wissenschaftlich fundierte Evaluationen können indessen nur auf genau formulierte Fragen Antworten geben (vgl. Kapitel 3.3.). Die Festlegung der Forschungsfrage ist deshalb der Angelpunkt jeder Studie. Im ersten Schritt einer Evaluation gilt es, ein politisches oder administratives Problem in eine wissenschaftlich bearbeitete Forschungsfrage umzugießen. Zu diesem Prozess der "Problembenennung" gehören die Eingrenzung der Fragestellung, die Prüfung der Machbarkeit, die Bestimmung des Evaluationskontextes und die Festlegung der Evaluationsdimension. Diese Aspekte hängen eng zusammen und müssen weitgehend gleichzeitig geklärt werden.

10.1.1. Eingrenzung der Fragestellung

Aus dem Blickwinkel der Evaluationsforschung kann an einer öffentlichen Politik grundsätzlich alles interessant sein. Im Rahmen der Analyse etwa von Subventionen in der Energiepolitik kann man sich fragen, welche Subventionen ausbezahlt werden, wer die Subventionen erhält, wer auch ohne Subvention gehandelt hätte, wieviel Energie eingespart worden ist und in welchem Masse die Umweltbelastung zurückgegangen ist. Jede Problemanalyse kann so weit ausgeweitet werden, dass alles als faktisch interdependent und relevant wahrgenommen wird. Alle Einflussfaktoren und Zusammenhänge lassen sich jedoch nicht gleichzeitig untersuchen. Eine Auswahl ist notwendig. Die Fragestellung muss eingegrenzt werden.

Dazu ist es einmal zweckmässig, zwischen *deskriptiv*, *normativ* und *kausal-orientierten* Fragestellungen zu unterscheiden (GAO 1991a: 7; Bussmann 1995a: 60). Bei deskriptiven Fragen steht die Beschreibung von Tätigkeiten, Objekten oder Sachverhalten im Zentrum (z.B. Wie hat sich die Förderung alternativer Energien in den letzten zehn Jahren verändert?). Unserer Ansicht nach handelt es sich allerdings bei dieser Art von Fragen nicht um eigentliche Evaluationen, weil deskriptive Fragestellungen nicht auf Wertungen abzielen. Evaluationen beinhalten dagegen definitionsgemäss (vgl. Kapitel 2.3.) Beurteilungen. Insbesondere bei umstrittenen Themen werden von auftraggebenden Instanzen jedoch oftmals keine explizit wertenden Aussagen zugelassen, so dass mit deskriptiven "Evaluationen" vorlieb genommen werden muss. Normative Fragestellungen vergleichen Ist- mit Soll-Zuständen (z.B. Werden die Ziele der Energiepolitik erreicht?). Bei kausal-orientierten Fragen werden Wirkungsketten verfolgt (z.B. Welchen Beitrag leisten Subventionen zur Verbreitung von Solaranlagen?). Ob eine Evaluationsfragestellung deskriptiven, normativen oder kausal-orientierten Charakter aufweist, hat grosse Auswirkungen auf das Design der Evaluation (vgl. Kapitel 11).

Im Zuge der Eingrenzung der Fragestellung geht es auch darum festzulegen, auf welchen *Stufen* die Politikumsetzung evaluiert werden soll. In früheren Kapiteln sind verschiedene Evaluationsgegenstände (Kapitel 4.2.) und Evaluationskriterien (Kapitel 5) auseinandergehalten worden. Die dort skizzierten Fragestellungen können kaum je in einer einzelnen Evaluation abgehandelt werden. Es ist deshalb zu entscheiden, ob eine bestimmte Policy auf ihre interne und externe Kohärenz untersucht, ob die Effizienz der Vollzugsorganisation geprüft, ob der Output analysiert, ob

die Impacts einer Effektivitätsprüfung unterzogen oder ob auch die Outcomes und damit die Wirksamkeit der Politik einbezogen werden sollen.

Schliesslich ist es für die Ausgestaltung des Untersuchungsdesigns entscheidend, ob eine Evaluation *ex ante* (im voraus) oder *ex post* (nach Einführung der Massnahmen) erfolgen soll.

Die Eingrenzung der Fragestellung wird vielfach durch die Tatsache erschwert, dass der eigentlichen Durchführung eines Evaluationsprojektes ein langer verwaltungsinterner Verhandlungsprozess vorangeht. In der Addition von verschiedenen Fragen liegt dann politisch vielfach die einzige Chance der Auftraggebers, Widerstände gegenüber Evaluationen zu umschiffen. Dadurch entsteht eine Menge von Fragen, die sich kaum mehr in einem Projekt bearbeiten lassen. Weitere Probleme können sich dadurch ergeben, dass die in der Fragestellung benutzten Begriffe derart schwammig oder mit einer Fülle unterschiedlicher Bedeutungsinhalte versehen sind, dass sich kaum darauf aufbauen lässt. Ausschlaggebend für das Gelingen einer Evaluation ist es deshalb, dass die Ausgangsfragen klar formuliert sind und dass realistische Ziele gesteckt werden. Da Evaluationen häufig als Auftragsforschung konzipiert sind, sind in dieser Arbeitsphase Kooperation und Verhandlungen mit dem Auftraggeber oder der Auftraggeberin besonders wichtig. Nur dadurch können falsche Erwartungen und damit Enttäuschungen vermieden werden.

10.1.2. Bestimmung des Evaluationskontextes und der Evaluationsdimension

Öffentliche Stellen, welche Evaluationen veranlassen, möchten damit meist ihre Kenntnisse über die Wirkungen von politischen Massnahmen verbessern. Dahinter steht der Wunsch nach Kontrolle und in der Regel nach Verbesserungsmöglichkeiten. Der implizite Widerspruch zwischen "Kontrollieren" und "Lernen" in öffentlichen Politiken bildet den Hintergrund jeder Evaluationsstudie. Hinter der Evaluation von energiepolitischen Massnahmen steht beispielsweise der in Artikel 24 des Energienutzungsbeschlusses vom 14. Dezember 1990 (SR 730.0) explizit formulierte Auftrag zur Rechenschaftsablage gegenüber dem Parlament. Die Überprüfung der Tätigkeit der Verwaltung durch die Parlamentarische Verwaltungskontrollstelle hat in erster Linie eine kontrollierende Funktion. Evaluationen, bei denen die Überprüfung im Vordergrund des Interesses steht, stossen oft auf Widerstände. Viele Bearbeiter von Studien

haben dies schon erfahren: ohne die Auskunftsbereitschaft der beteiligten Beamten lassen sich zahlreiche Evaluationsfragen nicht beantworten. Die Kontrollstellen haben sich daher im genannten Fall die Auskunftspflicht des Personals gesichert. Doch auch dies war der Kooperationsbereitschaft nicht immer förderlich.

Obwohl die Kontrolle von Verwaltungshandeln ein aus staatspolitischer Sicht absolut legitimes Ziel von Evaluationen ist, profitiert die Qualität der Ergebnisse davon, wenn die Untersuchung grosses Gewicht auf *Lern-effekte* bei allen Beteiligten (vgl. Kapitel 13) legt. Zu diesem Zweck hat die Evaluationsforschung verschiedene spezielle Verfahren entwickelt. Guba und Lincoln (1987: 1989) propagieren beispielsweise die *betroffenen-orientierte Evaluation*. Dabei werden von der Massnahme Betroffene in den Evaluationsprozess einbezogen und auf diese Weise Akzeptanz und Umsetzbarkeit der Ergebnisse verbessert. Ein gelungenes Beispiel für diese Art von partizipativer Evaluation ist die Untersuchung der Wirtschaftsförderung des Kantons Wallis, welche von Michel Rey, Luc Vodoz, Louis-M. Boulianne und Siegfried Alberton (Rey et al. 1993) durchgeführt wurde. Allerdings ergeben sich auch bei derartigen Zugängen Probleme. So sind die verschiedensten, oft divergierenden Ansprüche und Interessen zu berücksichtigen. Die Evaluatoren agieren nicht mehr als aussenstehende Beobachter, sondern teilweise bereits als Mediatoren (vgl. Kapitel 9), Methodologen und Experten. Sie laufen Gefahr, allzustark in den politischen Aushandlungsprozess integriert zu werden und ihre eigenen Ziele als Massstab zur Beurteilung des Programms heranzuziehen (vgl. Kapitel 7.2.). Diese besondere Situation ist bei der Entwicklung des Evaluationsdesigns gebührend zu berücksichtigen.

Eine ähnliche Intention wie die betroffenen-orientierte Evaluation verfolgt die *zielfreie Evaluation* (vgl. Kapitel 3.3.). Sie wurde beispielsweise von Scriven propagiert (Scriven 1980). Während die zielorientierte Evaluation mit Hilfe von Indikatoren den Zielerreichungsgrad misst und daher auf eindeutige und stabile Ziele angewiesen ist, beurteilt die zielfreie Evaluation eine Massnahme anhand des gesamten Sets von ausgelösten Wirkungen. Scrivens Motivation für zielfreie Evaluationen beruht darauf, dass die Vorgabe von Zielen notwendigerweise zu einer unvollständigen Erfassung der Programmkonsequenzen führen muss. Ausserdem werden nur die Interessen der anvisierten Beteiligten und Betroffenen berücksichtigt, nicht aber jene der Kreise, die unfreiwillig die positiven oder negativen Auswirkungen eines Programms zu spüren bekommen. Die Identifizierung der Ursprünge solcher Verzerrungen in der Beziehung zwi-

sehen der Evaluation und dem Programm ist für Scriven genauso von Bedeutung wie die Entwicklung von praktikablen Methoden zur Kontrolle der identifizierten Fehlerquellen.

10.2. MACHBARKEITSPRÜFUNG

Angesichts der vielfältigen Probleme der Abgrenzung von Forschungsfragen unter den Bedingungen eines gegebenen Evaluationskontextes drängt es sich zu Beginn einer Evaluation auf, deren Machbarkeit zu überprüfen. In der angelsächsischen Literatur wird in diesem Zusammenhang der Begriff "evaluability assessment" verwendet. Dieser Begriff hat jedoch keine einheitliche Bedeutung. Ein Protagonist des Konzepts, Wholey (1983), benutzt ihn im Sinne einer ausgedehnten Vorabklärung mit allen involvierten Stellen und Akteuren, namentlich im Hinblick auf die Ziele, die mit dem zu untersuchenden Programm verfolgt werden sollen sowie dem Wirkungsmodell, das ihm zugrundeliegt. Was im folgenden unter dem Titel "Machbarkeitsprüfung" abgehandelt wird, geht indessen von einem weniger weitreichenden Konzept aus.

Im Vordergrund der Machbarkeitsprüfung steht der Umstand, dass sich Evaluatorinnen und Evaluatoren in konfliktiven politischen Feldern bewegen und dort meist *Spannungen* ausgesetzt sind. So muss das Forschungsteam etwa folgende Fragen beantworten: Welches sind eigentlich die Motive für die Evaluation? Welche Rolle soll das Evaluationsteam übernehmen? Ist eine unabhängige Untersuchung überhaupt erwünscht oder sollen vorgefasste Meinungen bestätigt werden? Zur Klärung dieser und weiterer Fragen ist eine Machbarkeitsprüfung notwendig. Es handelt sich dabei um Überlegungen zur Planung und zur Durchführung der Untersuchung.

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

den? Welches ist die Rolle des Evaluationsteams? Ist es Ratgeber, Informationsbeschaffer oder Schiedsrichter?

Die *operationellen Abklärungen* betreffen dagegen etwa folgende Themen: Sind Widerstände gegenüber der Evaluation zu erwarten, welche eine systematische Bearbeitung des Themas beeinträchtigen oder gar verunmöglichen? Sind die Beteiligten kooperationsbereit? Entsprechen die zur Verfügung stehenden finanziellen Mittel und der festgelegte Zeitrahmen den Voraussetzungen für eine sachgerechte Bearbeitung der Fragestellung? Ist der Zugang zu den Daten, Materialien und Quellen gesichert? Sind die Beteiligten und Betroffenen kooperationsbereit?

Wesentlich ist in jedem Fall, dass im Vorfeld einer Evaluation abgeklärt wird, ob ein für eine Untersuchung geeignetes Klima besteht (vgl. Rossi und Freemann 1993: 406 f.). Dazu gehört auch die grundsätzliche Bereitschaft der Verantwortlichen, sich auf Lernprozesse einzulassen (vgl. Kapitel 13).

10.3. WIRKUNGSMODELL

Jeder politischen Massnahme liegen Vorstellungen über Wirkungszusammenhänge (vgl. dazu Kapitel 5.1.) zugrunde. Die Politik geht davon aus, dass mit dem Aufstellen und Umsetzen eines Programms etwas bewirkt wird.

Von der Formulierung einer Massnahme in einem Rechtserlass bis zur Erreichung der beabsichtigten Wirkung ist der Weg indessen meist lang (vgl. Kapitel 4.3.2.). Die Umsetzung eines Programms muss organisiert werden. Verantwortungen sind zu klären, Mittel zu sprechen, Abläufe festzulegen. Auch wenn Outputs objektiv feststellbar sind, können die Wirkungen noch auf sich warten lassen. Zuerst müssen die Politikadressaten die Massnahme kennen und akzeptieren. Erst wenn diese im erwarteten Ausmass darauf reagiert, können sich zielrelevante Impacts einstellen. Politische Massnahmen haben jedoch meist nicht allein Verhaltensänderungen zum Ziel. Die Installation von Solaranlagen ist beispielsweise kein Selbstzweck, sondern dient dazu, die Abhängigkeit der Schweiz von Energieimporten und von der Nuklearenergie zu reduzieren. Ob Impacts also auch die gewünschten gesellschaftlichen Outcomes bewirken, hängt vom Zutreffen der einer Politik zugrundegelegten Kausalhypothese bzw. ihres Wirkungsmodells ab.

Evaluationen haben u.a. das Zutreffen dieser von einer öffentlichen Politik postuliertem Wirkungszusammenhänge zu untersuchen (vgl. Kapitel 5.1.). Mit Hilfe der Analyse politischer Verlautbarungen, amtlicher Dokumente und ähnlicher Unterlagen lassen sich die Wirkungsüberlegungen des Gesetzgebers und der für das Programm Verantwortlichen rekonstruieren. Die Evaluationsteams können diese Vorstellungen als Grundlage für ihre Arbeit übernehmen, sie können sie erweitern oder ersetzen. Daraus entsteht das untersuchungsleitende Wirkungsmodell, das darstellt, wie der Untersuchungsgegenstand in den Augen der Evaluatoreninnen und Evaluatoren mit relevanten Einflussfaktoren zusammenhängt.

Das Wirkungsmodell hat den Charakter einer Hypothese. In unserem Beispiel könnte die Feststellung "Die Massnahme ist bei einem kleinen Teil der Adressaten bekannt" eine untersuchungsleitende Hypothese sein. Solche Hypothesen dienen als Leitlinien zum Abfragen des Gegenstandsbereichs². Sie brauchen nicht unbedingt induktiv aus den Überlegungen und Erwartungen der Akteure abgeleitet zu werden; vielfach ist es auch möglich, sie deduktiv aus theoretischen Vorstellungen über den Untersuchungsgegenstand herzuleiten. Bei der Erarbeitung eines Wirkungsmodells geht es auch darum abzuklären, ob aus der bisherigen Forschung Theorien vorliegen, welche für die Beantwortung der spezifischen Fragestellung nutzbringend verwendet werden können. Durch theoretische Bezüge wird die Hypothesenbildung angeregt und die Einordnung des Problems in einen grösseren Zusammenhang ermöglicht. Um der Gefahr einer selektiven Sichtweise auszuweichen, kann es sinnvoll sein, einem Wirkungsmodell verschiedene Hypothesen zugrunde zu legen. Nicht selten weichen ohnehin die induktiv gewonnenen Hypothesen von den theoretisch abgeleiteten ab. Verschiedene Wirkungspfade können die Evaluation unter Umständen auch auf unerwartete und unbeabsichtigte Outcomes führen. Bei komplexen Programmen ist die Ausarbeitung des Wirkungsmodells bzw. der Wirkungsmodelle deshalb von besonderer Bedeutung.

² Zum Vorgang, wie Hypothesen zu formulieren sind, vgl. von Alemann und Forndran, 1990, sowie Atteslander, 1993.

10.4. UNTERSUCHUNGSDESIGN UND FORSCHUNGSTECHNIK

Das *Untersuchungsdesign* ist das konzeptionelle Herz einer Evaluation. Darüber sind sich die meisten Evaluatorinnen und Evaluatoren einig. Weniger gross ist die Übereinstimmung hingegen betreffend der Frage, was in einem Evaluationsprojekt unter einem Untersuchungsdesign zu verstehen ist (z.B. Roth 1993: 711; Fitz-Gibbon und Morris 1987). Die verschiedenen Möglichkeiten und Forschungsansätze werden im Kapitel 11 eingehend erörtert. An dieser Stelle soll lediglich darauf verwiesen werden, dass die Wahl des Untersuchungsdesigns verschiedene Entscheidungen impliziert. Zunächst ist festzulegen, ob sich die Evaluation auf einen einzelnen Fall konzentrieren will oder ob Vergleiche unter mehreren Fällen vorgenommen werden sollen (Längsschnittanalyse eines einzelnen Gegenstandes oder Querschnittsvergleich mehrerer Untersuchungseinheiten). Im weiteren muss entschieden werden, ob alle möglichen Objekte zu erfassen sind oder ob eine Auswahl getroffen wird. Und schliesslich ist abzuklären, ob ein experimenteller, ein quasi-experimenteller oder ein nicht-experimenteller Ansatz gewählt werden soll.

Damit sind zwar noch nicht alle Entscheidungen über das Vorgehen bei der Evaluationsuntersuchung getroffen. Insbesondere ist noch nichts über die Wahl der Forschungsmethoden gesagt. Diese ordnen wir indessen einem nächsten Arbeitsschritt zu (vgl. Kapitel 12). Wie bei den anderen Arbeitsschritten gilt auch hier, dass die Abgrenzung zwischen Design und Techniken funktionell zu verstehen ist.

Die Auswahl des Evaluationsdesigns stützt sich auf verschiedene Faktoren. In erster Linie hat sie auf die eingegrenzte Forschungsfrage Rücksicht zu nehmen. Darüber hinaus spielen dabei die theoretischen Überlegungen zum Wirkungsmodell eine Rolle. Und schliesslich sind methodische Fragen der Forschungslogik für die Wahl des Designs ausschlaggebend.

In der Praxis spielen allerdings bei der Gestaltung eines geeigneten Designs auch noch andere Faktoren mit. Wichtig sind vor allem die Rahmenbedingungen, welche durch die zeitlichen, finanziellen und qualifikatorischen Ressourcen sowie durch den geplanten Verwendungszweck der Untersuchung vorgegeben werden. Die Wahl des geeigneten Untersuchungsdesigns ist deshalb eine höchst anspruchsvolle Aufgabe.

10.5. UMSETZUNG

Wissenschaftliche Forschung und politische Praxis stehen in einem viel-diskutierten Spannungsverhältnis. Die Politik stellt den Anspruch an die Evaluation, dass deren Ergebnisse in der Praxis umgesetzt werden können (vgl. Kapitel 13). Um diesem Anspruch zu genügen, müssen Evaluationen verschiedene Bedingungen erfüllen.

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

Mit der Forderung nach der *Nützlichkeit* von Evaluationsergebnissen für die politische Praxis geraten Studien oftmals mitten in politische Spannungsfelder (vgl. dazu Kapitel 6). Schon die Frage, welche Massnahmen evaluiert werden sollen, ist vielfach umstritten. Evaluatorinnen und Evaluatoren sollten sich darum Rechenschaft darüber geben, wessen Kriterien bei der Auswahl massgeblich waren. Auftraggebende Stellen sind nämlich aus ganz unterschiedlichen Motiven an Evaluationen interessiert.

In der Praxis lassen sich bei den Auftraggebern *vier Nutzungsinteressen* unterscheiden, die für die Umsetzung bedeutsam sind. Die erste Variante bildet ein technisches Interesse: Evaluationen sollen Wissen liefern, welches zur Optimierung staatlicher Interventionen eingesetzt werden kann. Evaluationen erfüllen in diesem Fall eine zweckmässige *instrumentelle Funktion*. Die Umsetzung erfolgt dadurch, dass sich Entscheidungsinstanzen und andere Interessengruppen direkt und konkret auf **F**orschungsergebnisse beziehen. Die zweite Möglichkeit besteht in einer eher *konzeptionellen Motivation*. Auch dieser Impuls ist für die Evaluation weitgehend unproblematisch. Die Evaluation reflektiert, stellt Fragen, schafft Begriffe, bringt Dinge zur Sprache und kann Lernprozesse auslösen. Die Untersuchungsergebnisse werden in diesem Falle als Denkanstösse bei der Weiterbearbeitung potentieller Lösungswege für bestimmte politische Probleme verwendet. Den dritten Motivationsstrang bilden *argumentative Absichten*. Evaluationen werden eingesetzt, um politische Argu-

mente zu erarbeiten, welche dann unter strategischen Gesichtspunkten umgesetzt werden. Viertens können Evaluationen auch veranlasst werden, um zur *Erhöhung der Akzeptanz* von Massnahmen und zu deren Rechtfertigung beizutragen. In diesem Fall dient Evaluation in erster Linie der Legitimierung. Vor allem bei den letzten beiden Formen besteht die Gefahr einer unzulässigen Instrumentalisierung der Evaluation durch die Politik. Nur die genehmen Resultate werden weiterverwendet (vgl. u.a. Rieder und Varone 1991).

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

KAPITEL 11: UNTERSUCHUNGSDESIGNS

Ulrich Klöti, Thomas Widmer

Das vorliegende Kapitel befasst sich mit der Wahl der Forschungsstrategie, mit der eine evaluative Frage angegangen wird. Auf die grosse Bedeutung dieses Arbeitsschrittes ist bereits im Kapitel 10.4. hingewiesen worden. Die bei dieser Strategiebestimmung zu treffenden Entscheidungen werden in einem sogenannten Evaluationsdesign zusammengefasst. Darunter verstehen wir die Festlegung der Form der empirischen Umsetzung einer evaluativen Fragestellung. Es geht somit bei der Erstellung eines Evaluationsdesigns darum, die Art und Weise, wie die an die Evaluation gestellte Fragestellung in eine empirische Untersuchungsanlage umzusetzen ist, in kontrollierbarer Weise zu bestimmen. Das Forschungsdesign soll namentlich die Projektmechanik (hypothesengesteuertes Zusammenwirken der zu erklärenden und der erklärenden Variablen) bestimmen, das gewählte Untersuchungskonzept darlegen und begründen sowie die zur konkreten Erfassung der Variablen ausgewählten Indikatoren angeben.

Der Abschnitt 11.1. enthält eine Beschreibung der verschiedenen Design-Dimensionen; diese werden anschliessend verglichen und kritisch diskutiert. Schliesslich werden anhand eines Flussdiagramms die verschiedenen Dimensionen zu einem Ganzen zusammengeführt und damit gleichzeitig eine Synopse der verschiedenen Design-Dimensionen sowie die Interdependenzen zwischen diesen dargestellt.

Im Abschnitt 11.2. geht es darum, anhand von Beispielen einige gängige Evaluationsdesigns vorzustellen, ihre Vor- und Nachteile zu erörtern und ihre Eignung für bestimmte evaluative Fragestellungen zu diskutieren.

Die Fragen der erkenntnistheoretischen Ausrichtung einer Evaluation sowie Beschreibungen und Erörterungen von qualitativen und quantitativen Untersuchungsverfahren finden sich im folgenden Kapitel 12.

11.1. DIMENSIONEN VON EVALUATIONSDESIGNS

Ein Evaluationsdesign beinhaltet die Umsetzung einer Evaluations-Fragestellung in eine Strategie für die empirische Forschung. Es hat nicht nur der Fragestellung, sondern vor allem auch dem zu evaluierenden Gegenstand (im folgenden Evaluandum genannt) Rechnung zu tragen. Die em-

pirische Sozialforschung hält dazu eine Reihe von Untersuchungskonzepten bereit, die anhand der nachfolgend angeführten zentralen Dimensionen unterschieden werden können:

- Einzelfalluntersuchungen und vergleichende Untersuchungsanlagen,
- Quer- und Längsschnittvergleiche,
- Auswahl und Vollerhebungen sowie
- Experimentelle, quasi-experimentelle und nicht-experimentelle Designs.

In der angeführten Liste fehlt ein Hinweis auf die qualitative bzw. quantitative Untersuchungsmethodik. In unserer Sicht handelt es sich dabei um eine Frage der *Forschungstechnik*; sie stellt somit eine verfahrenstechnische Entscheidung dar, die an anderer Stelle diskutiert wird (vgl. Kapitel 12). Sicherlich bestehen zwischen Untersuchungsanlage und Verfahrenstechnik mannigfaltige Interdependenzen. Trotzdem sind zwei unterschiedliche Dimensionen angesprochen, die wir - auch um einen Beitrag zur Versachlichung der Diskussion in diesem Bereich zu leisten - strikt trennen wollen. Entscheidungen auf der Ebene der Untersuchungsanlage bilden keine Präjudizien für die Wahl der Untersuchungstechnik(en). Verfahrenstechnische Fragen haben unseres Erachtens einen ganz klar nachgelagerten Charakter (vgl. dazu Cook und Reichhardt 1979; Bryman 1990).

Etwas anders präsentiert sich die Situation bezüglich der *erkenntnistheoretischen Ausrichtung* von Evaluationen. Auch diese Dimension fehlt in der oben angeführten Liste. Die paradigmatische Ausrichtung einer Evaluationsstudie steht bei der Festlegung des Evaluationsdesigns ebenfalls zur Diskussion. Die Entscheidung über deren Ausgestaltung wird in der Praxis weitgehend bestimmt durch die Ausrichtung der Fragestellung einerseits und die erkenntnistheoretische Position der Evaluatorin oder des Evaluators andererseits. Unseres Erachtens hat sich die erkenntnistheoretische Ausrichtung allerdings der Zielsetzung der Studie, wie sie in der Fragestellung zum Ausdruck kommt, unterzuordnen. Die paradigmatischen Positionen werden im Kapitel 12.1. ausführlicher diskutiert.

11.1.1. Einzelfalluntersuchungen und vergleichende Untersuchungsanlagen

11.1.1.1. Einzelfalluntersuchungen

Die am Einzelfall orientierte Untersuchungsanlage definiert den gesamten Untersuchungsbereich als einen Fall. Es findet kein Vergleich zu anderen Fällen statt. Die Fragestellung, die empirische Untersuchungsanlage, die Evaluationsergebnisse wie auch die darauf beruhenden Empfehlungen sind auf den einen, untersuchten Fall ausgerichtet. Im Grundsatz sind andere Fälle nicht von Interesse. Eine Generalisierung der Evaluationsergebnisse wird weder angestrebt noch ist eine solche Übertragung auf andere Fälle möglich. Denkbar ist höchstens, dass die Adäquatheit der Aussagen, die allenfalls übertragen werden sollten, genauestens überprüft wird, was - da sich die Evaluation auf einen Fall beschränkt - eine weitere Fallstudie erforderlich macht. Die empirisch-analytische Position (vgl. Kapitel 12.1.) lehnt die Generalisierbarkeit von Einzelfalluntersuchungen prinzipiell ab.

Der Ansatz hat den grossen Vorteil, dass der evaluierte Einzelfall detailliert in seiner Gesamtheit untersucht werden kann. Aus der vertieften Analyse eines Falles können für die Praxis relevante Ergebnisse resultieren, und es fallen unter Umständen auch Erkenntnisse von grossem theoretischem Interesse an. Vielfach wird in Abrede gestellt, dass Einzelfalluntersuchungen kausale Wirkungszusammenhänge eruieren könnten. Folgt man den Überlegungen, wie sie von Scriven unter der Bezeichnung "modus operandi" angeführt wurden, ist die Formulierung von kausalen Aussagen aufgrund von Einzelfalluntersuchungen jedoch durchaus möglich (vgl. Scriven 1976 und neuerdings Mohr 1995a: 248-273 und 1995b). In der Evaluationsforschung sind Einzelfalluntersuchungen eher selten; in Tat und Wahrheit kommen bei der Untersuchung der Wirkungen einer Massnahme auch in einem einzigen Fall meist vergleichende Untersuchungsanalysen zur Anwendung, weil meist ein Vorher/Nachher-Vergleich angestellt wird (Längsschnittanalyse).

Generelles Ziel einer Einzelfalluntersuchung bildet das Verständnis oder die Erklärung des Einzelfalles. Dies soll erreicht werden, indem der Fall möglichst detailliert und reich beschrieben wird, um aufgrund dieser Beschreibung weitergehende Aussagen zu ermöglichen (Stake 1995). In der Praxis besteht bei Einzelfalluntersuchungen die Gefahr, dass der deskriptive im Vergleich zum analytischen Aspekt zu viel Gewicht erhält, was den analytischen Gehalt der Evaluation reduzieren kann.

11.1.1.2. Vergleichende Untersuchungsanlagen

Unter dem Sammelbegriff "vergleichende Untersuchungsanlagen" werden verschiedene Konzepte subsumiert, die sich dadurch charakterisieren, dass sie auf einem Vergleich unterschiedlicher Untersuchungsgegenstände aufbauen. Dabei kann die Zahl der Objekte stark variieren. So baut eine sogenannte *Kontrollgruppenuntersuchung* auf der Untersuchung eines Objektes auf, das sodann mit einem oder mehreren (bis auf die zu evaluierende Massnahme oder andere deren Wirkung erklärende Variablen (vgl. Kapitel 5)) möglichst ähnlichen Objekten verglichen wird. Die Vergleichsobjekte werden typischerweise vom Evaluationsgegenstand (Evaluandum) nicht beeinflusst, im Gegensatz zur Untersuchungsgruppe. Aufgrund der beobachteten Differenzen wird dann auf die Wirkungen des zu evaluierenden Programms geschlossen. Eine "*vergleichende Fallstudie*" bezieht verschiedene Fälle gleichgewichtig in die Analyse mit ein. Auch hier wird aufgrund festgestellter Differenzen auf Programmwirkungen geschlossen, wobei bei diesem Untersuchungsansatz Schlüsse zu allen untersuchten Objekten angestrebt werden. Typischerweise führen vergleichende Fallstudien zu Ergebnissen, die sich auf die Gesamtheit eines Falls und nicht auf seine Elemente beziehen. Je nach Vorgehensweise bei der Auswahl der Fälle (vgl. unten Kapitel 11.1.3.) ist es auch möglich, über die konkret untersuchten Fälle hinausgehende Ergebnisse zu erhalten (vgl. Stake 1995: 7-8). Dies setzt jedoch voraus, dass die Auswahl der Fälle genügend breit abgestützt ist. Vielfach sind hier nur hypothetische Aussagen möglich.

Eine weitere Variante vergleichender Untersuchungsanlagen bilden *Längsschnittanalysen*, die einen oder mehrere Fälle über die Zeit hinweg analysieren. Verglichen wird hier zwischen den Zuständen eines Falls zu verschiedenen Zeitpunkten oder in verschiedenen Zeiträumen (vgl. dazu ausführlicher das nachfolgende Kapitel 11.1.2.).

Der letzte Typ von vergleichenden Untersuchungsanlagen, den wir hier erwähnen möchten, ist die *Breitenuntersuchung*. Hier wird eine verhältnismässig grosse Zahl von Vergleichseinheiten dazu verwendet, Rückschlüsse auf die Wirkungen eines Programms zu treffen. Auf eine vertiefte Analyse des Einzelfalls wird hier verzichtet, um eine grössere Zahl von Fällen untersuchen zu können. Weil sich die Untersuchung nur auf einzelne, sehr selektiv ausgewählte Variablen der Fälle bezieht, spricht man dabei auch von Variablenanalysen (Przeworski und Theune 1970, Kriesi 1994:38 ff.). Gesucht wird in derartigen Studien nach strukturellen

Regelmässigkeiten, die Rückschlüsse auf die Wirkungen des Evaluandums erlauben. Die Analyseeinheit bildet aber auch hier der einzelne Fall, und die Studie nimmt einen Vergleich der Fälle vor.

Die angeführten Typen treten oft in Kombination miteinander auf. Es werden aber auch *Mischformen* eingesetzt (so beispielsweise eine Breitenuntersuchung mit Kontrollgruppenvergleich). Vergleichende Untersuchungsanlagen haben den Vorteil, dass der komparative Ansatz eine erprobte und verhältnismässig aussagekräftige Vorgehensweise darstellt. Bei der Wahl des Typs vergleichender Untersuchungsanlagen ist zu beachten, dass es zwischen interner und externer Validität abzuwägen gilt. Sind generelle Erkenntnisse allgemeiner Art erwünscht (hohe externe Validität), ist eine möglichst grosse Zahl von zu berücksichtigenden Fällen zu wählen. Besteht die Absicht jedoch darin, für den Einzelfall möglichst gut fundierte Ergebnisse zu erhalten (hohe interne Validität), ist die Zahl der Fälle eher klein zu halten (vgl. dazu auch Kapitel 11.2.3.). Beide Eigenschaften (also hohe interne und externe Validität) gleichzeitig anzustreben, verbietet im Regelfall die Forschungsökonomie. Im Rahmen von Evaluationsprojekten ist es kaum möglich, die Vorteile einer Breitenuntersuchung mit jenen einer Einzelfallstudie zu kombinieren. Neben den finanziellen Restriktionen ist dabei auch zu bedenken, dass Ergebnisse von Evaluationsstudien innert nützlicher Frist vorzuliegen haben. Kompromisse sind praktisch immer erforderlich. Die Vor- und Nachteile der verschiedenen Vorgehensweisen sind umsichtig miteinander abzuwägen (vgl. dazu Cook und Campbell 1979: 37-94, besonders 82-85).

11.1.2. Quer- und Längsschnittanalysen

11.1.2.1. Querschnittanalysen

Querschnittanalysen vergleichen verschiedene Untersuchungsobjekte wie Gemeinden, Kantone oder Länder (Makro-Ebene), Organisationen, Verbände, Parteien, Unternehmen usw. (Meso-Ebene) oder Individuen (Mikro-Ebene). Charakteristisch ist, dass die genannten Untersuchungsobjekte zu einem bestimmten Zeitpunkt oder Zeitraum erfasst werden (synchrone Vergleiche). Ein Vorteil von Querschnittanalysen liegt darin, dass der weitere (generelle) Kontext, in dem sich die Objekte lokalisieren lassen, für alle Untersuchungsobjekte derselbe ist. So sieht sich eine privatwirtschaftliche Unternehmung im Kanton A den gleichen weltwirt-

schaftlichen Herausforderungen ausgesetzt, wie dies für ein Unternehmen im Kanton B der Fall ist. An diesem Beispiel lässt sich aber auch eine Problematik von Querschnittanalysen aufzeigen. Die Vergleichbarkeit der Untersuchungsobjekte ist oft nur schwer herzustellen. So muss beim Vergleich der beiden Unternehmen im obigen Beispiel Gewähr dafür bestehen, dass sie in den gleichen Märkten tätig sind, eine ähnliche Kostenstruktur und eine vergleichbare Grosse aufweisen, usw. Um diese Vergleichbarkeit herzustellen, bestehen verschiedene Ansätze. So können die Gruppen derart aus Objekten zusammengestellt werden, sodass ein Objekt aus Gruppe A mit einem Objekt aus Gruppe B ein möglichst ähnliches (Vergleichs-)Paar bildet. Eine andere Möglichkeit liegt darin, dass die Objekte nach dem Zufallsprinzip ausgewählt werden. Dies setzt jedoch voraus, dass eine genügend grosse Zahl von Vergleichseinheiten (Gesetz der grossen Zahl) und keine strukturellen Differenzen zwischen den zwei Gruppen bestehen. Weiter kann versucht werden, möglicherweise relevante Drittfaktoren zu kontrollieren.

Ein Vorteil von Querschnittanalysen bildet ihre verhältnismässig gute Handhabbarkeit. Im Gegensatz zu Längsschnittvergleichen sind Querschnittstudien methodisch einfacher durchzuführen und haben den in der Evaluationsforschung bedeutenden Vorteil, dass Aussagen aufgrund einer Einmal-Erhebung bereits möglich sind. Dies erlaubt, die Evaluationsergebnisse rechtzeitig vermitteln zu können. Im Gegensatz dazu eignen sich jedoch Querschnittanalysen weniger dazu, dynamische Prozesse zu erfassen. Sie bauen auf einem statischen Konzept auf. Deshalb empfiehlt sich bei der Untersuchung von stark volatilen Gegenständen der Beizug von Längsschnittansätzen.

Interkantonale und insbesondere internationale Querschnittvergleiche werden oft von unterschiedlichen Forschungsteams durchgeführt (Reisekosten, Ortskenntnisse etc.). Dabei besteht erfahrungsgemäss die Gefahr, dass bei den Fallstudien unterschiedliche Variablen und Indikatoren erhoben werden. In solchen Projekten ist eine genaue Verständigung über das gemeinsame Forschungsdesign und gegebenenfalls eine permanente Kontrolle über dessen Einhaltung von grosser Bedeutung (Weidner und Knoepfel 1983).

11.1.2.2. Längsschnittanalysen

Bei Längsschnittanalysen wird derselbe Gegenstand über die Zeit hinweg untersucht. Seine Entwicklung wird über genau identifizierbare Phasen

hinweg beobachtet (diachrone Vergleiche). In der Evaluationsforschung wird dabei typischerweise ein Gegenstand vor und nach der Einführung des zu evaluierenden Programms analysiert, um aufgrund auftretender Veränderungen auf Programmwirkungen zu schliessen. Aber nicht nur die Einführung eines Programms, auch Veränderungen an bestehenden Programmen oder Programmbeendigungen ("program termination", vgl. Brewer 1978; deLeon 1983 und 1987) können in dieser Form analysiert werden (vgl. Benninghoff 1995). In der Evaluationsforschung werden derartige Neuerungen als Interventionen bezeichnet. Da sich die Evaluationsforschung oft mit Veränderungshypothesen befasst, erscheinen solche Längsschnittdesigns oft als die adäquateren Ansätze als diachrone Untersuchungsanlagen; denn gesellschaftliche Veränderungen beanspruchen üblicherweise Zeit.

Längsschnittanalysen können in unterschiedlicher Weise ausgestaltet werden. Als klassisch zu bezeichnen ist das sogenannte "*pretest-posttest design*", bei dem der Zustand zu einem Erhebungszeitpunkt vor der Intervention mit jenem nach dieser verglichen wird. Im Regelfall werden dabei die Daten jeweils trotzdem nur in einer einzigen Feldbegehung erhoben. Es ist aber durchaus auch möglich, den Status zu mehreren Zeitpunkten vor und nach der Intervention zu erfassen. Dehnt man die Datenerhebung weiter aus, so dass sowohl in der Prä- wie auch in der Postinterventionsphase eine Vielzahl von Beobachtungen zur Verfügung stehen, spricht man von einer *Zeitreihenanalyse*. Mit dieser Untersuchungsanlage lässt sich eine höhere Verlässlichkeit der Ergebnisse erreichen, als dies bei einem einfachen "pretest-posttest design" der Fall ist (Widmer 1991: 123-125). Hingegen ist es vielfach aus praktischen Überlegungen nicht möglich, ein Zeitreihendesign umzusetzen, da Ressourcen zur Datenerhebung fehlen oder eine längerfristige Datenerhebung nicht möglich ist. Im Gegensatz dazu stellt das "pretest-posttest design" oft eine recht praktikable Lösung dar.

11.1.2.3. Kombination und Integration

Wie aus dem bereits Gesagten hervorgeht, haben sowohl Längs- wie Querschnittanalysen ihre spezifischen Vor- und Nachteile. Um die Leistungsfähigkeit der Designs zu steigern, bietet sich die Möglichkeit an, eine Kombination der beiden Untersuchungsanlagen zu wählen (Knoepfel 1995e: 49 ff.). Dies kann einerseits dadurch geschehen, dass zwei getrennte Analysen durchgeführt werden. Dadurch lassen sich die Analyse-

ergebnisse gegenseitig ergänzen, besser fundieren oder auch überprüfen. Hier spricht man von einer *Kombination* der Untersuchungsanlagen. Fliesen beide Ansätze in ein einziges Design ein, nennt man dies *Integration*.

Beide Vorgehensweisen haben, im Gegensatz zu der ausschliesslichen Verwendung eines Längs- oder Querschnittsdesigns, den grossen Vorteil, dass sie in der Lage sind, gleichermassen statische Phänomene wie auch dynamische Prozesse zu erfassen. Die dafür erforderlichen Aufwendungen verhindern jedoch in der Praxis der Evaluationsforschung häufig deren Verwendung (vgl. dazu Cook 1985; Denzin 1989: 234-247; Patton 1990: 464-472; Hakim 1992: 144-145).

11.1.3. Auswahl und Vollerhebung

Evaluationen können sich mit einer Auswahl oder mit allen bestehenden Objekten befassen.

11.1.3.1. Auswahl

Bei Evaluationen, die nur eine bestimmte Auswahl (auch als Stichprobe oder "sample" bezeichnet) untersuchen, sind vor allem zwei Punkte wichtig: die Vorgehensweise bei der Auswahl und die Gültigkeit der Ergebnisse.

Erstens besteht das Problem, *wie die Auswahl getroffen* werden soll, also wie aus der Menge aller Objekte (Grundgesamtheit oder auch Universum genannt) jene ausgewählt werden sollen, die in der Evaluation empirisch untersucht werden (vgl. dazu Henry 1990). Dabei sind grundsätzlich zwei verschiedene, systematische Vorgehensweisen zu unterscheiden. Die Wahrscheinlichkeits- oder Zufallsauswahl ("random sample") baut darauf auf, dass für alle Elemente der Grundgesamtheit dieselben Chancen bestehen, in die Stichprobe aufgenommen zu werden. Eine gezielte Auswahl ("purposive sample", "theoretical sample") versucht dagegen, aufgrund bestimmter Dimensionen die Eigenschaften der Elemente in der Grundgesamtheit in systematischer Weise in der Auswahl abzubilden. Derartige Stichproben sollen beispielsweise hinsichtlich gewisser Eigenschaften möglichst homogen oder heterogen sein. Auch das sogenannte Schneeballverfahren, bei dem ein Untersuchungsobjekt auf weitere Untersuchungsobjekte hinweist, lässt sich dieser Kategorie zuordnen. Die Vorgehensweise beim Treffen einer Auswahl hat massgebli-

eben Einfluss auf die Aussagekraft einer Evaluation (für eine ausführlichere Diskussion siehe Ackoff 1953:123-126).

Es ist daher nötig, das Design vor der Auswahl der Untersuchungseinheiten genau festzulegen und darin zu bestimmen, bei welchen Variablen Konstanz bzw. Varianz gesucht wird. Ausserdem ist es sinnvoll, in einer als Vollerhebung konzipierten Vorstudie ("Survey") einige leicht erhebbare Merkmale aller möglichen Untersuchungseinheiten der Grundgesamtheit zu analysieren. Vielfach ist eine solche Survey-Studie für die - spätere - Interpretation der Ergebnisse und für die genaue Situierung des Untersuchungsgegenstandes ebenso hilfreich wie für die Bereitstellung einer Argumentation gegenüber allfälligen Kritiken einer vergleichenden Evaluation (Argument des unzulässigen Vergleichs).

Zweitens entsteht bei Evaluationen die Frage, inwiefern die Untersuchungsergebnisse auf Gegenstände ausserhalb der empirisch untersuchten Objekte *übertragen* werden können (Generalisierung). Dies hängt wiederum sehr stark davon ab, wie die Stichprobe gewonnen wurde. Gehen wir von einer Wahrscheinlichkeitsauswahl aus, sind Schlüsse auf die nicht in der Stichprobe enthaltenen Elemente der Grundgesamtheit zulässig, sofern die Zahl der untersuchten Objekte ausreichend gross ist. Bei einer gezielten Auswahl ist die Generalisierung der Untersuchungsergebnisse auf theoretischer Ebene zu belegen. Insbesondere muss überzeugend nachgewiesen werden können, dass die zur Stichprobenziehung berücksichtigten theoretischen Dimensionen relevant sind und damit die Stichprobe alle für die Evaluation wichtigen Merkmale der Grundgesamtheit in adäquater Form abbilden.

Generell ist festzuhalten, dass empirische Aussagen nie auf Gegenständen übertragen werden können, die nicht Bestandteil der Grundgesamtheit sind. Generalisierungen dieser Art bedürfen weiterer Abklärungen.

11.1.3.2. Vollerhebung

Bei einer Vollerhebung werden alle Elemente einer definierten Grundgesamtheit in die empirische Untersuchungsanlage miteinbezogen. Grundgesamtheit und Stichprobe fallen damit zusammen. Die beim Treffen einer Auswahl relevanten kritischen Aspekte (Art der Stichprobenziehung und Verallgemeinerung der Resultate) entfallen. Trotzdem werden in der Evaluationsforschung häufig keine oder nur grobe Vollerhebungen (für die Fallauswahl) durchgeführt. Das liegt daran, dass normalerweise eine relativ grosse Zahl von Elementen in der Grundgesamtheit existieren. Der

vollumfängliche Einbezug aller Elemente würde vielfach den Rahmen der Evaluation sprengen; der dafür notwendige Aufwand könnte kaum gerechtfertigt werden.

Dennoch gibt es einige Situationen, bei denen eine Vollerhebung eingesetzt wird. An erster Stelle ist auf den Fall hinzuweisen, bei dem die Grundgesamtheit gerade aus einem Fall besteht. Hier wird gezwungenermaßen eine Vollerhebung durchgeführt. Aber auch in Situationen, in denen die Grundgesamtheit nur aus einigen wenigen Fällen besteht, wird häufig die Vollerhebung gewählt. Je nach Standardisierungsgrad der Erhebungs- und Analyseinstrumente kann auch bei einer grösseren Zahl von Elementen noch eine Vollerhebung ins Auge gefasst werden.

11.1.4. Experimentelle, quasi-experimentelle und nicht-experimentelle Designs

Wie weit ein Design als experimentell bezeichnet wird, hängt von der Vorgehensweise bei der Bestimmung der Vergleichsgruppe und der Art der Verteilung der Untersuchungsobjekte auf die Vergleichsgruppen ab. Hinsichtlich der Hypothese stellt die Bildung der Vergleichsgruppen im Prinzip die Zuordnung der Untersuchungsobjekte zu den Ausprägungen der unabhängigen oder der abhängigen Variable dar. In der Evaluationsforschung werden also beispielsweise zwei Vergleichsgruppen gebildet, wovon die erste aus Individuen besteht, die an dem zu evaluierenden Programm teilnehmen (Versuchsgruppe, "experimental group"), während die zweite Gruppe (Kontrollgruppe, Vergleichsgruppe, "comparison group") dies nicht tut. Aufteilungen dieser Art können vor (ex ante-Design) oder nach der Datenerhebung (ex post facto-Design) erfolgen. Wesentlich ist in allen Fällen, dass sich die beiden Gruppen - bis auf die Betroffenheit durch das Evaluandum oder das Explanandum - möglichst ähnlich sind (vgl. dazu und zum folgenden Campbell und Stanley 1963; Cook und Campbell 1979 und Mohr 1995a).

11.1.4.1. Experimentelle Designs

Experimentelle Untersuchungsanlagen gehen von der Überlegung aus, dass aufgrund einer Manipulation der unabhängigen Variablen - in der Evaluationsforschung also des Evaluandums - der Wirkungsnachweis erbracht werden kann, also kausale Zusammenhänge isoliert werden kön-

nen. Grundsätzlich sind zwei experimentelle Untersuchungsanlagen zu differenzieren: das Laborexperiment und das Feldexperiment. Während sich das Feldexperiment in einem "natürlichen" Kontext abspielt, wird beim Laborexperiment das Umfeld künstlich hergestellt. Experimente gehen davon aus, dass die oben beschriebene Zuordnung der Objekte zu den Untersuchungsgruppen durch eine Zufallsauswahl (sogenannte "Randomisierung") vor der Datenerhebung erfolgt. Diese Bedingung ist jedoch aus ethischen, politischen oder sozialen Gründen sehr oft nicht gegeben. So ist beispielsweise die Vergabe von Subventionen an zufällig ausgewählte Empfänger - während andere Personen in identischen Situationen diese nicht erhalten - nicht zulässig. Aus diesen Gründen sind echte Experimente mit Zufallsauswahl in der Praxis der Evaluation von öffentlichen Politiken nur selten anzutreffen.

11.1.4.2. Quasi-experimentelle Designs

Quasi-Experimente folgen - wie die echten Experimente - grundsätzlich derselben Logik, wobei die Anforderungen an Experimente (ex ante-Design, Randomisierung) nicht vollständig einzuhalten sind. Quasi-Experimente versuchen aber der experimentellen Vorgehensweise möglichst nahe zu kommen. Dazu behilft man sich verschiedener Instrumente, um den Verstoss gegen die experimentellen Anforderungen (zumeist die Randomisierung) teilweise auszugleichen. Häufig ist dabei das sogenannte paarweise Matching. Angestrebt wird hier, dass sich zu einem Untersuchungsobjekt in der Versuchsgruppe jeweils ein hinsichtlich allen relevanten Drittfaktoren möglichst ähnliches Objekt in der Vergleichsgruppe findet. Eine andere Möglichkeit besteht darin, Drittvariablen im Nachhinein durch die Verwendung von multivariaten, statistischen Verfahren zu kontrollieren (vgl. dazu unten Kapitel 12.3.). Diese Instrumente sind nicht in der Lage, die fehlende Randomisierung wettzumachen. Da aber sehr häufig ein echtes Experiment nicht realisiert werden kann, stellen Quasi-Experimente eine valable Alternative dar (Hedrick, Bickman und Rog 1993: 58).

11.1.4.3. Nicht-experimentelle Designs

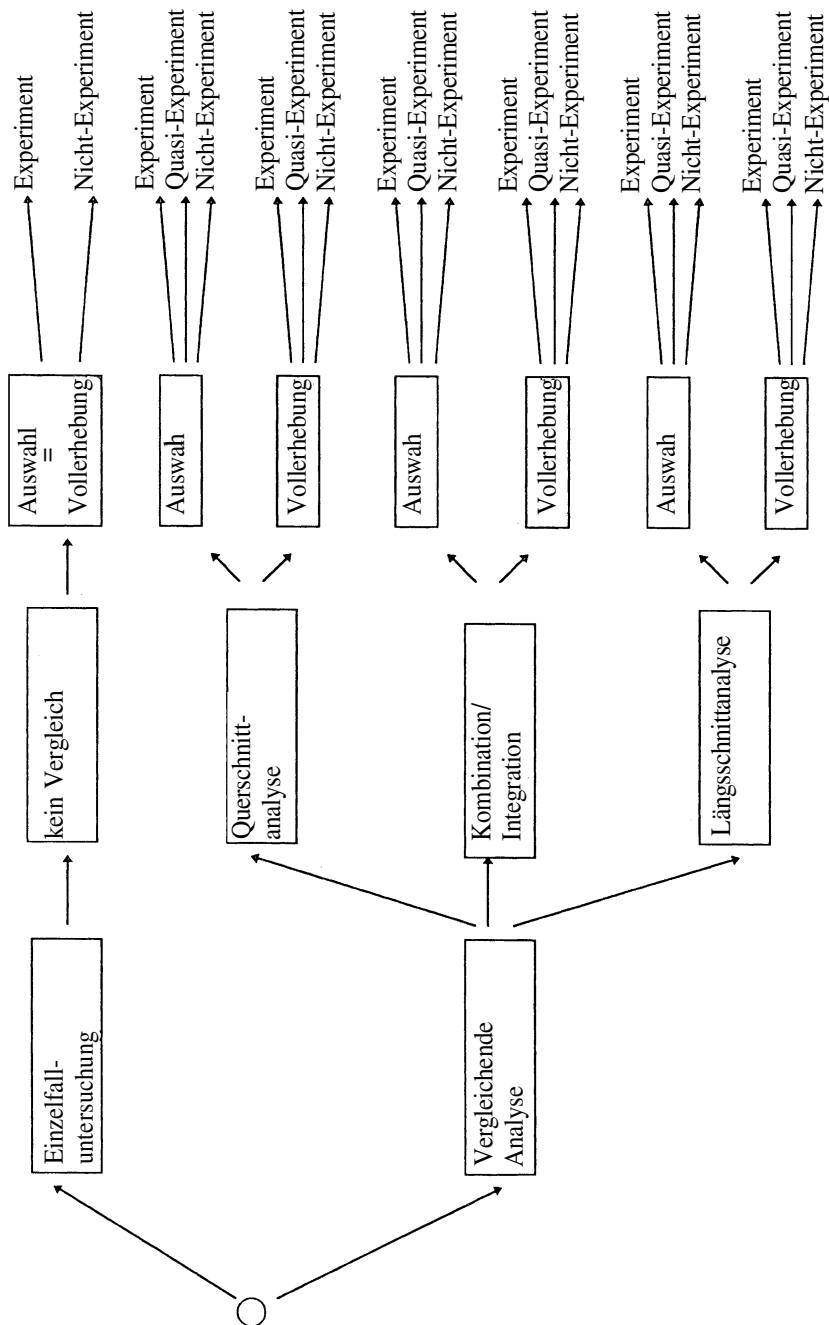
Unter der Bezeichnung der nicht-experimentellen Designs verstehen wir verschiedene Untersuchungsanlagen, die sich nicht auf eine experimentelle Logik stützen. Diese werden teilweise auch als "vorexperimentelle

Designs" bezeichnet, was darauf hinweisen soll, dass diese Untersuchungsanlagen im Rahmen der experimentellen Denkweise als wenig leistungsfähig eingestuft werden. Nicht-experimentelle Designs sind in der Evaluationsforschung relativ weit verbreitet, sind sie doch im allgemeinen die am einfachsten zu implementierende Designform. Beispiele sind etwa Evaluationen, die auf einer "One-shot case study" basieren, oder die klassischen ex post facto-Versuchsanlagen. Während Vertreter der experimentellen Schule diese als nichtaussagekräftig zurückweisen, wird von Seiten anderer Methodiker auf die spezifische Leistungsfähigkeit dieser Untersuchungsanlagen verwiesen. Insbesondere werden dabei die Möglichkeiten zur Hypothesengenerierung (im Gegensatz zur Absicht der Hypothesenüberprüfung bei experimentellen und quasi-experimentellen Designs), aber auch die Einfachheit der Untersuchungsanlage hervorgehoben. Die andere Stossrichtung der nicht-experimentellen Untersuchungsanlagen kommt auch darin zum Ausdruck, dass sie auch als deskriptive Designs bezeichnet werden (siehe etwa Hedrick, Bickman und Rog 1993: 44-51).

11.1.5. Entscheidungsbaum zur Ausgestaltung eines Evaluationsdesigns

Die vier hier diskutierten Dimensionen zur Definition eines Untersuchungsdesigns lassen sich recht weitgehend miteinander kombinieren, auch wenn in der Praxis gewisse Kombinationen sehr häufig oder nur sehr selten eingesetzt werden. Die nachfolgende Graphik soll dazu dienen, die Kombinationsmöglichkeiten in den unterschiedlichen Dimensionen zu veranschaulichen. Dabei haben wir die Darstellungsform eines Entscheidungsbaums gewählt, obwohl dies den falschen Eindruck erwecken könnte, dass diese Entscheide nacheinander in einer bestimmten Reihenfolge zu treffen wären. Dies ist aber in keiner Weise zutreffend. Die Definition eines Untersuchungsdesigns ist ein iterativer Prozess, bei dem keine bestimmte Reihenfolge einzuhalten ist und bei dem auf getroffene Entscheide (zumindest solange sich die Evaluation noch in der Planungsphase befindet) immer wieder zurückgekommen werden kann und oft auch soll.

Abbildung 4: Entscheidungsbaum zur Definition eines Untersuchungsdesigns



Aus der Darstellung wird deutlich, dass bestimmte Kombinationen der unterschiedlichen Dimensionen nicht denkbar sind. Dies gilt besonders im Bereich der Einzelfalluntersuchungen, bei denen Auswahl und Vollerhebung identisch sind und keine quasi-experimentellen Untersuchungsanlagen bekannt sind.

Im nachfolgenden Kapitel 11.2. werden nun die hier diskutierten Designformen anhand von typischen Kombinationsmöglichkeiten vertieft behandelt und mit Beispielen aus der Evaluationspraxis erläutert. Ebenso wird in diesem Abschnitt zu diskutieren sein, welche Designtypen in der Evaluationspraxis besonders häufig respektive besonders selten eingesetzt werden und welche Gründe für diese Differenzen verantwortlich sind.

11.2. HÄUFIG VERWENDETE UNTERSUCHUNGSDESIGNS

In der Evaluationspraxis ist es nicht immer möglich, eine in jeder Hinsicht optimale Untersuchungsanlage zu wählen. Aus Gründen der sicheren kausalen Zuordnung von Wirkungen zu bestimmten Programmen wäre es zwar z.B. wünschbar, systematisch angelegte Experimente mit einer Vollerhebung über längere Zeiträume durchzuführen. Aus forschungsökonomischen Erwägungen, weil für frühere Zeitpunkte keine Daten vorliegen oder erhoben werden können und wegen des Erfordernisses, eine Evaluationsstudie rasch abschliessen zu müssen, sehen sich Evaluationsteams indessen häufig veranlasst, auf aufwendige Designs zu verzichten und sich auf Momentaufnahmen eines Einzelfalles zu beschränken.

In diesem Optimierungskalkül, dem letztlich alle Evaluationen unterworfen sind, schälen sich immer wieder ähnliche *Typen* von Evaluationsdesigns als gangbare Lösungen heraus. Einige dieser häufig verwendeten Designs sollen in der Folge umschrieben, anhand von Beispielen erläutert und auf ihre Vor- und Nachteile hin untersucht werden. Dabei sollen folgende Typen zur Sprache kommen:

- Die Einzelfallstudie
- Der Fallstudienvergleich
- Der quasi-experimentelle Vorher/Nachher-Vergleich
- Die Zeitreihenanalyse
- Die Kombination von Design-Ansätzen.

Der Beschreibung der genannten Evaluationstypen ist vorauszuschicken, dass sich die einzelnen häufig verwendeten Designs nie in völlig reiner Form auf die in Grafik 1 unterschiedenen Dimensionen umlegen lassen. Vielmehr handelt es sich in der Praxis oft um Mischformen, die z.B. einen Querschnitt mit einem Längsschnitt verbinden, die nur teilweise experimentellen Charakter haben und die für gewisse Aspekte der Evaluation eine Vollerhebung wählen, für andere Aspekte indessen lediglich eine Auswahl der Untersuchungsobjekte analysieren. Diese Mischung von Elementen in einzelnen Untersuchungsdesigns wird nicht immer bewusst vorgenommen und schon gar nicht immer explizit gemacht. Dass sie aber vorkommt, braucht nicht weiter zu verwundern. Nicht selten wird nämlich ein gemischtes Untersuchungsdesign durch die Sachlage, durch Wünsche der auftraggebenden Stelle, durch begrenzte Mittel für die Evaluation und durch die Daten- und Quellenlage weitgehend vorgegeben.

11.2.1. Die Einzelfallstudie

Bei der Einzelfallstudie gilt der gesamte Untersuchungsbereich als einziger Fall. Dieser wird gesamtheitlich betrachtet. Der Fall, der für die Evaluation besonders interessante Aspekte aufweisen sollte, wird auf möglichst viele Dimensionen hin beobachtet, beschrieben und untersucht. Damit soll verhindert werden, dass ein Untersuchungsobjekt auf einige wenige Variablen reduziert wird. Dies würde dem Untersuchungsobjekt nicht gerecht, weil das Herausgreifen einzelner Merkmale einigermaßen willkürlich erscheint, die Komplexität des Gegenstandes verkürzt und die Individualität und Identität des zu Untersuchenden verletzt werden könnte. Tatsächlich geht es darum, ein ganzheitliches und damit realistisches Bild der sozialen Welt zu zeichnen. Mithin sind möglichst alle für das Untersuchungsobjekt relevanten Dimensionen in die Analyse einzubeziehen (Lamnek 1989: 5).

Mit dieser Umschreibung ist angedeutet, dass Einzelfallstudien häufig nicht von einem empirisch-analytischen, sondern von einem hermeneutischen Forschungsparadigma ausgehen (vgl. 12.1.). "Die Einzelfallstudie im qualitativen Paradigma strebt eine wissenschaftliche Rekonstruktion von Handlungsmustern auf der Grundlage von alltagsweltlichen, realen Handlungsfiguren an. Dabei versucht der Forscher nicht nur als alltagsweltlicher Handlungspartner die Figuren nachzuvollziehen, sondern diese in den wissenschaftlichen Diskurs zu überführen und Handlungsmuster zu

identifizieren, indem er allgemeinere Regelmässigkeiten vermutet" (Lamnek 1989: 16).

Die Einzelfallstudie verzichtet im Prinzip auf irgendeinen Vergleich. Sie orientiert sich weder an früheren Situationen (Längsschnitt), an anderen Fällen (Querschnitt) noch an einer vorgegebenen oder von den Forschern selbst entwickelten Sollgrösse. Wie bereits erwähnt (Kapitel 11.1.1.), ziehen freilich auch Einzelfallstudien oft implizit Vergleiche. So machen sie etwa die Aussage, auch nach der Einführung eines Arbeitsbeschaffungsprogramms seien in einer Region viele Arbeitslose anzutreffen. Sie meinen damit mehr Arbeitslose als früher, mehr als erwartet (Soll-Grosse) oder mehr als in anderen Regionen.

In der grossen Mehrheit arbeiten Einzelfallstudien mit qualitativen Forschungstechniken (vgl. 12.2.). Sie bevorzugen offene oder wenig strukturierte Interviews mit wenigen Betroffenen oder Fachleuten, interpretieren Dokumente und verzichten weitgehend auf quantitative Indikatoren. Angesichts ihrer methodischen Ausrichtung erhalten Einzelfallstudien häufig einen stark deskriptiven Anstrich.

Einzelfallstudien können allerdings durchaus grossen analytischen Gehalt haben. In welchem Ausmass dies der Fall ist, hängt stark von der Auswahl des untersuchten Falles ab. In einer Anleitung des GAO (1990) zur Durchführung von Fallstudien nimmt dieser Arbeitsschritt deshalb breiten Raum ein. Als Oberkriterien werden Bequemlichkeit, Zufallsauswahl und erwarteter Nutzen unterschieden. Das Bequemlichkeitskriterium kommt dann zum Zug, wenn jener Fall untersucht werden soll, über den am leichtesten Informationen erhältlich sind. Diesem pragmatischen Vorgehen steht die reine Zufallsauswahl gegenüber.

Nach unserer Auffassung ist allerdings die Auswahl des Falles streng an den Zielen der Evaluation zu orientieren. Das GAO unterscheidet dabei sieben verschiedene Selektionskriterien. Erstens ist es möglich, ein *Extrem* auszuwählen, um dadurch Hinweise auf die Ursachen für grosse Varianzen zu bekommen. Als zweites sind *Musterbeispiele* denkbar. Sie legen dar, was zum Erfolg in einzelnen Fällen beigetragen hat. Das Gegenbeispiel stellt der *kritische Fall* dar. Er weist auf spezifische Probleme hin. Weiter ist es möglich, einen *typischen Fall* auszuwählen. Er gibt Aufschluss darüber, wie die Wirkungskette im "Normalfall" zusammenhängt. Durch die Beschreibung von Fällen, welche ein spezielles Cluster vertreten, lassen sich verschiedene Programmtypen vergleichen. Ein *illustrativer Fall* kann dazu dienen, Probleme oder Chancen von Programmen besonders eindrücklich darzulegen. Als letzten Typ bezeichnet das GAO

Spezialfälle. Dabei wird ein herausragendes Ereignis untersucht. Dabei kann es sich beispielsweise um einen Fall handeln, der besonders viel Aufsehen erregt hat. Für eine solche Situierung des Falles sind, wie erwähnt (Kapitel 11.1.1.), vorgängige Survey-Studien hilfreich.

Ein gutes *Beispiel* für eine Einzelfallstudie ist die Arbeit von Michel Rey und seiner Equipe (Rey et al., 1993) zur *Wirtschaftsförderungspolitik des Kantons Wallis*. Die Autoren gehen vom Anreizcharakter dieses politischen Programms aus und leiten daraus ab, dass die Wirtschaftsförderung nur dann eine Wirkung entfalten könne, wenn sie eine Veränderung des Verhaltens der Adressaten auslöste. Dazu müssten die Adressaten die Politik nicht nur kennen und verstehen, sondern sie müssten auch eine positive Meinung darüber haben.

Die Evaluatoren führten deshalb mit rund 60 Walliser Akteuren Gespräche über verschiedene Fragen, die von einer zwölfköpfigen Evaluationsinstanz als problematisch bezeichnet worden waren. Aus den dabei gewonnenen qualitativen Aussagen haben die mit der Evaluation Beauftragten, die sich selbst als Experten bezeichnen, Schlussfolgerungen gezogen und Empfehlungen zuhanden der für die Wirtschaftsförderungspolitik Verantwortlichen gezogen. Angesichts der umstrittenen und enttäuschenden Ergebnisse sind im folgenden zwei Arbeitsgruppen eingesetzt worden, die in einem iterativen Prozess neue Ziele, Strategien, Vollzugs- und Koordinationsverfahren zu entwickeln hatten. Dieser Prozess der Weiterentwicklung der Politik sei fortzusetzen, schliessen die Autoren.

Das Beispiel zeigt die Vor- und Nachteile von derart konzipierten Einzelfallstudien auf. Als Vorteil ist der Umstand zu werten, dass ohne Vorurteile die verschiedenen möglichen Gründe für den (Miss-)Erfolg der Politik ermittelt werden konnten. Das Vorgehen bot auch die Möglichkeit, flexibel auf neue Situationen zu reagieren, Erhebungsdimensionen zu verändern und eine zweite Phase anzuschliessen. Das Untersuchungsdesign erwies sich auch deshalb als angebracht, weil die Wirkungszusammenhänge noch wenig bekannt waren und explorativ neue Zusammenhänge ermittelt werden konnten.

Umgekehrt ist das Vorgehen auch mit beträchtlichen Risiken verbunden. Die Verlässlichkeit der aufgezeigten Zusammenhänge bleibt teilweise ungesichert. Die Qualität der Resultate hängt stark von der sozialwissenschaftlichen (und sozialen!) Kompetenz des Evaluationsteams ab. Die Aussagen der Adressaten garantieren nicht, dass bei einer Änderung der Politik eine grössere Wirksamkeit zu erwarten ist. Eine Übertragbarkeit auf andere Fälle ist genau so wenig möglich, wie in anderen Kantonen mit

der entsprechenden Politik gewonnene Erfahrungen für den Untersuchungsfall genutzt werden können. Das Vorgehen ist somit aufwendig und in seinem Erfolg unsicher.

11.2.2. Vergleichende Fallstudien

Mit dem Vergleich mehrerer (wenigstens zweier) Fallstudien sollen die wichtigsten Nachteile der Einzelfallstudie behoben werden. Das Vorgehen bei der Erstellung der einzelnen Fallstudien ist in grossen Zügen dasselbe wie bei einer Einzelfallstudie. Im Hinblick auf die im Kapitel 11.1. angeführten Dimensionen ist zudem zu sagen, dass es sich auch bei den vergleichenden Fallstudien nicht um eine Art Experiment handelt und dass keine Vollerhebung angestrebt wird. Es geht im übrigen fast immer um Querschnittanalysen. Es ist allerdings auch hier zu betonen, dass die genannte Kombination von Merkmalen in den vier Design-Dimensionen keineswegs zwingend, wohl aber typisch ist und dass das Design der im Querschnitt vergleichenden Fallstudien auch nicht immer in seiner reinen Form zur Anwendung gelangt.

Solche vergleichend angelegte Fallstudien sind wohl das in der schweizerischen Evaluationsforschung am häufigsten verwandte Design. Es ist deshalb nicht ganz einfach, aus der Fülle der vorhandenen *Beispiele* das geeignetste auszuwählen. Solche Studien aus jüngster Zeit finden sich bei Terribilini (1995)¹, Gerheuser und Schmid (1993)², Knoepfel, Imhof und Zimmermann (1995)³, Knoepfel, Kissling-Näf und Marek (1997)⁴ oder Klöti, Haldemann und Schenkel (1993)⁵. Angesichts ihres hermeneutisch und qualitativ orientierten Vorgehens haben wir uns für die Arbeit von Willy Bierter und Hans-Martin Binder (1993) zur *wirtschaftlichen Innovationsförderung* entschieden. Die beiden Autoren gehen in dieser Untersuchung der Frage nach, wie weit staatliche Förderungsleistungen für den unternehmerischen Innovationsprozess wirksam sind. Dabei analysieren sie die Wirkungen der Instrumente des Bundes und der Kantone Solothurn und St. Gallen. Diese Auswahl der Fälle wird damit begründet, dass sie

¹ Verkehrsberuhigungsmassnahmen in ein einigen Schweizer Städten.

² Lohngleichheitspostulat in der Heimarbeit in verschiedene Kantone.

³ Massnahmen verkehrsbezogener Umweltpolitik.

⁴ Evaluation von Lernprozessen aus der Umwelt- und Gesundheitspolitik anhand von 28 Fallstudien.

⁵ Vergleich der Umwelt- und Verkehrspolitik einiger schweizerischer Grossstädte.

Varianz ermögliche, und zwar "sowohl auf der Ebene unterschiedlicher "Philosophien" von Wirtschafts- und Innovationsförderungs politik hinsichtlich Zielperspektiven, Strategien und Mitteleinsatz, als auch auf der Ebene unterschiedlicher Struktur-Merkmale (Wirtschaftsstruktur, regionales Umfeld usw.)" (Bierter und Binder 1993: 17). Die Studie stützt sich auf rund 40 Gespräche mit Vertretern von Bundesämtern, kantonalen Verwaltungen, Banken, betroffenen Unternehmungen, Verbänden sowie Forschungs- und Entwicklungsinstitutionen. Einer Runde von qualitativen Interviews folgten in einer zweiten Phase Gruppengespräche zur kritischen Überprüfung der ersten Befunde. Damit wird auch die Zeitdimension in die Untersuchung eingeführt, werden doch die in der späteren Phase von den früheren Aussagen abweichenden Haltungen der Akteure wenigstens teilweise auf die veränderten konjunkturellen und politischen Rahmenbedingungen zurückgeführt.

Das Vorgehen hat den Vorteil, dass es eine tiefe Durchdringung der drei Fälle erlaubt. Bei der Auswertung und bei den Schlussfolgerungen sind die Möglichkeiten, die der Vergleich bot, freilich nicht voll genutzt worden, da in den drei Fällen kein systematischer Zusammenhang zwischen den unterschiedlichen Ausgangslagen und den Wirkungen hergestellt wurde. Hingegen können die Autoren zu Recht darauf hinweisen, dass die für alle drei Fälle geltenden Schlussfolgerungen besser abgestützt werden konnten, als wenn lediglich ein Fall in die Untersuchung einbezogen worden wäre. Damit sind auch die wesentlichen Vorteile genannt, die vergleichende Fallstudien gegenüber der Einzelfallstudie bieten. Sie bieten die Chance, unterschiedliche Wirkungszusammenhänge aufzudecken, und sie sichern die Befunde, die sich in allen untersuchten Fällen erhärten lassen, besser ab. Damit soll aber keinen Verallgemeinerungen über die untersuchten Fälle hinaus das Wort geredet werden. Fallstudien haben ihre Gültigkeit nur für die untersuchten Fälle, auch wenn sie vergleichend angelegt sind.

Damit soll aus einer empirisch-analytischen Position indessen nicht ausgeschlossen werden, dass mit Fallstudien keine Hypothesen überprüft werden können. In dieser Forschungstradition wird unterschieden zwischen Fallstudien, die eher der Hypothesenbildung dienen, und solchen, welche in erster Linie die Hypothesenprüfung vor Augen haben (vgl. u.a. Eckstein 1975). *Hypothesenbildende Fallstudien* haben die Aufgabe, komplexe Zusammenhänge aufzuhellen und so zur Entwicklung empirisch begründbarer theoretischer Konzepte, Theorien, Hypothesen beizutragen (Lamnek 1989: 11). Sie dienen somit dem Kennenlernen und der Interpre-

tation von Problemzusammenhängen und der Formulierung von empirisch relevanten Hypothesen. Diese lassen sich in der nachfolgenden Hauptuntersuchung dann testen. Bei hypothesenbildenden Fallstudien werden meist typische Fälle untersucht, d.h. solche, welche den "Normalfall" repräsentieren. *Hypothesenprüfende Fallstudien* dagegen analysieren eher kritische Fälle. Damit sollen Hypothesen bestätigt, relativiert oder verworfen werden. Ausgangspunkt bilden Annahmen, welche das Eintreten eines bestimmten Ereignisses gewissermassen zwingend voraussagen. Zeigt eine Fallstudie nun ein Ausbleiben des unter den angegebenen Ausgangsbedingungen prognostizierten Ereignisses auf, so ist eine Modifikation oder eine Verwerfung der verwendeten Hypothese notwendig. Andernfalls gilt die Hypothese als vorläufig bestätigt. Ein Beispiel für hypothesentestende vergleichende Fallstudien aus der schweizerischen Evaluationsforschung findet sich bei Knoepfel, Imhof und Zimmermann (1995) im Rahmen der Evaluation der lufthygienischen Wirkungen der Massnahmenpläne.

11.2.3. Der quasi-experimentelle Vorher/Nachher-Vergleich

Vom Design der hypothesentestenden vergleichenden Fallstudien zum hier folgenden quasi-experimentellen Design des Vergleichs von Situationen "vorher" und "nachher" ist es nur noch ein kleiner Schritt. Das letzte Design geht von der Annahme aus, dass sich bei einem Fall, der einer staatlichen Massnahme ausgesetzt ist, verschiedene Merkmale anders entwickeln als bei dem anderen Fall, der nicht mit derselben Massnahme konfrontiert ist. Die Untersuchungsanlage ist denn auch relativ einfach. Man wähle mindestens zwei Fälle aus, die sich möglichst nur dadurch unterscheiden, dass im einen Fall ein Programm entwickelt und umgesetzt wird, im anderen dagegen nicht. Daraufhin vergleicht man in beiden Fällen die Situation vor der Einführung der Massnahme mit der Situation nachher. Diese Vorgehensweise geht aus der Abbildung 5 hervor.

Abbildung 5: Vorher/Nachher-Vergleich

Fälle (Kanton A und B)	Situation (Arbeitslosenrate)	
	vorher	nachher
Fall A (mit Massnahme)	5%	4%
Fall B (ohne Massnahme)	5%	5%

Hat sich die Situation im Fall A mit Massnahme verbessert, im Fall B ohne Massnahme dagegen nicht, dann gilt die Massnahme - im erwünschten Sinne - als wirksam. Ist also etwa im Kanton A aufgrund eines Arbeitsbeschaffungsprogramms die Arbeitslosigkeit von 5 auf 4 Prozent gesunken, im Kanton B ohne Arbeitsbeschaffungsprogramm dagegen bei 5 Prozent geblieben, so kann vorerst einmal vermutet werden, dass das Programm im Kanton A erfolgreich war.

Selbstverständlich liegen die Dinge in der Praxis nie so einfach. Die wichtigsten Schwierigkeiten dieses Designs sind die folgenden:

- Erstens ist die Ausgangslage kaum je in den beiden Fällen die gleiche. So ist es wahrscheinlich, dass z.B. Kantone mit einer hohen Arbeitslosenrate eher ein Arbeitsbeschaffungsprogramm einführen als andere. Die Ausgangslage im Kanton A ist deshalb vermutlich schlechter als jene im Kanton B.
- Zweitens dürfte sich in der Regel in beiden Fällen die Situation verändern, wenn auch nicht in gleichem Ausmass. Die Arbeitslosenrate kann im Kanton A von 5 auf 4 Prozent, im Kanton B von 4,2 auf 3,6 Prozent sinken. Die Bestimmung des Ausmasses der Wirkung einer Massnahme wird so deutlich schwieriger.
- Drittens bleibt unklar, ob die Verbesserung der Situation im Kanton A etwas mit der getroffenen Massnahme zu tun hat. Der Vergleich der Situationen vorher und nachher gibt allein noch keinen Hinweis auf die Kausalität.

Um diese Schwierigkeiten etwas zu mindern, kann die Strategie befolgt werden, mehrere Fälle mit und ohne Programm einzubeziehen. Wenn alle

Fälle mit Programm eine Verbesserung der Situation in der Zieldimension erreichen, nicht aber die Fälle ohne Programm, dann besteht etwas bessere Gewähr für die Wirksamkeit der Massnahme. Allerdings handelt sich der Forscher oder die Forscherin damit ein neues Problem ein. Es ist nämlich unwahrscheinlich, dass in allen Fällen das Programm zum gleichen Zeitpunkt eingeführt wird und dass es überall gleich ausgestattet ist. Damit werden zwingend unterschiedliche Vorher/Nachher-Situationen miteinander verglichen. Eine saubere und eindeutige Untersuchungsanlage wäre nur bei einem reinen Experiment möglich. Dieses ist aber im Rahmen öffentlicher Politiken meist nicht durchführbar (vgl. Abschnitt 11.1.4.). Eine Verbesserung des Evaluationsdesigns kann zwar mit dem Einbezug einer grösseren Zahl von Fällen erreicht werden, weil so gewisse Variablen, die das Bild verfälschen, besser kontrolliert werden können. Diesem Vorgehen sind aber meist natürliche Grenzen gesetzt. So lässt sich etwa die Zahl der Kantone nicht vermehren.

Als *Beispiel* dient deshalb eine Evaluation, bei der die Fallzahl nicht a priori beschränkt war. In einer Vollzugs- und Wirkungsevaluation der Instrumente der *eidgenössischen Wohneigentumsförderung* haben Schulz, Muggli und Hübschle (1993) nicht nur das Verhalten der Vollzugsinstanzen und der Adressaten auf die Übereinstimmung mit den beschlossenen Massnahmen überprüft (Vollzugskontrolle), sondern auch untersucht, ob das Hauptziel der Wohneigentumsförderung, nämlich die Ermöglichung des Erwerbs von Grundeigentum durch Unterstützung bei der Finanzierung in der Form von Bürgschaften, erreicht worden ist (Wirksamkeitsprüfung). Zu diesem Zweck wurden verschiedene Untersuchungsgruppen gebildet und befragt. Die zentrale Gruppe bildeten die neuen Eigentümer, die eine Förderung erfuhren (N=188). Diesen wurden drei Kontrollgruppen gegenübergestellt. Der ersten Gruppe gehören jene an, die seit 1976 Wohneigentümer geworden waren, ohne von der Wohneigentumsförderung zu profitieren (N=262). Die zweite Gruppe wurde von Mietern ohne realisierte Kaufabsichten gebildet (N=385). Einer dritten, gemischten Gruppe gehörten Personen an, die ein Gesuch um Wohneigentumsförderung zurückgezogen hatten oder deren Gesuch abgelehnt worden war (N=83). Die Analyse anhand von Interviews zeigt, dass sich die Gruppe der Geförderten in mehrfacher Hinsicht signifikant von den anderen unterscheidet. Dennoch rechnen die Autoren damit, dass ein Drittel bis zur Hälfte der Geförderten zu den "Mitnehmern" gehörten, die die Immobilie auch ohne Wohneigentumsförderung gekauft hätten. Da zudem kaum je mehr als 10 Prozent des neu gebauten Wohneigentums der Förde-

rung unterstand, darf auf Grund des Kontrollgruppenvergleichs angenommen werden, dass die Förderung des Wohneigentums ihr Ziel nur in einem sehr beschränkten Masse erreichte.

Das Beispiel zeigt eine zusätzliche Problematik des Vorher/Nachher-Vergleichs auf. Wenn die Situation im Zeitpunkt vor der Intervention nicht systematisch erfasst worden ist, wird es im Nachhinein äusserst schwierig, sie zu rekonstruieren. Hier wurde dies teilweise mit einer Befragung der Betroffenen versucht. Diesem Vorgehen sind freilich Grenzen gesetzt. Denn die Befragten können sich in der Regel nur schlecht an die frühere Lage erinnern. Ferner neigen sie dazu, diese Situation zu verklären, sei es in positiver oder negativer Richtung. Quasi-experimentelle Vorher/Nachher-Vergleiche eignen sich deshalb nur dann, wenn sich die Evaluation auf unproblematische Reihen von Aggregatdaten stützen kann oder wenn das Design vor der Durchführung der Massnahme geplant werden kann. Das ist etwa bei Schulversuchen (Einführung der Fünftagewoche, integrierte Oberstufe) der Fall. Eine solche Strategie wurde auch bei der Einführung von Tempo 50 innerorts in Testgemeinden gewählt (Arbeitsgruppe Verkehrssicherheit 1983). Rückwirkend ist es schwierig, das Vorher/Nachher-Design in reiner Form zu realisieren.

11.2.4. Die Zeitreihenanalyse

Die höchsten Anforderungen an die Datenqualität stellt die Zeitreihenanalyse. Hier geht man davon aus, dass sich die Wirksamkeit einer Massnahme daran zeigt, dass sich der Verlauf einer Reihe von Beobachtungen nach einer Intervention - über die üblichen zufälligen Schwankungen hinaus - gegenüber vorher verändert. Um trendbedingte sowie zufällige Entwicklungen in den Griff zu bekommen, beschränken sich entsprechende Untersuchungen nicht nur auf einen Vergleich der beiden Situationen vorher/nachher an zwei Stichdaten, sondern sie beobachten die Entwicklung kontinuierlicher während eines möglichst langen Zeitraums.

Soll die Zeitreihe mit statistischen Methoden analysiert werden, so sind je nach Ausweitungsverfahren mindestens 20 - 50 Beobachtungen erforderlich.

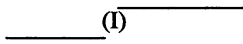
Schwierigkeiten bietet die Zeitreihenanalyse nicht zuletzt deshalb, weil in der Praxis die politisch-administrativen Interventionen nicht einmalig erfolgen. Vielmehr finden in der Regel mehrere relevante Interventionen

in nicht weit auseinanderliegenden Abständen statt. Es handelt sich um Prozesse mit sich überlagernden Einflüssen. Hinzu kommen zumeist exogene Faktoren, welche eine Zeitreihe ebenfalls beeinflussen. Die verschiedenen Wirkungen auf eine zeitliche Datenreihe analytisch auseinanderzuhalten, stellt höchste Anforderungen an die Qualität der Daten einerseits und an die Fähigkeiten der Forscherin oder des Forschers andererseits. Vielfach bleibt im Evaluationsalltag nur der pragmatische Weg, chronologische Entwicklungsverläufe einzelner Variablen in grafischen Darstellungen abzubilden und diese zu interpretieren.

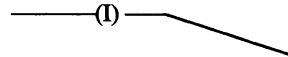
Das Untersuchungsdesign der Zeitreihe ist auch deshalb schwer zu handhaben, weil sehr verschiedene Wirkungsverläufe einer Intervention möglich sind und es nicht einfach ist, diese in klaren Hypothesen zu antizipieren. So kann eine Wirkung abrupt oder verzögert eintreten. Es kann auf der Zieldimension eine Verschiebung des Niveaus oder eine Veränderung der Richtung geben. Die Veränderung kann andauernd oder nur vorübergehend sein. Daraus lassen sich verschiedene Wirkungsverläufe kombinieren, wie sie im folgenden in Anlehnung an Glass, Willson und Gottman (1975) zusammengestellt werden.

Abbildung 6 Zeitpfade einer Wirkung (nach Glass, Willson und Gottman 1975: 44)

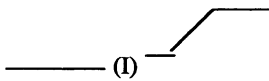
A. Abrupter Niveauwechsel



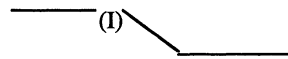
F. Verzögerter Richtungswechsel



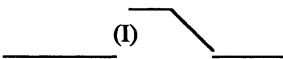
B. Verzögerter Niveauwechsel



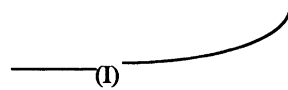
G. Vorübergehender Richtungswechsel



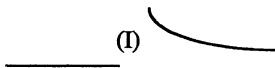
C. Vorübergehender Niveauwechsel



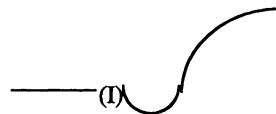
H. Beschleunigter Richtungswechsel



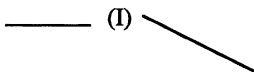
D. Verschwindender Niveauwechsel



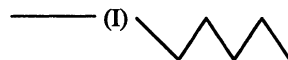
I. Entwicklungseffekt



E. Abrupter Richtungswechsel



J. Veränderung der Variabilität



Ein *Beispiel* für eine statistisch gestützte Zeitreihenanalyse ist die Evaluation verschiedener Massnahmen der Luftreinhaltepolitik in der Schweiz von Thomas Widmer (1991). In dieser Studie versucht der Autor, den Wirkungsverlauf von mehreren Massnahmen zur Reduktion der Luftbelastung durch Schwefeldioxid anhand einer Interventionsanalyse nach dem Ansatz von Box und Tiao (1975) zu erfassen und zu interpretieren. Die Interventionsanalyse ergab, dass weder die Richtlinien aus den Jahren 1972 und 1983 noch verschiedene private Normen einen signifikanten Einfluss auf den Schwefelgehalt von Heizölen und Diesel-Treibstoff geschweige denn auf die Schwefeldioxid-Immissionen hatten. Einzig von der Luftreinhalte-Verordnung von 1985 ging eine Wirkung auf diese Zieldimensionen aus.

11.2.5. Kombination und Triangulation

Angesichts der Schwierigkeiten, die entstehen, wenn mit einem einzigen Evaluationsdesign die Wirkungen einer Massnahme präzise erfasst werden sollen, sind in den letzten Jahren wieder vermehrte Anstrengungen unternommen worden, mit Hilfe des kombinierten Einsatzes verschiedener Designs die Gültigkeit der Aussagen von Evaluationsstudien zu erhöhen. In diesem Zusammenhang hat auch die sogenannte *Triangulation* neue Aktualität gewonnen. Der Begriff stammt ursprünglich aus dem Bereich der militärischen Navigation und bezeichnet eine Strategie, um, ausgehend von verschiedenen Referenzpunkten, die exakte Position eines Objekts festzustellen. In die Sozialwissenschaften wurde der Gedanke von Campbell und Fiske (1959) eingeführt. Deren Idee war es, eine Hypothese mit einer Serie komplementärer Testmethoden zu überprüfen. Dabei sollte der Forscher eine Vermutung auf der Basis der ersten Methode durch Befunde einer zweiten Methode absichern. Zudem wurde die Chance in Betracht gezogen, dass bei der Anwendung der zweiten Technik auf überraschende Erkenntnisse gestossen werden könnte. Durch Methodentriangulation sollten weiter in den einzelnen Methoden steckende Fehler entdeckt werden. Es erschien als äusserst unwahrscheinlich, dass sich mit mehreren Techniken dieselben falschen Resultate ergeben würden.

In den siebziger Jahren wurde das Prinzip der Triangulation von Denzin (1989) präzisiert und ausdifferenziert. Er unterschied fünf Typen von Design-Triangulation:

Als "*Daten-Triangulation*" bezeichnete er den Einbezug verschiedener Datenquellen in eine Untersuchung. Unter "*Investigator-Triangulation*" verstand er den Einbezug unterschiedlicher Beobachter oder Forscher, um aktorbedingte Verzerrungen der Ergebnisse zu vermeiden. Die Kombination verschiedener Erhebungstechniken benannte er "*methodologische Triangulation*". Den vierten Typ bezeichnete er als "*Theorien-Triangulation*". Hierbei werden die Daten mit verschiedenen Theorien konfrontiert und so die Erklärungskraft der einzelnen Ansätze geprüft (vgl. u.a. Flick 1992). Als fünften Typus schliesslich fügte er die sogenannte *multiple Triangulation* hinzu. Dieser Typus beinhaltet die Idee, dass eine Triangulation der Datenquellen, der Forschungsteams, der Methoden und der Theorien gleichzeitig vorgenommen wird. Der Ansatz ist in höchstem Masse anspruchsvoll, verlangt eine eigentliche übergeordnete Forschungsstrategie und ist derart aufwendig, dass er für eine Evaluation nur ausnahmsweise in Frage kommen dürfte.

Auf andere Typen der Kombination von verschiedenen Forschungsansätzen kann die Evaluationsforschung hingegen sehr wohl zurückgreifen. Besonders Michael Patton (1990) hat auf ihre Chancen hingewiesen. Eine besondere Form der Triangulation kann auch in der *Kombination der Vergleichsebenen* liegen. Balthasar und Knöpfel (1994) bezeichnen sie als "*konzeptionelle Triangulation*". Sie besteht darin, dass der gleiche Gegenstand, d.h. dasselbe Evaluandum mit verschiedenen Analysekonzepten angegangen wird. Dieses Vorgehen hat zwei Vorteile: Einmal können aus den verschiedenen Teiluntersuchungen Ergebnisse hervorgehen, welche als Fragestellungen, als Datengrundlage oder als Randbedingung in andere Teile einfließen können. Dann ist zu erwarten, dass - ähnlich wie in der Fotografie - auch in der Evaluation zusätzliche "Lampen" einen Gegenstand besser auszuleuchten und unerwünschte Schatten zu beseitigen vermögen (vgl. dazu auch Balthasar und Knöpfel 1994: 166). Durch unterschiedliche Zugangsweisen wird die Abhängigkeit der Ergebnisse von einzelnen Analysekonzepten relativiert.

Siegfried Lamnek (1989) hebt als Vertreter eines alternativen Ansatzes allerdings hervor, dass Triangulationen keine "Hauptwirklichkeit" erkennen lassen, sondern zu komplementären Bildern der Realität führen. Triangulation sollte demnach in dieser Sicht weniger der Konstruktion einer "objektiven" Wahrheit dienen, als bewusst die Breite und die Tiefe der Analyse erweitern. Es sind von den einzelnen Untersuchungsteilen nicht deckungsgleiche, sondern komplementäre Ergebnisse zu erwarten, welche

sich ineinander fügen und sich ergänzen können, aber nicht kongruent sein müssen (Lamnek 1988: 236).

Als *Beispiel* für eine Studie mit mehrfachem Zugang dient uns die Evaluation von Balthasar und Knoepfel (1994). Sie befasst sich mit der Analyse der *innovationsfördernden Auswirkungen* eines Programms der schweizerischen Umweltpolitik. Drei *luftreinhaltepolitische Massnahmen* im Bereich der Hausfeuerungen werden daraufhin untersucht, ob sie zu technischen Innovationen führen, welche die geforderte Abgasqualität ermöglichen lassen. Die beiden Autoren bearbeiten die Thematik mit unterschiedlichen Vorgehensweisen, die sie als "Analysekonzepte" bezeichnen. Die Auswirkungen der Typenprüfung für Ölbrenner und Heizungskessel werden anhand einer Marktanalyse untersucht. Bei den Wirkungen der Feuerungskontrolle werden eine Umfrage und eine multiple Regression durchgeführt. Bei der Massnahme der Grenzwertsetzung stützt sich die Evaluation auf eine Fallstudie. Die Untersuchung des gesamten dreiteiligen Programms erfolgt über eine statistische Zeitreihenanalyse. Die verschiedenen Ansätze ergänzen sich gut und erlauben es, vertiefte Erkenntnisse über die Wirkungszusammenhänge zu gewinnen. Um eine "Triangulation" im eigentlichen Sinne handelt es sich allerdings nicht, da nur jede Teilmassnahme aufgrund eines eigenen Designs auf ihre Wirkungen untersucht, nicht aber ein einzelnes Evaluandum mit je verschiedenen Forschungsansätzen analysiert wird.

Dies war bei einem anderen *Beispiel* der Fall. Die "Arbeitsgruppe Gesetzesevaluation (AGEVAL)" des Eidgenössischen Justiz- und Polizeidepartements gab zwei Forschungsequipen den Auftrag, mit je einem anderen Untersuchungsdesign die Wirkungen der *beruflichen Vorsorge* auf den Arbeitsmarkt zu analysieren. Gerheuser (1991) packte die Aufgabe im Sinne einer Fallstudie an und stützte seine Analyse auf die Aussagen von Personalverantwortlichen von 42 Unternehmen und (halb-) öffentlichen Institutionen; die Gespräche fanden im Rahmen von elf Hearings in verschiedenen Regionen der Schweiz statt. Schaetti (1990) stützte sich dagegen auf Daten der Zentralen Ausgleichsstelle der AHV in Genf und "verglich die Häufigkeit und Wahrscheinlichkeit von Arbeitslosigkeit, Stellenwechseln und gering entlöhnten Arbeitsverhältnissen vor und nach der Einführung des Obligatoriums" ("Arbeitsgruppe Gesetzesevaluation" 1991) der beruflichen Vorsorge. Diese beiden Designs zur Analyse der Wirkungen derselben Massnahme ergänzten sich und brachten die erwünschten komplementären Ergebnisse. Das Vorgehen entspricht einer

gleichzeitigen Triangulation der "Methoden" und der "Investigatoren" und hat in diesem Sinne für die Schweiz Modellcharakter.

Aus diesen Beispielen ist freilich nicht der Schluss zu ziehen, Evaluationen seien nur dann sinnvoll und liessen nur dann gültige Schlüsse zu, wenn sie sich auf Studien stützen, denen eine Kombination von mehreren Untersuchungsdesigns zugrunde liegt. Ein solches Vorgehen nach dem Vorbild der Triangulation ist aufwendig und kostspielig; es sollte nur dann zum Einsatz gelangen, wenn viel auf dem Spiel steht und von der Evaluation weitreichende Entscheidungen abhängen. Bei weniger umstrittenen Massnahmen und bei Programmen, die mit einfachen Instrumenten arbeiten, genügen durchaus Untersuchungsdesigns, wie wir sie vorher in diesem Kapitel geschildert haben. In jedem Falle aber gilt, dass die Wahl des Evaluationsdesigns immer zuerst von der Fragestellung abhängig zu machen ist. Als zweites Kriterium kommen die zur Verfügung stehenden Mittel in Betracht. Die persönlichen Präferenzen der Forschungsteams sollten eine untergeordnete Rolle spielen.

KAPITEL 12: FORSCHUNGSMETHODEN

Thomas Widmer, Hans-Martin Binder¹

Jede Evaluation stützt sich auf Informationen, die bei den Forschungsteams zunächst nicht vorhanden sind. Das vorliegende Kapitel befasst sich mit der Frage, wie diese Materialien und Daten, die für die Beurteilung der Wirkung einer politischen Massnahme unentbehrlich sind, erhoben und ausgewertet werden sollen. Es geht somit um die Forschungstechniken.

Die Auswahl der Forschungsverfahren steht meist in einem Zusammenhang mit der erkenntnistheoretischen Position der Forscherinnen und Forscher. Im Abschnitt 12.1. unterscheiden wir diesbezüglich das empirisch-analytische Paradigma von interpretativ-hermeneutischen Ansätzen. Die Darstellung der eigentlichen Forschungstechniken ist auf zwei Abschnitte verteilt. Zunächst werden (12.2.) die qualitativen und anschliessend (12.3.) die quantitativen Verfahrensweisen erörtert. In beiden Abschnitten geht es um die Erfassung der Materialien und Daten, um ihre Auswertung sowie um die Fragen der Gültigkeit und der Verlässlichkeit der daraus gezogenen Schlüsse. Schliesslich enthält das Kapitel einige Hinweise auf die Probleme der Darstellung der Ergebnisse von Evaluationen.

12.1. FORSCHUNGSPARADIGMEN

In den Sozialwissenschaften gibt es seit langem eine Kontroverse zwischen zwei grundlegenden erkenntnistheoretischen Positionen. Die eine ist an den Naturwissenschaften, die andere an den Geisteswissenschaften orientiert. Die Auseinandersetzung zwischen diesen beiden Richtungen hat sich schon früh auch auf die Evaluationsforschung übertragen (Campbell 1979). Es drängt sich deshalb auf, an dieser Stelle auf den damit verbundenen Methodenstreit kurz einzugehen.

Das an den Naturwissenschaften orientierte *empirisch-analytische Paradigma* bildet in der Evaluationsforschung immer noch den Mainstream. Es geht von einer rationalistischen Sicht der sozialen Welt aus. Diese kann in Teilerscheinungen oder Elemente aufgeschlüsselt werden, zwi-

¹ Hans-Martin Binder: Ausschliessliche Verantwortung für den Abschnitt 12.2.; für die übrigen Teile liegt die alleinige Verantwortung bei Thomas Widmer.

schen denen Zusammenhänge bestehen. Diese werden durch die Bildung von Variablen erfasst. Aufgrund von theoretischen Annahmen werden Hypothesen postuliert, die es zu überprüfen gilt. Können die Hypothesen nicht falsifiziert werden, so gelten sie vorläufig als gesichert. Gültige Hypothesen werden zu neuen Theorien zusammengefügt, die es erlauben, die soziale Welt zu erklären. Die Evaluationsforschung, die sich diesem Ansatz verpflichtet fühlt, neigt dazu, mit experimentellen oder quasi-experimentellen Untersuchungsdesigns zu arbeiten. Sie stützt sich dabei auf deduktiv hergeleitete Wirkungsmodelle, die sie in der Wirklichkeit einem rigorosen Test unterzieht.

Im Gegensatz dazu steht das geisteswissenschaftliche *interpretativ-hermeneutische Paradigma*. Dieses geht davon aus, dass die soziale Welt nur in ihrer Totalität erfassbar ist. Zu deren Entschlüsselung ist ein verstehendes Vorgehen erforderlich. Im strengen Sinne gibt es keine objektive gesellschaftliche Wirklichkeit, sondern lediglich verschiedene subjektive Rekonstruktionen davon. Erkenntnisse sind deshalb letztlich nur über Kommunikation zu gewinnen. Die Interpretation dieser perspektivenabhängigen Sichtweisen erfordert zwar Vorkenntnisse und Vorstrukturierungen des Untersuchungsfeldes; sie zwingt indessen auch zu einer Offenheit gegenüber neuen Informationen, welche die ursprünglichen Annahmen im Laufe des Forschungsprozesses verändern können. In der Evaluationsforschung bedeutet dies, dass vorwiegend das Design der Fallstudie gewählt wird. Dieses erlaubt es, möglichst alle mit einer Massnahme zusammenhängenden Erscheinungen in dichter Weise zu beschreiben und so die Grundlagen für die Beurteilung von Wirkungen bereitzustellen.

Die beiden geschilderten Ansätze haben selbstverständlich auch gewisse *Affinitäten* zu den in den kommenden Abschnitten vorzustellenden *Forschungstechniken*. So sind Untersuchungsdesigns, die auf dem empirisch-analytischen Ansatz basieren, mindestens teilweise auf quantitative Daten angewiesen. Umgekehrt sind mit dem interpretativ-hermeneutischen Paradigma fast ausschliesslich qualitative Forschungsv erfahren verbunden.

Abbildung 7 Verbindung von Paradigma und Forschungsverfahren

Paradigma	Forschungsverfahren	
	quantitativ	qualitativ
empirisch-analytisch	Normalfall	gelegentlich
interpretativ-hermeneutisch	selten	Normalfall

Unsere Übersicht will indessen zeigen, dass die Verbindungen von "empirisch-analytisch" mit "quantitativ" sowie von "hermeneutisch" mit "qualitativ" lediglich den Normalfall darstellen, dass sie aber keineswegs zwingend sind. Es wird somit keineswegs postuliert, dass quantitative Verfahren der Datenerhebung und -auswertung nur im Rahmen des empirisch-analytischen Ansatzes zur Anwendung gelangen können, auch wenn in der Forschungspraxis Exponenten des interpretativ-hermeneutischen Paradigmas nur selten mit quantitativen Daten arbeiten. In der Praxis ist es umgekehrt recht häufig, dass empirisch-analytisch orientierte Forscherinnen und Forscher auch qualitative Daten in ihre Untersuchungen einbeziehen. Keine Kombination ist auszuschließen. Wenn demnach im folgenden für die Darstellung der quantitativen und der qualitativen Verfahren jeweils die Kombination mit dem Paradigma gewählt wird, das den Normalfall bildet, so geschieht dies nicht aus systematischen, sondern aus didaktischen Gründen und zur Vereinfachung der Darstellung.

12.2. QUALITATIVE VERFAHREN

Die in diesem Abschnitt diskutierten Forschungstechniken unterstehen einem hermeneutischen, interpretativen Paradigma. Das entsprechende Denkmodell geht davon aus, dass Menschen zwar in einem von aussen beschreibbaren kulturellen, sozialen, politischen und wirtschaftlichen Umfeld handeln, dass sie aber stets ihre Situation und ihre Interaktionen deuten müssen. In Anlehnung an die Theorien des Symbolischen Interaktionismus (siehe dazu Blumer 1976) und der Ethnomethodologie fordert Wilson (1976), dass jede soziale Interaktion als interpretativer Prozess

aufzufassen sei. Menschen müssten jede soziale Situation und jeden Prozess in ihrem eigenen Umfeld für sich deuten, müssten ihre eigenen Rollenzuschreibungen erfassen, und ihre Verhaltens- und Handlungsoptionen beurteilen (vgl. Mayring 1993). Methodologisch kann diesem interpretativen Paradigma mit einem hermeneutischen Ansatz entsprochen werden. Dabei werden Texte im Sinne des Verstehens interpretiert. "Hermeneutisches Verstehen bezieht sich auf das Erfassen menschlicher Verhaltensäußerungen und Produkte" (Lamnek 1988: 65).

Hermeneutische Interpretation ist indessen keinesfalls eine regellose "Freistil"-Methodologie. Dieser sozialwissenschaftliche Verstehensprozess hat sich auch an gängigen Güte-Kriterien wie Validität und Reliabilität zu messen. Die Philosophen Schleiermacher (1768-1834) oder Dilthey (1833-1911) haben unter dem Begriff des "hermeneutischen Zirkels"² entsprechende wissenschaftliche Standards gesetzt (Lamnek 1988: 67 f.). Hermeneutisch interpretative Forschungsansätze haben demzufolge im wesentlichen zwei Prinzipien zu genügen: dem Prinzip der Offenheit und dem Prinzip der Kommunikation.

Das Prinzip der Offenheit beinhaltet als Forderung, "den Wahrnehmungstrichter empirischer Sozialforschung so weit als möglich offen zu halten, um dadurch auch unerwartete, aber dafür umso instruktivere Informationen zu erhalten" (Lamnek 1988: 22). Diese Forderung hat zur Konsequenz, dass theoretische Strukturierungen und Hypothesen, wie auch methodische Verfahren im Verlauf des Forschungsprozesses erweitert, modifiziert und revidiert werden können. Diese Offenheit sowohl gegenüber den Untersuchungspersonen und ihren Kontexten als auch gegenüber den angewendeten Methoden zeigt, dass diese Art qualitativer Sozialforschung eine starke "Explorationsfunktion" aufweist: "Qualitative Sozialforschung legt das Schwergewicht auf die explorierende Felderkundung, die bei Anwendung standardisierter Techniken ohne vorherige Erkundungsphase meist sträflich vernachlässigt wird." (Lamnek 1988: 22; nach Hoffmann-Riem 1980)

² Es werden zwei hermeneutische Zirkel unterschieden (Lamnek 1988: 260):

- I. Ein zunächst rudimentäres Vorverständnis ist zum Verstehen eines Textes und dessen Auslegung notwendig. Das durch das Textverstehen angeeignete Wissen trägt zur Erweiterung und Korrektur des Vorverständnisses bei, womit sich der Text wiederum besser verstehen lässt.
- II. Die Einzelelemente eines Textes sind nur aus dem Gesamtzusammenhang verständlich, der sich wiederum nur aus den Einzelteilen ergibt."

Gemäss dem Prinzip der *Kommunikation* ist davon auszugehen, dass Evaluatorinnen und Evaluatoren in ihrem Forschungsprozess nicht unabhängig sind von den Daten und Informationen, die sie durch die an der Evaluation beteiligten Akteure erheben. Dieses wechselseitige Aufeinanderbezogene von Forscherinnen und Forschern und ihrem Evaluandum ist konstitutives Element im Rahmen eines interaktiven Forschungsprozesses. Daten sind nur über Kommunikations- und Verstehensprozesse zu gewinnen. Damit unterliegen sie aber auch subjektiven Deutungen, die im Ablauf bestimmter Interaktionsprozesse erfolgen: Akteure strukturieren und deuten mit ihren Definitions- und Interpretationsleistungen die von ihnen erfasste Wirklichkeit. Demzufolge ist die dargestellte Sicht der Wirklichkeit stets perspektivenabhängig; und mit dem Wechsel der subjektbezogenen Perspektive ändert sich auch die dargestellte Wirklichkeit (Schütze 1978).

Das Prinzip der Offenheit und das Prinzip der Kommunikation haben vorgehensprägende Implikationen auf die Methodik und den konkreten Forschungsablauf einer Evaluation. Vor allem Guba und Lincoln (1989) haben unter der Bezeichnung "Fourth Generation Evaluation" den Ansatz der interaktiven Evaluation stark geprägt. Dieser weist im übrigen zahlreiche Bezugspunkte zum Konzept lernorientierter Evaluationen auf (vgl. Kapitel 9 und 13.3.).

12.2.1. Interaktive Evaluation

Folgende Konzept-Elemente stehen bei diesem Forschungsansatz im Vordergrund:

Evaluation muss alle beteiligten Akteure ("stakeholders") mit ihren Wert- und Zielvorstellungen in einen Prozess der Aushandlung einbeziehen, um deren unterschiedliche Wert- und Zielvorstellungen und die unterschiedlichen Sichtweisen und Realitätskonstrukte miteinander zu konfrontieren. Ziel ist es, ein möglichst facettenreiches Bild des Evaluations-Gegenstandes zu erhalten.

Diesem Konzept zufolge folgt die interaktive Evaluation in ihrem methodologischen Ansatz einer *konstruktivistischen Sicht der Realität*: Es gibt nicht eine "objektive" Realität. Die Beschreibung der Realität ist vielmehr ein Konstrukt des Beobachters bzw. der Beobachterin und ist das

Resultat von Interaktionen zwischen den Beobachtenden und den Beobachteten (Bierter und Binder 1993: 197).

Die im Rahmen eines solchen Evaluationsansatzes zur Anwendung gelangenden methodischen Instrumente müssen sicherstellen,

- dass alle im Zusammenhang mit dem Evaluations-Gegenstand involvierten Akteure identifiziert werden,
- dass deren Wert- und Zielvorstellungen erfasst werden, und
- dass erfasst wird, wie die Akteure ihre Wert- und Zielvorstellungen mit dem Gegenstand und Kontext der Evaluation in Verbindung bringen.

In einem hermeneutischen Auswertungs- und Interpretationsprozess müssen die verschiedenen Wahrnehmungen und Realitätskonstrukte der beteiligten Akteure verstanden und miteinander konfrontiert werden. Über einen Aushandlungsprozess unter Beteiligung der einbezogenen Akteure muss schliesslich versucht werden, einen Konsens im Verstehen und über die gegebenen Interpretationen zu erreichen, aus dem heraus konkrete Handlungsperspektiven zu entwickeln sind (Guba und Lincoln 1989: 72 ff.). Das Ziel dieses Interaktionsprozesses soll darin bestehen, "die Realitätssicht möglichst vieler Beteiligter adäquat zu repräsentieren, divergente Positionen deutlich zu machen und die Basis für Schlussfolgerungen und Empfehlungen klar zu benennen" (Kraus 1991: 413).

Dieser Evaluationsansatz zielt klar auf den notwendigen Nutzenaspekt jeder Evaluation. Eine noch explizitere Nutzenorientierung umfasst der Ansatz der "Utilization-Focused-Evaluation" von Patton (1978), der die Zielintention einer Evaluation noch weit stärker auf die Erarbeitung und Entwicklung von konkreten Handlungsstrategien und -Optionen für einen definierbaren Kreis von Akteuren fokussiert.

Allen Ansätzen der qualitativen, interaktiven Evaluationsforschung ist gemeinsam, aus einem Interaktionsprozess von Wissenschaft und politischer Praxis, auf der Basis erkannter und kommunizierter Realitätskonstrukte neue Handlungsperspektiven zu eröffnen. "So verstanden, wäre Evaluation der Versuch, die Kluft zwischen Wissenschaft und Praxis aufzuheben. Die Diskursivierung der Handlungsprogramme aller Beteiligten, wie sie in der qualitativen Evaluationsforschung stattfindet, könnte dann, konzeptuell integriert in eine organisatorische Praxis, idealerweise kollektive Lernprozesse ermöglichen, mindestens jedoch verfestigte Strukturen aufbrechen und eingefrorene Kompromisse verflüssigen." (Kraus 1991:

415). Solche interaktive Evaluation ist v. a. ein wesentliches Merkmal von Mediationsprozessen (vgl. Kapitel 9).

12.2.2. Gütekriterien qualitativer Evaluationsforschung

Auch im Rahmen des interpretativen Paradigmas haben sich die methodischen Ansätze mit den Anforderungen an die Gültigkeit und Verlässlichkeit wissenschaftlicher Messprozesse auseinanderzusetzen. In der methodologischen Diskussion indessen werden Sinn, Adäquanz und Machbarkeit von Gütetests je nach Methodenansatz sehr unterschiedlich beurteilt.

12.2.2.1. Reliabilität

Eine hohe Zuverlässigkeit der Messinstrumente und der Datenerhebungsverfahren setzt - streng genommen - eine identische Wiederholbarkeit des Forschungsergebnisses zu unterschiedlichen Zeitpunkten und die Unabhängigkeit der Messung von der Datenerhebungsperson voraus. Wenn wir aber davon auszugehen haben, dass die Datenerhebung im Rahmen der qualitativen Evaluationsforschung als soziale Interaktion zwischen dem forschenden und dem zu befragenden Akteur zu verstehen ist, dann ist dieser Prozess einer bestimmten Dynamik unterworfen, und demzufolge durch vielfaltigen Merkmalswandel gekennzeichnet. Die Stabilität des Messergebnisses bei wiederholter Durchführung ist deshalb - unter positivistischen Prämissen - nicht gewährleistet. Rossi und Freemann (1993: 230) bezeichnen diesen Umstand als gravierenden Mangel, weil unscharfe Messinstrumente reale Wirkungen verwischen oder verschleiern würden; es wäre demzufolge weniger wahrscheinlich, dass eine erfolgreiche Intervention entdeckt werden könne. Demgegenüber attestiert Lamnek dem interaktiven Evaluationsansatz eine nicht zu unterschätzende Stärke durch den Einbezug subjektiver Anteile in diesem kommunikativen Forschungsprozess: "An die Stelle von Replizierbarkeit von Untersuchungsbedingungen und Forschungsergebnis tritt im interpretativen Paradigma die Betonung der "situativen Kontextgebundenheit" von Datenerhebungs- und Auswertungsergebnissen." (Lamnek 1988: 162).

Auch Guba und Lincoln bezeichnen in ihrem Ansatz der Fourth-Generation-Evaluation die Reliabilität im positivistischen Sinn als nicht adäquat für einen konstruktivistischen Ansatz. Sie schlagen mit dem Begriff der "Dependability" ein paralleles Kriterium zur Beurteilung der Stabilität von

Datenerhebungen und Messverfahren vor. Dieses Konzept der Verlässlichkeit muss gewährleisten, dass Aussenstehende den Forschungsprozess und die im Verlaufe dieses Prozesses getroffenen Entscheidungen und Interpretationen des Evaluators nachvollziehen können: "The technique for documenting the logic of process and method decisions is the dependability audit." (Guba und Lincoln 1989: 242).

12.2.2.2. Validität

Die im Rahmen einer interaktiven und interpretativen Evaluationsstrategie hauptsächlich zur Anwendung gelangenden Methoden (Interviews, Gruppendiskussionen) umfassen zum grossen Teil Interaktions- und Kommunikationsprozesse. Unter wahrnehmungs- und erkenntnistheoretischen Aspekten (Benseler, Hejl und Köck 1980; Varela und Thompson 1992; Watzlawick und Krieg 1992) geht es nicht darum, "objektive" Informationen von Befragten zu erhalten; vielmehr interessiert die Forscher und Forscherinnen das vom befragten Akteur kognitiv strukturierte Feld des Handelns und seiner entsprechenden Umwelt (Kontextbezogenheit) sowie seine Begründungen (seine "impliziten Theorien"). Unter dieser Prämisse ist deshalb jede Äusserung der Befragten eo ipso "gültig" (Bierter, Binder und Rüegg 1985: 49 ff.). Das im positivistischen Validitätsverständnis wichtige Korrespondenz-Kriterium (Übereinstimmung einer Behauptung mit der "objektiven" Welt) wird von einer hermeneutisch und konstruktivistisch orientierten qualitativen Evaluationsforschung abgelehnt. Mit einem solchen Perspektivwechsel von der notwendigen Übereinstimmung mit einer "objektiven" Welt hin zu einer Vorstellung von der sozialen Wirklichkeit findet auch eine Verschiebung von der alles prägenden Beobachtung der sozialen Wirklichkeit hin zu einer Auseinandersetzung über sie statt. Ein solcher Wandel erfordert ein kommunikatives und pragmatisches Validierungskonzept (Kvale 1991: 427 ff.).

In Anlehnung an Volmery (1983: 124 ff.) gilt es im Rahmen des interpretativen Paradigmas neue qualitative Gültigkeitskriterien zu entwickeln. Drei solche Validierungen stehen im Vordergrund:

Ökologische Validierung basiert auf der Überzeugung, dass gültige Informationen über Forschungsgegenstände und Untersuchungspersonen nur in deren "natürlichem Lebensraum" gewonnen werden können. Dieser natürliche Kontext darf nicht durch künstliche oder standardisierende Elemente eingeengt oder entfremdet werden. Ökologische Validierung beschränkt sich nicht auf den Prozess der Datenerhebung, sondern fordert

auch den Einbezug der Kontextbedingungen in den hermeneutischen Interpretationsprozess.

Bei der *kommunikativen Validierung* werden Interpretationsergebnisse durch erneutes Befragen der Akteure auf Stimmigkeit und Gültigkeit hin überprüft. Die Forscherinnen und Forscher vergewissern sich durch diesen Validierungsprozess, ob sie die befragten Akteure richtig verstanden haben und ob diese ihren Interpretationen zustimmen. Kommunikative Validierung bedeutet demnach "die dialogförmige Überprüfung der Validität von Erkenntnisansprüchen" (Kvale 1991:429). Sie muss die "Interpretationsgemeinschaft" auf die untersuchten Personen ausdehnen, denn nur so kann die notwendige Zustimmung der Befragten zu den Interpretationen des Evaluators - und damit "die Einbeziehung der Öffentlichkeit in die Interpretationsgemeinschaft" - sichergestellt werden (Kvale 1991: 430).

Argumentative Validierung bedeutet, dass das Forschungsteam seine (theoriegeleiteten) Vorannahmen gegenüber den einbezogenen Akteuren offenlegen muss; seine Interpretationen müssen vor diesem gemeinsamen Hintergrund auf ihre Stimmigkeit überprüft werden. Durch diese Transparenz des Verstehens- und Interpretationsprozesses wird dieser auch notwendigerweise kritisierbar.

Diese Kriterien, die in ihrer Ausrichtung etwa dem konventionellen Kriterium der "internen Validität" entsprechen, fassen Guba und Lincoln unter dem Begriff der "Credibility" zusammen und entwickeln sechs methodische Zugänge zur Überprüfung dieser Glaubwürdigkeit (Guba und Lincoln 1989: 236 ff.).

In Korrespondenz zur "externen Validität" steht das Konzept der "Transferability" (Übertragbarkeit) nach Guba und Lincoln (1989: 241 f.) oder, nach Kvale (1991: 430 f.), dasjenige der "pragmatischen Validität". Die Validierung besteht dabei in der "Verifizierung im wörtlichen Sinne einer Wahrheitsprobe": Die Wirksamkeit der Erkenntnisgewinnung, die "Wahrheit der Schlussfolgerungen" muss sich in der Praxis, im konkreten Handeln beweisen. "Ein pragmatisches Konzept der Wahrheit ist dadurch charakterisiert, dass es uns dabei hilft, Handlungen in die Wege zu leiten, die zu den gewünschten Zielen führen" (Kvale 1991: 431).

12.2.3. Möglichkeiten und Grenzen qualitativer Evaluationsforschung

Qualitative, am interpretativen Paradigma orientierte Evaluationsforschung will in erster Linie dem Anspruch der Problemorientiertheit im Hinblick auf die Nützlichkeit genügen (vgl. Kapitel 13.3.). Ausgehend von praktischen Problemstellungen des Evaluationsbereiches sollen die Ergebnisse auf eine Optimierung des Vollzuges und eine höhere Wirksamkeit des Programms ausgerichtet sein. Die aus dieser Perspektive im Rahmen von Evaluationsstudien zu erarbeitenden Handlungsstrategien sollen die beteiligten und betroffenen Akteure auf einen politik- und sozialinnovatorischen Lernprozess hinführen.

Einzelfallbezogenheit, Problemorientierung und Kontextbezogenheit implizieren, dass die in die Evaluation miteinbezogenen Forschungssubjekte in ihrem natürlichen, alltäglichen Handlungskontext untersucht werden. Die Möglichkeit einer Generalisierbarkeit der Forschungserkenntnisse und -ergebnisse muss deshalb in jedem Fall argumentativ begründet werden. In diesem Sinn kann qualitative Evaluationsforschung dem positivistischen Anspruch auf schlüssige Verallgemeinerbarkeit nicht in der Art genügen, wie dies einzelne quantitative Ansätze (allerdings auch nur unter nicht sonderlich komplexen Bedingungen) zu tun vermögen. Der besondere Wert verstehensorientierter und interaktiver Evaluationsforschung liegt aber gerade in der Fähigkeit, Programme in einer hohen Komplexität ihrer Strukturen, Abläufe, Verfahren und Kontextsituationen sowie ihrer Entwicklungsdynamik zu erfassen und sie mit Hilfe der beteiligten Akteure zu interpretieren. In diesem Sinn legt dieser Evaluationsansatz sein Schwergewicht auf die problem- und einzelfallbezogene Nützlichkeit und nicht so sehr auf die Generalisierbarkeit der Evaluationsergebnisse.

12.2.4. Datenerhebung

12.2.4.1. Dokumentenanalyse

Nachdem der Rahmen des Evaluationsauftrages präzise definiert ist (Auftraggebende und deren Rolle im Evaluationsprozess, Definition des Evaluandums mit entsprechender Definition der Fragestellung, Auflistung der notwendigerweise einbezogenen Akteure), steht am Anfang jeder Evaluation eine seriöse Dokumentenanalyse. Sie verschafft das notwen-

dige Kontextwissen über die vorhandenen (und dem Forschungsteam zur Verfügung stehenden) Materialien, das es überhaupt erst erlaubt,

- die historischen Zusammenhänge der Fragestellung in ihren Grundzügen zu erfassen;
- das notwendige Fachwissen zum Gegenstandsbereich zu erlangen;
- als Forscher und Forscherin das eigene Vorverständnis gegenüber dem Gegenstandsbereich und seinen Kontexten zu entwickeln.

Das typische Merkmal der Dokumentenanalyse ist "die intensive, persönliche Auseinandersetzung mit dem Dokument, welches in seiner Einmaligkeit möglichst umfassend durchleuchtet und interpretiert wird" (Atteslander 1971: 67).

Da Dokumente als Datenmaterial nicht erhoben werden müssen, weil sie ja bereits fertig vorliegen, entstehen bei der *Datenerhebung* keine Schwierigkeiten, wohl können aber bei der *Datenverfügbarkeit* und *-aus-sicht* Probleme auftreten, die im Evaluationsbericht unbedingt transparent gemacht werden müssen:

- Aus politischen oder aus Gründen der Vertraulichkeit kann der Zugang zu wichtigen Dokumenten (oder bereits das Wissen um das Bestehen solcher Dokumente) verwehrt bleiben.
- Welche Dokumente in die Untersuchung einbezogen werden, ist ein subjektiver Entscheidungsakt der Evaluierenden.

Das Textmaterial kann etwa aus folgenden Dokumentenarten bestehen: gesetzliche Grundlagen; Weisungen; Strategie- und Konzeptpapiere; Sitzungs- und Tagungsprotokolle; Zeitungs- und Zeitschriftenberichte oder andere Publikationen. Auf der Basis der genau definierten Fragestellung muss vom Evaluator oder der Evaluatorin entschieden werden, welche Dokumente allein dazu dienen, das notwendige und möglichst umfassende Kontextwissen zu erlangen, das für eine schlüssige Interpretation der übrigen Daten dient, und welche Dokumente in einem theoriegeleiteten interpretativen oder quantitativen Analyse- und Auswertungsverfahren (quantitative und qualitative Inhaltsanalyse) zu erschliessen sind. (Eine ausführliche Darstellung der Theorie, der Methoden und der Praxis der Inhaltsanalyse findet sich in Merten 1983.)

12.2.4.2. *Das problemzentrierte Interview*

Das nicht-standardisierte qualitative Interview weist verschiedene Formen und Ausprägungen auf. Diese unterscheiden sich vor allem auf der Ebene der Strukturierung des Fragerasters.

Das *narrative Interview* z.B. zeichnet sich aus durch eine völlig offene Gesprächsführung; die Befragten sollen durch Erzählungen "Orientierungsmuster des Handelns" und eine "retrospektive Interpretation" des erzählten Handelns offenlegen (Lamnek 1989: 71). In diesem Sinn stellt das narrative Interview eine klar induktive Vorgehensweise dar; das Forschungsteam beginnt die Datenerhebung, ohne über - die Themenbereiche des Interviews strukturierende - wissenschaftliche Konzepte zu verfügen (Lamnek 1989: 74). Dies jedoch entspricht nicht einer idealen und realistischen Ausgangslage für eine Evaluation politischer Programme.

Eine zweite Form des qualitativen Interviews stellt das *fokussierte Interview* (nach Merton und Kendall 1984) dar. Diese Interviewform stützt sich zwar - wie beim problemzentrierten Interview auch - auf einen Interviewleitfaden. Dieser ist aber das Resultat einer detaillierten Theorie- und Hypothesenbildung durch die Forschungsequipe. Beim fokussierten Interview geht es um die Überprüfung von Hypothesen. Diese Interviewform ist deshalb stärker an einer positivistischen Forschungslogik orientiert (Lamnek 1989: 78).

Die dritte Form, auf welcher wir uns in Bezug auf die Evaluationsforschung konzentrieren, stellt das *problemzentrierte Interview* (Witzel 1985) dar. Seine wesentliche Eigenschaft besteht darin, dass es sich auf eine vordefinierte Problemstellung ausrichtet, die vom Interviewer in seinen wichtigsten Aspekten bereits im voraus erfasst (durch Literaturstudium und Dokumentenanalyse, durch Vorgespräche mit Experten im Untersuchungsfeld) und zu einem theoretischen Konzept verdichtet wurde. Diese Problemstellung wird im Verlauf des Interviews entlang eines Leitfadens im Rahmen eines kontinuierlichen Verstehensprozesses strukturiert, kommuniziert und interpretiert. Die partielle Standardisierung durch einen theoriegeleiteten Gesprächsleitfaden ermöglicht es, im Rahmen der Datenauswertung eine grössere Vergleichbarkeit der einzelnen Interviews zu erreichen.

Das problemzentrierte Interview basiert - trotz seiner themenbezogenen Strukturierung durch den Leitfaden - auf dem Erzählprinzip, d.h. die Auswahl, Reihenfolge und Sprachform dessen, worüber berichtet, was bewertet und interpretiert wird, bleibt den Befragten überlassen. In dieser

Interviewform bleibt auch mit einer gewissen Leitfadenbezogenheit viel Raum, in dem sich durch das frei assoziierende Erzählen wichtige Handlungszusammenhänge und zentrale kognitive Strukturen der Interviewpartner herauschälen. In diesem Sinn ist das problemzentrierte Interview nahe verwandt mit dem narrativen Interview.

Wie gross die Qualität und Dichte der Information ist, die im Rahmen eines Interviews ermittelt wird, hängt einerseits davon ab, wie gross die Fähigkeit des Befragers ist, den Interviewpartner zum Erzählen zu animieren und ihm den dazu notwendigen Raum zuzugestehen; von Bedeutung ist aber auch die Bereitschaft der Befragten, überhaupt erzählen zu wollen.

Kohli (1978) nennt als wichtigste Bedingung für den korrekten und ziel führenden Verlauf dieses Prozesses die Offenheit: Der Interviewte muss frei antworten können, ohne dabei durch vorgegebene Antwortkategorien eingeschränkt zu werden. Durch diese *Offenheit* kann erreicht werden, dass

- erkennbar wird, ob die Befragten verstanden haben, worum es geht und was der Interviewer erfahren möchte;
- die individuellen Sinngehalte und Bedeutungszuschreibungen, die kontextbezogenen kognitiven Strukturen und Interpretationen der Befragten und deren Sicht der Zusammenhänge innerhalb des erfragten Gegenstandsbereiches voll zum Ausdruck kommen können;
- den Befragten ein grösserer Aktivitätsraum zugestanden wird und diese dadurch die Möglichkeit zur eigenen Steuerung des Gesprächs haben. Offene Fragen erlauben den Befragten eigene Antwortformulierungen und lassen ihnen Raum zu längeren, komplexeren und assoziationsreichen Antworten;
- die Situation (räum-, zeit- und personenbezogen), in der das Interview stattfindet, von den Beteiligten thematisiert und in der nachfolgenden Interpretation durch das Evaluationsteam reflektierend miteinbezogen werden kann.

In partizipativen akteurbezogenen (stakeholder-based) Evaluationen sind problemzentrierte Interviews deshalb ein sehr adäquates Datenerhebungsinstrument, weil es im Rahmen des interpretativen Paradigmas und einem exploratorischen Untersuchungsdesign gerade darum geht, von den einzelnen Akteuren Situationsdeutungen, Handlungsmotive, Zweck-

Mittel-Relationen und entsprechende Selbstinterpretationen zu erfragen. Durch die "diskursive Verständigung über Interpretationen" sind mit offenen und teilstandardisierten Interviews die Möglichkeiten einer "empirischen Umsetzung handlungstheoretischer Konzeptionen" gegeben (Hopf 1991: 180).

Dem Forschungsteam dient zur Strukturierung des Gespräches ein an der Fragestellung orientierter *thematischer Leitfaden*. Dieser legt als eine Art Checkliste fest, welche Themenbereiche im Verlauf des Interviews anzusprechen sind. Die Einstiegsfragen in die einzelnen Themenbereiche werden im voraus formuliert und dienen im wesentlichen dazu, dass die Befragten - ohne sich unter einem intellektuellen Leistungsdruck zu fühlen - an ihren Alltagserfahrungen anknüpfen können. Die jeder Befragungssituation innewohnende "Künstlichkeit" (die beteiligten Personen sind einander fremd, ihre motivationale Beziehung ist asymmetrisch) kann dadurch etwas abgebaut und ein lockeres Interaktions- und Kommunikationsklima geschaffen werden. Bei Expertengesprächen wird von den Interviewten sehr häufig gewünscht oder verlangt, dass ihnen der Leitfaden vor dem Gesprächstermin schriftlich zugestellt wird. Diesem Begehren kann ohne methodische Bedenken entsprochen werden. Meistens hat diese Vorinformation bei Expertinnen und Experten auch zur Folge, dass diese sich inhaltlich auf die Befragung vorbereiten, was tendenziell zu umfassender und vollständiger Information führt.

Die Interviewer sollten sich in jedem Fall die sprachliche Ungebundenheit erhalten, die Fragen in einer der jeweils konkreten Interviewsituation und dem Interviewpartner angepassten Formulierung zu stellen. Nachfragen, i.S. von Verständnis- und Präzisierungsfragen, sollen auf die unmittelbar vorangegangenen Äusserungen der Befragten Bezug nehmen. Interviewer müssen sich in ihrer Gesprächsführung flexibel den Befragten, ihrem Erfahrungsbereich und ihren Sprachgewohnheiten anpassen. Die Reaktionsmöglichkeiten der Befragten sollten möglichst gross sein; diese müssen sich frei und ungezwungen äussern können. Müssen die Befragten mit offensichtlichen Widersprüchen und Ungereimtheiten konfrontiert werden, so sollte dies - im Hinblick auf die Erhaltung einer angenehmen Gesprächssituation - mit der nötigen Zurückhaltung erfolgen. Um authentische Äusserungen von hoher Verlässlichkeit zu erhalten, müssen Befragte über hinreichende Sach- und Sprachkompetenz verfügen, damit sie ihre Sicht der Realität, der Handlungen und der Kontexte hinreichend differenziert formulieren können.

In aller Regel werden Interviews auf *Tonband* aufgenommen. Dies ermöglicht einerseits, dass sich die Evaluatorinnen und Evaluatoren voll auf das Gespräch und dessen Ablauf konzentrieren können, ohne daneben durch das Erstellen von Notizen absorbiert zu sein. Andererseits gewährleisten Tonbandaufzeichnungen vollständige, authentische und von verschiedenen Personen bearbeitbare Gesprächsprotokolle, womit die Datenaufbereitung (Transkription) und Datenauswertung wesentlich erleichtert wird. Darüber hinaus kann eine Tonbandaufzeichnung später bei grösseren Differenzen in der Interpretation als Beweismittel dienen. In jedem Fall darf ein Gespräch nur mit dem ausdrücklichen Einverständnis der Befragten auf Tonband aufgenommen werden. Den Interviewten ist Vertraulichkeit bei der Verwendung der Tonbänder zu garantieren.

Die *Auswahl der Interviewpartner und -Partnerinnen* richtet sich in erster Linie nach der expliziten Fragestellung der Evaluation - einschliesslich ihres thematischen, institutionellen und räumlichen Geltungsbereiches. Aus diesem Kontext ergeben sich die relevanten Akteure. Handelt es sich dabei um Einzelpersonen, so stellt sich die Frage der Auswahl nicht (allenfalls die Frage der Anzahl). Handelt es sich um eine Institution, so besteht möglicherweise ein Spielraum, welche der für den fraglichen Handlungsbereich repräsentativen Personen für eine Befragung auszuwählen sind³. Im Rahmen dieses Spielraumes ist es notwendig und legitim, jene Personen auszuwählen, die grundsätzliche Bereitschaft (wenn nicht gar Interesse) für ein solches Interview zeigen und die über die notwendige Sach- und Sprachkompetenz verfügen. Über die Kriterien und den Prozess zur Auswahl der Gesprächspartnerinnen und -partner muss Transparenz bestehen. Dabei ist durchaus zulässig, dass im Sinne eines iterativen Forschungsprozesses im Verlauf der Evaluation aufgrund gemachter Erfahrungen und des sich entwickelnden Erkenntnisstandes auch neue Personen einbezogen werden.

12.2.4.3. Gruppendiskussion

Bei der Gruppendiskussion als Methode der Datenerhebung geht es nicht um eine Erfragung von Einstellungen und Meinungen einzelner Personen, sondern um eine prozessorientierte - also situationsabhängige - Entwicklung von Meinungen und Interpretationen durch die Auseinandersetzung

³ Bei Politikevaluationen kommt es recht häufig vor, dass innerhalb einer Institution mehrere Akteure mit unterschiedlichen Aufgaben und Positionen zu einzelnen Massnahmen vorzufinden sind (vgl. Kapitel 4).

mit anderen Akteuren (Lamnek 1989: 136). Meinungen und Einstellungen sind sehr oft stark an gesellschaftlichen Zusammenhänge gebunden und können durch die Beteiligten demzufolge besser in sozialen Situationen - also in der Gruppe - offengelegt und erfasst werden. Durch Gruppendiskussionen ist es möglich, an "kollektive Einstellungen" heranzukommen (Mayring 1993: 54).

Zwar ist die Methode der Gruppendiskussion in der qualitativen Sozialforschung vorrangig aus dem Forschungsfeld der Meinungs- und Einstellungsforschung hervorgegangen (hier v.a. Pollock 1955); grösstenteils wurde mit Realgruppen gearbeitet (so etwa Niessen 1977). Dreher und Dreher schreiben dem Begriff "Diskussion" - in einer methodologisch weiter fassenden Art - die folgende Bedeutung zu: "Diskutieren wird methodologisch als valide Handlung erachtet, bei der durch den argumentativen Austausch von Begründungen und Bewertungen die Klärung eines Sachverhalts angestrebt wird, d.h. diskutieren ist exklusiv das Verfahren zur Erweiterung und Klärung von individuellen Sinnperspektiven in der Auffassung von Realität" (Dreher und Dreher 1991:187). In dieser Anwendungsperspektive der Gruppendiskussion als "Basis edukativer Intervention" wird diese einerseits zur Informationsquelle für die Forschung, aber auch zum Lernprozess für die daran Beteiligten (Dreher und Dreher 1991: 187 f.).

Im Rahmen von Evaluationen kann die Methode der Gruppendiskussion eine wichtige und wertvolle *Ergänzung* zu den Einzelinterviews bilden. Sowohl mehrfach übereinstimmende als auch sich widersprechende oder kontrastierende Äusserungen und Interpretationen der einzelnen Befragten können der "Gruppe" der Akteure vorgelegt werden mit dem Ziel, im argumentativen Diskurs Differenzierungen, Ergänzungen und Begründungen der übereinstimmenden oder divergierenden Einzelaussagen und -Interpretationen zu erreichen.

Innerhalb einer Evaluationsstudie steht die Methode der Gruppendiskussion nicht als alleiniges Datenerhebungsverfahren im Vordergrund. Die Gruppendiskussion eignet sich aber als Instrument für eine kommunikative und argumentative Validierung der in den Einzelgesprächen erhobenen Daten. So haben beispielsweise Bierter und Binder (1993) die in zahlreichen Einzelbefragungen gewonnenen Informationen sowie ihre Auswertung und Interpretation den beteiligten Akteuren im Rahmen von verschiedenen Workshops vorgelegt. Durch diese Auseinandersetzung mit den Forschungsergebnissen und den daraus ableitbaren künftigen Handlungsstrategien hatten die befragten Akteure die Möglichkeit, die

Interpretationen und Argumentationen der Forscher zu kritisieren und zu ergänzen. Die Gruppendiskussion wird damit zu einem zusätzlichen wichtigen Stein im Methoden-Mosaik, der es ermöglicht, den Untersuchungsgegenstand und sein Umfeld aus mehreren Blickwinkeln anzugehen und auszuleuchten. Diese Konzeption entspricht auch den Vorstellungen von Köckeis-Stangl (1980: 363) von einem Forschungsprozess im Sinne einer "mehrperspektivischen Triangulation", die ermöglicht, als Ergebnis kein einheitliches, sondern ein "kaleidoskopartiges" Bild des Untersuchungsgegenstandes zu erhalten. Damit gewinnt die Triangulation (vgl. Kapitel 11.3.) zu ihrem Wert als Validierungsstrategie zusätzlich die Bedeutung, dass durch sie in einer Untersuchung bewusst unterschiedliche Forschungsperspektiven miteinander verbunden und unterschiedliche Aspekte des Untersuchungsgegenstandes thematisiert werden (Flick 1991: 433).

12.2.5. Datenauswertung

Im Rahmen qualitativer Analyseverfahren geht es in der Phase der Datenauswertung darum, das in einer offenen und exploratorischen Form erhobene Datenmaterial sozialwissenschaftlich korrekt und praktikabel auszuwerten. Wie bereits im Abschnitt über das problemzentrierte Interview ausgeführt wurde, ist es angezeigt, Befragungen auf Tonband aufzunehmen. Das auszuwertende Datenmaterial aus Interviews und Gruppendiskussionen liegt also meistens in Form zahlreicher Tonbandprotokolle vor. Es stellt sich vorerst die Frage, wie die Fülle an Material hinsichtlich Form und Umfang verarbeitbar gemacht werden kann.

Das Verfahren der Textinterpretation muss so konzipiert sein, dass, von der ursprünglichen Fragestellung geleitet, das Datenmaterial in einem hermeneutischen Interpretationsprozess zusammengefasst, strukturiert und verstanden werden kann - und das mit einem Aufwand, der zeitlich wie personell den Ressourcen des Evaluationsauftrages entspricht. Die qualitative Sozialforschung hat in der jüngeren Vergangenheit verschiedene Auswertungsverfahren entwickelt (v.a. aus dem Bereich der Sozialpsychologie), die den Verstehensprozess von Sinnstrukturen mit hoher Transparenz und Präzision entwickeln und abbilden können. Es handelt sich dabei aber zumeist um Verfahren, die einen enormen Arbeitsaufwand erfordern und die im Falle von politikorientierten Evaluationen ein zu starkes Gewicht auf die detaillierte Analyse von Sprachcodes und deren

kontextbezogene Bedeutung legen. In diese Verfahrenskategorie gehören etwa die Ansätze der objektiven Hermeneutik (Oevermann 1979) und die Textinterpretationen nach Soeffner (1979). Wir werden uns deshalb im folgenden auf ein Auswertungsverfahren konzentrieren, das dem Anspruch der Machbarkeit genügt und das vorhandene Textmaterial entsprechend dem Erkenntnisinteresse politikorientierter Evaluation unter den erforderlichen Gutekriterien zu interpretieren vermag. In der Praxis hat sich in dieser Hinsicht das Verfahren der "qualitativen Inhaltsanalyse" nach Mayring (1983) bewährt. Diese Konzeption der Datenauswertung umfasst drei Grundcharakteristiken: Zusammenfassung, Explikation und Strukturierung. Mayring (1993: 86) fasst diese Merkmale wie folgt zusammen:

"Zusammenfassung: Ziel der Analyse ist es, das Material so zu reduzieren, das die wesentlichen Inhalte erhalten bleiben, durch Abstraktion ein überschaubares Korpus zu schaffen, das immer noch Abbild des Grundmaterials ist.

Explikation: Ziel der Analyse ist es, zu einzelnen fraglichen Textteilen (Begriffen, Sätzen, ...) zusätzliches Material heranzutragen, das das Verständnis erweitert, das die Textstelle erläutert, erklärt, ausdeutet.

Strukturierung: Ziel der Analyse ist es, bestimmte Aspekte aus dem Material herauszufiltern, unter vorher festgelegten Ordnungskriterien einen Querschnitt durch das Material zu legen oder das Material aufgrund bestimmter Kriterien einzuschätzen." (Mayring 1993: 86).

Die Datenauswertung wird in mehreren Schritten durchgeführt. Als erster Schritt erfolgt die *Datenaufbereitung* in Form einer zweckmässigen *Transkription* der Tonbandprotokolle. Schon in diesem Verfahrensschritt geht es darum, die vorhandene Materialfülle zu reduzieren:

- Der abgehörte Text wird sprachlich bereinigt und geglättet: Übertragung vom Dialekt in die Schriftsprache; Satzbau- und Stilfehler werden behoben.
- Mehrfachnennungen, Wiederholungen werden zusammengefasst in ein einziges Zitat.
- Textstellen, die nichts mit der Fragestellung zu tun haben oder sehr nebensächlich sind, werden weggelassen.

In einem zweiten Schritt wird das gesamte transkribierte *Textmaterial* einem beschreibenden Kategoriensystem zugeordnet, d.h. einem Auswertungsraster, der aufgrund der Fragestellung entwickelt wird (Themenbe-

reiche, Evaluationskriterien). Dadurch erfolgt eine begründete Differenzierung der Fragestellung und eine Definition von Analysekatgeorien, die letztlich bestimmen, welches Material für die Interpretation Bedeutung hat. Die Definition dieser Kategorien muss so genau und differenzierend sein, dass alle Textstellen, die im Transkript aus dem ersten Auswertungsschritt stehen, schlüssig einer Kategorie zugeordnet werden können.

Durch die Selektion und Bündelung von Aussagen (Paraphrasen, Zitate) wird das Textmaterial unter den vorgegebenen Auswertungskategorien zusammenfassend geordnet. Dadurch erfolgt gleichzeitig eine fallbezogene Reduktion und eine inhaltliche Strukturierung.

In einem nächsten Schritt geht es darum, die einzelfallbezogene Darstellung fallübergreifend zu *paraphrasieren* und zu *generalisieren*, um so zu einer Gesamtdarstellung zu gelangen. Wichtig ist, dass im Anschluss daran die betroffenen Akteure dazu befragt werden, ob sie mit dieser paraphrasierenden Interpretation einverstanden sind, ob sie sich in den entstandenen Aussagen wiedererkennen. Durch diese Kommunikation zwischen den Befragten und den Evaluationsteams kann die Gültigkeit der erfolgten Interpretation überprüft werden. Diese kommunikative Validierung bietet gleichzeitig die Gelegenheit zur Explikation, die ein integrierter Bestandteil der qualitativen Inhaltsanalyse ist: Durch direkte persönliche Nachfragen bei den Befragten können zu interpretationsbedürftigen Aussagen und Textstellen zusätzliche Erklärungen und Erläuterungen in die Analyse miteinbezogen werden.

Der Schritt der *kommunikativen Validierung* kann im Rahmen einer Gruppendiskussion (im Rahmen von strukturierten Workshops - siehe oben) erfolgen. Wenn die Evaluatorinnen und Evaluatoren ihre Interpretationen und ihre zugrundeliegenden Vorannahmen und theoretischen Bezüge gegenüber den befragten Personen offenlegen, findet eine argumentative Validierung statt: Durch die Präsentation nachvollziehbarer Interpretationen, Generalisierungen und Argumentationen werden diese kritisierbar. Eine weitere Form der kommunikativen und argumentativen Validierung ist der Diskurs innerhalb der Forschergemeinschaft sowie durch den breiteren Einbezug der Öffentlichkeit in die Interpretationsgemeinschaft (Kvale 1991: 430). Ein solcher Einbezug der Öffentlichkeit bedingt natürlich eine entsprechende Berichterstattung und Präsentation der Evaluationsergebnisse.

Partizipative und interaktive Validierung einer von den Forscherinnen und Forschern angestellten Interpretation von Daten ist nicht gleichzusetzen mit einer Art demokratischer "Beschlussfassung über den Schlussbe-

richt" oder über die Evaluation selbst. Die Evaluation basiert auf kommunikativ validierten Wirklichkeitsinterpretationen; sie ist damit indessen nicht zu verwechseln. Denn die Evaluation ist und bleibt eine mit nachvollziehbarem Design und kommunikativ validierten Daten erstellte Bewertung von Programm- und Massnahmenwirkungen, die einzig und allein die beauftragten Forschungsteams zu verantworten haben. Die partizipative Erhebungsmethode ist *ein* mögliches Mittel zur Erhebung gesellschaftlicher Wahrheit über Wirkungszusammenhänge, die dem konstruktivistischen Paradigma verpflichtet ist und daher eine relativ grosse Chance hat, von den Akteuren akzeptiert zu werden (auch wenn sie sich später gegebenenfalls als "falsch" herausstellen sollte) (vgl. Kapitel 13.3.).

12.3. QUANTITATIVE VERFAHREN

Dieser Abschnitt geht davon aus, dass eine Politikevaluation streng am rationalistischen Paradigma orientierte, hypothesenüberprüfende **Vor**gehensweise verfolgt. Dies impliziert einige Annahmen: Es bestehen bereits ausreichende Kenntnisse über die Strukturierung des Untersuchungsfeldes. Es liegt zudem ein deduktiv hergeleitetes Modell **vor**, das sich aus verschiedenen Hypothesen über die Wirkungszusammenhänge zusammengesetzt. Die Festlegung dieser Prämissen erfolgt aus didaktischen Gründen, um das Kapitel von Themen zu entlasten, die in diesem Band an anderen Stellen behandelt werden (vgl. Kapitel 12.1.).

Es kann in diesem Abschnitt nicht darum gehen, eine Einführung in die quantitativen Datenerhebungs- und Auswertungsverfahren zu geben, wie sie bereits in einer grossen Vielfalt in diversen Lehrbüchern zur sozialwissenschaftlichen Methodik vorliegen. Vielmehr soll es darum gehen, auf die spezifisch bei der Umsetzung dieser Verfahren in der Evaluationsforschung bestehenden Fragen einzugehen.

12.3.1. Datenerhebung

Die Erhebung quantitativen Datenmaterials kann einerseits so erfolgen, dass bereits bestehendes Material für eine Evaluation zugänglich gemacht wird. Anderenfalls sind die quantitativen Daten im Rahmen der Evalua-

tion eigens zu erheben. Im ersten Fall spricht man von einer Sekundärerhebung, im zweiten von einer Primärerhebung.

Für *Sekundärerhebungen* können unterschiedliche Datenquellen beigezogen werden. Dabei kann zwischen folgenden Typen unterschieden werden:

- prozessgenerierte Datenbestände,
- amtliche Statistiken und
- forschungsgenerierte Datenbestände.

Unter *prozessgenerierten Datenbeständen* versteht man Daten, die im politischen Prozess entstehen, die also beispielsweise im Rahmen des Vollzugs politisch-administrativer Massnahmen anfallen. Dazu gehören Informationen, welche die vollziehende Stelle für ihre eigenen Bedürfnisse bewusst und absichtlich herstellt (etwa im Bereich der Vollzugskontrolle). Daten können aber auch beiläufig als Nebenprodukt der Verwaltungstätigkeit anfallen. Prozessgenerierte Datenbestände weisen als besonderen Vorteil auf, dass sie - die Kooperation der Verwaltungsstelle vorausgesetzt und unter Vorbehalt der Datenschutzbestimmungen - oft für eine längere Zeitperiode verfügbar sind. Sie sind oft auch in einer nicht aggregierten Form zugänglich und zudem im Normalfall recht verlässlich. Ihre Nachteile liegen darin, dass sie nicht nach den Bedürfnissen einer Evaluation, sondern aufgrund der Erfordernisse der vollziehenden (oder der planenden) Verwaltung gesammelt und aufgezeichnet werden. Weiter ist damit zu rechnen, dass das Datenmaterial oft nicht für längere Zeitperioden anhand derselben Erfassungsregeln festgehalten wurde, sondern dass verschiedentlich Modifikationen auftreten, welche die Vergleichbarkeit der Daten in Frage stellen können. Dies gilt nicht nur für Vergleiche über die Zeit, sondern ebenso für Querschnittvergleiche. Oft sind die Differenzen der Erfassungsart der Daten zur selben Thematik zwischen verschiedenen öffentlichen Institutionen sehr gross. Teilweise lassen sich zwar derartige Mängel durch adäquate Umformungen umgehen, häufig ist dies jedoch nicht möglich. Weiter ist damit zu rechnen, dass Daten aus früheren Jahren oft nicht mehr existieren oder lediglich in einer Form vorliegen, die einen grossen Aufwand zur Aufarbeitung erfordert.

Unter Informationen aus *amtlichen Statistiken* versteht man Datenmaterialien, die durch spezielle statistische Dienste verwaltungsintern und in

Ausnahmefällen auch verwaltungsextern erstellt werden. Diese Daten sind üblicherweise frei zugänglich und im Normalfall auch seit Erfassungsbeginn verfügbar. Auch hier besteht zwar oft die Schwierigkeit, dass im Datenmaterial über die Zeit hinweg Strukturbrüche auftreten können, welche auf einen Wechsel der Erfassungsmodalitäten oder der Fragestellung zurückzuführen sind. Im Gegensatz zu den prozessgenerierten Daten sind aber hier oft genauere Informationen zu den Umstellungen erhältlich, die eine Rekonstruktion erleichtern können. Nachteilig auswirken kann sich bei Daten aus amtlichen Statistiken, dass die Informationen oft nur in aggregierter Form vorliegen und sich damit - je nach Fragestellung der Evaluation - als wenig nützlich erweisen. Weiter ist darauf hinzuweisen, dass die amtliche Statistik in der Schweiz, trotz der massiven Verbesserungen in den letzten Jahren, im internationalen Vergleich nach wie vor als rückständig bezeichnet werden muss. Die Botschaft des Bundesrates zum Beitritt der Schweiz zum Europäischen Wirtschaftsraum (EWR) sah vor, die Verwaltung im Bereich der amtlichen Statistik um 157 Stellen aufzustocken, um in diesem Bereich "europäisches Niveau" zu erreichen (BB1 1992 IV: 527-528). Zu verschiedensten Themen bestehen auch heute noch erhebliche Lücken. Vorteilhaft zeigen sich die Daten aus amtlichen Statistiken hinsichtlich der Form der Verfügbarkeit. Nicht nur ist der Zugang normalerweise leicht zu bewerkstelligen, auch stehen die Daten oft in maschinenlesbarer Form zur Verfügung, was - je nach Datenmenge - für die Evaluation grosse Erleichterungen bringen kann.

Als *forschungsgenerierte Datenbestände* bezeichnet man Datenmaterial, das im Zusammenhang mit einem Forschungs- oder Evaluationsprojekt entstanden ist. Diese Datenbestände sind somit auf eine spezifische Fragestellung zugeschnitten. Es kann sich jedoch als zweckmässig erweisen, dieses Datenmaterial im Rahmen einer Evaluation einer sekundären Nutzung zuzuführen. Dafür müssen die Daten jedoch zumindest Teile der aktuellen evaluativen Fragestellung abdecken. Meistens ist diese Bedingung aber nicht vollumfänglich erfüllt, so dass die forschungsgenerierten Datenbestände mit weiteren Informationen ergänzt werden müssen. Es kann sich aber als sehr nützlich erweisen, auf diese Daten zurückzugreifen, da damit ressourcenintensive Erhebungen entfallen können. Problematisch ist bei dieser Datenquelle aber - leider - sehr oft der Zugang. Viele Primärforscherinnen und -forscher geben eigens erhobenes Datenmaterial nur sehr ungern an Dritte weiter. Oder sie tun dies erst zu einem Zeitpunkt, in dem das Material deutlich an Aktualität verloren hat. Dies-

bezüglich interessant sind die Datenvermittlungsdienste, welche der "Schweizerische Informations- und Daten-Archivdienst für die Sozialwissenschaften" (SIDOS) anbietet. Es ist zu hoffen, dass dadurch die Sekundärnutzung von forschungsgenerierten Datenbeständen (auch in der Schweiz) zunehmend erleichtert wird (vgl. dazu SIDOS 1994).

Sehr oft erweisen sich die zur Verfügung stehenden Daten aus den obengenannten Quellen nicht als ausreichend, um die Evaluation zu bearbeiten, so dass sich eine *Primärerhebung* aufdrängt. Dazu stellt die sozialwissenschaftliche Methodenlehre verschiedenste Instrumente zur Verfügung. Für Zwecke der Evaluationsforschung bieten sich dafür insbesondere folgende Erhebungsverfahren an:

- standardisierte Befragung
- Beobachtung / Messung
- quantitative Inhaltsanalyse
- Indizes und Skalen

Während bei den beiden erstgenannten Techniken eine eigentliche Datenproduktion erfolgt, zielen die beiden anderen darauf ab, bereits bestehende Quellen derart aufzuarbeiten, dass die Informationen einer quantitativen Datenauswertung zugeführt werden können. Die obige Aufzählung ist in keiner Weise als vollständig zu betrachten. Für weitere Erhebungstechniken, die sich ebenfalls für Evaluationen eignen können, wird auf die einschlägige Literatur verwiesen (bspw. Atteslander 1993; Friedrichs 1984; Kromrey 1994; Schnell, Hill und Esser 1993).

Die *standardisierte Befragung* ist wohl das bekannteste Verfahren zur sozialwissenschaftlichen Datenerhebung und auch in der Evaluationsforschung weit verbreitet (Weiss 1975). Standardisiert meint, dass allen Befragten die gleichen Fragen vorgelegt werden. Um die Befragungen vergleichbar zu machen, versucht man dabei die Bedingungen möglichst konstant zu halten. Zur Durchführung von standardisierten Befragungen stehen verschiedene Formen zur Verfügung:

- persönliche Befragungen
- telefonische Befragungen
- schriftliche Befragungen

Bei einer persönlichen Befragung werden die Daten in direkter Interaktion zwischen Interviewer und Befragten erhoben. Der Interviewte sucht dabei üblicherweise den Befragten zu Hause auf. Es kommt jedoch auch vor, dass solche Interviews an anderen Orten - also etwa auf der Strasse - stattfinden. Beim telefonischen Interview wird die Person am Telefon befragt. Dabei kommt in letzter Zeit häufig ein "Computer assisted telephone interviewing"-System, abgekürzt CATI, zum Einsatz. Solche Systeme erleichtern es den Interviewern, sich im Befragungsprozess zurechtzufinden, und ermöglichen den Einsatz komplexerer Fragebogenkonstruktionen. Ein weiterer Vorzug der CATI-Technik besteht darin, dass der Interviewer im Laufe des Telefongesprächs die Antworten in maschinenlesbarer Form erfassen kann, womit ein sonst notwendiger Arbeitsschritt eingespart werden kann. Bei einer schriftlichen Befragung werden die Fragebogen den Befragten - normalerweise per Post - zugestellt. Diese füllen die Fragebogen ohne die Unterstützung eines Interviewers selbst aus und retournieren den ausgefüllten Fragebogen. Ferner können diese Befragungsformen auch kombiniert werden, etwa so, dass im Anschluss an eine persönliche Befragung die befragte Person darum gebeten wird, einen ergänzenden schriftlichen Fragebogen auszufüllen. Es empfiehlt sich, die Auswahl der Befragungsform überlegt zu treffen, da mit den verschiedenen Formen unterschiedlichste Vor- und Nachteile verbunden sind (für genauere Erläuterungen siehe Weiss 1975; Holm 1991; GAO 1991b: 11-17).

Die Befragungstechnik ist nicht nur sehr weit verbreitet, sondern es wurden auch eine Vielzahl verschiedener Regeln geschaffen, um Befragungen valider und reliabler zu gestalten (siehe etwa die Hinweise in Atteslander 1993: 126-200; Schnell et al. 1993: 328-389; Holm 1991). Diese Regeln sollten beachtet und vor allem auch konsequent umgesetzt werden.

Beobachtung und Messung werden in der Evaluationsforschung relativ häufig eingesetzt, wobei deren Einsatz oft zu wenig reflektiert wird. Aufgrund des interdisziplinären (oder zumindest multidisziplinären) Charakters der Evaluationsforschung befassen sich Evaluationsstudien oft mit Untersuchungsgegenständen, welche im strengen Sinn nicht oder nur teilweise sozialwissenschaftlicher Art sind. Vor allem Fragestellungen aus den Gegenstandsbereichen der Naturwissenschaften sind häufig in Evaluationen anzutreffen. Man denke dabei etwa an die verschiedenen Projekte zur Abklärung der Wirksamkeit staatlicher Umweltschutzmassnahmen (so Brandner et al. 1995; Balthasar und Knöpfel 1994; Knoepfel und

Zimmermann 1987; Widmer 1991). Aber auch die Ingenieurwissenschaften sind - etwa im Bereich der Energiepolitik - stark vertreten (siehe etwa Linder, Looser, Trombitas und Wittwer 1990, Aebischer et al. 1993). Das gleiche gilt für die Medizin, die bei Wirksamkeitsüberprüfungen von gesundheitspolitischen Massnahmen wichtige Beiträge einbrachte (siehe bspw. Dubois-Arber, Jeannin, Meystre-Agustoni, Gruet und Paccaud 1993; Bahy 1986). Durch die Art der in der Evaluationsforschung zu bearbeitenden Fragestellungen drängen sich daher oft Verfahren wie Beobachtung oder Messung auf. Auch hier gilt, dass es sich um Verfahrenstechniken handelt, die eine lange Entwicklung hinter sich haben, in deren Verlauf das Instrumentarium verfeinert wurde. Dabei sind eine Vielzahl von Verfahrensregeln entstanden, die berücksichtigt werden sollten. Während die Beobachtung in einigen sozialwissenschaftlichen Disziplinen eine recht weite Verbreitung gefunden hat, sind eigentliche Messtechniken doch eher im Bereich der exakten Wissenschaften anzutreffen. Reichen dafür die Kenntnisse des Evaluators selbst nicht aus, ist es zweckmässig, für Datenerhebungen dieser Art externe Expertise einzuholen. Im Idealfall wird ein multidisziplinär zusammengesetztes Projektteam gebildet.

Die *quantitative Inhaltsanalyse* erlaubt eine Umsetzung verbaler Daten in quantitative Angaben, welche eine quantitative Auswertung erlauben. Als Rohdaten dienen jegliche Quellen in Textform, wie etwa Dokumente, Verwaltungsakten, Zeitungsartikel, Transkripte von Reden, Interviews, Diskussionen usw. Im Gegensatz zur qualitativen Inhaltsanalyse (vgl. Abschnitt 12.2.) geht es bei der quantitativen Inhaltsanalyse also darum, aus qualitativen Informationen quantitative Angaben zu generieren. Dies erfolgt zum Beispiel in der Form, dass ein gegebener Text entlang eines bestehenden theoretischen Codierschemas, welches die relevanten Kategorien umfasst, durchgearbeitet wird. Dabei werden die den Kategorien zugehörigen Begriffe, Sätze, Sequenzen oder andere Textteile (bewertet und) ausgezählt, womit numerische Angaben entstehen, die einer Datenauswertung zugeführt werden können. Das Ziel besteht darin, eine quantitative Erfassung eines Textes zu erreichen, welche intersubjektiv zu identischen Ergebnissen führt. In diesem Prozess gehen natürlich im Text enthaltene Informationen, wie etwa die Position einer bestimmten Aussage im Text, also der Kontext einer Aussage, verloren. In dieser Art lassen sich sowohl kürzere wie auch längere, zahlreiche wie auch nur wenige oder gar ein einziger Text analysieren. Für diese Zwecke existieren bereits verschiedene Computerprogramme, welche die Arbeit - je nach Fragestellung und Analyseform - massgeblich erleichtern können (siehe

dazu Weber 1990). Zur quantitativen Inhaltsanalyse besteht auch bereits eine ansehnliche Literatur, die vorteilhafterweise vor dem Einsatz dieser Technik zu konsultieren ist (Krippendorff 1980; Merten 1983; Lisch und Kriz 1978).

Indizes und Skalen erlauben es dem Evaluator, bestehendes quantitatives Datenmaterial zu inhaltlich miteinander verbundenen Indikatoren in einer Dimension zu erfassen. Das heisst, es werden verschiedene quantitative Indikatoren in einen Index oder eine Skala überführt, welche(r) die Informationen der Rohdaten möglichst gut repräsentieren soll. Der Unterschied zwischen Indizes und Skalen sei im folgenden kurz erläutert. Ein *Index* versucht, Indikatoren zu verschiedenen Teilaspekten zu einem Indikator (Index genannt) zusammenzufassen. Es liegen also empirische Angaben zu verschiedenen Teilkonzepten eines Konstrukts vor. Der resultierende Index soll das Gesamtkonstrukt möglichst gut darstellen. Die verschiedenen Einzelindikatoren werden dabei zumeist additiv, seltener multiplikativ miteinander verbunden. Das in der Schweiz wohl prominenteste Beispiel stellt der sogenannte Landesindex der Konsumentenpreise (Koch 1990: 283-297) dar, der aufgrund der Preisangaben verschiedenster Konsumprodukte (aus dem sogenannten Warenkorb) aufgrund einer festgelegten Gewichtung errechnet wird. Eine *Skala* hingegen misst ein Konstrukt zum Beispiel in einem Fragebogen - anhand von einzelnen Fragen (Items genannt), die jeweils das *gleiche* Konstrukt in unterschiedlich starken Ausprägungen erfassen. Zur Erstellung von Skalen stehen eine grosse Anzahl von unterschiedlichen Skalierungsverfahren zur Verfügung (Mokken-, Guttman-, Likert-Skala etc.; vgl. Bortz 1984: 149-159; Schnell et al. 1993: 193-221; Friedrichs 1984: 172-188). Für ähnliche Fragestellungen bestehen - neben Indizes und Skalen - weitere Verfahren wie die Cluster-Analyse oder die Faktorenanalyse, die auf komplexeren statistischen Modellen beruhen (Aldenderfer und Blashfield 1984; Backhaus, Erichson, Flinke und Weiber 1994; Bortz 1989: 684-705 bzw. 613-683; Kim und Mueller 1979).

Mit den hier kurz beleuchteten Erhebungsverfahren lassen sich quantitative Daten unterschiedlicher Art erheben. Von Bedeutung sind hier besonders die folgenden Eigenschaften:

- *Messniveau*: Das Messniveau gibt die Qualität der Messskala an, mit der die Indikatoren gemessen werden. Es wird (grob) unterschieden zwischen Nominal-, Ordinal- und Kardinalskalen. Bei nominalskalierten Daten ist lediglich die Unterscheidung der verschiedenen Ausprä-

gungen möglich (wie der Eigenname, die Parteizugehörigkeit, der Wohnort oder die Telefonnummer einer Person). Ordinale Skalen zeichnen sich dadurch aus, dass zusätzlich die verschiedenen Werte in eine Reihenfolge gebracht werden können (wie Schulnoten oder die soziale Schichtzugehörigkeit). Die Klassen der nominalen und ordinalen Messniveaus werden auch als kategorial bezeichnet. Kardinalskalierte Daten dagegen erlauben zudem, die Differenzen zwischen den verschiedenen Ausprägungen miteinander zu vergleichen (wie Alter, Einkommen, Temperatur). Diese Skalentypologie kann weiter differenziert werden. So umfasst die Kategorie der kardinalen Skalen sowohl Intervallskalen (wie Temperatur in °C), wo kein absoluter Nullpunkt besteht, wie auch Ratio-Skalen, welche über einen absoluten Nullpunkt verfügen (wie Einkommen, Temperatur in Grad Kelvin). Die Messniveaus der erfassten Indikatoren sind deshalb wichtig, weil sich die Interpretationsmöglichkeiten je nach Skalentyp stark unterscheiden. Weiter sind sie auch bei der Auswahl der adäquaten Analyseverfahren von grosser Bedeutung (vgl. Abschnitt 12.2.). Das Messniveau erhobener Daten kann zwar, beispielsweise mit Kategorisierungen, verändert werden, jedoch nur in einer Richtung. Es ist möglich, aus einer Kardinalskala eine Ordinal- oder Nominalskala zu generieren (was jedoch stets mit einem Informationsverlust verbunden ist). Es besteht jedoch ohne zusätzliche Informationen keine Möglichkeit, aus einer Nominalskala nachträglich eine Ordinal- oder Kardinalskala zu produzieren, da die im empirischen Material enthaltenen Informationen dazu nicht ausreichen.

- *Erhebungsebene:* Vor der Datenerhebung ist die Erhebungseinheit zu definieren. Es muss festgelegt werden, worauf sich die Variablen beziehen sollen. Es wird dabei grob in Individualdaten und Aggregatdaten unterschieden. Bei Individualdaten bildet ein Individuum (Person, Entscheidung, Parzelle etc.) die Erhebungseinheit (Mikroebene), bei Aggregatdaten dagegen eine Gruppe von Individuen, eine Institution (Mesoebene), eine Gemeinde oder ein Staat (Makroebene). Auch dieser Entscheid ist bereits in der Phase der Datenerhebung zu treffen und bestimmt massgeblich die Interpretationsmöglichkeiten. Aufgrund der Analyse von Aggregatdaten sind keine Interpretationen auf individueller Ebene möglich (sog. ökologischer Fehlschluss, vgl. dazu Robinson 1950; Erbring 1990). Es ist hingegen möglich, Individualdaten auf der Meso- oder Makro-Ebene zu aggregieren. Anders ausge-

drückt bedeutet dies, dass die Erhebungseinheit sehr wohl kleiner aber niemals grösser als die Analyse- oder die Aussageeinheit sein kann.

- *Erhebungsumfang:* Zu Beginn einer Datenerhebung ist zu entscheiden, ob Datenmaterial für alle möglichen Untersuchungseinheiten oder nur für eine Auswahl (eine sog. Stichprobe; vgl. Kapitel 11.1.) erhoben werden soll. Insbesondere wenn eine Grundgesamtheit mit einer grösseren Anzahl von Elementen vorliegt, wird in der quantitativen Sozialforschung sehr häufig mit Stichproben gearbeitet (siehe Kaplitza 1991; GAO 1992a; Bortz 1984: 239-362). Es bestehen verschiedene Verfahren zur Ziehung von Stichproben. Weiter muss festgelegt werden, wie gross die Stichprobe sein soll. Auch diese Entscheide haben einen Einfluss auf die Analyse des Datenmaterials und auf deren Interpretation. So sind bei Interpretationen von Stichprobendaten immer auch die statistischen Fehler im Auge zu behalten. Zudem erhalten die in der Analyse resultierenden Signifikanzen (zur Erläuterung vgl. Abschnitt 12.3.2.) von Analyseergebnissen in diesen Fällen eine weitreichende Bedeutung. In seltenen Fällen kommt es vor, dass nach der Durchführung einer Vollerhebung eine Stichprobenziehung erfolgt (z.B. Klöti und Risi 1991: 267-283).
- *Erhebungszeitraum:* Je nach gewähltem Untersuchungsdesign (vgl. dazu Kapitel 11) sind Daten zu nur einem Zeitpunkt oder zu verschiedenen Zeitpunkten (oder Zeiträumen) zu erheben. Der Zeitpunkt bzw. der Zeitraum sind präzise zu definieren. Soll eine zeitkontinuierliche Betrachtung erfolgen, ist zu entscheiden, in welchem Zeitintervall Daten erhoben werden sollen (Stunden-, Tages-, Wochen-, Monats-, Quartals- oder Jahresdaten). Die Auswahl des Erhebungszeitraums soll an die Fragestellung angepasst und adäquat zum gewählten Untersuchungsdesign erfolgen. Weiter ist darauf zu achten, dass kurzfristige Veränderungen (innert Tagen, Wochen oder Monaten) nur in unzureichender Form mit längerfristig angelegten Daten (Jahresdaten) erfasst werden können. In kurzfristigen Intervallen erhobene Daten können auf längere Intervalle aggregiert werden (z.B. Widmer 1991: 93-101).

Auch die Frage, welche Erhebungsart(en) im spezifischen Evaluationsprojekt zu wählen ist (sind), kann nur aufgrund der konkreten Fragestellung, der gegebenen Rahmenbedingungen und der empirischen Datenlage

beurteilt werden. Trotzdem lassen sich aber einige grundlegende Leitlinien formulieren.

- Bestehendes Datenmaterial, das zur Klärung der Fragestellung nützlich sein und mit vergleichsweise kleinem Aufwand erhoben werden kann, sollte berücksichtigt werden.
- Es sollte immer darauf geachtet werden, dass für die Auswertung des erhobenen Datenmaterials ausreichend Mittel zur Verfügung stehen. Es macht keinen Sinn, "Datenberge" zu produzieren, für deren Auswertung die Ressourcen nicht ausreichen.
- Datenerhebungen sollten immer in Abstimmung mit der Fragestellung erfolgen. Es hat keinen Sinn, Informationen zu erheben, die keinen Beitrag zur Analyse leisten können.
- Sekundärerhebungen sind im Regelfall günstiger und schneller durchführbar als Primärerhebungen.
- Primärerhebungen haben den Vorteil, dass sich die Datenerhebung exakt auf die Bedürfnisse der Evaluation ausrichten lassen. Bei Sekundärerhebungen ist dies nicht beeinflussbar.
- Auch bei Sekundärerhebungen ist der Entstehungszusammenhang des Materials bei der Datenauswertung und Interpretation zu berücksichtigen.

Erhobenes Datenmaterial ist hinsichtlich seiner Qualität mittels Gütekriterien (wie Validität und Rehabilität, vgl. dazu Schnell et al. 1993: 156-176; Cedro und Gualtierotti 1996; mit einem breiteren Ansatz Patton 1984) zu beurteilen. Dies gilt sowohl für Primär- wie auch für Sekundärerhebungen. Ist die Datenqualität aufgrund dieser Überprüfung nicht in ausreichender Weise gegeben, ist dies bei Analyse und Interpretation zu berücksichtigen. Im Extremfall muss auf eine Auswertung der Daten gänzlich verzichtet und auf anderes Datenmaterial ausgewichen werden. Weiter ist eine systematische Überprüfung des Datenmaterials durchzuführen (Plausibilität, Ausschliessen fehlerhafter Codes etc.), um zu vermeiden, dass die Evaluation fehlerhafte Angaben enthält. Zweckmäßigerweise erfolgt diese Überprüfung auch nach der Erfassung der Daten in maschinenlesbarer Form, weil diese Übertragung ebenfalls eine Fehlerquelle darstellen kann. Sind die Daten in einem EDV-System erfasst, kann die Datenüberprüfung mit EDV-Unterstützung erfolgen.

12.3.2. Datenauswertung

Es ist nicht möglich, in diesem Kapitel einen nur im Ansatz vollständigen Überblick über die diversen quantitativen Auswertungsverfahren zu bieten (vgl. dafür die einschlägige Einführungsliteratur, wie Bortz 1989; Wonnacott und Wonnacott 1990; Kriz 1983; GAO 1992b und die darin enthaltenen Literaturverweise). Zur Wahl der adäquaten Datenanalysetechnik hat sich eine Broschüre als äusserst hilfreich erwiesen, welche mittels eines Entscheidungsbaums zum passenden Verfahren führt (Andrews, Klem, Davidson, O'Malley und Rodgers 1981). Im folgenden soll es darum gehen, einige häufig eingesetzte Analysetechniken exemplarisch herauszugreifen und gewisse mit diesen Verfahren in Zusammenhang stehende Aspekte aufzugreifen. Weiterhin wird versucht, jeweils beispielhaft auf praktische Umsetzungen dieser Verfahren in der (schweizerischen) Evaluationsforschung hinzuweisen. Zur Strukturierung des Kapitels verwenden wir folgende drei Dimensionen in der angeführten Hierarchie:

- *Vergleichsdimension*, also Querschnitt-, Längsschnittanalysen und kombinierte Verfahren,
- *Dimensionalität der Analyse*, also univariate, bivariate und multivariate Analysen und
- *Messniveau des Datenmaterials*, also Verfahrenstechniken zu nominalem, ordinalem und kardinalen Datenmaterial⁴.

Diese Dimensionen sind in keiner Weise als erschöpfend zu betrachten, genügen aber für die vorgesehene kurze Übersicht. Die nachfolgende Tabelle zeigt mögliche Verfahren bei Querschnittanalysen:

⁴ Hierzu drängt sich noch eine Zusatzbemerkung auf. Ist eine Analyse mit Daten vorgesehen, welche über unterschiedliche Messniveaus verfügen (bspw. ordinale und kardinale Variablen), ist für die Wahl des Analyseverfahrens im Regelfall das "niedrigste" Messniveau (in unserem Beispiel die Ordinalskala) ausschlaggebend.

Abbildung 8 Dimensionalität und Messniveaus ausgewählter Verfahren in Querschnittanalysen

	<i>nominal</i>	<i>ordinal</i>	<i>kardinal</i>
<i>univariat</i>	Häufigkeitsauszählung (mit Prozentuierung), Modalwert		Masse der zentralen Tendenz, Dispersionsmasse
<i>bivariat</i>	Kreuztabellen- analyse, ϕ (Phi) bzw. Cramers V und λ (Lambda) u.a., χ^2 (Chi-Quadrat)	Kreuztabellen- analyse, Yule's Q bzw. Gamma (γ), Kendalls tau-b, Stuarts tau-c, Somers d, χ^2 (Chi-Quadrat)	Korrelations- analyse (r), bivariate Regressions- analyse (R^2)
<i>multivariat</i>	Log-lineare Modelle, Latente Klassenanalyse (LCA), Lineare Strukturgleichungsmodelle (LISREL, Version 8)		multiple Regression, Kausalanalyse, Lineare Struktur- gleichungsmodelle (LISREL)

Wie aus der Tabelle ersichtlich wird, grenzen die Dimensionalität der Analyse und die Messniveaus der erhobenen Daten die Auswahl an möglichen Analyseverfahren bereits stark ein. Dies sei an der *univariaten Analyse* mittels Massen der zentralen Tendenz (wie arithmetisches Mittel, Median oder Modalwert) erläutert. Wie aus der Tabelle hervorgeht, ist der Modalwert bei ordinalen oder nominalen Daten das einzige hier zulässige Mass. Die Berechnung des arithmetischen Mittels ist hier also zum Beispiel nicht zulässig. Verwendet man zur Analyse von Häufigkeitsauszählungen prozentuierte Angaben, ist darauf zu achten, dass die zugrunde liegenden Häufigkeiten ausreichend gross sind. Ein absolutes MUSS besteht dabei in der Angabe der Fallzahl (N resp. n), der sog. Prozentuierungsbasis. Zur Analyse der Streuung empirischer Verteilungen auf kardinaler Ebene bietet sich eine Vielzahl verschiedener Masse an. Genannt sind hier exemplarisch die Varianz (s^2), die Standardabweichung (s), verschiedene Quantile (Quartile, Zentile etc.), die Streubreite ("ränge") und der Variationskoeffizient (CV). Diese Masse haben jeweils

ihre spezifische Charakteristik. So dient s üblicherweise dazu, zu beurteilen, wie gross die Streuung im Verhältnis zu den absoluten Werten ist. Sie hat gegenüber anderen Massen den Vorzug, dass sie dieselbe Einheit aufweist wie die absoluten Werte. Möchte man dagegen zwei empirische Verteilungen bezüglich ihrer Dispersion miteinander vergleichen, ist dies mit s nur möglich, wenn die Mittelwerte der beiden Verteilungen gleich gross sind. In den übrigen Fällen ist der Vergleich verzerrt. Deshalb wird für derartige Vergleiche üblicherweise der CV verwendet. Dieser Koeffizient ist sog. mittelwertbereinigt, d.h., das Mass ist um die Mittelwerte der Verteilungen korrigiert. Die univariaten Verfahren können im Kontext der Evaluationsforschung einerseits als eigenständige Analyseverfahren eingesetzt werden (für Beispiele dazu siehe etwa Gerheuser und Schmid 1993: 12-18; Schulz, Muggli und Hübschle 1993: 73-119; Balthasar und Knöpfel 1994: 104-118). Andererseits ist es sehr empfehlenswert, die erhobenen Daten vor einer bi- oder multivariaten Analyse einer ausführlichen univariaten Analyse zu unterziehen. Erstens können damit Erkenntnisse gewonnen werden, die in der mehrdimensionalen Analyse verborgen bleiben, andererseits bildet die univariate Analyse eine notwendige Vorarbeit für weitergehende Analyseschritte (Beispiel: Überprüfung der Verteilungsannahmen).

Mit den *bi- und multivariaten Verfahren* werden Zusammenhänge zwischen verschiedenen Variablen untersucht. Es sei an dieser Stelle deutlich darauf hingewiesen, dass (alleine) mit empirischen Datenauswertungen nicht auf *kausale* Zusammenhänge geschlossen werden kann. Die empirische Analyse ist lediglich in der Lage, eine theoretisch hergeleitete Kausalität zu überprüfen. Die *Kreuztabellenanalyse* ist im Prinzip eine zweidimensionale Häufigkeitsauszählung. Es wird überprüft, ob zwischen kategorialen (also nominalen bzw. ordinalen) Variablen ein empirischer Zusammenhang vorliegt. In der vorangehenden Abbildung 8 finden sich die für das jeweilige Messniveau adäquaten Assoziationskoeffizienten. Diese erlauben Aussagen über die Stärke eines Zusammenhangs. Zur Feststellung der Signifikanz eines Zusammenhangs wird üblicherweise ein sogenannter χ^2 -Test durchgeführt. Dieser lässt die Beurteilung zu, ob ein bestimmter Zusammenhang auf Zufall beruht oder ob eine strukturelle Regelmässigkeit vorliegt. Aussagen dieser Art sind nie "sicher", sondern immer einer bestimmten Irrtumswahrscheinlichkeit ausgesetzt. Die Grosse dieser Fehlerwahrscheinlichkeit, welche man in Kauf zu nehmen bereit ist, wird mit dem sog. Signifikanzniveau festgelegt. Im sozialwissenschaftlichen Bereich wird üblicherweise von fünf- oder einprozentigen

Irrtumswahrscheinlichkeiten ausgegangen (in Kurzschreibweise $p < .01$ oder $p < .05$). Handelt sich bei dem untersuchten Datenmaterial um eine Zufallsstichprobe (oder annähernd um eine solche) und erweist sich ein Zusammenhang als signifikant, kann das Analyseergebnis mit der festgelegten Irrtumswahrscheinlichkeit auch auf die Grundgesamtheit übertragen werden. Analoge Signifikanztests - wie das χ^2 -Mass bei Kreuztabellen - bestehen auch bei den anderen Verfahren zur Überprüfung von Zusammenhängen (bei der Korrelationsanalyse wird üblicherweise eine t-Test-Statistik berechnet).

Die *Korrelationsanalyse* verfolgt im Prinzip den gleichen Zweck wie die Kreuztabellenanalyse. Auch hier geht es um die Überprüfung bivariater Zusammenhänge, allerdings bei kardinalen Daten. Die sog. Produkt-Moment-Korrelation nach Pearson ist das am weitesten häufigsten verwendete Verfahren in der Korrelationsanalyse. Es resultiert ein Korrelationskoeffizient, der Pearsons r genannt wird. Dieser Koeffizient lässt Aussagen über die Stärke des empirischen Zusammenhangs zu. Die Korrelationsanalyse geht von der Annahme aus, dass der Zusammenhang eine lineare Form aufweist. Das bedeutet, dass sich die Werte der beiden untersuchten Variablen proportional zueinander verhalten. Dies kann mittels eines Streudiagramms optisch überprüft werden. Der resultierende Punkteschwarm, ein Punkt steht für eine Beobachtung, sollte annäherungsweise linear verlaufen. Auf die Überprüfung der Linearitätsannahme sollte keinesfalls verzichtet werden. Nicht-lineare Zusammenhänge sind im Bereich der Evaluationsforschung relativ häufig anzutreffen und die Auswirkungen eines Verstosses gegen diese Annahme verzerrt die Resultate massiv.

Quasi eine Erweiterung der Korrelationsanalyse bildet die *bivariate Regressionsanalyse*. Mit Hilfe dieser Methode versucht man eine sog. Regressionsgerade zu schätzen, welche den Punkteschwarm im Streudiagramm möglichst weitgehend repräsentiert. Aus einer Regressionsanalyse resultiert ein sogenannter Determinationskoeffizient R^2 , welcher Aussagen darüber erlaubt, zu welchem Anteil die Varianz in einer Variable (als abhängige (oder endogene) Variable (y) bezeichnet) durch die andere Variable (als unabhängige (oder exogene) Variable (x) bezeichnet) "erklären" kann. Es handelt sich auch hierbei nicht um eine kausale, sondern lediglich um eine statistische Erklärung. Weiter wird auch die Steigung (mit b bezeichnet) der Geraden geschätzt. Anhand des Schätzwertes b lassen sich Folgerungen formulieren, wie beispielsweise: "Wenn um eine Einheit ansteigt, verändert sich y um den Wert b ." Bivariate Regressionsanalysen lassen sich auch durchführen, wenn eine nicht-lineare Be-

ziehung vorliegt, oder wenn andere Verstösse gegen die Grundannahmen einer Regressions-schätzung vorliegen. Das konventionelle Schätzverfahren (Methode der kleinsten Quadrate ("ordinary-least-squares", abgekürzt ÖLS) ist dazu jedoch weniger gut geeignet als andere, komplexere Schätzverfahren (siehe dazu Gujarati 1988; Stewart 1991; für Ambitionierte: Johnston 1972). Es bestehen auch Weiterentwicklungen der Regressionsmethodik, die überdies die Auswertung nicht-kardinaler Daten ermöglichen (so etwa die Logit-Analyse, siehe Urban 1993; Demaris 1992). Ein Beispiel für die Anwendung höher entwickelter Regressionsverfahren findet sich in einer Evaluationsstudie zu den Auswirkungen der beruflichen Vorsorge auf den Arbeitsmarkt (Schaetti 1990).

Im Bereich der *multivariaten Analyse* wird hier nur gerade die multiple Regressionsanalyse erläutert. Für die übrigen Verfahren sei auf die Literatur verwiesen (Backhaus et al. 1994; Log-lineare Modelle: Fienberg 1980; Knoke und Burke 1980; Fingleton 1984; Ishii-Kuntz 1994; für Anwendungen vgl. Keller 1991; Widmer und Buri 1992; Latente Klassenanalyse (LCA): McCutcheon 1987, für eine Anwendung siehe Keller 1991; Lineare-Strukturgleichungsmodelle (LISREL): Bollen 1989; Bollen und Long 1993; Kühnel 1993; Saris und Stronkhorst 1984). Die multivariate (oder multiple) Regressionsanalyse dehnt das bivariate Regressionsmodell weiter aus, indem nicht nur eine unabhängige Variable (x), sondern mehrere exogene Faktoren ($x_1, x_2, \dots, x_n$) in das Modell integriert werden. In der Evaluationsforschung können damit Analysen durchgeführt werden, welche nicht nur den Zusammenhang zwischen einer staatlichen Massnahme (bspw. x_1 und einer Zieldimension (y) untersuchen, sondern auch Dritt-Variablen ($x_2, x_3, \dots, x_j$, die ebenfalls einen Einfluss auf die Zieldimension ausüben könnten, in die Analyse zu integrieren (vgl. Kapitel 3.3.). Das einfachste Regressionsmodell geht von einer relativ grossen Anzahl von Annahmen aus, welche die Anwendungsmöglichkeiten in sozialwissenschaftlichen Evaluationsforschungen deutlich einschränken. Hinzuweisen ist hier auch auf das Erfordernis, dass zur Schätzung von Regressionsgleichungen eine ausreichende Zahl von Beobachtungen zur Verfügung steht. Die Zahl der Freiheitsgrade (df) muss im absoluten Minimum 20 betragen, was bedeutet, dass für ein Modell mit gerade drei exogenen Faktoren (x_1, x_2 und x_3) mindestens 24 Fälle zur Verfügung stehen müssen. Ein Beispiel für eine Anwendung des multivariaten Regressionsmodells in der Evaluation findet sich in der Evaluationsstudie zu den technologischen Auswirkungen der umweltpolitischen Vorschriften im Bereich der Hausfeuerungen, die durch Andreas Balthasar und Carlo

Knöpfel (1994: 99-114) erarbeitet wurde. Weitere evaluative Anwendungen finden sich in zwei Arbeiten zur Wirkungsüberprüfung der Kehrrechtsackgebühren (Hausheer 1991; Schwarz und Brozovic 1992).

Bei *Längsschnittanalysen* sind im Prinzip dieselben Verfahren wie bei der Querschnittanalyse einsetzbar, wobei sich verschiedene der diskutierten Verfahren besser im Querschnittbereich umsetzen lassen. Die Verfahrensauswahl wird im Bereich von Längsschnittanalysen durch einige, eigens für Untersuchungen dieser Art entwickelte Verfahren ergänzt. Die Analyse in der Längsschnittdimension beinhaltet zusätzlich zu den Problemen, die auch im Querschnittbereich auftreten, weitere Schwierigkeiten (siehe dazu ausführlicher Menard 1991). An erster Stelle ist dabei darauf hinzuweisen, dass es sich bei Daten, die entlang der Zeitachse orientiert sind, um Beobachtungen handelt, welche nicht unabhängig voneinander sind. So besteht oftmals ein sehr starker Zusammenhang von Datenwerten zum Zeitpunkt t mit den Angaben zum vorhergehenden Zeitpunkt ($t-1$) oder zur nachfolgenden Beobachtung ($t+1$). Diesem Umstand können viele der oben beschriebenen Querschnittverfahren zu wenig Rechnung tragen. Diese gehen oft davon aus, dass die Daten aus voneinander unabhängigen Beobachtungen stammen. Bei einem Verstoss gegen die Annahmen führen diese Verfahren zu verzerrten Resultaten. Abbildung 9 enthält eine Zusammenstellung einiger Analyseverfahren zur Längsschnittanalyse. Auch hier ist deutlich darauf hinzuweisen, dass die in der Tabelle enthaltene Zusammenstellung nur eine kleine Auswahl der möglichen Verfahren darstellt.

Abbildung 9 Datenerfordernisse ausgewählter Verfahren in Längsschnittanalysen

<i>Verfahren</i>	<i>Anzahl Datenpunkte</i>	<i>Messniveau</i>	<i>Fehlende Werte</i>
<i>graphische Analyse</i>	frei	frei	möglich
<i>Vorher/Nachher-Vergleich</i>	mind. 1 Prä- und 1 Postintervention	frei	möglich
<i>Lebensverlaufs-analyse</i>	vollständiger Lebensverlauf	kategorial (nominal & ordinal)	nicht möglich
<i>Regressionssätze</i>	über 20 Datenpunkte (df min. 20)	kardinal/ (kategorial)	problematisch
<i>Box-Jenkins-Ansätze</i>	mind. 30 Prä- und 30 Postintervention	kardinal	problematisch
<i>Pooled Time Series</i>	NxT über 30	kardinal/ (kategorial)	problematisch

Die Tabelle enthält zu jedem der genannten Verfahren die Anforderungen, die an das Datenmaterial gestellt werden: die Anzahl der Datenpunkte, welche für die Analyse zur Verfügung stehen müssen, das erforderliche Messniveau der abhängigen Variable (y), also der Zieldimension, sowie die Flexibilität des Verfahrens in Hinblick auf den Umgang mit fehlenden Werten im Datenmaterial.

Ambivalent einsetzbar und für jede weitergehende Analyse unverzichtbar ist die optische Auswertung der vorhandenen Daten. Aufgrund einer *graphischen Auswertung* sind vielfach schon interessante Erkenntnisse möglich. Bei der graphischen Darstellung von Daten ist darauf zu achten, dass die zur Anwendung gelangenden Massstäbe keine verzerrende Wirkung aufweisen. Insbesondere ist wenn immer möglich auf logarithmische Skalen und auf die Aussparung von Skalenteilen zu verzichten. Anwendungsbeispiele zu diesem Verfahren finden sich in Evaluationsstudien sehr häufig. Exemplarisch sei hier auf die Studie zur Wohneigentumsförderung

des Bundes hingewiesen, welche diese Technik einsetzt (siehe Schulz, Muggli und Hübschle 1993).

Vorher/Nachher-Vergleiche mit oder ohne Kontrollgruppe(n) sind in Evaluationen häufig anzutreffen (siehe etwa INFRAS 1990). Aufgrund der geringen Datenerfordernisse - es muss lediglich eine Angabe vor der Einführung einer Massnahme und eine für die Zeit nach der Einführung vorliegen - ist auch dieses Verfahren oft verwendbar. Es besteht die Möglichkeit, eine Vorher/Nachher-Untersuchung mittels statistischer Analysen durchzuführen. Dabei stehen Verfahren im Vordergrund, welche testen, inwiefern die Differenzen von der Vor-zur Nachinterventionsperiode signifikant sind. Diese Analysen können sowohl das Niveau (Mittelwertvergleiche) als auch die Streuung (Varianzvergleiche) der abhängigen Variablen über die Zeit hinweg betreffen. Problematisch sind derartige Untersuchungen vor allem dann, wenn nur wenige Datenpunkte vorliegen. Insbesondere bei stark volatilen Datenreihen oder wenn ein Trend vorliegt, kann diese Untersuchungsanlage zu verzerrten Resultaten führen. Empfehlenswert zur Erhöhung der Reliabilität und Validität ist sicherlich immer der Einbezug von Kontrollgruppen in unterschiedlichster Form (vgl. dazu die ausführliche Darstellung in Cook und Campbell 1979: 95-146; zur Analyse vgl. Reichardt 1979).

Die *Lebensverlaufsanalyse* (Ereignis(daten)analyse, "event history analysis") ermöglicht die Analyse von Längsschnittdaten mit kategorialer abhängiger Variable. Verbreitet ist diese Untersuchungsform bisher vor allem im Bereich der quantitativ ausgerichteten, soziologischen Biographieforschung. Für weitere Hinweise sei auf die Literatur verwiesen (Blossfeld, Hamerle und Mayer 1986; Diekmann und Mitter 1984; Yamaguchi 1991).

Der *Regressionsansatz* ist nicht nur bei Querschnitt-, sondern auch in Längsschnittuntersuchungen einsetzbar. Für die Schätzung von Zeitreihenregressionen kann jedoch zumeist nicht auf die OLS-Schätzprozeduren zurückgegriffen werden, sondern es sind andere Verfahren zu verwenden. In der Ökonometrie ist eine Vielzahl verschiedener Techniken zur unverzerrten Analyse von Zeitreihendaten entwickelt worden. Diese setzen jedoch zumeist ein tiefergehendes Verständnis der Schätzmodelle voraus, die in der Literatur ausführlich erläutert sind (Ostrom 1990; Gujarati 1988; Stewart 1991; Johnston 1972). Als Anwendungsbeispiel hierzu sei auf die Evaluationsstudie zu den Auswirkungen der Ausländerpolitik auf den Strukturwandel in der Schweiz hingewiesen (Stalder, Frick, Gaillard und Salzgeber 1994).

Ein reines Zeitreihenverfahren, das kein Äquivalent im Querschnittbereich kennt, stellt die Interventionsanalyse im Rahmen der *Box-Jenkins-Ansätze* dar. Diese, auch unter der Bezeichnung ARIMA-Modelle (autoregressive, integrierte, moving-average-Modelle) bekannten Analyseverfahren gehen von den in Zeitreihen enthaltenen Strukturen, wie Trends, Zyklen und Saisonalitäten aus, und versuchen so die Effekte einer Intervention zu erfassen (vgl. dazu Box und Jenkins 1976; McCleary und Hay 1980; Schmitz 1989; Wei 1990). Diese Methodik kann sowohl in bivariater (Interventionsanalyse) wie auch in multivariater Form (Transferfunktionsanalyse) eingesetzt werden. Der grosse Vorteil dieser Methode besteht darin, dass auch komplexere Interventionen und Wirkungsverläufe erfasst werden können und dass, im Gegensatz zu anderen Verfahren, die spezielle Charakteristik von Zeitreihendaten beachtet wird. Nachteilig wirken sich vor allem zwei Faktoren aus, die den Einsatz dieses Verfahrens oft verhindern. Erstens sind die Erfordernisse, welche an das empirische Datenmaterial gestellt werden, sehr hoch (vgl. Abbildung 9). Zweitens ist das Verfahren in der Umsetzung doch recht anspruchsvoll und zeitintensiv. Als Beispiele für die Anwendung dieses Verfahrens sei auf eine Studie zur Evaluation luftreinhaltepolitischer Massnahmen in der Schweiz (Widmer 1991) verwiesen.

Abbildung 9 enthält ein weiteres Verfahren, das jedoch kein reines Längsschnittmodell darstellt. Es handelt sich dabei um die *"pooled time series"*, ein Verfahren, das eine kombinierte Längsschnitt-Querschnittanalyse ermöglicht (vgl. dazu Stimson 1985; Sayrs 1989). Aufbauend auf dem Regressionsansatz hat dieses Verfahren den Vorteil, dass in Situationen, in denen nur in einer Vergleichsdimension keine befriedigende statistische Analysemöglichkeit bestehen würde, trotzdem eine statistische Analyse ermöglicht wird. Sehr häufig zeigt sich nämlich in der Praxis das Problem, dass keine ausreichende Zahl von Fällen (im Querschnitt) oder von Beobachtungszeitpunkten (im Längsschnitt) zur Verfügung stehen, um z.B. eine Regressionsanalyse durchzuführen. Durch eine Kombination des Datenmaterials der beiden Vergleichsdimensionen ist die Zahl der Beobachtungen jedoch ausreichend gross, um eine zuverlässige Schätzung zu ermöglichen. Auch diese Methodik ist in der Umsetzung nicht immer einfach. Durch die dimensionale Erweiterung der Analyse erhöht sich auch deren Komplexitätsgrad. Ein Beispiel für die Umsetzung findet sich in einem Artikel, der sich mit den Auswirkungen der Arbeitslosigkeit auf den Wertewandel befasst (Clarke und Dutt 1991).

Die Vergleiche in der Zeitdimension haben, trotz allen Schwierigkeiten bei deren Implementation, einige gewichtige Vorteile (vgl. dazu Menard 1991: 5-21), die gerade in der Evaluationsforschung von grösstem Interesse sein können. Erwähnt sei an dieser Stelle etwa die analytische Trennung von Alters- und Zeiteffekten, welche nur in der Zeitdimension zu realisieren ist, die Bedeutung der zeitlichen Ordnung bei der Analyse von Wirkungsketten oder die Dynamik sozialer Prozesse, die mit Querschnittanalysen nicht ausreichend erfasst werden kann. Die Evaluationsforschung untersucht zumeist Veränderungshypothesen, und Veränderungen sind wohl am zweckmässigsten im Zeitverlauf analysieren.

Die hier vorgestellten Analyseverfahren stellen sehr unterschiedliche Anforderungen an das vorhandene Datenmaterial und an die Kenntnisse des Analytikers. Weiterhin ist die Verfahrensauswahl auf die Fragestellung und auf den Themenbereich abzustimmen. Es ist im Auge zu behalten, dass Analysetechniken keineswegs neutrale Werkzeuge darstellen. Praktisch alle quantitativen Verfahren gehen von mehr oder weniger engen Prämissen aus, welche in der Interpretation nicht vergessen gehen sollten. Evaluationen entstehen häufig unter engen zeitlichen und finanziellen Rahmenbedingungen. Aus dieser Konstellation heraus besteht die Gefahr, dass Analyseverfahren nicht sachgerecht eingesetzt werden oder dass die Ergebnisse der Analysen über- oder gänzlich falsch interpretiert werden. Beherrschen die Evaluatoren oder die Evaluatorinnen eine Verfahrenstechnik, sind sie versucht, diese Kompetenz auch in nachfolgenden Evaluationen einzusetzen, auch wenn dies von der Sache her wenig zweckmässig erscheint. Dadurch kann die Verfahrenstechnik zum Selbstzweck werden. Die Wahl der Evaluierenden bestimmt damit die eingesetzte Methodik und nicht - wie dies zu wünschen wäre - die zu untersuchende Fragestellung. Andererseits ist auch ein Evaluationsteam nicht in der Lage, alle dargestellten Verfahren (und die grosse Zahl darüber hinaus bestehender Techniken) zu beherrschen. Weil kaum Zeit besteht, sich für eine spezifische Evaluation in eine gänzlich neue Methodik einzuarbeiten und diese dann auch noch in hoher Qualität durchzuführen, scheint es oftmals angebracht, sich das nötige Know-how durch Beizug einer Expertenperson zu beschaffen.

Zum Abschluss seien noch vier Empfehlungen formuliert:

- Die Qualität des Methodeneinsatzes ist entscheidend. Im Zweifelsfall scheint es deshalb empfehlenswert, sich für die "einfachere" Analyse-

technik zu entscheiden und sich dabei um eine möglichst hochstehende Umsetzung zu bemühen.

- Sehr oft bauen quantitative Datenanalyseverfahren aufeinander auf, sind also hierarchisch organisiert. Es empfiehlt sich, nicht direkt die komplexeste Analysetechnik anzusteuern, sondern zuerst die zugrundeliegenden Verfahren zu verwenden, um anschliessend die vorgesehene komplexe Methodik umzusetzen. Beispielsweise ist es von Vorteil, zuerst eine gründliche univariate Analyse des Datenmaterials vorzunehmen, bevor man sich auf ein multivariates Verfahren konzentriert.
- Lassen die Rahmenbedingungen dies zu, ist es zweckmässig, die Evaluation nicht nur auf eine Auswertungsmethodik auszurichten, sondern verschiedene Analyseverfahren einzusetzen.
- Bei der Formulierung der Schlussfolgerungen einer Evaluation sollte man sich immer vergewissern, ob man dabei alle vorgenommenen Annahmen berücksichtigt und alle empirischen Unzulänglichkeiten, welche jede Evaluationsstudie enthält, einbezogen hat.

12.3.3. Darstellung der Ergebnisse

Die Evaluationsforschung hat verschiedene Qualitätskriterien entwickelt, welche Normen zur Durchführung und zur Darstellung evaluativer Untersuchungen enthalten (siehe z.B. Joint Committee 1994; vgl. dazu auch Widmer 1996). Im Zusammenhang mit der Darstellung der Anwendung quantitativer Verfahrenstechniken und deren Ergebnisse sind dabei vor allem die folgenden Aspekte zu beachten:

- *Klarheit des Berichts:* Evaluationsteams sollten darauf bedacht sein, den Evaluationsbericht in einer Form abzufassen, die es erlaubt, die verwendeten Verfahren zu identifizieren und einzuschätzen. Die verwendeten Techniken müssen detailliert und eindeutig beschrieben werden. Zur Entlastung des Berichts kann dies auch in einem methodischen Anhang geschehen. Es sollte das Ziel eines Evaluationsberichts sein, die durchgeführten Analysen für den Leser nachvollziehbar auszugestalten. Dies bedingt, dass Entscheide, welche im Untersu-

chungsverlauf getroffen werden, nicht nur dargelegt, sondern auch explizit begründet werden.

- *Offenheit in der Darlegung:* Bei der Darstellung der Evaluation sollten auch Probleme oder Schwierigkeiten, die im Evaluationsprozess entstanden, zur Sprache kommen. Insbesondere gilt es auch auf die Konsequenzen derartiger Aspekte hinzuweisen. Ebenso ist es unumgänglich, dass der Evaluationsbericht klare Hinweise auf die Grenzen der Aussagekraft der Evaluation enthält.
- *Begründung der Folgerungen:* Der Evaluationsbericht sollte es dem Leser ermöglichen, einen Bezug zwischen empirischer Analyse und angebrachten Schlussfolgerungen herzustellen. Enthält eine Evaluation weitergehende Folgerungen und Empfehlungen, welche sich nicht in direkter Weise mit den empirischen Analysen begründen lassen, ist dies deutlich hervorzuheben.

Diese Hinweise gelten natürlich nicht nur für schriftliche Evaluationsberichte, sondern sind in analoger Weise auch bei einer Berichterstattung in anderer Form (mündliche Berichterstattung, Video etc.) zu beachten (vgl. dazu Morris, Fitz-Gibbon und Freeman 1987; Wottawa und Thierau 1990: 156-164; Passow 1987).

Die Anwendung quantitativer Verfahren im Bereich der Evaluationsforschung steht in einem Spannungsfeld zwischen methodischer Präzision und Anschlussfähigkeit bei den Adressaten. Gemeint ist damit die Problematik, dass höher entwickelte und vielleicht adäquatere oder aussagekräftigere statistische Verfahren gegenüber der Praxis schwerer mitteilbar werden (vgl. Kapitel 10). Daraus ergibt sich die Frage, inwiefern Evaluationsverfahren durch die Adressaten überhaupt nachvollziehbar sein sollen. Hier bestehen in der Literatur unterschiedliche Auffassungen. Für das Primat der Anschlussfähigkeit sprechen teilweise die Erkenntnisse aus der Forschung zur Nutzung von Evaluationsergebnissen oder von sozialwissenschaftlichem Wissen im allgemeinen (siehe etwa Alkin 1990: 19-78; Beck und Bonss 1989; Wittrock 1991; Chen 1994; Shadish 1994; und natürlich: Weiss 1981; 1987; 1993). Dagegen sprechen die Erfahrungen aus anderen, nicht-sozialwissenschaftlichen Disziplinen mit anwendungsorientierter Charakteristik. Aber auch dort wird vermehrt eine "Kontrolle" der Forschungstätigkeiten eingefordert.

Evalutationsberichte haben sich in jedem Fall darum zu bemühen, den Prozess, die Prozeduren und Verfahren und die Ergebnisse einer Evaluation möglichst klar, offen und nachvollziehbar darzulegen, also Transparenz zu schaffen. Nur so kann auch erreicht werden, dass der Evaluation die ihr zukommende Beachtung geschenkt wird. Dadurch wird weiter auch die Möglichkeit, dass Evaluationen missbraucht werden, eingeschränkt. Die Qualität der Berichterstattung ist sowohl durch sprachliche, formale, als auch durch inhaltliche und methodische Aspekte geprägt. Die Möglichkeiten für eine übersichtliche Gestaltung des Berichts (Stichworte: graphische Darstellungen, tabellarische Übersichten, Verzeichnisse, Querverweise usw.) sollten genutzt werden. Die Anfertigung eines qualitativ hochstehenden Evaluationsberichts braucht Zeit und Sachkenntnis. Dafür sind entsprechende Mittel einzuplanen. Auch die gekonnteste Evaluation ist wenig hilfreich, wenn es nicht gelingt, die Evaluation (und nicht nur ihre Ergebnisse) zu kommunizieren.

TEIL V: LERNORIENTIERTE EVALUATIONEN

KAPITEL 13: UMSETZUNG VON LERNEN DURCH EVALUATIONEN

Ingrid Kissling-Näf, Peter Knoepfel, Werner Bussmann

Wie, wo und in welchem Ausmass fliessen wissenschaftliche Erkenntnisse in den politischen Prozess? Können Evaluationen Lernprozesse induzieren? Fragen nach der direkten Anwendung von Forschungsergebnissen waren für die Sozialwissenschaften in den vergangenen drei Jahrzehnten besonders bezüglich öffentlicher Politiken immer relevant. Während die sogenannte Verwendungsforschung ("utilization research") die Interaktion von Wissen und öffentlichen Politiken zu beleuchten versuchte, haben neuere kognitionstheoretische Ansätze im Bereich der Politikanalyse Wissen und Ideen als Schlüsselvariablen für die Erklärung des Policy-Change wiederentdeckt und thematisiert. Obwohl der Übergang von der Verwendungsforschung zu den neueren lerntheoretischen Ansätzen fliegend ist, unterscheiden sich beide Richtungen in ihren Anliegen und müssen darum getrennt behandelt werden. Bei der Verwendungsforschung stand von Anfang an die Frage im Vordergrund, wie Evaluationen oder Wissensbestände auf den politischen Prozess einwirken können. Im Gegensatz dazu stiess die Politikanalyse auf der Suche nach erklärenden Faktoren für Änderungen der Grosswetterlage in einzelnen Politikbereichen (Instrumente, Paradigmen) auf das interpretative Paradigma.

Im folgenden werden die umsetzungsorientierte Forschung (Kapitel 13.1.) wie auch die neueren kognitionstheoretischen Ansätze (Kapitel 13.2.) mit Erklärungsmodell und Resultaten vorgestellt. Anschliessend wird die Ausrichtung der Evaluationsmethodik auf die Förderung von Lernprozessen (Kapitel 13.3.) behandelt. Schliesslich versuchen wir, auf der Basis vorliegender Studien und Resultate des NFP 27 Erfolgsbedingungen für die Verwendung von Wissen im politischen Prozess aufzuführen (Kapitel 13.4.)

13.1. DIE UMSETZUNGSORIENTIERTE LERNFORSCHUNG

"Es zeigte sich bald, dass die "Umsetzung" wissenschaftlicher Erkenntnisse in die politische Praxis voller Tücken ist: Entweder brachten die Projekte nichts, was von der Praxis als nützlich erachtet wurde, oder sie kamen mit ihren Erkenntnissen zur Unzeit, oder die Zusammenarbeit mit den Interessenten scheiterte an widrigen Umständen, oder die Resultate führten gar zu politischen Streitigkeiten." (Freiburghaus und Zimmermann 1985:1)

Ausgangspunkt für das neue Forschungsfeld der umsetzungsorientierten Forschung war in den sechziger und siebziger Jahren die Tatsache, dass wissenschaftliche Erkenntnisse nur sehr beschränkt in der Praxis genutzt wurden. Die rationalistischen wissenschaftlichen Konzepte waren zum damaligen Zeitpunkt geprägt von der etwas naiven Idee, dass die Sozialwissenschaften wissenschaftliche Inputs liefern, die direkt in Planungs- und Politikformulierungsprozesse eingebaut werden können. Die Unter- und Fehlnutzung von Wissen machte jedoch sehr schnell deutlich, dass dieses Modell der Realität nicht angepasst war. Die *Verwendungsforschung* kann darum als Reaktion auf die rationalistische Planungsrevolution angesehen werden. Sie wurde sehr stark von den Erfahrungen im Zusammenhang mit der Umsetzung von Evaluationen geprägt (Alkin, Daillak und White 1979; Cousins und Leithwood 1986; Leviton und Hughes 1981). Dabei wurde das Modell einer direkten instrumentellen Nutzung wissenschaftlicher Ergebnisse schon sehr bald in Frage gestellt. Weiss hat schon sehr früh darauf hingewiesen, dass Evaluationen nur einer unter mehreren Inputs für die Entscheidungsträger sind. "Diejenigen, die glauben, Evaluierung würde das Politische aus dem Entscheidungsprozess herausnehmen, müssen enttäuscht werden. Innerhalb jeder Organisation werden Entscheidungen durch Verhandlungen und Anpassung getroffen, durch Politik. Das ist das System, welches wir für die Verknüpfung von *Werten* mit Tatsachen haben." (Weiss 1974: 22). Obwohl die Verwendungsforschung auch heute noch nicht über eine tragfähige Theorie verfügt, hat sich schliesslich das sogenannte *Enlightenment-Modell* (vgl. Weiss und Bucuvalas 1980; Wittrock 1991: 336 f.; Wittrock und Wagner 1992: 227) durchgesetzt. Dadurch beeinflusst die Nutzung von wissenschaftlichen Resultaten in erster Linie die konzeptuel-

len Vorstellungen über Sachverhalte und bringt damit langfristige, ungeplante und diffuse Wirkungen hervor¹.

Der Einfluss der Sozialwissenschaften auf den modernen Staat hing, wie Vertreter dieses Ansatzes zeigten, von der geschichtlichen Entwicklung und den *wissensvermittelnden Institutionen* einer Nation ab. Erstere hat sich auf die Etablierung der Sozialwissenschaften ausgewirkt, die das Agenda-Setting und die Problemlösungs- und Lernkapazität des politischen Apparates nicht unberührt gelassen haben. Auf der institutioneilen Ebene wiederum ist es von Bedeutung, ob die Sozialwissenschaften schon früh universitär verankert wurden und ob es Organe gab, die Verwaltungen und Wissenschaftler in Kontakt brachten. In den Vereinigten Staaten z.B. wurden die Sozialwissenschaften, ganz im Gegensatz zur Schweiz, bereits im 19. Jahrhundert in die Universitäten aufgenommen, Schulen für öffentliche Politiken wurden bereits Anfang der sechziger Jahre dieses Jahrhunderts gegründet (Weiss und Wittrock 1991: 356 ff.).

Mit der Schaffung von Institutionen verbindet sich die Speicherung, die Weiterentwicklung und die mögliche spätere Diffusion von Wissen. Wichtig für die Verbreitung sozialwissenschaftlichen Wissens war v.a. die OECD, die aufgrund ihrer internationalen Bedeutung und ihrer Funktion als "think tank" in manche nationalen Politiken hineingewirkt haben dürfte. Im nationalen Kontext wiederum hat es eine Rolle gespielt, ob sogenannte "Diskurskoalitionen" zwischen der Politik und der Wissenschaft aufgebaut und institutionalisiert werden konnten². In der Schweiz stellen z.B. die ausserparlamentarischen Kommissionen einen möglichen Ort der Begegnung zwischen Forschung und Politik dar. Wie Germann (1991) gezeigt hat, war (und ist) deren Zusammensetzung aber primär auf die Repräsentation wichtiger Interessen ausgerichtet, und sozialwissenschaftliches Wissen konnte (und kann) kaum eingebracht werden. Die wachsenden Bedürfnisse der Verwaltung nach mehr Steuerungswissen führten in der Schweiz Mitte der siebziger Jahre zur Einführung der Nationalen Forschungsprogramme, die praxisnahe Forschung im nationalen Interesse bezweck(t)en. Daneben stützen sich Dienststellen der Verwaltung (v.a. in

¹ "Over a span of time and much research, ideas... filter into the consciousness of policy-making officials and attentive publics. They come to play a part in how policy-makers define problems and the options they examine for coping with them. ... At this point in their development, "enlightenment" may be the wisest use of the social sciences." Weiss 1978: 77 f. zitiert nach Wittrock 1991: 337.

² Gute Beispiele dafür sind Grossbritannien und die Niederlande. Vgl. Blume u.a. 1991 und Smith 1991.

der Bildungspolitik, Wirtschaftspolitik, Landwirtschaft, Umweltpolitik, Forstwirtschaft, Sicherheitspolitik) in zunehmendem Masse auf die Resultate von Forschungsanstalten, von -Instituten und von Universitäten ab bzw. bauen Brücken zu diesen Institutionen auf (vgl. Kapitel 2).

Es lässt sich nicht in jedem Fall eruieren, wann welche Forschungsergebnisse Einfluss auf den Policy-Prozess genommen haben. Als Eintrittsstellen des wissenschaftlichen Wissens kommen je nach Phase im Policy-Prozess und im sozialen Lernpfad die Öffentlichkeit, Parteien und Verbände, das Parlament oder auch die Verwaltung selbst in Frage. Die Öffentlichkeit dürfte vor allem bei der Problemerkennung zu Beginn des Policy-Cycle oder in der eigentlichen Umbruchphase eines Paradigmas in Erscheinung treten. Ansonsten dürfte ein Grossteil des Wissens über Verwaltungsakteure in das Politikfeld aufgenommen werden (vgl. Freiburghaus und Zimmermann 1985: 15; Hall 1993).

Kingdon (1984) hat in seinem viel rezipierten Werk zum Agenda-Setting die theoretische Grundlegung der Interaktionsweise zwischen den Sphären der Politikproduktion und der Wissenschaft weiterentwickelt. Wissenschaftliche Ideen werden seiner Meinung nach nur dann genutzt, wenn sogenannte "*Policy Windows*" (Gelegenheitsfenster) auftreten. Er bleibt der Enlightenment-Idee verpflichtet und macht über die Bündelung der verschiedenen Variablen Faktorkonstellationen aus, unter denen ein Wissenstransfer stattfinden kann. In Anlehnung an das Garbage-Can-Modell (Cohen, March und Olson 1972) versucht er ausgehend von drei Strömen zu zeigen, warum politische Ideen plötzlich an Relevanz gewinnen. Nebeneinander und unabhängig voneinander fliessen der Strom der *Probleme*, der über die Dringlichkeit und Wichtigkeit von Fragestellungen informiert, und der *Policy-Strom*, der Lösungen, Vorschläge und angehäuften Wissen umfasst. Als dritte Flussgrösse sind die *politischen Ereignisse* wie z.B. Wahlen oder neue Mehrheiten im Parlament ins Modell aufgenommen. Eine Handlungsgelegenheit ergibt sich nur dann, wenn die drei Ströme konvergieren. Ein Policy Window öffnet sich, wenn sich auf der Politikebene (Ereignisse) oder auf der Problemebene weitreichende Änderungen ergeben. Die Policy-Entrepreneurs müssen die entsprechende Situation nutzen und als Interface zwischen den Strömen neues Wissen in den Prozess eingeben. So schnell wie sich ein Fenster öffnet, kann es sich auch wieder schliessen.

Die Prämissen der neueren Ansätze, die die Auswirkungen der Wissenschaft auf die Praxis thematisieren, sind aber ebenfalls nicht unumstritten; denn die wissenschaftliche Erkenntnis ist nur eine gesellschaftliche Pro-

blemlösungsmethode ("social problem solving", Lindblom und Cohen 1979). Wittrock (1991: 344 ff.; 350) schlägt darum vor, das erkenntnisleitende Interesse des Enlightenment-Ansatzes beizubehalten, aber den Approach zusätzlich um weitere Interaktionsweisen zwischen den Sphären der Politikproduktion und der Wissenschaft zu ergänzen.

13.2. LERNEN ALS NEUES PARADIGMA

Neuere kognitionsorientierte Ansätze der Politikforschung betonen die Bedeutung von Wissen und Ideen als erklärende Faktoren für den Policy-Change und die Gestaltung von Politiken wieder stärker. Man geht davon aus, dass Innovationen und inkrementale Veränderungen nicht mehr nur durch externe Faktoren ausgelöst werden, sondern als Ergebnis von kollektiven Lernprozessen zu interpretieren sind.

Die Politikanalyse hat während langer Zeit den Ideen und dem Wissen als erklärenden Variablen bei der Generierung und Umsetzung von Politiken keine grosse Bedeutung beigemessen. Im Gegensatz dazu gehen die neueren kognitionstheoretischen Ansätze davon aus, dass Ideen Einfluss auf die Weltbilder und Wertvorstellungen der Individuen nehmen und ebenfalls in den kollektiven Sinngabungsprozess einfließen. Diese stärkere Auseinandersetzung mit Ideen, Ideologien und Wissensbeständen kann als kognitive Wende der Policy-Forschung verstanden werden (vgl. Nullmeier 1994). Auf der Ebene des Politikfeldes werden damit Lernprozesse weniger im Zusammenhang mit spezifischen Evaluationen analysiert. Die Beschreibung der Politikverläufe wird in der Literatur vielmehr anhand von politisch dominanten Deutungsmustern vorgenommen, die deren Entwicklung prägen und einschränken³.

Interpretative bzw. konstruktivistische Ansätze gehen nicht von einem sozialtechnologischen Modell der Politikberatung aus, bei dem ein direkter Wissenstransfer stattfindet. Vielmehr entstehen kognitive Strukturen im sozialen Prozess, dort werden auch Problemdefinitionen und Situationsdeutungen erarbeitet. Realität wird immer als sozial ausgehandelt verstanden, sie ist das Resultat von komplexen Interpretationskämpfen und

Dem interpretativen Paradigma sind Arbeiten von Politikwissenschaftlern wie Majone 1993, Sabatier 1993, Schön und Rein 1994, Fischer 1994, Fischer und Forester 1993, Singer 1993 oder auch Nullmeier 1993 verpflichtet. Eine gute Einführung in die Theorieentwicklung und Bedeutung des interpretativen Ansatzes in der Politikwissenschaft findet sich bei Nullmeier 1994.

von je neuen Konstruktionen der Realität. "Dieser Prozess der kollektiven Sinngebung strukturiert das Handlungswissen der Akteure. [...] Zweitens wird mit der Allokation des Wissens der Handlungsrahmen re-konstruiert: Policy-Debatten verändern damit zugleich den kognitiven Kontext." (Singer 1993: 158). Ideen können darum als eigenständige Faktoren auf den Policy-Diskurs einwirken. Soziales Lernen wird dort gleichgesetzt mit der Weiterentwicklung und Veränderung des Policy-Paradigmas.

Heclo (1974) als eigentlicher Vorläufer hat in einer sehr frühen Arbeit, einer historisch-vergleichenden Untersuchung zur britischen und schwedischen Sozialpolitik, darauf hingewiesen, dass makroökonomische Faktoren nur partiell die Richtung der sozialpolitischen Entscheidungen zu erklären vermögen. Ebenso wichtig waren Interaktionen unter Politikfeldspezialisten sowie die geleistete Interpretationsarbeit. Sabatier und Jenkins-Smith (1993) haben mit ihrem Konzept der "Advocacy Coalition" versucht, Lernprozesse als wichtigen Faktor für den politischen Wandel zu verstehen. Innerhalb der Politikfelder machen sie Koalitionen mit divergierendem ideologischem Hintergrund aus. Verschiebungen auf der Ebene der *Grundüberzeugungen* ("Belief System") erklären in ihrem Modell inhaltliche Richtungsänderungen. Mit der Berücksichtigung der ideologischen Variablen und der Wissensbestände sollte zugleich die Engführung der Institutionenanalyse überwunden werden.

Vor allem Hall (1993) hat mit der Erforschung des ordnungspolitischen Wechsels der Wirtschaftspolitik in Grossbritannien in den achtziger Jahre auf die Bedeutung der ideologischen Vorstrukturierungen der Politikfelder und den Einfluss von Wissen und Deutung hingewiesen. Die Abkehr vom keynesianischen Interventionsmodell vollzog sich in England schrittweise. Hall unterscheidet in diesem Zusammenhang Lernprozesse mit unterschiedlicher Tiefe. Mit Veränderungen auf der *ersten Ebene* sind kleinere inkrementelle Änderungen gemeint, wie sie sich bei der Anpassung des Budgets an neuere Entwicklungen ergeben. Zugleich wurden in Grossbritannien zwischen 1970 und 1989 auch "*Second-order-changes*" vorgenommen, die als Reaktion auf Misserfolge die makroökonomische Interventionspalette anpassten. Dazu gesellte sich mit der Wahl von Margaret Thatcher eine radikale Umorientierung der Politik und die Praktizierung eines neuen wirtschaftspolitischen Paradigmas ("*Third-order-change*"). Gesellschaftliches Lernen kann aber nicht nur nach der Tiefenschärfe der Lernprozesse, sondern auch phasenspezifisch aufgeschlüsselt werden: "The period of an entrenched paradigm has two stages: one in which the paradigm largely is unchallenged, and one in which challenges

begin to build up. The period of transition involves three stages in which the challenges lead to some tentative or experimental changes, a stage in which the experts disagree openly with each other, and a stage in which experts go public and the relevant policy community is enlarged dramatically. Finally, there is an extended period in which the new paradigm is institutionalized and members of the relevant policy community come to accept the hegemony." (Howlett 1994: 634).

Damit wurde deutlich, dass der Konsensbereich mehrschichtig ist und dass je nach politischer Phase andere Inhalte (grundlegende Werte, sekundäre Elemente) zur Diskussion stehen. Ausmass und Reichweite der Veränderung und Neubewertungen sagen zudem etwas über den Lerntyp aus. Ein Wandel des Konsenses und die Wirkungen von Ideen lassen sich jedoch anhand der Diskurse nur aufzeigen, wenn ein entsprechend grosser Zeitraum (mindestens zehn Jahre) für die Untersuchung gewählt wird. "Dies (ist) auch (nötig), weil die "enlightenment"-Funktion (Aufklärungsfunktion) der beratenden Policy-Analyse realistischereweise damit rechnen muss, dass ein kumulativer Effekt von Forschungsbefunden nur allmählich die Problemsicht der politischen Entscheider und der Öffentlichkeit modifiziert" (Héritier 1993a: 18).

Wie Singer (1993: 167 ff.) zeigt, dürfte das Einfließen von neuen Ideen zusätzlich noch von den kontingenten *Rahmenbedingungen des Policy-Diskurses* abhängen. So ist der politische Prozess eher Veränderungen und Neubewertungen zugänglich, wenn der Handlungsraum durch eine prinzipielle Unsicherheit gekennzeichnet ist⁴. Als Beispiel dafür führt Singer das Konzept der Währungsintegration bei der deutschen Wiedervereinigung an. Obwohl von allen wichtigen Akteuren zu Beginn abgelehnt, wurde es in einer Situation von Unsicherheit akzeptiert. Im Gegensatz dazu sind wahrscheinlich hoch konfliktuelle und geschlossene Foren einer Erweiterung des Argumentationsspielraums und der Integration von neuen Ideen eher abgeneigt. Eher positiv dürften sich institutionalisierte Formen der Politikberatung, die auf einen dauerhaften Kontakt zwischen Personen aus der Forschung und aus dem Politikfeld angelegt sind, auswirken. Inwieweit es den "epistemic communities" jeweils gelingt, politikrelevantes Wissen zur Verfügung zu stellen, ist empirisch zu überprüfen.

⁴ Dies wurde bereits von Freiburghaus und Zimmermann (1985: 88 f.) hervorgehoben, welche "Latenzphasen" als für die Wissenschaft "komfortable" Situationen charakterisierten.

Der Einbezug von Ideen und Wissensbeständen und der Rückgriff auf Lerntheorien zur Erklärung von Innovationen und inkrementalen Veränderungen im Bereich der öffentlichen Politiken kann als eigentliche Neuerung des Policy-Forschung bezeichnet werden. Es erstaunt darum auch nicht, dass sich in den vergangenen Jahren zunehmend mehr Politikwissenschaftler mit individuellen und kollektiven Lernprozessen befasst haben, die den Policy-Prozess beeinflussen (z.B. Bennett und Howlett 1992; Duncan und Weiss 1979; Kissling-Näf 1996; Rose 1993).

13.3. ERFAHRUNGEN MIT LERNORIENTIERTEN EVALUATIONEN

Die kognitive Wende bei der Erklärung des Policy-Change läuft parallel zu neueren Entwicklungen in der Evaluationsforschung. Mit interaktiven und diskursiven Methoden (vgl. Kapitel 12.2.1.) sollen die Voraussetzungen geschaffen werden, um Lernprozesse auszulösen und zu begünstigen. Getreu dem Motto "The medium ist the message" wird dabei besondere Aufmerksamkeit auf die prozesshaften Elemente von Evaluationen gelegt, denen ebenso grosse Bedeutung wie den eigentlichen Ergebnissen zugemessen wird (vgl. auch Kapitel 9). Es wurde eine Vielzahl von Evaluationskonzepten entwickelt: responsive (Stake 1980, Beywl 1988), verwendungsorientierte (Patton 1978), kreative (Patton 1981), entwicklungsfördernde (Patton 1994), "ermächtigende" (Fettermann 1994, 1995, 1996), pluralistische (Monnier 1987, Viveret 1989), akteurorientierte bzw. partizipative (Rey et al. 1993), interaktive (Bierter und Binder 1993, Binder 1994) sowie Evaluationen der vierten Generation (Guba und Lincoln 1989). Als gemeinsamen Nenner dieser Konzepte kann man die *Interaktion* (Steuerung des Forschungsprozesses und Erarbeitung der wichtigsten Forschungsfragen durch Forschungsteams *und* durch von der Evaluation betroffene Akteure), den *Wertpluralismus* (Berücksichtigung verschiedener Wertvorstellungen im Hinblick auf die Beurteilung öffentlicher Politiken) und den epistemologischen *Konstruktivismus* ("Realität" als Konstrukt, bei welchem alle Beteiligten ihren Gesichtspunkt zur Geltung bringen können) aufführen. Ziel solcher Evaluationen ist nicht, *die* Wahrheit zu ermitteln und *die* Bewertung einer öffentlichen Politik zu erreichen, sondern eine Verständigung und einen Konsens über die verschiedenen Gesichtspunkte im Hinblick auf deren Beurteilung zu erzielen. Da in der Regel qualitative Forschung flexibler und somit leichter an sich verändernde Forschungsprioritäten anzupassen ist (vgl. Arbeitsgruppe

"Gesetzesevaluation" (AGEVAL) 1991:36 f.), erstaunt nicht, dass massgebliche Protagonisten umsetzungs- oder lernorientierter Evaluationskonzepte vor allem qualitative Forschungstechniken verwenden (vgl. etwa Patton 1990). Eine solche Koppelung ist indessen nicht wesensnotwendig, und es ist durchaus denkbar, auch die Durchführung quantitativer Evaluationen stärker interaktiv und lernorientiert auszugestalten (vgl. Kapitel 12.2.1.).

Lernorientierte Evaluationen können sich auf unterschiedliche Kontexte und Konfigurationen beziehen. Es können *personenorientierte* und *policyorientierte* Ansätze unterschieden werden⁵. Bei ersteren werden der Evaluationsprozess und die Evaluation auf eine oder mehrere Personen ausgerichtet. Die Evaluation hat einen klar bestimmbar Adressaten, und für die Umsetzung ist wichtig, dass dessen Bedürfnisse möglichst weitgehend abgedeckt werden. Die Evaluation wickelt sich in einem präzise definierten Beratungskontext ab. Der Einbezug einer engeren oder weiteren Öffentlichkeit ist weder absolut erforderlich noch möglicherweise erwünscht. Für solche lernorientierten Evaluationen können die aus der Management- und Organisationsberatung entwickelten reichhaltigen Rezepte (z.B. Unterstützung durch Top-Management erforderlich, klare Vorgaben nötig, Fairness im Ablauf sicherstellen, Umsetzung aktiv begleiten) übernommen werden, welche hier keiner besonderen Erwähnung bedürfen. In der Evaluationsforschung hat sich vor allem Patton (1978, 1981, 1990, 1994) diesem Ansatz verschrieben.

Der *politikorientierte Ansatz* lernfördernder Evaluationen richtet sich nach den Interaktionszusammenhängen öffentlicher Politiken. Er ist deshalb durch einen eher diffusen Kreis von möglichen Adressaten gekennzeichnet. Letztere können auf unterschiedlichen Ebenen angesiedelt sein: Angehörige von Parlamenten, Behördenmitglieder, Vollzugsverantwortliche, Politikbetroffene, Medien, breitere Öffentlichkeit. Wer sich vom Problem oder vom Policy-Anliegen betroffen und vom Evaluationsprozess bzw. dessen Resultaten angesprochen fühlt, ist zu Beginn schwer abschätzbar. Es ist deshalb sehr schwierig, weil von kontingenten Bedingungen abhängig, solche Evaluationen in der Weise auszugestalten, dass Lernprozesse in Gang gesetzt werden können. Im Rahmen des Nationalen Forschungsprogramms "Wirksamkeit staatlicher Massnahmen" (NFP 27) wurden zwei solcher Evaluationen durchgeführt. Die erste (Rey et al. 1993) war "mono-policy"-orientiert und hatte die Walliser Wirtschaftsför-

⁵ Vgl. etwa das Votum Patton in Alkin 1990: 27.

derung zum Gegenstand (vgl. dazu auch Kapitel 11.2.2.). Ein akteurorientierter ("évaluation par les acteurs") bzw. partizipativer Ansatz wurde vor allem deshalb gewählt, weil der Erfolg der auf Finanzhilfen und Beratungsmassnahmen gestützten Wirtschaftsförderung von der Mitwirkung zahlreicher Partner abhängig ist. M.a.W. handelt sich um eine stark partnerschaftlich orientierte Politik. Das *Vorgehen* war durch den starken Einbezug der wichtigen Handlungsträger gekennzeichnet. Nachdem der Evaluationsgegenstand in etwa bestimmt war, wurden aufgrund einer Grobanalyse die wichtigsten Akteure ermittelt und diese anschliessend im "Schneeballverfahren" nach weiteren Akteuren, welche am Politikprozess beteiligt sind, befragt. Gestützt auf die Akteuranalyse wurde eine Evaluationsbegleitgruppe ("instance d'évaluation") bestehend aus einem Dutzend Personen zusammengestellt. Diese war zusammen mit dem Evaluationsteam ("chargé d'évaluation") das eigentliche Steuerungsorgan der Evaluation, mit folgenden Aufgaben:

1. Vorarbeiten: (Re)definition (Einengung, Ausweitung) des Evaluationsgegenstandes, Änderung (z.B. Ergänzung) der Zusammensetzung der Begleitgruppe, Festlegung der "Spielregeln" innerhalb der Begleitgruppe (Mehrheitsentscheidungen, Konsenslösungen, Verhältnis zwischen Begleitgruppe und formellem Auftraggeber).
2. Festlegung von Forschungsprioritäten und Auswahl der Evaluationskriterien.
3. Diskussion der Forschungsstrategie, der -methoden und des Arbeitsprogramms: Die Mitglieder der Begleitgruppe mussten die verschiedenen Möglichkeiten des Vorgehens kennen, damit sie sich über die zu wählende Option aussprechen konnten. Dasselbe galt für das Arbeitsprogramm. Das Einverständnis der Begleitgruppe zu den Fragen (1) - (3) bildete eine wichtige Vorbedingung für die spätere Akzeptanz der Resultate.
4. Begleitung der Evaluation: periodische Besprechung der Zwischenresultate, Nutzung der Kenntnisse der Mitglieder der Begleitgruppe bei der Durchführung der Evaluation.
5. Gutheissung der Evaluationsresultate, Mitwirkung an deren Umsetzung: Um die Akzeptanz der Resultate und die Mitwirkung an der Um-

Setzung zu erhöhen, wurde der Begleitgruppe ein Entwurf des Schlussberichts zur Stellungnahme unterbreitet.

Durch die Beteiligung an einem gemeinsamen Forschungsprozess, bei welchem sowohl der formelle Auftraggeber wie auch das Evaluationsteam das Steuer teilweise aus der Hand geben, werden die Erfolgchancen für eine Umsetzung der Resultate verbessert. Das Evaluationsteam hat nur noch eine begrenzte Verantwortung für die Steuerung des Evaluationsprozesses, aber spielt weiterhin eine bedeutsame Rolle, (1) indem es methodische und sachbereichsspezifische *Expertisen* einbringt, (2) die *Animation* der Begleitgruppe sicherstellt und über die Einhaltung der vereinbarten Spielregeln wacht, (3) Konflikte zur Sprache bringt und neue Beurteilungen und Perspektiven entstehen lässt ("*Hebammenfunktion*"). Soweit möglich (d.h. forschungsökonomisch verkraftbar) sind diese unterschiedlichen Funktionen auch durch unterschiedliche Personen des Evaluationsteams wahrzunehmen.

Die zweite Evaluation (Bierter und Binder 1993; Binder 1994) war räumlich (Bund, Kantone Solothurn und St. Gallen) und inhaltlich (innovationsfördernde Politiken) sehr viel breiter gefasst als die erste und bewegte sich damit im "*Interpolicy*"-Feld (vgl. auch Kapitel 11.2.2.). Das Projekt verfügte nicht über eine Begleitgruppe. Es verfolgte indessen wie das ersterwähnte eine interaktive Evaluationsstrategie, indem nach einer ersten Befragungsrunde die Zwischenresultate und Thesen in Seminaren mit den Akteuren eingehend diskutiert und weiterentwickelt wurden. Dadurch wurde ein Diskurs zwischen privaten und staatlichen Akteuren sowie bei letzteren zwischen unterschiedlichen Dienststellen sowie zwischen unterschiedlichen Ebenen (Bund, Kantone) zur Frage der bisherigen Resultate und künftigen Möglichkeiten der Innovationsförderung initiiert.

Neuere (inter)policy-orientierte Evaluationsformen wie die beiden Beispiele fügen sich in die komplexe Mechanik moderner Gesellschaften mit ihren Subsystemen (Wirtschaft, Kultur, Erziehung u.a.m.), parzellierten Politiken (Landwirtschaftspolitik, Umweltpolitik usw.) und fragmentierten Staatsebenen (Bund, Kantone, Gemeinden, Zweckverbände, Regionsgebilde). Ihr Ziel ist, möglichst alle argumentativ bedeutsamen Gesichts-

punkte der Akteure⁶ einzubeziehen. Evaluationen sollen das gegenseitige Verständnis ("Empathie", vgl. Papadopoulos 1995: 129-131) der Akteure aus unterschiedlichen gesellschaftlichen Subsystemen bzw. unterschiedlichen staatlichen Ebenen und Dienststellen fördern und Verhandlungen (negotiation: Guba und Lincoln 1989) in Gang bringen bzw. erleichtern⁷. Dies soll ermöglichen, Politik nicht als Nullsummenspiel zu betrachten, sondern im Sinne einer "Verbesserungsevaluation" ("formative evaluation") Spielräume für Lösungen auszuloten, die allen Akteuren einen Nutzen bringen (z.B. durch Erhöhung des Potentials für einen Ressourcenaustausch, vgl. Kapitel 13.4.). Namentlich im "Interpolicy"-Bereich sollen durch Evaluationen Akteure aus verschiedenen Politiken im Hinblick auf gemeinsame (Querschnitts-)Probleme ins Gespräch gebracht und die Bildung von Netzwerken (vgl. Kapitel 7.2.) begünstigt werden. Daraus werden Impulse erwartet, die über den eigentlichen Prozess der Evaluation hinausgreifen sollen (vgl. Binder 1994: 83).

Der Einsatz lernorientierter Evaluationen beruht auf verschiedenen Voraussetzungen. Es braucht dazu erstens - soweit die Evaluation nicht in freier Forschung erfolgt - das *Einverständnis des Auftraggebers*, die Steuerung des Forschungsprozesses teilweise aus der Hand zu geben. Erforderlich ist zweitens die Bereitschaft aller involvierter Akteure, sich auf einen gemeinsamen Prozess einzulassen. Diese Bereitschaft dürfte in Zeiten grosser politischer Konflikte nicht oder nur zum Teil vorhanden sein. Drittes Erfordernis ist eine grosse methodische, soziale und fachliche *Kompetenz der beauftragten Evaluаторinnen und Evaluatoren*, damit ausreichende materielle Grundlagen zur Beurteilung der untersuchten Politik(en) geschaffen bzw. faire Bedingungen für alle Akteure beim Evaluationsprozess sichergestellt werden können.

Wir möchten nicht versäumen, die möglichen Nachteile und *Defizite* lernorientierter, pluralistischer, partizipativer Evaluationen zu erwähnen (vgl. v.a. Papadopoulos 1995: 132). Der kostenmässige Zusatzaufwand aufgrund der Offenheit des Forschungsprozesses und der Animation des Diskurses verdient dabei noch am wenigsten Erwähnung. Denkbar ist

Es geht beim Einbezug der Akteure somit nicht um das Kriterium ihrer Repräsentativität, sondern darum, die Vielfalt möglicher Argumentationsformen einzufangen. Im Diskursprozess können gerade auch die Argumente politisch peripherer Akteure sehr wertvoll sein. Der Einbezug weiterer Akteure bzw. der Forschungsprozess kann dann abgebrochen werden, wenn sich die/deren Argumente wiederholen (mündl. Hinweis H.-G. Soeffner, s. auch Gerheuser 1991: 121).

Insoweit sind die Grenzen zwischen solchen lernorientierten Evaluationen und der Mediation (vgl. Kapitel 9) fließend.

auch, dass die Evaluation bloss symbolische Funktion hat, mit dem Ziel, dass etwas gemacht wird, ohne dass wirklich etwas passiert⁸. Bedeutsamer sind allerdings die Bedenken, dass die Evaluation ihre kritische, *reflexive Funktion verlieren* kann, wenn sie stärker an den Prozess der Policy-deliberation und der Policy-decision angebunden wird. Es scheint, dass zwischen dem Ausmass der Fähigkeit zur Selbstreflexion und dem Grad der Handlungsorientierung ein gewisser Trade-off besteht (vgl. Soeffner 1989: 23-29). Die Tendenz geht jedenfalls in die Richtung, dass stark lernorientierte Evaluationen die von Scriven (1972: 125) mit Nachdruck hervorgehobene Funktion der kritischen Bestandesaufnahme und Bilanzierung ("summative evaluation") vernachlässigen und sich "bloss" auf Lernprozesse konzentrieren. Dies kann dazu führen, dass Massnahmen und Politiken, die aufgrund ihrer geringen Wirksamkeit und Wirtschaftlichkeit eingestellt werden sollten, weiterentwickelt werden. Zudem besteht auch bei lernorientierten Evaluationen die Möglichkeit, dass der Diskurs durch unterschiedliche *Machtpotentiale* der involvierten gesellschaftlichen Gruppen stark beeinflusst werden kann: Das höhere Sanktionspotential einzelner Akteure kann durchaus auf den Argumentationsprozess durchschlagen, sofern das Evaluationsteam dieser Tendenz nicht bewusst entgegensteuert. Insbesondere besteht die Gefahr, dass die spezifischen Policy-Interessen der staatlichen Verwaltung sowie der anvisierten gesellschaftlichen und wirtschaftlichen Gruppen (= Adressaten) gegenüber den Allgemeininteressen der Bürger und Steuerzahler zu stark gewichtet werden und dass eine kritische Überprüfung der Tauglichkeit der eingeschlagenen Politik umgangen wird. Als letztes ist zu erwähnen, dass mit dem Hinweis auf intensive Lernprozesse und Gespräche auch der Diskurs in einer breiteren politischen Öffentlichkeit über die zur Debatte stehenden Politiken abgeblockt werden kann. Die Erwähnung dieser Gefahren soll zeigen, dass lernorientierte Evaluationen ein Instrument mit Licht- und Schattenseiten sind, das sinnvoll verwendet, aber auch missbraucht werden kann.

Mark und Shotland (1986: 144) und Weiss (1984: 265) weisen auf den fundamentalen Widerspruch hin, demokratische Prozesse auf von oben gelenktem Weg einzuführen, und scheuen sich nicht, den Begriff "pseudo-empowerment" zu verwenden.

13.4. VORAUSSETZUNG FÜR LERNPROZESSE IN ÖFFENTLICHEN POLITIKEN

Das Phänomen lernorientierter, interaktiver Evaluationen ist ein Indiz für die Bedeutung der Lernthematik. Diese steht heute im Rampenlicht der Managementliteratur (z.B. Senge 1990), und in der öffentlichen Verwaltung bzw. in öffentlichen Politiken könnte dem Begriff eine ähnliche Karriere bevorstehen. Dafür sind, um die Begriffe von Luhmann zu verwenden, die Komplexität und Kontingenz moderner Gesellschaften verantwortlich. Nach dem interpretativen Ansatz (vgl. Kapitel 13.2.) sind Probleme nicht einfach gegeben, sondern Konstrukte. Probleme, Lösungen und Gelegenheiten für politisches Handeln (vgl. den Ansatz von Kingdon in Kapitel 13.2.) können auf sehr vielfältige Weise miteinander verbunden werden, und "Lernen" kann gerade darin bestehen, "intelligente" Verknüpfungen vorzunehmen. Der Ansatz des Lernens gewinnt auch deshalb an Bedeutung, weil seit dem Verschwinden des Eisernen Vorhangs die grossen ideologischen Debatten über die grundsätzlichen Funktionen des Staats an Gewicht verloren haben und weil die Frage nach der Art und Weise der staatlichen Einflussnahme (Kontextsteuerung, Anreizstaat, Steuerstaat, Dienstleistungsstaat u.a.m.) ebenso bedeutsam geworden ist wie die Frage nach "wieviel Staat" (vgl. Kapitel 6.5.2.). Zudem stehen angesichts wirtschaftlicher und gesellschaftlicher Globalisierung sowohl Firmen als auch öffentliche Verwaltungen unter erhöhtem Wettbewerbsdruck, und "Lernen" wird als Strategie angesehen, um diesen Herausforderungen zu begegnen. Im Unterschied zur "Umsetzung" deutet der Ausdruck des "Lernens" zudem semantisch an, dass es nicht einfach darum geht, Wissensbestandteile als solche zu übernehmen und anzuwenden, sondern (kreativ) in den Handlungskontext einzupassen ("Verwandlung" und nicht "Verwendung" von Wissen, vgl. Beck und Bonss 1989). Damit zollt auch der Begriff "Lernen" zu einem gewissen Teil den systemtheoretischen Ansätzen (selbstreferentielle Systeme) Tribut.

Es ist verständlich, dass nach Rezepten gesucht wird, um Lernprozesse in öffentlichen Politiken zu begünstigen. Entsprechende Handlungsanleitungen können sich auf einzelne Personen (vgl. die personenbezogenen Evaluationsansätze in Kapitel 13.3.), auf einzelne Organisationen (wie in der lernorientierten Managementliteratur) oder auf öffentliche Politiken bzw. Interpolices oder auf Kombinationen davon beziehen. Wir werden in der Folge vor allem den (Inter)policy-Aspekt beleuchten.

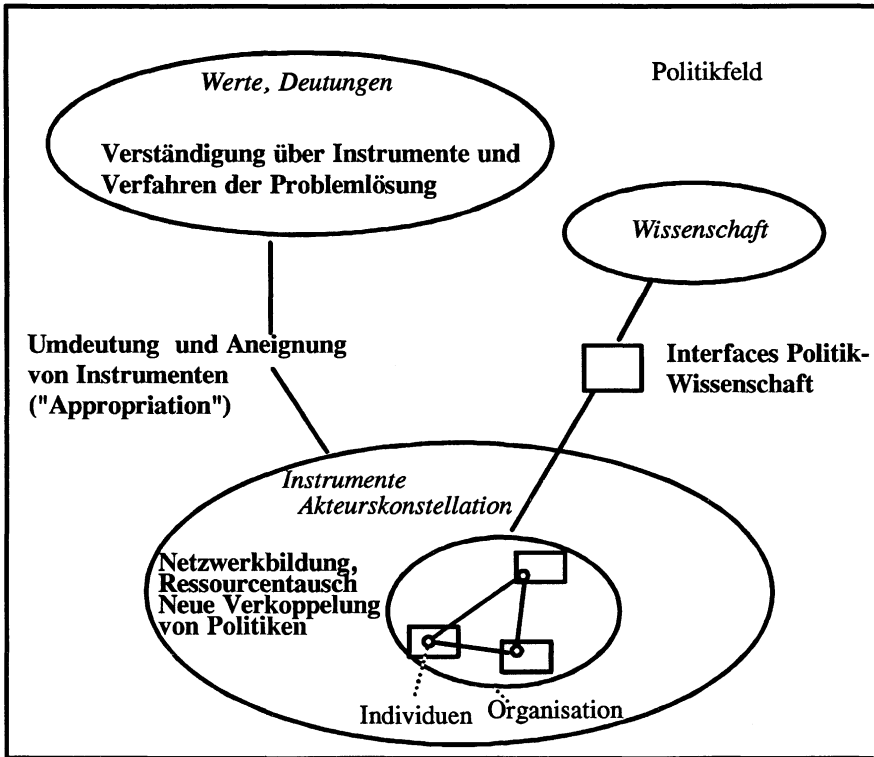
Dabei ist vor vorschnellen Erwartungen an umsetzbare Rezepte für die Förderung von Lernprozessen zu warnen. Namentlich der Ansatz der "Policy-Windows" (vgl. Kapitel 13.2.) legt die Vermutung nahe, dass sich öffentliche Politiken aufgrund von wenig steuerbaren Faktoren entwickeln. Auch eine noch so sorgfältige Abstimmung des Forschungsplanes einer Evaluation auf bevorstehende Entscheide kann in der Regel nicht verhindern, dass unerwartet eintretende Ereignisse die politischen Prioritäten in eine ganz neue Richtung lenken können. Noch grössere Zurückhaltung gegenüber naiven Erwartungen ist namentlich aufgrund der systemtheoretischen Überlegungen von Luhmann und Willke angebracht.

Trotz dieser einschränkenden Bemerkungen legen im Rahmen des Nationalen Forschungsprogramms "Wirksamkeit staatlicher Massnahmen" durchgeführte empirische Forschungen über Lernprozesse (Knoepfel, Kissling-Näf und Marek 1997⁹; Urio, Bari und Glauser 1996¹⁰) als erste "Annäherungen" an die Lernthematik eine Reihe von Faktoren offen, welche Lernprozesse begünstigen oder benachteiligen bzw. die Richtung der Lernprozesse beeinflussen können. Dabei stehen politikfeldspezifische Eigenschaften im Vordergrund. Politikfelder - verstanden als alltags-sprachliche Nominalkategorien - bezeichnen durch institutionelle und sachliche Zusammengehörigkeit geprägte Gegenstandsbereiche von Politiken (vgl. Windhoff-Héritier 1987: 21 f.). In diesem Sinne konstitutiv für ein Politikfeld sind die einem Politikbereich zugrundegelegten *Deutungsmuster und Werte*, der Typ der Interventionspalette (*Instrument*), die *Akteurkonstellationen und Kooperationsformen* sowie - wenn vorhanden - ein ausgebautes *Wissenssystem* (Abb. 10). Entlang dieser Variablen können die folgenden Erfolgsbedingungen für wissenschaftliche Inputs und kollektive Lernprozesse identifiziert werden.

⁹ Dabei wurden im Rahmen von 28 Fallstudien in vier Bereichen (Umweltverträglichkeitsprüfung, Luftreinhaltung und Verkehr in Vorortsgemeinden, ökologische Direktzahlungen in der Landwirtschaft, Innovationen im Gesundheitswesen) Lernprozesse nachgezeichnet, mit Hypothesen verglichen und entlang verschiedener Variablen ausgewertet. Als "Lernen" wurden Änderungen im Politikoutput, Netzwerk und Konsensbereich definiert, die von den Akteuren bewusst auch als kognitive Änderungen wahrgenommen wurden.

¹⁰ In der Studie wurden die Ursachen und Wirkungen von Evaluations- und Lernbemühungen in vier kantonalen Gesundheitsförderungsprogrammen (Bern, Zürich, Genf, Tessin) analysiert.

Abbildung 10 Lernprozesse in öffentlichen Politiken



Aus diesen Forschungen lassen sich folgende Erkenntnisse ableiten:

- Lernprozesse, welche sich in einem verbesserten Politik-Output niederschlagen, setzen nicht notwendigerweise einen Konsens über grundlegende Wertvorstellungen sowie über Problemsicht und Handlungsbedarf, sondern nur eine *Verständigung über Instrumente und Verfahren der Problemlösung* voraus. Im Rahmen des Projekts Knoepfel, Kissling-Näf und Marek (1997) wurden in zahlreichen Fallstudien Lernprozesse beobachtet, ohne dass sich die Grundüberzeugungen der am Netzwerk beteiligten Akteure einander angenähert hätten. Dieser Befund liegt auf der von Sabatier und Jenkins-Smith (Sabatier 1993, Jenkins-Smith und Sabatier 1993) vertretenen Linie, wonach sich Policy-Koalitionen ("advocacy-coalitions") überwiegend auf Überein-

Stimmung in den oberen (Instrumente, Verfahren) und nicht unbedingt in den tieferliegenden (grundlegende Werte und Problemsicht) wertmassigen Schichten stützen. Ob sich eine Verständigung über den einzuschlagenden Wege (Instrumente, Verfahren) anschliessend auf die tieferliegenden Schichten fortsetzt, konnte wegen der zu kurzen Beobachtungsfrist (3 Jahre) nicht festgestellt werden. Relativierend ist zudem anzumerken, dass in die entsprechende Untersuchung keine Politikbereiche mit gravierenden sozioethischen Konflikten (z.B. Abtreibung oder Humangenetik) einbezogen wurden, bei denen eine politische Deblockierung sehr viel schwieriger zu erreichen ist (vgl. Papadopoulos 1995: 121-123).

- Das im jeweiligen Politikbereich *dominierende Instrument* (regulative Politik, finanzielle Anreize usw.) setzt den Grundrahmen für die Art und Weise des Lernens (= Lernform). In der Untersuchung von Knoepfel, Kissling-Näf und Marek (1997) kamen nicht in allen Politikbereichen die gleichen Lernformen vor¹¹. Anbefohlenes Lernen fand sich - wesensnotwendig - ausschliesslich in regulativen Politiken, und instrumentelles Lernen (Verhaltensänderungen aufgrund negativer oder positiver Anreize) - ebenso wesensnotwendig - in Politikbereichen, welche finanzielle Instrumente einsetzten. Die Lernform des "trial and error" wurde wohl auch nicht ganz überraschend vor allem in Politiken gefunden, die sich durch eine hohe Komplexität auszeichnen (Abfallanlagen mit Umweltverträglichkeitsprüfung (UVP), ökologische Direktzahlungen, Gesundheitspolitik). Die Form des "Lernens am Modell" (Beobachtung und Übertragung von Lösungen, die zu anderen Zeitpunkten oder anderenorts erfolgreich eingesetzt wurden; "lesson-drawing" nach Rose 1993) wurde dagegen in allen vier Politikbereichen gefunden; sie dürfte, da freiwillig und nur mit Informationskosten verbunden, die attraktivste Lernform darstellen. Von einer Institutionalisierung sozialer Lernprozesse kann bei experimentellen Gesetzgebungen (Lernen im "Laboratorium") gesprochen werden, wo, wie etwa in der Drogenpolitik Effizienz und Wirksamkeit der Intervention begleitend evaluiert werden.

¹¹ Unterschieden wurden folgende Lernformen: anbefohlenes Lernen, instrumentelles Lernen, trial and error, Lernen am Modell ("lesson-drawing"), Lernen im "Laboratorium" (Pilotversuche, Sozialexperimente).

- Policy-orientierte Lernprozesse setzen den Austausch mit der Umwelt und den zentralen Akteuren voraus. Medium dafür ist das *Netzwerk* (vgl. Kapitel 7.2.). Lernprozesse und Veränderungen der Netzwerkstruktur gehen Hand in Hand. Netzwerke sind in metaphorischer Weise der "Ort" für Mechanismen der Interaktion und Kooperation zwischen verschiedenen Akteuren im Hinblick auf die Lösung von Policy-Problemen. Sie bilden ein Medium für Lernprozesse und zwar im doppelten Sinne des Wortes: Erstens drücken sich - nach den Ergebnissen des Projekts Knoepfel, Kissling-Näf und Marek (1997) - leistungsstarke Lernprozesse in relativ starken Veränderungen der Netzwerkstruktur aus, d.h. beides geht Hand in Hand (ohne nachweisbare kausale Beeinflussung in die eine oder andere Richtung). Zweitens - dies zeigt eine detaillierte historische Analyse in den Fallstudien des erwähnten Projekts - erleichtert eine Aktivierung und bewusste Zusammensetzung des Netzwerks Lernprozesse. Öffentliche "Entrepreneurs" haben in mehreren Fällen als aktive "Networker" zusätzliche Personen um einen Tisch versammelt. Die Erweiterung des Netzwerks hat den Aufbau neuer Kommunikationszusammenhänge begünstigt und gleichzeitig auch neue Potentiale des Ressourcentauschs (siehe weiter unten) eröffnet. Dadurch wurden Politikinnovationen überhaupt erst möglich.

Die Notwendigkeit der Öffnung von Organisationen gegenüber der Umwelt als Voraussetzung für Lernprozesse wird auch von Urio, Bari und Glauser (1995: 187-189) betont, ohne dass dabei auf den Netzwerkbegriff zurückgegriffen wird. Diese Autoren weisen darauf hin, dass der Erfolg eines Programms auch davon abhängt, inwieweit es gelingt, die Teilmassnahmen in eine kohärente Gesamtpolitik einzubetten und dafür die Unterstützung der Hauptakteure zu gewinnen. Eine solche kohärente Gesamtstrategie braucht eine ausreichend lange Vorlaufzeit. Sie kann im Rahmen von Debatten innerhalb der Organisation und zwischen dieser und den externen Akteuren erarbeitet werden. Eine kohärente Strategie bzw. eine gemeinsame Sicht des Politikprogramms innerhalb der Organisation muss dabei die Öffnung gegenüber der Umwelt und Lernprozesse nicht beeinträchtigen; beobachtet wurde vielmehr das Gegenteil (Urio, Bari und Glauser 1995: 195). Dieser scheinbar paradoxe Befund lässt sich dadurch erklären, dass erfolgreiche Interaktionen mit der Umwelt und Lernprozesse systemtheoretisch eine interne Stabilisierung und gleichzeitig eine Fokussierung voraussetzen (vgl. Bussmann 1994: 58 f.).

- Lernprozesse sind oft begleitet von einem *Ressourcentausch*; neue (z.B. über den Politikbereich hinausgreifende) Möglichkeiten des Ressourcentauschs können Voraussetzung für erfolgreiche Lernleistungen sein. Für beide Seiten vorteilhafte Austauschbeziehungen (z.B. von Informationen, von Legitimation) sind das zentrale Element von Netzwerken. Netzwerkbildung und Ressourcentausch¹² sind deshalb zwei Seiten der gleichen Münze. Ressourcentausch kann zugleich aber an erweiterte Voraussetzungen, insbesondere an den erfolgreichen Abschluss von Verhandlungen gebunden sein. Die kommunikativen Austauschbeziehungen zwischen den Akteuren des Netzwerks fördern die "Empathie", d.h. das Verständnis für die gegenseitige Verhandlungslogik, und bilden deshalb eine geeignete Voraussetzung für solche Verhandlungen. Letztere können als die dominante Logik von Netzwerken bezeichnet werden (Mayntz 1993b: 11).

Ein erweitertes Netzwerk kann insbesondere neue Partner in Kommunikationsbeziehungen bringen und neue Potentiale für den Ressourcentausch, z.B. über mehrere Politiken (Interpolicy) oder über mehrere Partner (z.B. Ringtausch) ermöglichen. Im Projekt Knoepfel, Kissling-Näf und Marek (1997) wurde beispielsweise die Zustimmung einer Gemeinde zum Bau einer Kehrrichtverbrennungsanlage auf ihrem Gebiet mit der Verpflichtung des Kantons zum Bau einer Umfahrungsstrasse verbunden, und in der Landwirtschaftspolitik übernahmen die Produzentenorganisationen Vollzugsaufgaben bei den ökologischen Direktzahlungen und erhielten dafür als Gegenleistung Mitspracherechte bei der Ausformulierung ihrer Bedingungen (Tausch - Vollzugsleistungen gegen Mitsprache). Über den bisherigen Politikbereich hinausgreifende Austauschbeziehungen (z.B. zwischen Landwirtschaft und Umweltschutz) können, sofern auf Dauer angelegt, *Politiken neu verknüpfen*, was die Palette der Möglichkeiten für Lösungen erweitert. Letztlich können dadurch die Grenzen der Politikarenen neu abgesteckt werden (was indessen vermutlich Veränderungen in den grundlegenden Wertvorstellungen voraussetzt).

- Lernprozesse laufen auch über Einzelpersonen bzw. werden durch öffentliche "*Entrepreneurs*" wesentlich begünstigt. Obwohl nicht im

¹² Knoepfel, Kissling-Näf und Marek (1997) unterscheiden als grundlegende Ressourcen: Konsens, Legitimation, Geld, Information und Zeit. In dieser Studie nicht thematisiert wurden die beiden anderen Ressourcen "politische Macht" und "Personal"; vgl. dazu Kapitel 4.1.

Zentrum der Untersuchungen von Knoepfel, Kissling-Näf und Marek (1997) und Urio, Bari und Glauser (1996) trat in den Fallstudien die bedeutsame Rolle von Persönlichkeiten für Lernprozesse deutlich hervor. In einzelnen Fällen (v.a. Gesundheitspolitik) war bereits das Netzwerk stark auf einen oder mehrere Pole ausgerichtet ("Zentralität" des Netzwerks) und damit in hohem Masse personenbezogen bzw. -abhängig (Gentile 1995).

Der "persönliche Faktor" kann Lernprozesse auf unterschiedliche Art und Weise begünstigen. Eine Möglichkeit für "Entrepreneurs" besteht etwa darin, die Fäden nur indirekt zu ziehen. Dies geschieht etwa dadurch, dass weitere Partner an den Tisch gebracht und das Netzwerk erweitert wird ("*networking*"). Solches bewusstes Networking erlaubt die Herstellung erweiterter Kommunikationsbeziehungen zwischen Partnern, deren Handeln sich gegenseitig beeinflusst. Die Beziehungen innerhalb des Netzwerks und der Erfolg des darin stattfindenden Austauschs hängen durchaus auch von der *Bereitschaft* zentraler Akteure zur *Kooperation* ab. In der Evaluation von Rey et al. (1993) weigerte sich beispielsweise ein wichtiger Akteur, sich auf den Evaluationsprozess einzulassen. Es gelang ihm dadurch allerdings nicht, die Evaluation völlig zu torpedieren.

Der "persönliche Faktor" spielt aber auch bei der Weitergabe kognitiver Wissensbestände und für die Glaubwürdigkeit von Wissen eine bedeutende Rolle. Das Ausmass der beruflichen Weiterbildung und deren Form (Bibliothek, Dokumentation, Kolloquien, informelle Kontakte zwischen Wissenschaft und Verwaltung usw.) können durch initiative administrative Führungspersönlichkeiten wesentlich beeinflusst werden (Urio, Bari und Glauser 1996: 195-197). Wichtig ist auch, in welcher Form das Wissen übermittelt wird. Je grösser der Nutzen im Vergleich zu den Kosten ist, desto höher sind die Chancen für eine Weiterverwendung (vgl. Nagel und Mills 1993: 89 ff.). Dient das Wissen zudem der Sicherung der eigenen Position, oder können Forschungsergebnisse zur Legitimation des eigenen Handelns verwendet werden, erhöht sich die Wahrscheinlichkeit für deren Weiterverwendung. Es hat sich auch gezeigt, dass die Verwaltungsmodernisierung und die Einführung von neuen Methoden oft an Neu- und Umbe-setzungen gebunden sind (vgl. Freiburghaus und Zimmermann 1985: 248).

In der Organisationellen Lernforschung werden die obigen Aussagen bestätigt. In den Schlussfolgerungen zu einer internationalen Studie

zum Lernen von Regierungen formuliert Rist (1994: 193-204) Vorbedingungen für kollektive Lernprozesse. Er zeigt, dass Organisationen internes Material leichter und stärker in ihre Überlegungen aufnehmen als externe Informationen; die Akzeptanz der Informationen korreliert zudem positiv mit der Glaubwürdigkeit der Quelle. Wichtig scheint auch, ob der interne Empfänger glaubwürdig ist und die Untersuchung in den Augen anderer Institutionen legitim ist; der Diffusion von Wissen generell förderlich sind zudem informelle Kontakte.

- *Die Umdeutung von Instrumenten* kann deren *Aneignung* ("Appropriation") erleichtern und dadurch weitreichende Veränderungen auslösen. In mehreren Fallstudien des Projekts Knoepfel, Kissling-Näf und Marek (1997) wurden Situationen beobachtet, in welchen Akteure neue Anliegen und Politikfragen aktiv aufgriffen und die neuen *Argumente in ihre Position einbauten*. Sie konnten dadurch mögliche Nachteile begrenzen und ihre Interessen wahren. So beteiligten sich an den ökologischen Direktzahlungen viele Bauern aufgrund wirtschaftlicher Überlegungen und nicht primär aufgrund ihres ökologischen Bewusstseins. Man könnte auf den ersten Blick meinen, dadurch werde ein blosses "Scheinelernen" ausgelöst, das sich in einer Neubenennung bisheriger Sachverhalte äussere. Es können indessen Lernprozesse induziert werden, die darüber hinausgreifen. Einerseits werden in Bezug auf Instrumente neue Argumentationsketten aufgebaut, welche den Wertewandel beeinflussen können. Andererseits werden Akteure und Netzwerke für neue Zwecke genutzt ("den Bock zum Gärtner machen"), innerhalb derer sie neue Verhaltensmöglichkeiten entdecken können. Dadurch werden Chancen zur Aneignung ("Appropriation") von neuen Instrumenten geschaffen. Knoepfel, Kissling-Näf und Marek (1997, Kapitel 10.2., Punkt 1) schreiben denn auch: "Die Notwendigkeit von Aneignungsprozessen bei einem Instrumentenwechsel legt die Vermutung nahe, dass solche Situationen (Umdeutung des Instruments, A.d.A.) eigentliche Sternstunden für kommunikative Lernprozesse in den Netzwerken öffentlicher Politiken darstellen. Im Laufe ihrer Aneignung erfahren nicht nur die Instrumente und die Outputs, sondern wegen intensivierter Lernprozesse auch die einzelnen Akteure (...) Veränderungen."
- Ein gut ausgebautes *Wissenssystem*, insbesondere geeignete Interfaces begünstigen den Wissenstransfer. Aus systemtheoretischer Sicht ist auf

die unterschiedlichen Referenzsysteme von Wissenschaft (Wahrheit) und Politik (Macht und Machbarkeit) hinzuweisen (vgl. Freiburghaus 1989). Beide Systeme bedürfen daher geeigneter Interfaces, um sinnvoll miteinander kommunizieren zu können. Im Agrarbereich (vgl. Bussy 1995) wurden mit den landwirtschaftlichen Schulen solche Interfaces zwischen Forschung und Praxis geschaffen. Wird das Know-how zusätzlich noch an eigenen öffentlichen Forschungsanstalten weiterentwickelt, ergeben sich für die Diffusion und Zugänglichkeit optimale Bedingungen (Zimmermann 1990). Die im Rahmen des Projekts Urio, Bari und Glauser (1997) untersuchten Fälle lassen zudem erkennen, dass die Einführung von Evaluationssystemen durch Chefbeamte begünstigt wurde, die über eine solide akademische Ausbildung und eine gewisse Affinität zum Wissenschaftsbereich verfügten. Im Rahmen der Tessiner Massnahmen zur Verhütung kardio-vaskulärer Krankheiten wurde im Rahmen wissenschaftlicher Kolloquien der Dialog Wissenschaft-Praxis bewusst gesucht und erfolgreich abgewickelt (Urio, Bari und Glauser 1995: 184-186).

Die Ausdifferenzierung des Wissenssystems dürfte, wie ein Vergleich verschiedener Politiken gezeigt hat, eine direkte Folge des Alters einer Politik sein. Allerdings haben auch jüngere Politiken wie die Umweltpolitik in relativ kurzer Zeit starke Brücken zwischen Wissenschaft und Politik aufgebaut.

13.5. IMPLIKATIONEN FÜR DIE UMSETZUNG VON EVALUATIONEN

Die Darstellung von Faktoren, welche Voraussetzung für Lernprozesse sind bzw. damit einhergehen, kann geeignet sein, die Frage der Umsetzung (vgl. auch Kapitel 6) von bzw. des Lernens aus Evaluationen aus steril gewordenen Verknüpfungen zu lösen. Die Diskussion konzentrierte sich bisher vor allem auf die Frage, ob Evaluationen instrumentell oder zur "Aufklärung" ("enlightenment") benutzt würden. Die Beobachtung von Lernprozessen in öffentlichen Politiken lässt dagegen die Vermutung aufkommen, dass diese Unterscheidung nicht dermassen bedeutsam ist bzw. dass viele Lernprozesse quer dazu ablaufen.

Während sich die frühere Verwendungsforschung dafür interessierte, inwieweit die *Resultate* von Evaluationen messbar umgesetzt werden, legt die am Lernen in öffentlichen Politiken interessierte Forschung die Ver-

mutung nahe, dass der *Prozess* der Evaluation auf jeden Fall auch in die Überlegungen und in die empirischen Untersuchungen einbezogen werden muss (vgl. King 1995). Schon aus dem Evaluationsprozess selbst können Erkenntnisse und Verhaltensänderungen bei den Akteuren resultieren, die mit den eigentlichen Resultaten der Evaluation (wie sie etwa im Schlussbericht formuliert werden) nur in einem lockeren Zusammenhang stehen.

Evaluationen als Prozess können Veränderungen in der *Netzwerkstruktur* induzieren. Ein neuer Akteur tritt auf den Plan: die Wissenschaft/Beratung. Durch die Durchführung einer oder die Beteiligung an einer Evaluation können Evaluatorinnen oder Evaluatoren an Glaubwürdigkeit gewinnen (bzw. gegebenenfalls auch verlieren). Nicht nur ihre Resultate können Einfluss nehmen, sondern auch ihre weiteren Äusserungen, ja ihre ganze Herangehensweise an das Problem. Im Gegensatz zu den meisten anderen evaluationsbeteiligten Akteuren sind sie vom praktischen Handlungszwang entlastet und können als reflektierende Beobachter neue Gesichtspunkte einbringen. Dabei können durchaus auch Argumente und Standpunkte in den Prozess einfließen, welche nicht unbedingt auf die empirischen Ergebnisse der Evaluation gestützt sind. Der Einbezug der Wissenschaft kann durchaus die *Qualität der Kommunikationsbeziehungen im Netzwerk* beeinflussen, beispielsweise indem vermehrt darauf geachtet wird, dass Behauptungen anhand von Beispielen belegt werden. Im Falle lernorientierter, partizipativer Evaluationen werden im Rahmen der Evaluation zudem oft weitere Akteure einbezogen, was eine Veränderung des Netzwerks zur Folge hat.

Wenn Evaluationen nicht als Black-box-Evaluationen ausgestaltet sind, sondern Einblick in die Policy-Dynamik geben, können sie Faktoren ausleuchten, die hemmende oder fördernde Funktion für den Vollzug und die Wirksamkeit von Massnahmen haben. Dabei werden sehr oft neue, bisher ausgeblendete Akteure identifiziert, deren Kooperation für den Erfolg einer Politik bedeutsam oder notwendig ist. Evaluationen schaffen dadurch die Informationsvoraussetzungen, um diese Akteure ins Netzwerk einzubeziehen. Wenn sie detaillierten Einblick in deren Handlungslogik geben, erleichtern sie die Voraussetzungen für den *Ressourcentausch*, welcher es ermöglicht, eine aktivere Mitwirkung der für den Vollzug bedeutsamen Akteure einzuhandeln.

Beziehen sich Evaluationen auf Interpolicy-Bereiche (z.B. Ökologie und Landwirtschaft), was bei der Ermittlung (intendierter oder nicht intendierter) Nebenwirkungen in der Regel unausweichlich ist, können sie bisher unberücksichtigte Querbeziehungen (z.B. Beitrag der landwirt-

schaftlichen Beratung zur ökologischen Sensibilisierung der Bauern) ans Licht bringen. Sie schaffen neue Argumentationsketten und induzieren damit kognitive Änderungen. In diesem Kontext kann vor allem die Aufklärungsfunktion von Evaluationen gesehen werden. Letztlich können Evaluationen einen Beitrag dazu leisten, unterschiedliche Politiken gegenseitig neu zu "verknüpfen". In etwas bescheidenerer Dimension können Evaluationen aber auch die Aneignung von Instrumenten ("Appropriation") erleichtern und dadurch potentiell Änderungen auslösen, die über das entsprechende "Scheinelernen" hinausgreifen.

Eine ganz bedeutsame Funktion haben Evaluationen im Zusammenhang mit dem "*Lernen am Modell*" ("lesson-drawing"). Dies wurde weiter oben als die am weitesten verbreiteten Lernform dargestellt. Sie hat den Vorteil, dass sie bloss Informationskosten über die anderenorts bzw. früher gemachten Erfahrungen mit Instrumenten verursacht. Evaluationen können eine massgebliche Quelle für solches Lernen sein. Ergebnisse von bzw. Lehren aus spezifischen Evaluationen sind nicht nur für die untersuchte Massnahme von Bedeutung. Sie haben einen potentiell sehr grossen Anwendungsbereich, da sie auf ganz unterschiedliche Kontexte übertragen werden können (z.B. vom kommunalen auf den kantonalen Kontext). Im Rahmen von *Evaluationssynthesen* können die Schwächen einzelner Studien eliminiert werden. Dadurch wird eine höhere Glaubwürdigkeit der Resultate erzielt, indem beispielsweise Studien über einzelne kommunale/kantonale Massnahmen oder Untersuchungen über die Anwendung eines bestimmten Instruments (wie Informationskampagnen) in verschiedenen Politikbereichen aggregiert werden.

Die Überlegungen zum *persönlichen Faktor* lassen zudem vermuten, dass Evaluationen, auch wenn sie politikorientiert angelegt sind, im Hinblick auf Lernprozesse davon profitieren können, dass sie bei einer im Politiknetzwerk bedeutsamen und zudem der Wissenschaft aufgeschlossenen Person "angedockt" werden. Denn neue Interfaces zwischen Politik und Wissenschaft erleichtern die Umsetzung von Evaluationsstudien beträchtlich.

LITERATURVERZEICHNIS

- Ackoff, Russell L. 1953. *The design of social research*. Chicago: University of Chicago Press.
- Aebischer, B. et al. 1993. *Evaluation der Bewilligungspflicht für elektrische Widerstandsheizungen*, Bern: Bundesamt für Energiewirtschaft, Schriftenreihe Evaluation/ EDMZ.
- Aldenderfer, Mark S. und Blashfield, Roger K. 1984. *Cluster Analysis*. Beverly Hills: Sage.
- Alkin, Marvin C. 1990. *Debates on evaluation*. University of California Newbury Park: Sage.
- Alkin, Marvin C., Daillak, R. und White, P. 1979. *Using evaluation: Does evaluation make a difference?* Beverly Hills, CA: Sage.
- Amy, D. 1987. *The politics of environmental mediation*. New York.
- Amy, D. 1983. The politics of environmental mediation, *Ecology Law Quarterly* 11, 1: 1-19.
- Andrews, Frank M., Klem, Laura, Davidson, Terrence N., O'Malley, Patrick M. und Rodgers, Willard L. 1981. *A guide for selecting statistical techniques for analyzing social science data*. 2nd. ed.. Ann Arbor: University of Michigan, Institute for Social Research.
- Anwander, Sibyl u.a. 1990. *Direktzahlungen an die Berglandschaft. Eine Gesetzesevaluation*. Zürich: Verlag der Fachvereine.
- "Arbeitsgruppe Gesetzesevaluation" 1991. *Die Wirkungen staatlichen Handelns besser ermitteln: Probleme, Möglichkeiten, Vorschläge*, Schlussbericht an das Eidgenössische Justiz- und Polizeidepartement. Bern: EDMZ (Artikel Nr. 407.570 d).
- "Arbeitsgruppe Verkehrssicherheit" 1983. *Versuch Tempo 50*. Schlussbericht, Bern: EDMZ.
- Arbeitspapier Nr. 1, in: *Organisationsentwicklung 3/1996*, Chemnitz-Zwickau: Technische Universität.
- Ashford, Dag (ed.) 1990. Discretionary Politics: Intergovernmental Social Transfers in Eight Countries, *International Review of Comparative Public Policy*. Greenwich (Ct.): JAI Press.
- Ashford, Dag 1978. The Structural Analysis of Public Policy or Institutions Really Do Matter, in: Ashford, Dag (ed.), *Comparing Public Policies*. Beverly Hills/London: Sage: 81-98.
- Atteslander, Peter 1993. *Methoden der empirischen Sozialforschung*, 7. Auflage. Berlin: de Gruyter.

- Aucoin, Peter 1988. Contraction, Managerialism and Decentralization in Canadian Government, in: *Governance. An International Journal of Policy and Administration* 1: 144-161.
- Backhaus, Klaus, Erichson, Bernd, Flinke, Wulff und Weiber, Rolf 1994. *Multivariate Analysemethoden*, 7. Auflage, Berlin: Springer.
- Bahy, M. 1986. *Evaluation des traitements de substitution (cures à la méthadone) dans le canton de Genève de 1981 à 1984*. Rapport du médecin cantonal. Juin 1986. (mimeo)
- Baitsch, Christof 1993. *Was bewegt Organisationen? Selbstorganisation aus psychologischer Perspektive*. Frankfurt a. M.
- Baitsch, Christof, Knoepfel, Peter und Eberle, Armin 1996. Prinzipien und Instrumente organisationalen Lernens. Dargestellt an einem Fall aus der öffentlichen Verwaltung, *Organisationsentwicklung* (Zürich) 3: 4-21.
- Balthasar, Andreas et al. 1995. *Der Vollzug von Bundespolitiken durch die Kantone. Eine Metaevaluation. Schlussbericht zuhanden der Parlamentarischen Verwaltungskontrollstelle*. Bern/Luzern (noch unveröffentlicht).
- Balthasar, Andreas und Knöpfel, Carlo 1994. *Umweltpolitik und technische Entwicklung*. Basel: Helbing & Lichtenhahn (Reihe Ökologie & Gesellschaft, Bd. 8).
- Baroni, Daniela, Brunner, Verena, Knoepfel, Peter und Moor, Pierre 1991: *Strassenverkehrsrecht im Lichte des Umweltrechts - Le droit de la circulation routière face à la législation de l'environnement*. Basel: Helbing & Lichtenhahn (Reihe Ökologie & Gesellschaft, Bd. 5).
- Baslé, Maurice 1994. *L'évaluation des politiques publiques et des contrats de plan Etat-région en France: la situation à fin 1994*, Cahiers économiques de Bretagne no 4/1994, Rennes: Faculté des sciences économiques de l'Université de Rennes.
- Bassand, Michel, Joye, Dominique und Schuler, Martin (Hrsg.) 1988. Les enjeux de l'urbanisation: Agglomerationsprobleme in der Schweiz. *Société suisse d'études pour l'organisation de l'espace et la politique régionale (OEPR/Schweizensche Studiengesellschaft für Raumordnung und Regionalpolitik (ROREP)*. Bern etc.: Lang.
- Beck, Ulrich und Bonss, Wolfgang (Hrsg.) 1989. Verwissenschaftlichung ohne Aufklärung, in: Beck, Ulrich und Bonss, Wolfgang, *Weder Sozialtechnologie noch Aufklärung? Analysen zur Verwendung sozialwissenschaftlichen Wissens*. Frankfurt a.M.: Suhrkamp: 7-45.

- Bennett, C.J. und Howlett M. 1992. The lessons of learning: Reconciling theories of policy learning and policy change, *Policy Sciences* 25: 275-294.
- Benninghoff, Martin 1995. *D'un processus législatif à une réorganisation administrative: l'Office fédéral de l'agriculture face aux enjeux écologiques*, Cahiers de l'IDHEAP no 149, Chavannes-près-Renens: IDHEAP.
- Benseler, F., Hejl, P.M. und Köck, W.K. (Hrsg.) 1980. *Autopoiesis, Communication and Society: The Theory of Autopoietic System in the Social Sciences*. Frankfurt/New York.
- Benz, A. 1995. Verhandlungen, Politiknetzwerke in der Horizontalen Politikverflechtung, in: Jansen, Dieter und Schubert, Karl (Hrsg.), *Netzwerke und Politikproduktion: Konzepte, Methoden, Perspektiven*, Marburg: 185-204.
- Berger, Peter und Luckmann, Thomas 1966. *Die gesellschaftliche Konstruktion der Wirklichkeit*. 3. Aufl., Frankfurt: Fischer.
- Beywl, Wolfgang 1988. *Zur Weiterentwicklung der Evaluationsmethodologie. Grundlegung, Konzeption und Anwendung eines Modells der responsiven Evaluation*. Bern: Lang.
- Bierter, Willy und Binder, Hans-Martin 1993. *Unternehmerische Innovationsprozesse. Die Wirksamkeit staatlicher Förderungsmassnahmen - eine Evaluation*. Chur/Zürich: Rüegger.
- Bierter, Willy, Binder, Hans-Martin und Rüegg, E. 1985. *Energiepolitische Handlungsmöglichkeiten auf kommunaler und kantonaler Ebene*. Liestal.
- Binder, Hans-Martin 1994. Lernprozess als Ziel einer interaktiven Evaluationsstrategie, in: Bussmann, Werner (Hrsg.), *Lernen in Verwaltungen und Policy-Netzwerken*, Chur/Zürich: Rüegger: 75-84.
- Bingham, G. 1986. *Resolving environmental disputes: A decade of experience*. Washington D. C.
- Blossfeld, Hans-Peter, Hamerle, Alfred und Mayer, Karl Ulrich 1986. *Ereignisanalyse*. Frankfurt: Campus.
- Blume, St., Hagendijk, R. und Prins, A. 1991. Political culture and the policy orientation in Dutch social science, in: Wagner, P. und Weiss, C.H. (eds.), *Social Sciences and Modern States. National Experiences and Theoretical Crossroads*. Cambridge: Cambridge University Press: 168-190.
- Blumer, H. 1976. Der methodologische Standort des Symbolischen Interaktionismus, in: Arbeitsgruppe Bielefelder Soziologen (Hrsg.), *All-*

- tagswissen, *Interaktion und gesellschaftliche Wirklichkeit, Band 1: Symbolischer Interaktionismus und Ethnomethodologie*, Reinbek b/Hamburg, 3. Auflage: 80-146.
- Bobrow, Davis B. und Dryzek, John S. 1987. *Policy Analysis by Design*. Pittsburgh: University of Pittsburgh Press.
- Boggio, Yann et al. 1995. *Le point de vue des victimes sur l'application de la LAVI*. Genève: Université de Genève, Centre d'Etude, de Technique et d'Evaluation Législatives
- Bohnet, Iris und Frey, Bruno S. 1994. Direct-Democratie Rules: The Role of Discussion. *Kyklos* 3: 341-354.
- Bollen, Kenneth A. 1989. *Structural equations with latent variables*. New York: Wiley.
- Bollen, Kenneth A. und Long, Scott J. 1993. *Testing structural equation models*. Newbury Park: Sage.
- Bortz, Jürgen 1989. *Statistik. Für Sozialwissenschaftler*. 3. Auflage. Berlin: Springer.
- Bortz, Jürgen 1984. *Lehrbuch der empirischen Forschung für Sozialwissenschaftler*. Berlin: Springer.
- Bossel, H. 1989. *Simulation dynamischer Systeme*. Braunschweig/Wiesbaden.
- Box, George E.P. und Jenkins, Gwilym M. 1976. *Time series analysis. Forecasting and control*. Rev. ed. Oakland: Holden Day.
- Brandner, Balthasar et al. 1995. *Skitourismus. Von der Vergangenheit zum Potential der Zukunft*. Chur: Rüegger.
- Brewer, Garry D. 1978. Termination: Hard Choices, Harder Questions, *Public Administration Review*, Vol. 38, No. 3: 338-344.
- Brewer, Garry D. und DeLeon, Peter 1983. *The Foundations of Policy Analysis*. Chicago: The Dorsey Press.
- Bryk, A.S. (Hrsg.) 1983. *Stakeholder-based evaluations*. San Francisco: New Directions for Program Evaluation, No 17, Jossey-Bass.
- Bryman, Alan 1990. *Quantity and Quality in Social Research*. London: Unwin Hyman.
- Buckle, L.G. und Thomas-Buckle, S.R. 1986. Placing environmental mediation in context: Lessons from "failed" mediations. *Environmental Impact Assessment Review* 6: 55-70.
- Busch, A. und Merkel, W. 1992. Staatshandeln in kleinen Staaten: Schweiz und Österreich, in: Abromeit, H. und Pommerehne, W., *Staatstätigkeit in der Schweiz*, Bern: 193-219.

- Bussmann, Werner 1996a. Evaluationen in Regierung und Verwaltung sinnvoll einsetzen, in *Schriftenreihe der Schweizerischen Gesellschaft für Verwaltungswissenschaft*. Bern (im Erscheinen).
- Bussmann, Werner 1996b. Democracy and Evaluation's Contribution to Negotiation, Empowerment and Information: Some Findings from Swiss Democratic Experience. *Evaluation* 2(3): 307-319.
- Bussmann, Werner 1995a. *Evaluationen staatlicher Massnahmen erfolgreich begleiten und nutzen. Ein Leitfaden*. Chur/Zürich: Rüegger.
- Bussmann, Werner 1995b. Evaluations and grassroots politics: The case of Switzerland. *Knowledge and Policy*, Fall, 8/3: 85-98.
- Bussmann, Werner 1995c. *Accompagner et mettre à profit avec succès les évaluations des mesures étatiques, guide de réflexion*. Genève: Edition Georg.
- Bussmann, Werner 1995d. Instrumente der Erfolgskontrolle, *VOP*: 345-351.
- Bussmann, Werner 1995e. Wirkungsorientierte Verwaltung, NPM und Evaluationen, in: Hablützel Peter et al. (Hrsg.), *Umbruch in Politik und Verwaltung - Ansichten und Erfahrungen zum New Public Management in der Schweiz*. Bern: Haupt: 367-378.
- Bussmann, Werner (Hrsg.) 1994. *Lernen in staatlichen Verwaltungen und Policy-Netzwerken*. Chur/Zürich: Rüegger.
- Bussmann, Werner 1993. Was haben Vollzugsstudien der Verwaltungspraxis gebracht? *Schweizerisches Jahrbuch für Politische Wissenschaft*, 33: 261-275.
- Bussmann, Werner 1992a. Rechtsetzung unter Unsicherheit, *Gesetzgebung heute*, 2: 39-52.
- Bussmann, Werner 1992b. Der schweizerische Föderalismus: Anachronismus oder Modell für Europa? *Schweizerisches Jahrbuch für Politische Wissenschaft*, 32: 107-119.
- Bussmann, Werner 1989. Von der Doppelbödigkeit des Verhältnisses zwischen Wissenschaft und Politik, *Schweizerisches Jahrbuch für Politische Wissenschaft*, 29: 17-30.
- Bussmann, Werner 1986. *Mythos und Wirklichkeit der Zusammenarbeit im Bundesstaat*. Bern/Stuttgart: Haupt.
- Bussmann, Werner 1981. Vollzugsprobleme des Gewässerschutzes auf kantonaler Ebene, *Schweizerisches Jahrbuch für Politische Wissenschaft* 21: 241-254.
- Bussmann, Werner 1980. *Gewässerschutz und kooperativer Föderalismus in der Schweiz*. Bern: Haupt.

- Bussy, Claire 1995. *Apprentissage dans les politiques publiques. Application de l'an. 31b de la Loi sur l'agriculture dans les cantons de Neuchâtel, Zoug et Nidwald*, Cahiers de l'IDHEAP no 152, Chavannes-près-Renens: IDHEAP.
- Busterud, J. 1990. Mediation: The State of the Art, *The Environmental Professional* 2: 34-39.
- BUWAL 1991. *Zur Lage der Umwelt in der Schweiz*. Bern.
- Campbell, Donald T. und Stanley, J.C. 1963. Experimental and quasi-experimental designs for research on teaching, *Handbook of Research on Teaching*, N.L.Gage. Chicago: Rand McNally: 171-246.
- Carnevale, P. 1986. Strategie choice in mediation, *Negotiation Journal* 2, 2: 41-56.
- Castle, F. 1993. *Is Expenditure enough? On the nature of the dependent variable in comparative public policy Analysis*. Québec (unv. MS).
- Cedro, Lorenzo und Gualtierotti, Antonio F. 1996. Une "étude de cas" dans le domaine des sondages par simulation à l'aide du logiciel Minitab, *Etude de cas de l'IDHEAP*, no 9, Chavannes-près-Renens: IDHEAP.
- C.E.O.P.S. (Conception d'évaluations pour les organisations et les politiques publiques / Ecole nationale des travaux publics et de l'Etat) 1990. *Evaluation et décision*, Séminaire de méthode, Paris: Plan urbain.
- Chelimsky, Eleanor 1987. Linking Program Evaluation to User Needs, in: Palumbo, Dennis J. (Hrsg.), *The Politics of Program Evaluation*. Newbury Park: Sage: 72-99.
- Chelimsky, Eleanor 1985. *Program evaluation: Pattern and directions*. Washington: ASPA.
- Chen, Huey-Tsyh 1994. Current trends and future directions in program evaluation, *Evaluation Practice*, Vol. 15, No. 3: 229-238.
- Chen, Huey-Tsyh 1990. *Theory-Driven Evaluations*. Newbury Park/London etc.: Sage.
- Chen, Huey-Tsyh und Rossi, Peter (Hrsg.) 1992. *Using theory to improve program and policy evaluations*. New York: Greenwood Press.
- Clarke, Harold D. und Dutt, Nitish 1991. Measuring value change in western industrialized societies: The impact of unemployment, *American Political Science Review*, Vol. 85, No. 3: 905-920.
- Claus, Frank und Wiedemann, Peter (Hrsg.) 1994. *Umweltkonflikte. Vermittlungsverfahren zu ihrer Lösung*. Taunusstein: Blottner

- Cohen, Michael D., March, James, G. und Olson, Johan P. 1972. A Garbage Can Model of organizational Choice, *Administrative Science Quarterly* 17: 1-25.
- Commissariat général du Plan 1989. *Evaluer les politiques publiques*, Paris: La documentation française.
- Conseil scientifique de l'évaluation 1993, 1994 et 1995. *L'évaluation en développement*, rapport annuel sur l'évolution des politiques publiques, Paris: La documentation française.
- Cook, Thomas et al. (Hrsg.) 1992. *Meta-analysis for explanation: A case book*. New York: Russell Sage Foundation.
- Cook, Thomas D. 1985. Postpositivist Critical Multiplism, in: Shotland, R. Lance and Mark, Melvin M. (ed.), *Social Science and Social Policy*, Beverly Hills: Sage: 21-62.
- Cook, Thomas D. und Campbell, Donald T. 1979. *Quasi-experimentation. Design & analysis issues for field settings*. Boston: Houghton Mifflin.
- Cook, Thomas D. und Reichhardt, Charles S. 1979. *Qualitative and Quantitative Methods in Evaluation Research*. Beverly Hills: Sage.
- Cormick, G. und Knaster, A. 1986. Mediation and scientific issues, *Environment* 28: 6-30.
- Cousins, J.B. und Leithwood, K. 1986. Current empirical research on evaluation utilization, *Review of Educational Research*, 56: 331-364.
- Crowfoot, J.E. und Wondolleck, J.M. 1991. *Environmental disputes. Community involvement in conflict resolution*. Washington D.C.
- Daily, A., Weidner, H. und Fietkau, H.-J. (Hrsg.) 1994. *Mediation als politischer und sozialer Prozess. Loccumer Protokolle 73/93*. Rehburg-Loccum.
- Damkowski, Wulf und Precht, Claus 1995. *Public Management: neuere Steuerungskonzepte für den öffentlichen Sektor*. Stuttgart: Kohlhammer.
- deLeon, Peter 1987. Policy Termination as a Political Phenomenon, in: Palumbo, Dennis J., *The Politics of Program Evaluation*, Newbury Park: Sage: 173-199.
- deLeon, Peter 1983. Policy Evaluation and Program Termination, *Policy Studies Review*, Vol. 2, No. 4: 631-647.
- Demaris, Alfred 1992. *Logit Modeling. Practical Applications*. Newbury Park: Sage.

- Dente, Bruno 1989: *Politiche pubbliche e pubblica amministrazione*. Rimini: Maggioli Queste Istituzioni Ricerche.
- Denzin, Norman K. 1989. *The Research Act*. 3rd ed. Englewood Cliffs: Prentice Hall.
- Derlien, Hans-Ulrich 1994. Evaluation zwischen Programm und Budget, in: Hofmeister, Albert (Hrsg.), *Möglichkeiten und Grenzen der Programmsteuerung: Controlling und Evaluation. Verwaltungspraxis in Ost und West in Zeiten des Wandels*. Schriftenreihe der Schweizerischen Gesellschaft für Verwaltungswissenschaft, Band 21: 43-61.
- Derlien, Hans-Ulrich 1990. Genesis and Structure of Evaluation Efforts in Comparative Perspective, in: Rist, Ray C. (Hrsg.), *Program Evaluation and the Management of Government*. New Brunswick: Transaction Publishers: 146-176.
- Derlien, Hans-Ulrich 1976. *Die Erfolgskontrolle staatlicher Planung. Eine empirische Untersuchung über Organisation, Methode und Politik der Programmevaluation*. Baden-Baden: Nomos.
- Diekmann, Andreas und Mitter, Peter 1984. *Methoden zur Analyse von Zeitverläufen*. Stuttgart: Teubner.
- Dierkes, Meinolf, Weiler, Hans N. und Berthoin Anthal, Ariane 1987 (eds.) 1987. *Comparative Policy Research. Learning from Experience*. New York: St. Martin's Press.
- Döhler, M. 1994. Lernprozesse in Politiknetzwerken, in: Bussmann, W., *Lernen in Verwaltungen und Policy-Netzwerken*. Chur/Zürich: Rüegger: 39-54.
- Döhler, M. 1990. *Gesundheitspolitik nach der "Wende". Policy-Netzwerke und ordnungspolitischer Strategiewechsel in Grossbritannien, den USA und der Bundesrepublik Deutschland*. Berlin.
- Dreher, M. und Dreher, E. 1991. Gruppendiskussionsverfahren, in: Flick, U. et al. (Hrsg.), *Handbuch Qualitative Sozialforschung*. München: 186-188.
- Dryzek, John S. 1990. *Discursive democracy*. Cambridge.
- Dryzek, John S. und Ripley, Brian 1988. The Ambitions of Policy Design, *Policy Studies Review*, Vol. 7, No. 4: 705-719.
- Dubois-Arber, Françoise, Jeannin, André, Meystre-Agustoni, Giovanna, Gruet, Florence und Paccaud, Fred 1993. *Evaluation de la stratégie de prévention du SIDA en Suisse sur mandat de l'office fédéral de la santé publique. Quatrième rapport de synthèse 1991-*

1992. Cahiers de recherches et de documentation no 82. Lausanne: Institut universitaire de médecine sociale et préventive Lausanne.
- Dunn, William N. 1994. *Public Policy Analysis: An Introduction*. Second Edition. Englewood Cliffs: Prentice-Hall.
- Dunn, William N. und Kelly, Rita Mae (eds.) 1992. Advances in Policy Studies since 1950. *Policy Studies Review Annual*, Vol. 10. New Brunswick: Transaction.
- Duran, Patrice 1990. Le savant et la politique, pour une approche raisonnée de l'analyse des politiques publiques, *L'année sociologique* 40.
- ECOPLAN 1993. *Schneeräumung von Urner Passstrassen*. Evaluation im Auftrag der Baudirektion des Kantons Uri. Altdorf/Bern: Bauamt des Kantons Uri.
- Edelman, Murray 1964. *The symbolic uses of politics*. Chicago/London.
- Eichenberger, Reiner und Frey, Bruno S. 1994. Bessere Politik durch Föderalismus und direkte Demokratie, in: Hermann-Pillath, Carsten et al. (Hrsg.), *Marktwirtschaft als Aufgabe*. Stuttgart: Fischer: 774-787.
- Eidgenössisches Personalamt (Hrsg.) 1996. *Alarm um die Städte I Les villes et l'urgence d'agir*, EPA-Kolloquium 1995. Bern: Schriftenreihe des Eidg. Personalamtes.
- EJPD 1975. *Auswirkungen von Tempo 100/130*. Schlussbericht der vom EJPD eingesetzten Arbeitsgruppe "Tempo 100". Bern.
- Emery, Yves 1996. Le management de la qualité dans les administrations publiques: une des pierres angulaires du New Public Management, in: Emery, Yves (Hrsg./ed.), *Total Quality Management und ISO-Zertifizierung in der öffentlichen Verwaltung der Schweiz I Total Quality Management et certification ISO dans les administrations publiques en Suisse*, Schriftenreihe der schweizerischen Gesellschaft für Verwaltungswissenschaften (SGVW) / publication de la Société suisse des sciences administratives (SSSA), Band/volume 34. Bern: Verlag SGVW: 35-82.
- Erbring, Lutz 1990. Individuals write large. An epilogue on the "ecological fallacy", in: Stimson, James A., *Political Analysis*, Vol. 1. Ann Arbor: University of Michigan Press: 235-269.
- ERS Standards Committee 1982. Evaluation Research Society Standards for Program Evaluation, in: Rossi, Peter H. (ed.), *Standards of Evaluation Practice*. New Directions for Program Evaluation, Vol. 15. San Francisco: Jossey-Bass.

- Etheredge, L.S. und Short J. 1983. Thinking about government learning, *Journal of Management Studies* 20: 41-58.
- Faganini, H.P. 1991. *Föderalistischer Aufgabenverbund in der Schweiz*. Bern: Haupt.
- Fetterman, David (Hrsg.) 1996. *Empowerment evaluation. Knowledge and Tools for Self-Assessment and Accountability*. Newbury Park: Sage.
- Fetterman, David 1995. In Response to Dr. Daniel Stufflebeam's Empowerment Evaluation, Objectivist Evaluation and Evaluation Standards, *Evaluation Practice*, Vol.16, June: 179-199.
- Fetterman, David 1994. Empowerment Evaluation, *Evaluation Practice*. Vol. 15, February: 1-15.
- Fienberg, Stephen E. 1980. *The analysis of cross-classified categorical data*. Cambridge: MIT.
- Fietkau, H.-J. und Weidner, H. (unter Mitarbeit von Holzinger, K., Lackmann, B. und Pfingsten, K) 1996. *Umweltverhandeln*. Berlin (i.E.).
- Fietkau, H.-J. 1994. *Leitfaden Umweltmediation. Hinweise für Verfahrensbeteiligte und Mediatoren*. Schriften zu Mediationsverfahren im Umweltschutz Nr. 8, Wissenschaftszentrum Berlin für Sozialforschung FS II. Berlin: 94-323.
- Fietkau, H.-J. und Weidner, H. 1992. Mediationsverfahren in der Umweltpolitik, *Aus Politik und Zeitgeschichte* B 39-40: 24-34.
- Fingleton, Bernard 1984. *Models of category counts*. Cambridge: Cambridge University Press.
- Fischer, Frank 1994. *Evaluating public policy*. Chicago.
- Fischer, Frank 1992. Participatory expertise: Toward the democratization of policy science, in: Dünn, W. und Kelly, R. (eds.), *Advances in policy studies since 1950*. New Brunswick: 351-376.
- Fischer, Frank und Forester, John (ed.) 1993. *The argumentative turn in policy analysis und planning*. Durham/London.
- Fisher, R., Ury, W. und Patton, B. 1991. *Getting to yes: Negotiating agreement without giving in*. New York.
- Fitz-Gibbon, Carol Taylor und Morris, Lynn Lyons 1987. *How to Design a Program Evaluation*. Newbury Park: Sage.
- Flick, Uwe 1991. Triangulation, in: Flick, Uwe et al. (Hrsg.), *Handbuch Qualitative Sozialforschung*. München: 430-434.

- Flora, Peter und Heidenheimer, Arnold J. (eds.) 1981. *The Development of Welfare States in Europe and America*. New Brunswick/London: Transaction Books.
- Freiburghaus, Dieter 1989. Interfaces zwischen Wissenschaft und Politik, *Schweizerisches Jahrbuch für Politische Wissenschaft*, 29: 267-277.
- Freiburghaus, Dieter 1988. *Kleines Glossar zur Evaluationsforschung*. Lausanne (unv. MS).
- Freiburghaus, Dieter 1987. *Präventivmassnahmen gegen die Arbeitslosigkeit in der Schweiz*. Bern: Haupt.
- Freiburghaus, Dieter und Zimmermann, Willi 1985. *Wie wird Forschung politisch relevant? Erfahrungen in und mit den Schweizerischen Nationalen Forschungsprogrammen*. Bern/Stuttgart: Haupt.
- Freiburghaus, Dieter, Zimmermann, Willi und Balthasar, Andreas 1990. *Evaluation der Förderung praxisorientierter Forschung (KWF)*. Bern: Schriftenreihe des Bundesamts für Konjunkturfagen, Studie Nr. 12.
- Frey, Bruno S. und Bohnet, Iris 1994. The Swiss Experience with Referenda and Federalism. *Idioma, Revue de linguistique et de traductologie*: 147-160.
- Frey, Bruno S. und Kirchgässner, Gebhard 1993. Diskursethik, Politische Ökonomie und Volksabstimmungen. *Analyse & Kritik* 15: 129-149.
- Frey, René L. 1991. *Verteilung der Finanzierungslasten*. Vortrag am Kolloquium vom 8. und 9. November in Saarbrücken (unv. MS).
- Friedrichs, Jürgen 1984. *Methoden empirischer Sozialforschung*. 12. Auflage. Opladen: Westdeutscher Verlag.
- Fülgraff, Georges M. 1994. Das Mediationsverfahren zum Abfallwirtschaftskonzept des Kreises Neuss, in: Daily, A., Weidner, H. und Fietkau, H.-J. (Hrsg.), *Mediation als politischer und sozialer Prozess. Loccumer Protokolle 73/93*. Rehburg-Loccum.
- Fürst, D., Roggendorf, W., Scholles, F. und Stahl, R. 1994. Umweltinformationssysteme - ein effektiver Beitrag zum vorsorgenden Umweltschutz oder Ausdruck symbolischer Politik?, *UVP-report* 1: 33-38.
- GAO (ed.) 1992a. *Using Statistical Sampling*. May 1992. GAO/PEMD-10.1.6. Washington, D.C.
- GAO (ed.) 1992b. *Quantitative Data Analysis: An Introduction*. June 1992. GAO/PEMD-10.1.11. Washington, D.C.

- GAO (ed.) 1992c. *The Evaluation Synthesis*. March 1992. GAO/PEMD-10.1.2. Washington, D.C.
- GAO (ed.) 1991a. *Designing Evaluations*. May 1991. GAO/PEMD-10.1.4. Washington, D.C.
- GAO (ed.) 1991b. *Using Structured Interviewing Techniques*. July 1991. GAO/PEMD-10.1.5. Washington, D.C.
- GAO (ed.) 1990. *Case Study Evaluations*. November 1990. Transfer Paper 10.1.9. Washington, D.C.
- Generalsekretariat EVED (Hrsg.) 1993. *Die sozialen Kosten des Verkehrs in der Schweiz: eine Gesamtrechnung nach Verkehrsträgern*. Bern: Generalsekretariat EVED.
- Gentile, Pierre 1995. *Lernprozesse in Verwaltungen. Etude cas sur trois politiques sanitaires en Suisse*, Cahiers de l'IDHEAP no 142. Chavannes-près-Renens: IDHEAP.
- Gerheuser, Frohmüt W. 1991. *Die Wirkungen der beruflichen Vorsorge auf den Arbeitsmarkt*. Chur/Zürich: Rüegger.
- Gerheuser, Frohmüt W. und Schmid, André 1993. *Bedingt wirksam: Das Lohnungleichheitspostulat des Heimarbeitsgesetzes*. Bern: Nationales Forschungsprogramm "Wirksamkeit staatlicher Massnahmen". Serie Schlussberichte.
- Germann, Raimund E. 1996. *L'appareil étatique et le gouvernement, Administrations publiques en Suisse*, vol. 1. Berne: Haupt.
- Germann, Raimund E. 1991. Arenas of interaction: social science and public policy in Switzerland, in: Wagner, P. und Weiss, C. H. (eds.), *Social Sciences and Modern States. National Experiences and Theoretical Crossroads*. Cambridge: Cambridge University Press: 191-206.
- Germann, Raimund E. 1986. Die Beziehung zwischen Bund und Kantonen im Verwaltungsbereich, in: Germann, Raimund E., Weibel, Ernest, Graf, Hans Peter (Hrsg.), *Handbuch politischer Systeme der Schweiz*, Band 3: Föderalismus. Bern: Haupt: 343-370.
- Germann, Raimund E. et al. 1979. *Fédéralisme en action: l'aménagement du territoire*. St. Saphorin.
- Glass, Gèné V., Willson, Victor L. und Gottman, John M. 1975. *Design and Analysis of Time-Series Experiments*. Boulder: Colorado Associated University Press.
- Goldberg, S.B., Sander, F.E.A. und Rogers, N.H. 1992. *Dispute resolution. Negotiation, mediation and other processes*. Boston u.a.

- Grawe, Klaus, Donati, Ruth und Bernauer, Frederike 1994. *Psychotherapie im Wandel: von der Konfession zur Profession*. Göttingen: Hogrefe.
- Gray, Andrew, Jenkins, William und Segsworth, Robert (eds.) 1993. *Budgeting, Auditing and Evaluation*. New Brunswick/London: Transaction Publishers.
- Grolimund, Peter und Peter, Kathrin 1994. *Integrierte ökosystembezogene Umweltbeobachtung: Konzept für die Einführung eines Beobachtungssystems*. Zürich: vdf-Hochschulverlag AG an der ETHZ.
- Guba, Egon G. und Lincoln, Yvonna S. 1989. *Fourth Generation Evaluation*. Newbury Park: Sage.
- Guba, Egon G. und Lincoln, Yvonna S. 1987. The countenances of fourth-generation evaluation: Description, judgement and negotiation, in: Palumbo, D., *The politics of program evaluation*. Newbury Park u.a.: Sage: 202-234.
- Gujarati, Damodar N. 1988. *Basic econometrics*. New York: McGraw-Hill.
- Hablützel, Peter 1995. New Public Management als Modernisierungschance - Thesen zur Entbürokratisierungsdiskussion, in: Hablützel, Peter, Haldemann, Theo, Schedler, Kuno und Schwaar, Karl (Hrsg.), *Umbruch in Politik und Verwaltung. Ansichten und Erfahrungen zum New Public Management in der Schweiz*. Bern: Haupt: 499-508.
- Hakim, Catherine 1992. *Research Design. Strategies and Choices in the Design of Social Research*. London: Routledge.
- Hall, P. 1993. Policy paradigms, social learning and the State, *Comparative Politics*, 25: 275-296 (= Hall, P. 1994. Policy paradigms, social learning and the State, *Policy Studies Journal*, 22: 631-649).
- Hanser, Christian, Küster, Jürg und Cavelti, Guido 1994. *Hotellerieförderung durch Bund und Kantone, Evaluation der Auswirkungen in der Hotellerie*. Bern: Bundesamt für Industrie, Gewerbe und Arbeit, Beiträge zur Tourismuspolitik, Nr. 3.
- Hausheer, Othmar 1991. *Die Kehrriechtsackgebühr - ein wirkungsvolles umweltpolitisches Instrument? Eine empirische Untersuchung*. Chur: Rüegger.
- Heclo, Hugh 1978. Issue networks and the executive establishment, in: King A. (ed.), *The new American political System*. Washington: 87-124.

- Heclo, Hugh 1974. *Modern social politics in Britain and Sweden*. New Haven.
- Hedrick, Terry E., Bickman, Leonard und Rog, Debra J. 1993. *Applied research design, apractical guide*. Newbury Park: Sage.
- Heidenheimer, Arnold J., Heclo, Hugh und Teich-Adams, Caroline 1990. *Comparative Public Policy. The Politics of Social Choice in America, Europe and Japan*. Third edition. New York: St. Martin's Press.
- Heinelt, Hubert 1993. Policy und Politics. Überlegungen zum Verhältnis von Politikinhalt und Politikprozessen. In: Héritier, Adrienne (Hrsg.): Policy-Analyse. Kritik und Neuorientierung. *Politische Vierteljahresschrift Sonderheft 24*: 307-327.
- Hellstern, Gerd-Michael und Wollmann, Helmut (Hrsg.) 1984. *Evaluierung und Erfolgskontrolle in Kommunalpolitik und -Verwaltung*. Basel.
- Hellstern, Gerd-Michael und Wollmann, Helmut 1983. *Evaluierungsforschung: Ansätze und Methoden - dargestellt am Beispiel des Städtebaus*. Basel: Birkhäuser Verlag.
- Hengartner, Thomas und Merki, Christoph Maria 1993. Heilmittel, Genussmittel, Suchtmittel: Veränderungen in Konsum und Bewertung von Tabak in der Schweiz. *Schweizerische Zeitschrift für Geschichte*, 43: 375-418.
- Henry, G. 1990. *Practical Sampling*. Newbury Park: Sage.
- Henry, G. und Dickey, K. 1993. Implementing performance monitoring: a research and development approach, *Public Administration Review*, 53: 203-212.
- Héritier, Adrienne 1993a. Einleitung. Policy-Analyse, *Politische Vierteljahresschrift Sonderheft 24*: 9-36.
- Héritier, Adrienne 1993b. Policy-Netzwerkanalyse als Untersuchungsinstrument im europäischen Kontext: Folgerungen aus einer empirischen Studie regulativer Politik, *Politische Vierteljahresschrift Sonderheft 24*: 432-447.
- Hesse, Joachim Jens und Zöpel, Christoph (Hrsg.) 1990. *Der Staat der Zukunft*. Baden-Baden: Nomos.
- Hofferbert, Richard I. 1974. *The Study of Public Polic*. Indianapolis/New York: The Bobbs-Merrill Company Inc.
- Hoffmann-Riem, C. 1980. Die Sozialforschung einer interpretativen Soziologie, *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, Jg. 32: 339-372.

- Hoffmann-Riem, W. und Schmidt-Assmann, E. 1990. *Konfliktbewältigung durch Verhandlungen* (2 Bände). Baden-Baden.
- Hofmeister, Albert (Hrsg.) 1994. *Möglichkeiten und Grenzen der Programmsteuerung: Controlling und Evaluation*. Bern: Schriftenreihe der Schweizerischen Gesellschaft für Verwaltungswissenschaften, Bd. 21.
- Holm, Kurt (Hrsg.) 1991. *Die Befragung L Der Fragebogen - Die Stichprobe*. 4. Auflage. Tübingen: Francke.
- Holznagel, B. 1990. *Konfliktlösung durch Verhandlungen*. Baden-Baden.
- Honeyman, C. 1993. A consensus on mediators' qualifications, *Negotiation Journal* 9: 293-308.
- Hopf, Ch. 1991. Qualitative Interviews in der Sozialforschung. Ein Überblick, in: Flick, U. (Hrsg.), *Handbuch Qualitative Sozialforschung*. München: 321-370.
- Hoppe, R. 1993. Political judgment and the policy cycle: The case of ethnicity policy arguments in the Netherlands, in: Fischer, F. and Forester, F., *The argumentative turn in analysis and planning*. Durham/London: 77-100.
- Howlett, M. 1994. Policy paradigms and policy change: lessons from the old and new canadian policies towards aboriginal people, *Policy Studies Journal* 22: 631-649.
- Howlett, Michael und Ramesh, M. 1995. *Studying Public Policy: Policy Cycles and Policy Subsystems*. Toronto: Oxford University Press.
- Hunter, J.E. und Schmidt, F.L. 1989. *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*. Newbury Park: Sage.
- INFRAS 1990. *Evaluation Kehricksackgebühr*. Schriftenreihe Umwelt Nr. 136, Bern: BUWAL.
- Inhaber, H. 1992. Of LULUs, NIMBYs and NIMTOOs, *The Public Interest* 107: 52-64.
- Ishii-Kuntz, Masako 1994. *Ordinal log-linear models*. Thousand Oaks: Sage.
- Jänicke, Martin 1993. Vom Staatsversagen zur politischen Modernisierung?, Ein System aus Verlegenheitslösungen sucht seine Form, in: Bohret und Wewer (Hrsg.), *Zwischen Globalisierung und Regionalisierung*. Opladen: 63-77.
- Jänicke, Martin 1986. *Staatsversagen - Die Ohnmacht der Politik in der Industriegesellschaft*. München.
- Jänicke, Martin und Weidner, Helmut (eds.) 1995. *Successful environmental policy. A critical evaluation of 24 cases*. Berlin: Sigma.

- Jansen, Dieter und Schubert, Karl (Hrsg.) 1995. *Netzwerke und Politikproduktion. Konzepte, Methoden, Perspektiven*. Marburg.
- Jenkins-Smith, Hank C. und Sabatier, Paul 1993. The dynamics of policy-oriented learning, in: Sabatier, Paul und Jenkins-Smith, Hank C., *Policy change and learning*. Boulder u.a.: 41-58.
- Jobert, Bruno und Müller, Pierre 1987: *L'Etat en action. Politiques publiques et corporatisme*. Paris: Presses universitaires de France.
- Johnston, J. 1972. *Econometric methods*. 2nd ed. Tokyo: McGraw-Hill.
- Joint Committee on Standards for Educational Evaluation 1994. *The Program Evaluation Standards: How to Assess Evaluations of Educational Programs*. 2nd ed. Thousand Oaks: Sage.
- Joint Committee on Standards for Educational Evaluation 1981. *Standards for Evaluations of Educational Programs, Projects and Materials*. New York: McGraw-Hill.
- Jordan, G. und Schubert, K. 1992. A preliminary ordering of policy network labels, *European Journal of Political Research* 21: 7-27.
- Kaplitza, Gabriele 1991. Die Stichprobe, in: Holm, Kurt, *Die Befragung*, 1. 4. Auflage. Tübingen: Francke: 136-186.
- Katzenstein, P.J. 1985. *Small states in world markets*. London.
- Keller, Felix 1991. Autoritärer Populismus und soziale Lage, in: Bornschier, Volker, *Das Ende der sozialen Schichtung?*. Zürich: Seismo: 274-303.
- Kenis, P. und Schneider, V. 1991. Policy networks and policy analysis: Scrutinizing a new analytical toolbox, in: Marin, B. und Mayntz, R., *Policy networks. Empirical evidence and theoretical considerations*. Frankfurt a. M.: 25-59.
- Kim, Jae-On und Mueller, Charles W. 1979. *Introduction to Factor Analysis. What it is and how to do it*. Beverly Hills: Sage.
- Kindermann, Harald 1988. Symbolische Gesetzgebung, in: Gesetzgebungstheorie und Rechtspolitik, *Jahrbuch für Rechtssoziologie und Rechtstheorie*. Band 13. Opladen: 222-245.
- King, Gary, Keohane, Robert O. und Verba, Sidney 1994. *Designing Social Inquiry. Scientific Inference in Qualitative Research*. Princeton: Princeton University Press.
- King, J.A. 1995. Involving practitioners in evaluation studies: How viable is collaborative evaluation in schools?, in: Cousins, J.B. and Earls, L.M. (eds.), *Participatory evaluation in education: Studies in evaluation use and organizational learning*. London: Falmer Press: 86-102.

- Kingdon, J. 1984. *Agendas, alternatives, and public policies*. Michigan.
- Kissling-Näf, Ingrid 1996. *Lernprozesse und Umweltverträglichkeitsprüfung. Staatliche Steuerung über Verfahren und Netzwerkbildung im Entsorgungsbereich* (Diss. HSG).
- Kissling-Näf, Ingrid und Knoepfel, Peter 1994. Politikorientierte Lernprozesse. Konzeptuelle Überlegungen zu Lernprozessen, in: Bussmann W. (Hrsg.), *Lernen in Verwaltungen und Policy-Netzwerken*. Chur/Zürich: Rüegger: 99-129.
- Kissling-Näf, Ingrid und Knoepfel, Peter 1992. Politikverflechtung dank zentralstaatlichem Immobilismus? in: Abromeit, Heidrun und Pommerehne, Werner, W. (Hrsg.), *Staatstätigkeit in der Schweiz*. Bern: Haupt: 43-69.
- Kissling-Näf, Ingrid und Knoepfel, Peter en collaboration avec Bussy, Claire, Gentile, Pierre und Marek, Daniel 1993. *Politikorientierte Lernprozesse. Konzeptuelle Überlegungen zu Lernprozessen in Verwaltungen*, Cahiers de l'IDHEAP no 120. Chavannes-près-Renens: IDHEAP.
- Kissling-Näf, Ingrid, Marek, Daniel und Gentile, Pierre 1994. *Politikorientierte Lernprozesse. Analysekonzept zur empirischen Erhebung im Feld*, Cahiers de l'IDHEAP no 131. Chavannes-près-Renens: IDHEAP.
- Kissling-Näf, Ingrid und Wildi-Ballabio, Elena 1993. Kontrollinstrumente zur erfolgreichen Implementation von Politiken: Impulse aus der Umweltbeobachtung für ein integriertes Policy-Monitoring, *Schweizerisches Jahrbuch für Politische Wissenschaft*. Bern: 277-294.
- Klöti, Ulrich und Risi, Franz-Xaver 1991. *Politische Bildung Jugendlicher. Ergebnisse der Rekrutenbefragung 1988*. Aarau: Sauerländer.
- Klöti, Ulrich, Haldemann, Theo und Schenkel, Walter 1993. *Die Stadt im Bundesstaat - Alleingang oder Zusammenarbeit?* Chur/Zürich: Rüegger.
- Klöti, Ulrich und Schneider, Gerald 1989. *Die Informationsbeschaffung des Gesetzgebers*. Grüsch: Rüegger.
- Klok, Peter Jan 1995. A Classification of Instruments for Environmental Policy, in: Dente, Bruno (ed.), *Environmental Policy in Search of New Instruments*, Dordrecht/Boston/London: Kluwer Academic Publishers: 21-36.

- Knoepfel, Peter 1996a: Plädoyer für ein tatsächlich wirkungsorientiertes Public Management, *Swiss Political Science Review* 2 (1): 151-164.
- Knoepfel, Peter 1996b: *Total Quality Management et fédéralisme: Le point de vue de l'analyse des politiques publiques / Total Quality Management und Föderalismus: Betrachtungen aus der Sicht der Politikanalyse*, Cahiers de l'IDHEAP no 159. Chavannes-près-Renens: IDHEAP.
- Knoepfel, Peter 1995a. New Public Management: Vorprogrammierte Enttäuschungen oder politische Flurschäden - eine Kritik aus der Sicht der Politikanalyse, in: Hablützel, Peter et al. (Hrsg.), *Umbruch in Politik und Verwaltung - Ansichten und Erfahrungen zum New Public Management in der Schweiz*. Bern: Haupt: 453-470.
- Knoepfel, Peter (Hrsg., im Auftrag der SAGUF) 1995b. *Die Lösung von Umweltkonflikten durch Verhandlung / La solution de conflits environnementaux par la négociation*. Basel: Helbing & Lichtenhahn (Reihe Ökologie & Gesellschaft, Bd. 10).
- Knoepfel, Peter 1995c. Le New Public Management, est-ce la panacée?, *Swiss Political Science Review* 1 (4): 130-138.
- Knoepfel, Peter 1995d. New Institutional Arrangements for a New Generation of Environmental Policy Instruments: Intra- and Inter-policy-Cooperation, in: Dente, Bruno (ed.): *Environmental Policy in Search of New Instruments*, European Science Foundation, Dordrecht: Kluwer Academie Publishers: 197-233.
- Knoepfel, Peter 1995e. *Politiques publiques comparées* (Notes de cours, dossier 11, version octobre 1995). Chavannes-près-Renens: IDHEAP.
- Knoepfel, Peter 1992. Umweltbeobachtung und Umweltberichterstattung in der Schweiz, in: Weidner, H., Zieschank, R. und Knoepfel, P. (Hrsg.), *Umwelt-Information*. Berlin: 250-285.
- Knoepfel, Peter 1989. Wissenschaftliche Beratung im Politikvollzug: Gefahr für die Demokratie? Einige Thesen. *Schweizerisches Jahrbuch für Politische Wissenschaft*, 29: 81-102.
- Knoepfel, Peter 1979. *Öffentliches Recht und Vollzugsforschung, Beiträge der sog. Implementationsforschung zur Staats- und verwaltungsrechtlichen Parlamentsdiskussion*. Bern: Haupt (Reihe Staat und Politik, Bd. 24).

- Knoepfel, Peter 1977. *Demokratisierung der Raumplanung*. Berlin: Dunker & Humboldt.
- Knoepfel, Peter, Baitsch, Christof und Eberle, Armin 1995. *Überprüfung der Aufbauorganisation des Amtes für Umweltschutz des Kantons St. Gallen*, Chavannes-près-Renens: IDHEAP und St. Gallen: Baudepartement.
- Knoepfel, Peter, Imhof, Rita und Zimmermann, Willi 1995. *Luftreinhaltepolitik im Labor der Städte. Der Massnahmenplan - Wirkungen eines neuen Instruments der Bundespolitik im Verkehr*. Basel, Helbing & Lichtenhahn (Reihe Oekologie & Gesellschaft, Bd. 9).
- Knoepfel, Peter und Kissling-Näf, Ingrid 1993. Transformation öffentlicher Politiken durch Verräumlichung - Betrachtungen zum gewandelten Verhältnis zwischen Raum und Zeit, *Politische Vierteljahresschrift Sonderheft 24*: 267-288.
- Knoepfel, Peter, Kissling-Näf, Ingrid und Marek, Daniel 1997. *Lernen in öffentlichen Politiken*. Basel: Helbing & Lichtenhahn (Reihe Oekologie & Gesellschaft, Bd. 12) (im Erscheinen).
- Knoepfel, Peter und Weidner, Helmut 1982. Formulation and Implementation of Air Quality Control Programmes: Patterns of Interest Consideration, *Policy and Politics*, Bristol, vol. 10, Nr. 1: 85-110.
- Knoepfel, Peter und Zimmermann, Willi (unter Mitarbeit von Sailer, G., Matafora, E.) 1992. *Evaluation des BUWAL. Expertenbericht zur Evaluation der Luftreinhaltung, des ländlichen Gewässerschutzes und der UVP des Bundes*, Schlussbericht. Bern: EDMZ. (zitiert in: BBl. 1992III: 1502 und FF 1992III: 1424 gemäss Bericht der Geschäftsprüfungskommission des Nationalrates vom 18. Mai 1992).
- Knoepfel, Peter und Zimmermann, Willi 1987. *Oekologisierung von Landwirtschaft. Fünf Geschichten und eine Analyse*. Aarau, Frankfurt am Main, Salzburg: Sauerländer.
- Knoke, D. und Kuklinski, J.H. 1982. *Network analysis*. London.
- Knoke, David und Burke, Peter J. 1980. *Log-linear Models*. Beverly Hills: Sage.
- Koch, Dieter 1990. Preisstatistik, in: Bohley, Peter und Jans, Armin, *Einführung in die Wirtschafts- und Sozialstatistik der Schweiz*. Bern: Haupt: 281-322.
- Kohli, M. 1978. Offenes und geschlossenes Interview, *Soziale Welt*, 1.
- Kraus, W. 1991. Qualitative Evaluationsforschung, in: Flick, U. et al. (Hrsg.), *Handbuch Qualitative Sozialforschung*. München: 412-415.

- Kriesi, Hanspeter 1995. *Le système politique suisse*. Paris: Economica.
- Kriesi, Hanspeter 1994. *Les démocraties occidentales. Une approche comparée*. Paris: Economica.
- Kriesi, Hanspeter (Hrsg.) 1993. Vollzugsprobleme/Problèmes de la mise en œuvre de politiques publiques, *Schweizerisches Jahrbuch für Politische Wissenschaften* 33. Bern: Haupt.
- Kriesi, Hanspeter 1991. *The Political opportunity structure. New social movements, its impact on the mobilisation*. Berlin.
- Kriesi, Hanspeter 1980. *Entscheidungsstrukturen und Entscheidungsprozesse in der Schweizer Politik*. Frankfurt u.a.: Campus.
- Krippendorff, Klaus 1980. *Content analysis. An introduction to its methodology*. Beverly Hills: Sage.
- Kriz, Jürgen 1983. *Statistik in den Sozialwissenschaften. Einführung und kritische Diskussion*. 4. Auflage. Opladen: Westdeutscher Verlag.
- Kromrey, Helmut 1994. *Empirische Sozialforschung*. 6. Auflage. Opladen: Leske + Budrich.
- Kühnel, Steffen 1993. Lassen sich ordinale Daten mit linearen Strukturgleichungsmodellen analysieren?, *ZA-Information*, Bd. 33: 29-51.
- Kvale, Steinar 1991. Validierung: Von der Beobachtung zu Kommunikation und Handeln, in: Flick, Uwe et al. (Hrsg.), *Handbuch Qualitative Sozialforschung*, München: 427-431.
- Lamnek, Siegfried 1989. *Qualitative Sozialforschung. Band 2 Methoden und Techniken*. München.
- Lamnek, Siegfried 1988. *Qualitative Sozialforschung. Band 1 Methodologie*. München/Weinheim.
- Larson, Allan L. 1980: *Comparative Political Analysis*. Chicago: Nelson-Hall Inc. Publishers.
- Lasswell, Harold Dwight 1971. *Apreview of policy sciences*. New York: American Eisevier.
- Lasswell, Harold Dwight 1966. *The Analysis of Political Behaviour: An Empirical Approach*. London: Routledge.
- Lasswell, Harold Dwight 1951. The policy orientation. In: Lerner, Daniel und Lasswell, Harold Dwight (eds.) *The policy sciences*. Stanford: Stanford UP.
- Leviton, Laura C., und Hughes, E.F.X. 1981. Research on the utilization of evaluations: A review and synthesis. *Evaluation Review*, 5: 525-548.

- Lichtenecker, R. 1994. Umweltinformationssysteme, in: Bartel, R. und Hackl, F., *Einführung in die Umweltpolitik*, München: 63-79.
- Lijphart, Arend 1984. *Democracies. Patterns of Majoritarian and Consensus Government in Twenty-One Countries*. New Haven/London: Yale University Press.
- Lindblom, Charles E. und Cohen, David K. 1979. *Usable Knowledge, Social Science and Social Problem Solving*. New Haven and London: Yale University Press.
- Lindblom, Charles E. und Woodhouse, Edward J. 1993. *The Policy-Making Process*. 3rd ed. Englewood Cliffs: Prentice Hall.
- Linder, Stephen H. und Peters, Guy B. 1991. "The Logic of Public Policy Design: Linking Policy Actors and Plausible Instruments." *Knowledge and Policy*, Vol. 4, Nos. 1&2: 125-151.
- Linder, Stephen H. und Peters, Guy B. 1990. "Policy Formulation and the Challenge of Conscious Design." *Evaluation and Program Planning*, Vol. 13:303-311.
- Linder, Stephen H. und Peters, Guy B. 1989a. Implementation as a guide to policy formulation: a question of "when" rather than "whether", *International Review of Administrative Sciences*, Vol. 55: 631-652.
- Linder, Stephen H. und Peters, Guy B. 1989b. Instruments of Government: Perceptions and Contexts, *Journal of Public Policy* 9, 1: 35-58.
- Linder, Stephen H. und Peters, Guy B. 1988. The Analysis of Design or the Design of Analysis, *Policy Studies Review* 7, No. 4: 738-750.
- Linder, Wolf 1995. *Swiss Democracy*. New York: St. Martin's Press.
- Linder, Wolf 1989. Wissenschaftliche Beratung der Politik, *Schweizerisches Jahrbuch für Politische Wissenschaft*, 29. Bern: Haupt.
- Linder, Wolf 1988. *Politische Entscheidung und Gesetzesvollzug in der Schweiz*. Bern: Haupt/Linder, Wolf, *La décision politique en Suisse - Genèse et mise en oeuvre de la législation*. Lausanne: Réalités sociales.
- Linder, Wolf et al. 1990. *Verbrauchsabhängige Heizkostenabrechnung, Förderung erneuerbarer Energien*. Bern: Forschungszentrum für schweizerische Politik.
- Linder, Wolf, Hotz, Beat und Werder, Hans 1979. *Planung in der schweizerischen Demokratie*. Bern/Stuttgart: Haupt.
- Lisch, Ralf und Kriz, Jürgen 1978. *Grundlagen und Modelle der Inhaltsanalyse*. Reinbek: Rowohlt.

- Love, Arnold J. 1991. *Internal evaluation: Building Organisations from Within*. Newbury Park: Sage.
- Lowi, Theodore J. 1972. Four Systems of policy, politics and choice. *Public Administration Review* 32: 298-310.
- Luhmann, Niklas 1973. *Zweckbegriff und Systemrationalität*. Frankfurt a.M.: Suhrkamp
- Madauss, B. 1991. *Handbuch Projektmanagement*. Stuttgart.
- Mader, Luzius 1993. La législation expérimentale en Suisse, in: Morand, Charles-Albert (éd.), *Evaluation législative et lois expérimentales*. Aix-en-Provence: Presses universitaires d'Aix-Marseille: 221-242.
- Mader, Luzius 1988. Experimentelle Gesetzgebung, in: Grimm, Dieter und Maihofer, Werner (Hrsg.), *Gesetzgebungstheorie und Rechtspolitik*. Opladen: Westdeutscher Verlag: 211 - 221.
- Mader, Luzius 1985. *L'évaluation législative: Pour une analyse empirique des effets de la législation*. Lausanne: Collection Juridique Romande.
- Majone, G. 1993. Wann ist Policy-Deliberation wichtig?, *Politische Vierteljahresschrift Sonderheft* 24: 97-115.
- Malik, F. 1984. *Strategie des Managements komplexer Systeme*. Bern/Stuttgart.
- Mark, Melvin M. und Shotland, Lance R. 1986. Stakeholder-Based Evaluation and Value Judgments, *Evaluation Studies Review Annual*, Vol. 11: 131-152.
- Mastronardi, Philippe 1995. Staatsrecht und Verwaltungsorganisation, Reflexionen am Beispiel des New Public Managements, *Aktuelle juristische Praxis* 12/95: 1541-1553.
- Mastronardi, Philippe 1991. Experimentelle Rechtsetzung im Bund, *Zeitschrift für Schweizerisches Recht*, NF Bd. 110, 1. Halbband, Heft 4/5: 449 - 469.
- Mayntz, Renate 1993a. Policy-Netzwerke und die Logik von Verhandlungssystemen, *Politische Vierteljahresschrift Sonderheft* 24: 39-56.
- Mayntz, Renate 1993b. Modernization and the Logic of Interorganizational Networks, *Knowledge and Policy*, Vol. 6, No. 1: 3-16.
- Mayntz, Renate (Hrsg.) 1983. *Implementation politischer Programme II. Ansätze zur Theoriebildung*. Opladen: Westdeutscher Verlag.
- Mayntz, Renate (Hrsg.) 1980. *Implementation politischer Programme*. Empirische Forschungsberichte. Königstein/Ts.

- Mayring, Ph. 1993. *Einführung in die qualitative Sozialforschung*. Weinheim (2. Auflage).
- McCleary, Richard und Hay, Richard A. 1980. *Applied time series analysis for the social sciences*. Beverly Hills: Sage.
- McCutcheon, Allan L. 1987. *Latent Class Analysis*. Newbury Park: Sage.
- Menard, Scott 1991. *Longitudinal Research*. Newbury Park: Sage.
- Merten, Klaus 1983. *Inhaltsanalyse. Einführung in Theorie, Methode und Praxis*. Opladen: Westdeutscher Verlag.
- Merton, R.K. und Kendall, P. 1984. Das fokussierte Interview, in: Hopf, C. und Weingarten, E. (Hrsg.), *Qualitative Sozialforschung*. Stuttgart (2. Auflage): 171-204.
- Meylan, Jean-Pierre 1983. *Evaluation von Innovationen im Bereich der Primarschule*. Bern/Stuttgart: Haupt.
- Meylan, Jean-Pierre 1981. *Innovation und Evaluation von Schulversuchen*. Bern/Stuttgart: Haupt.
- Mény, Yves und Thöning, Jean-Claude 1989. *Politiques publiques*. Paris: Presses universitaires de France.
- Miller, Delbert C. 1991. *Handbook of Research Design and Social Measurement*, 5th ed. Newbury Park: Sage.
- Mohr, Lawrence B. 1995a. *Impact Analysis for Program Evaluation*, 2nd ed. Thousand Oaks: Sage.
- Mohr, Lawrence B. 1995b. *The Qualitative Method of Impact Analysis*. Paper prepared for delivery at the annual meeting of the American Evaluation Association, Vancouver, November 1-5., 1995. Ann Arbor: University of Michigan. (mimeo)
- Monnier, Eric 1987. *Evaluations de l'action des pouvoirs publics*. Paris: Economica.
- Moor, Pierre 1994. Les fondements généraux, *Droit administratif*, vol. I, 2ème édition, revue et mise à jour. Berne: Stämpfli.
- Moor, Pierre 1992. L'organisation des activités administratives. Les biens de l'Etat, *Droit administratif*, vol. II. Berne: Stämpfli.
- Morand, Charles-Albert (éd.) 1993. *Evaluation législative et lois expérimentales*. Aix-en-Provence: Presses universitaires d'Aix-Marseille.
- Morris, Lynn Lyons, Fitz-Gibbon, Carol Taylor und Freeman, Marie E. 1987. *How to communicate evaluation findings*. Newbury Park: Sage.
- Muller, Pierre 1990. *Les politiques publiques*. Paris: Presses universitaires de France.

- Nagel, Stuart S. (ed.) 1994. *Encyclopedia of Policy Studies*. 2nd edition. New York: Marcel Dekker.
- Nagel, St. und Mills, M. 1993. *Professional developments in policy studies*. Westport/London.
- Naschold, Frieder 1995. *Ergebnissteuerung, Wettbewerb, Qualitätspolitik: Entwicklungspfade des öffentlichen Sektors in Europa*. Berlin: ed. Sigma.
- Naschold, Frieder 1993. *Modernisierung des Staates. Zur Ordnungs- und Innovationspolitik des öffentlichen Sektors*. Berlin.
- Niessen, M. 1977. *Gruppendiskussion. Interpretative Methodologie, Methodenbegründung, Anwendung*. München.
- Nullmeier, F. 1994. *Interpretative Ansätze in der Politikwissenschaft*. Tagungspapier der deutschen Vereinigung für politische Wissenschaft. Konstanz.
- Nullmeier, F. 1993. Wissen und Policy-Forschung. Wissenspolitologie und rhetorisch-dialektisches Handlungsmodell, *Politische Vierteljahresschrift Sonderheft* 24: 175-196.
- Nüssli, Kurt 1985. *Föderalismus in der Schweiz: Konzepte, Indikatoren, Daten*. Grüsch: Rüegger.
- O'Toole, L.J. 1993. Multiorganizational policy implementation: some limitations and possibilities for rational-choice contributions, in: Scharpf, F.W., *Garnes in hierarchies and networks*, Frankfurt a. M.: 27-64.
- Oevermann, Ulrich et al. 1979. Die Methodologie einer "objektiven Hermeneutik" und ihre allgemeine forschungslogische Bedeutung in den Sozialwissenschaften, in: Soeffner, Hans-Georg (Hrsg.), *Interpretative Verfahren in den Sozial- und Textwissenschaften*. Stuttgart: 352-434.
- Offe, Claus 1987. Die Staatstheorie auf der Suche nach ihrem Gegenstand. Beobachtungen zur aktuellen Diskussion, *Jahrbuch zur Verwaltungswissenschaft*. Baden-Baden: 309-320.
- Offe, Claus und Wiesenhal, Helmut 1980. Two Logics of Collective Action - Theoretical Notes on Social Class and Organisational Form, in: Zeitlin, Maurice (Hrsg.), *Political Power and Social Theory*. Greenwich (Ct.): JAI Press: 65-115.
- Organe parlementaire de contrôle de l'administration 1994. *Evaluation de la législation extra-parlementaire en matière de prévoyance professionnelle*. Berne: Centrale de documentation de l'Assemblée fédérale.

- Osborne, D. und Gaebler, T. 1993. *Reinventing government. How the entrepreneurial spirit is transforming the public sector*. New York.
- Ostrom, Charles W. 1990. *Time series analysis. Regression techniques*. 2nd ed. Newbury Park: Sage.
- Palumbo, Dennis J. und Calista, Donald J. 1990. *Implementation and the policy process. Opening up the black-box*. New York: Greenwood Press.
- Palumbo, Dennis J. und Oliverio, Annamarie 1989. Implementation theory and the theory-driven approach to validity, *Evaluation and Program Planning*, Vol. 12: 337-344.
- Papadopoulos, Yannis 1995. *Complexité sociale et politiques publiques*. Paris: Montchrestien.
- Parsons, Wayne 1995. *Public Policy. An Introduction to the Theory and Practice of Policy Analysis*. Aldershot: Edward Elgar.
- Passow, A. Harry 1987. Reporting the results of evaluation studies, *International Journal of Educational Research*, Vol. 11, No. 1: 115-123.
- Patton, Michael Quinn 1994. Developmental Evaluation. *Evaluation Practice*, Vol. 15, No.3: 311-319.
- Patton, Michael Quinn 1990. *Qualitative Evaluation and Research Methods*, 2nd ed. Newbury Park: Sage.
- Patton, Michael Quinn 1984. Data Collection: Options, strategies, and cautions, in: Rutman, Leonard, *Evaluation research methods: A basic guide*, 2nd ed. Newbury Park: Sage: 39-63.
- Patton, Michael Quinn 1981. *Creative evaluation*. Beverly Hills: Sage.
- Patton, Michael Quinn 1978. *Utilization-Focused Evaluation*. Beverly Hills/London: Sage.
- Pfingsten, K. und Fietkau, H.-J. 1995. *Das Neusser Mediationsverfahren aus Sicht der Beteiligten. Ergebnisdarstellung der schriftlichen Befragung*. Schriften zu Mediationsverfahren im Umweltschutz Nr. 9, Wissenschaftszentrum Berlin für Sozialforschung FS II 95-302. Berlin.
- Piazza, Cecilia 1997. *Le rôle de l'Office régional de placement dans la mise en oeuvre des mesures actives prévues par l'assurance-chômage*, Travaux de cours et mémoires de l'IDHEAP no 1, Chavannes-près-Renens.
- Pictet, Jacques 1996. *Dépasser l'évaluation environnementale: procédure d'étude et insertion dans la décision globale*. Lausanne: Presses polytechniques et universitaires romandes.

- Pollitt, Christopher 1995. Justification by Works or by Faith? Evaluating the New Public Management. *Evaluation*. 1/2: 133-154.
- Pollock, F. 1955. *Gruppenexperiment*. Frankfurt am Main.
- Pressman, Jeffrey und Wildavsky, Aaron 1984. *Implementation*. 3rd ed., Berkeley: University of California Press.
- Pressman, Jeffrey und Wildavsky, Aaron 1973. *Implementation*, Berkeley: University of California Press.
- Przeworski, Adam und Theune, Henry 1970: *The Logic of Comparative Social Enquiry*, New York: Wiley-Interscience.
- Pruitt, D. und Kressel, K. 1985. The mediation of social conflict: An introduction, *Journal of Social Issues* 41, 2: 1-10.
- Reichard, Ch. 1988. Der Dritte Sektor. Entstehung, Funktion und Problematik von Non-Profit-Organisationen aus verwaltungswissenschaftlicher Sicht, *Die Öffentliche Verwaltung*, Heft 9: 363-370.
- Reichardt, Charles S. 1979. The statistical analysis of data from nonequivalent group designs, in: Thomas, Cook, D. und Campbell, Donald T., *Quasi-experimentation. Design and analysis issues for fieldsettings*. Boston: Houghton Mifflin: 147-205.
- Reinermann, Heinrich 1975. *Programmbudgets in Regierung und Verwaltung*. Baden-Baden: Nomos.
- Renn, Ortwin 1995. Der kooperative Diskurs: Theorie und praktische Erfahrungen mit einem Deponieprojekt im Kanton Aargau, in: Knoepfel, Peter (Hrsg., im Auftrag der SAGUF), *Die Lösung von Umweltkonflikten durch Verhandlung / La solution de conflits environnementaux par la négociation*, Basel: Helbing & Lichtenhahn (Reihe Oekologie & Gesellschaft, Bd. 10): 191-243.
- RESOLVE (Hrsg.) 1994. *The cutting edge: environmental dispute resolution for the nineties*. Symposium summary. Washington D.C.
- Rey, Michel 1995. La gestion du processus de consultation et de décision: un nouvel enjeu en aménagement du territoire, in: Knoepfel, Peter (Hrsg. im Auftrag der SAGUF), *Die Lösung von Umweltkonflikten durch Verhandlung / La solution de conflits environnementaux par la négociation*, Basel: Helbing & Lichtenhahn (Reihe Oekologie & Gesellschaft, Bd. 10): 245-260.
- Rey, Michel et al. 1993. *La politique valaisanne de promotion économique: une évaluation privilégiant le point de vue des acteurs*. Berne: Programme national de recherche 27.

- Rieder, Stefan und Varone, Frédéric 1992. *Schlussbericht des Vorprojektes Vollzugsevaluation des Aktionsprogrammes "Energie 2000"*. Bern: Forschungszentrum für schweizerische Politik der Universität Bern.
- Rist, Ray C. 1994. The preconditions for learning: Lessons from the public sector, in: Leeuw, F. L., Rist, R.L. und Sonnichsen, R.C., *Can governments learn? Comparative perspectives on evaluation and organizational learning*. New Brunswick: Transaction: 328-358.
- Rist, Ray C. 1990a. *Program Evaluation and the Management of Government*. New Brunswick: Transaction Publishers.
- Rist, Ray C. 1990b. On the Application of Program Evaluation Designs - Sorting Out Their Use and Abuse. *Knowledge in Society*, 1989-90, Vol. 2, Number 4: 76-96.
- Robinson, W.S. 1950. Ecological correlations and the behavior of individuals, *American Sociological Review*, Vol. 15, No. 3: 351-357.
- Rosé, Richard 1993. *Lesson-drawing in public policy. A guide to learning across time and space*. Chatham.
- Rosé, Richard 1984. *Understanding Big Government: The Programme Approach*. London: Sage.
- Rosenthal, Robert 1991. *Meta-Analytic Procedures for Social Research*. Rev. ed. Newbury Park: Sage.
- Rossi, P.H. und Freeman, H.E. 1993. *Evaluation: a Systematic Approach. Fifth Edition*. Newbury Park: Sage.
- Rossi, P.H., Freeman, H.E. und Hofmann, G. 1988. *Programm-Evaluation. Einführung in die Methoden angewandter Sozialforschung*. Stuttgart.
- Roth, Erwin 1993. *Sozialwissenschaftliche Methoden. Lehr- und Handbücher der Sozialwissenschaften*. München/Wien: Oldenbourg.
- Ruch, Luzian 1993. BAGE: Ein variantenorientierter Ansatz, in: Frey, F. et al. (Hrsg.), *Die kompetente Organisation*. Stuttgart: Schäffel Poeschel: 255-280.
- Sabatier, Paul A. 1993. Policy change over a decade or more, in: Sabatier, Paul A. und Jenkins-Smith, Hank C., *Policy change and learning*, Boulder u.a.: 13-39.
- Sabatier, Paul A. und Jenkins-Smith, Hank C. 1993. The advocacy coalition framework: Assessment, revisions, and implications for scholars and practitioners, in: Sabatier, Paul A. und Jenkins-Smith, Hank C., *Policy change and learning*, Boulder u.a.: 211-235.

- Salisbury, Robert H. 1968. The Analysis of Public Policy: A Search for Theory and Roles, in: Ranney, Austin (ed.), *Political Science and Public Policy*. Chicago: Markham Publishing: 154-172.
- Saris, Willem und Stronkhorst, Henk 1984. *Causal modelling in nonexperimental research*. Amsterdam: Sociometric Research Foundation.
- Sayrs, Lois W. 1989. *Pooled time series analysis*. Newbury Park: Sage.
- Schaetti, Stephan 1990. *Berufliche Vorsorge und Arbeitsmarkt*. Basel: WWZ-Studien Nr. 18.
- Scharpf, F.W. (Hrsg.) 1993. *Garnes in hierarchies and networks*. Frankfurt a. M.
- Scharpf, Fritz W. 1983. Interessenlage der Adressaten und Spielräume der Implementation bei Anreizprogrammen, in: Mayntz, Renate (Hrsg.) 1983, *Implementation politischer Programme, Ansätze zur Theoriebildung*, Opladen: Westdeutscher Verlag: 89-116.
- Scharpf, Fritz W., Reissert, Bernd und Schnabel, Fritz 1976. *Politikverflechtung: Theorie und Empirie des kooperativen Föderalismus in der Bundesrepublik*. Kronberg.
- Schedler, Kuno 1995. *Ansätze einer wirkungsorientierten Verwaltungsführung*. Bern: Haupt.
- Schmidt, Kurt und Eberhard, Wille 1970. *Die mehrjährige Finanzplanung. Wunsch und Wirklichkeit*. Tübingen: Mohr/Siebeck.
- Schmidt, Manfred G. 1996. Parteien und Policy-Output, in: Gabriel, Otto, Niedermeyer, Otwin und Stöss, Robert (Hrsg.), *Parteidemokratie in Deutschland*. Bonn: Bundeszentrale für politische Bildung.
- Schmidt, Manfred G. (Hrsg.) 1988. Staatstätigkeit: International und historisch vergleichende Analysen, *Politische Vierteljahresschrift Sonderheft* 19, Opladen: Westdeutscher Verlag.
- Schmidt, Manfred G. 1982. *Wohlfahrtsstaatliche Politik unter bürgerlichen und sozialdemokratischen Regierungen. Ein internationaler Vergleich*. Frankfurt am Main: Campus.
- Schmitz, Bernhard 1989. *Einführung in die Zeitreihenanalyse*. Bern: Huber.
- Schneider, Stefan, Häberling, Do und Keiser, Stephan 1995. *Erfolgskontrolle Blaue Zone Bern Kirchenfeld*. Zürich: Planungsbüro Jud.
- Schneider, Stefan et al. 1992. *Parkierungsbeschränkungen mit Blauer Zone und Anwohnerparkkarte: Empfehlungen für die Einführung*. Zürich: Planungsbüro Jud.

- Schneider, Stefan et al. 1990. *Pilotversuch Zürich Hottingen*. Zürich: Planungsbüro Jud.
- Schneider, Volker 1994. *Zur Handlungs- und Problemlösungsfähigkeit von Politiknetzwerken*. Vortrag am Schweizerischen Politologentag in Balsthal
- Schneider, Anne L. und Ingram, Heien 1990. Policy Design: Elements, Premises, and Strategies. In: Nagel, Stuart S. (ed.): *Policy Theory and Policy Evaluation*. New York: Greenwood Press: 77-101.
- Schnell, Rainer, Hill, Paul B. und Esser, Elke 1993. *Methoden der empirischen Sozialforschung*. 4. Auflage. München: Oldenbourg.
- Schön, D. und Rein, M. 1994. *Frame Reflection. Toward the resolution of intractable policy controversies*. New York.
- Schulz, Hans-Rudolf, Muggli, Christoph und Hübschle, Jörg 1993. *Wohneigentumsförderung durch den Bund, Die Wirksamkeit des Wohnbau- und Eigentumsförderungsgesetzes (WEG)*. Bern: Schriftenreihe Wohnungswesen, Band 55.
- Schütze, Fritz 1978. Was ist "kommunikative Sozialforschung"?, in: Gärtner, A. und Hering, S. (Hrsg.), *Modellversuch "Soziale Studiengänge" an der GH Kassel*, Materialien 12: Regionale Sozialforschung, Kassel: 117-131.
- Schwarz, Heinrich und Brozovic, Adrian 1992. *Erfolgskontrolle Sackgebühr*. Im Auftrag des Gewässerschutzamtes des Kantons Bern. Solothurn: BSB + Partner, (mimeo).
- Schweizerischer Nationalfonds zur Förderung der wissenschaftlichen Forschung, Expertenkommission des Schwerpunktprogrammes "Zukunft Schweiz" 1996. *Ausführungsplan*. Bern
- Schweizerischer Wissenschaftsrat 1993. *Revitalising swiss social sciences. Evaluation report*. Bern: Schweizerischer Wissenschaftsrat/FOP 13.
- Sciarini, Pascal und von Holzen, Madeleine 1994. *GATT - Europe: la Suisse face à ses paysans*. Genf und Lausanne: Journal de Genève et Gazette de Lausanne.
- Scott, J. 1991. *Social network analysis. A handbook*. London u.a.
- Scriven, Michael 1991. *Evaluation Thesaurus*. 4th ed. Newbury Park: Sage.
- Scriven, Michael 1980. *The Logic of Evaluation*. Inverness: Edgepress.

- Scriven, Michael 1976. Maximizing the Power of Causal Investigations: The Modus Operandi Method, *Evaluation Studies Review Annual*, Vol. 1: 101-118.
- Scriven, Michael 1973. Goal-free Evaluation, in: House Ernest, R., *School Evaluation: The Politics and Process*, Berkeley: McCutchan: 319-328.
- Scriven, Michael 1972. The Methodology of Evaluation, in: Weiss, Carol H. (Hrsg.), *Evaluating social Programs*, Boston: Allyn & Bacon: 123-136.
- Senge, Peter 1990. *The Fifth Discipline. The Art and Practice of Learning Organisations*. New York.
- Shadish, William R. 1994. Need-based evaluation theory: What do you need to know to do good evaluation? *Evaluation Practice*, Vol. 15, No. 3: 347-358.
- Shadish, William R., Cook, Thomas D. und Leviton, Laura 1991. *The Foundations of Program Evaluation, Theories of Practice*. Newbury Park: Sage.
- SIDOS (Hrsg.) 1994. *Sozialwissenschaftliche Forschung*. Zürich: Seismo.
- Simeon, R. 1993. *What is to be explained: Some observations of measuring government activity*. Québec (unv. MS).
- Singer, Otto 1993. Policy Communities and Diskurs-Koalitionen: Experten und Expertise in der Wirtschaftspolitik, in: *Politische Vierteljahresschrift Sonderheft 24*: 149-174.
- SKUB (Schweizerische Kommission für Umweltbeobachtung) 1993. *Bericht der Schweizerischen Kommission für Umweltbeobachtung zur Einrichtung einer integrierten ökosystembezogenen Umweltbeobachtung*. Bern: Schweizerische Akademie für Naturwissenschaften (SANW).
- Smith, C. 1991. Networks of influence: the social sciences in the United Kingdom since the war, in: Wagner, P. und Weiss, C.H. (eds.), *Social Sciences and Modern States. National Experiences and Theoretical Crossroads*. Cambridge: Cambridge University Press: 131-147.
- Soeffner, Hans-Georg 1989. Alltagsverstand und Wissenschaft, in: Soeffner, Hans-Georg, *Alltag der Auslegung: Auslegung des Alltags*. Frankfurt a.M.: Suhrkamp: 10-50.
- Spöhring, W. 1989. *Qualitative Sozialforschung*. Stuttgart.
- Stadt St. Gallen 1992. *Umweltbericht 1992*. St. Gallen

- Stadt Zürich 1990. *Umweltbericht 1989/90*. Zürich: Umweltschutzfachstelle der Stadt Zürich.
- Stake, Robert E. 1995. *The Art of Case Study Research*. Thousand Oaks: Sage.
- Stake, Robert E. 1994. Case Studies, in: Denzin, Norman K. und Lincoln, Yvonna S. *Handbook of Qualitative Research*. Thousand Oaks: Sage: 236-247.
- Stake, Robert E. 1980. Program evaluation, particularly responsive evaluation, in: Dockrell, W.B. und Hamilton, D. (Hrsg.), *Rethinking educational Research*. London: Hodder & Stoughton: 72-87.
- Stalder, Peter, Frick, Andreas, Gaillard, Serge und Salzgeber, Renate 1994. *Auswirkungen der Ausländerpolitik auf den Strukturwandel in der Schweiz*. Bern: Nationales Forschungsprogramm "Wirksamkeit staatlicher Massnahmen" (NFP 27), Serie Schlussberichte.
- Statistisches Jahrbuch der Schweiz 1992. Zürich 1993.
- Stewart, Jon 1991. *Econometrics*. New York: Philip Allan.
- Stimson, James A. 1985. Regression in space and time: A statistical essay, *American Journal of Political Science*, Vol. 29: 914-947.
- Stufflebeam, Daniel, L. und Shinkfield, Anthony 1985. *Systematic Evaluation*. Boston et al.: Kluwer-Nijhoff Publishing.
- Subirats, J. 1995. Policy instruments, public deliberation and evaluation processes, in: Dente Bruno (ed.), *Environmental policy in search of new instruments*, Dordrecht u.a.: 143-157.
- Susskind, L. und Cruikshank, J. 1987. *Breaking the impasse. Consensual approaches to resolving disputes*. New York.
- Susskind, L. und Ozawa, C. 1985. Mediating public disputes: Obstacles and possibilities, *Journal of Social Issues* 41, 2: 145-159.
- Terribilini, Serge *De la distributivité des politiques régulatrices: Discrimination socio-spatiales en matière de modération du trafic. Constat et causes*, Cahiers de l'IDHEAP no 151. Chavannes-près-Renens: IDHEAP.
- Tsuru, Shigeto und Weidner, Helmut (eds.) 1989. *Environmental policy in Japan*. Berlin.
- Universität Zürich: Gerichtlich-Medizinisches Institut 1977. *Unfalluntersuchung Rücksitzpassagiere und Kinder*. Im Auftrag der Eidg. Polizeiabteilung, EJPD. Von Hansjörg Sprenger und Felix Walz. Bern: Eidg. Polizeiabteilung.
- Urban, Dieter 1993. *Logit-Analyse*. Stuttgart: Fischer.

- Urio, Paolo, Bari, Jean-Paul und Glauser, Christoph 1996. *L'évaluation comme instrument d'apprentissage et de changement organisationnels: L'exemple des programmes cantonaux de promotion de la santé*. Berne: Programme national de recherche 27.
- Van den Daele, W. 1993. Sozialverträglichkeit und Umweltverträglichkeit. Inhaltliche Mindeststandards und Verfahren bei der Beurteilung neuer Technik, *Politische Vierteljahresschrift Sonderheft 24*: 219-248.
- Van der Knaap, Peter 1995. Policy Evaluation and Learning: Feedback, Enlightenment or Argumentation. *Evaluation*, Vol. 1(2): 189-216.
- Van Waarden F. 1992. Dimensions and types of policy networks, *European Journal of Political Research*, Vol. 21: 29-52.
- Varela, F.J. und Thompson, E. 1992. *Der Mittlere Weg der Erkenntnis, Die Bezeichnung von Ich und Welt in der Kognitionswissenschaft - der Brückenschlag zwischen wissenschaftlicher Theorie und menschlicher Erfahrung*. Bern.
- Viveret, Patrick 1989. *L'évaluation des politiques et des actions publiques. Rapport au Premier ministre*. Paris: La Documentation Française.
- Volmerg, Ute 1983. Validität im interpretativen Paradigma. Dargestellt an der Konstruktion qualitativer Erhebungsverfahren, in: Zedler, P., Moser, H. (Hrsg.), *Aspekte qualitativer Sozialforschung. Studien zu Aktionsforschung, empirischer Hermeneutik und reflexiver Sozialtechnologie*. Opladen: 124-143.
- Volvo Car Corporation und Swedish Road Safety Office 1980. *Injury-reducing effect of seat belts on rear seat passengers*. Göteborg: Volvo AB.
- Von Prittwitz, Volker 1993. Reflexive Modernisierung und öffentliches Handeln, in: Von Prittwitz, Volker (Hrsg.), *Umweltpolitik als Modernisierungsprozess*, Opladen: 36-49.
- Wagner, Peter, Weiss, Carol Hirschon, Wittrock, Björn und Wollmann, Helmut (Hrsg.) 1991. *Social sciences and modern states. National experiences and theoretical crossroads*. Cambridge: Cambridge University Press u.a.
- Wagner, Peter und Wollmann, Helmut 1986. Fluctuations in the development of evaluation research: do "regime shifts" matter? *International Social Science Journal*: 205-218
- Wälti, Sonja 1997. *Federalism in Switzerland: Policy Making and Policy Implementation*. Chavannes-près-Renens: IDHEAP (nicht veröf-

- fentlichter Beitrag im Auftrag der Giovanni Agnelli Foundation, Italien).
- Beitrag im Auftrag der Giovanni Agnelli Foundation, Italien)
- Wasserman, St. und Faust, K. 1994. *Social network analysis. Methods and application*. Cambridge.
- Waterman, R. und Wood, D. 1993. Policy monitoring and policy analysis, *Journal of Policy Analysis and Management*, 12: 685-699.
- Watzlawick, P. und Krieg, P. (Hrsg.) 1991. *Das Auge des Betrachters. Beiträge zum Konstruktivismus*. München.
- Weaver, R. Kent und Rockman, Bert A. (eds.) 1993. *Do Institutions Matter? Government Capabilities in the United States and Abroad*. Washington D.C.: The Brookings Institution.
- Weber, Robert P. 1990. *Basic content analysis*. 2nd ed. Newbury Park: Sage.
- Wei, William W.S. 1990. *Time series analysis*. Redwood: Addison-Wesley.
- Weidner, Helmut 1993. Der verhandelnde Staat. Minderung von Vollzugskonflikten durch Mediationsverfahren?, in: Kriesi, H. (Hrsg.), *Vollzugsprobleme/Problèmes de la mise en œuvre de politiques publiques*, *Schweizerisches Jahrbuch für Politische Wissenschaften* 33. Bern u.a.: 225-244.
- Weidner, Helmut und Fietkau, H.-J. 1995. *Environmental Mediation: The Mediation procedure on the waste management concept in the district of Neuss, North Rhine-Westfalia*. Initial results of the companion social science research, Schriften zu Mediationsverfahren im Umweltschutz Nr. 10, Wissenschaftszentrum Berlin für Sozialforschung FS II. Berlin: 95-303.
- Weidner, Helmut und Knoepfel, Peter (Hrsg.) 1984. *Luftreinhaltepolitik in städtischen Ballungsräumen, ein internationaler Vergleich*. Frankfurt a.M./New York: Campus, Bd. 1.
- Weidner, Helmut und Knoepfel, Peter 1983. Innovation durch international vergleichende Politikanalyse dargestellt am Beispiel der Luftreinhaltepolitik, in: Mayntz, Renate (Hrsg.), *Implementation politischer Programme II Ansätze zur Theoriebildung*. Opladen: Westdeutscher Verlag: 221-255.
- Weiss, Carol Hirschon 1993. Mega-lessons from U.S. evaluation experience, *Politika*, Vol. 25, No. 1: 64-76.

- Weiss, Carol Hirschon 1987. Where politics and evaluation research meet, in: Palumbo, Dennis James, *The politics of program evaluation*. Newbury Park: Sage: 47-70.
- Weiss, Carol Hirschon 1984. Toward the Future of Stakeholder Approaches in Evaluation, *Evaluation Studies Review Annual*, Vol. 9: 255-268.
- Weiss, Carol Hirschon 1981. Measuring the use of evaluation, in: Ciarlo, James A., *Utilizing evaluation*. Beverly Hills: Sage: 17-33.
- Weiss, Carol Hirschon 1978. Improving the linkage between social research and public policy, *Knowledge and Policy*: 23-81.
- Weiss, Carol Hirschon 1975. Interviewing in evaluation research, in: Struening, Eimer L. und Guttentag, Marcia, *Handbook of evaluation research*, Volume I, Beverly Hills: Sage: 355-395.
- Weiss, Carol Hirschon 1974. *Evaluierungsforschung. Methoden zur Einschätzung von sozialen Reformprogramme*. Opladen.
- Weiss, Carol Hirschon und Bucuvalas, Michael J. 1980. *Social Science Research and Decision-Making*. New York: Columbia University Press.
- Weiss, Carol Hirschon und Wittrock, Björn 1991. Summing up: social sciences and modern states, in: Wagner, P. und Weiss, C. H. (eds.), *Social Sciences and Modern States. National Experiences and Theoretical Crossroads*. Cambridge: Cambridge University Press: 356-368.
- Wheeler, M. 1994. Negotiating NIMBYs: Learning from the failure of the Massachusetts Siting Law, *Yale Journal on Regulation* 11: 241-290.
- Wholey, Joseph S. 1983. *Evaluation and Effective Public Management*. Boston: Little Brown.
- Wholey, Joseph S. 1979. *Evaluation: Promise and Performance*. Washington DC: The Urban Institute.
- Widmer, Thomas 1996. *Meta-Evaluation: Kriterien zur Bewertung von Evaluationen*. Bern: Haupt.
- Widmer, Thomas 1991. *Evaluation von Massnahmen zur Luftreinhaltepolitik in der Schweiz*. Zürich: Rüegger.
- Widmer, Thomas und Buri, Christof 1992. Brüssel oder Bern: Schlägt das Herz der "Romands" eher für Europa?, *Schweizerisches Jahrbuch für Politische Wissenschaft* 32: 363-387.

- Widmer, Thomas, Rothmayr, Christine und Serdült, Uwe 1996. *Kurz und gut? Qualität und Effizienz von Kurzevaluationen*. Zürcher Beiträge zur Politikwissenschaft, Band 19, Chur/Zürich: Rüegger.
- Wille, Eberhard 1977. Mittel- und langfristige Finanzplanung, in: Fritz Neumark et al. (Hrsg.), *Handbuch der Finanzwissenschaft*, Bd. I, 3. Aufl. Tübingen: 427-474.
- Willke, Helmut 1992. *Ironie des Staates*. Frankfurt: Suhrkamp.
- Willke, Helmut 1983. *Entzauberung des Staates. Überlegungen zu einer sozialen Steuerungstheorie*. Königstein/Taunus: Athenäum.
- Wilson, T.P. 1976. Theorien der Interaktion und Modelle soziologischer Erklärung, in: Arbeitsgruppe Bielefelder Soziologen (Hrsg.), *Alltagswissen, Interaktion und gesellschaftliche Wirklichkeit*, Band 1: Symbolischer Interaktionismus und Ethnomethodologie, 3. Auflage. Reinbek b/Hamburg: 54-79.
- Windhoff-Héritier, Adrienne 1989. Wirksamkeitsbedingungen politischer Instrumente, *Jahrbuch zur Staats- und Verwaltungswissenschaft*, Band 3: 89-118.
- Windhoff-Héritier, Adrienne 1987. *Policy-Analyse: Eine Einführung*. Frankfurt/New York: Campus.
- Wingens, Matthias 1988. *Soziologisches Wissen und politische Praxis. Neuere theoretische Entwicklungen der Verwendungsforschung*. Frankfurt/New York: Campus.
- Wittrock, Björn 1991. Social knowledge and public policy: eight models of interaction, in: Wagner, P. und Weiss, C.H. (eds.), *Social Sciences and Modern States. National Experiences and Theoretical Crossroads*. Cambridge: Cambridge University Press: 333-353.
- Wittrock, Björn und Wagner, Peter 1992. *History and context in comparative public policy*. Pittsburgh/London: 227-246.
- Witzel, A. 1985. Das problemzentrierte Interview, in: Jüttemann, G. (Hrsg.), *Qualitative Forschung in der Psychologie. Grundfragen, Verfahrensweisen, Anwendungsfelder*. Weinheim: 227 ff.
- Wonnacott, Thomas H. und Wonnacott, Ronald J. 1990. *Introductory statistics*. 5th ed. New York: Wiley.
- Wottawa, Heinrich und Thierau, Heike 1990. *Lehrbuch Evaluation*. Bern: Huber.
- Yamaguchi, Kazuo 1991. *Event history analysis*. Newbury Park: Sage.
- Zieschank, R. 1992. Einleitung, in: Weidner, H., Zieschank, R. und Knoepfel, P., *Umwelt-Information*. Berlin: 7-31.

- Zillessen, H. und Barbian, T. 1992. Neue Formen der Konfliktregelung in der Umweltpolitik, *Aus Politik und Zeitgeschichte*, Bände 39/40: 14-23.
- Zillessen, H., Dienel, P. und Strubelt, W. (Hrsg.) 1993. *Die Modernisierung der Demokratie. Internationale Ansätze*. Opladen.
- Zimmermann, Willi 1990. Klärschlammanalysen, Boden- und Düngeberatung: Ökologische landwirtschaftliche Forschung auf dem Weg ins Hinterland der landwirtschaftlichen Praxis, in: Knoepfel, Peter (Hrsg.), *Landwirtschaftliche ökologische Beratung - ein Modell für die allgemeine Umweltberatung?*. Basel: Helbing & Lichtenhahn (Reihe Oekologie & Gesellschaft, Bd. 4): 106-116.
- Zimmermann, Willi, Kolly, Doris und Mahon, Pascal 1989. *Die wissenschaftliche Evaluation von Bundespolitik*. Bern: EDMZ.

STICHWORTVERZEICHNIS

Adressaten s. *Politikadressaten*

AGEVAL s. *Arbeitsgruppe "Gesetzesevaluation"*

Arbeitsgruppe "Gesetzesevaluation" 26, 69, 105, 148, 212, 265

ARIMA 251

Aufklärung 258, 278

Aufklärungsfunktion 263

von Evaluationen 280

Befragung 207, 227-228; 236-237

Begleitung 28, 29, 120, 266

Behördenarrangement 63, 72, 75, 85, 88-89, 93, 95-98, 107, 126, 136-138

Beobachtung 154-155, 221, 237-238, 273, 278

Beratung 13, 16-17, 121-122, 126, 156, 184, 261, 265, 279, 280

Betroffene s. *Politikbetroffene*

Bewertung 45-48, 116, 125, 152, 164, 168, 233, 264

Black-box-Evaluation 21, 24, 134, 279

Controlling 21-22, 39, 45, 58, 76, 106, 147, 150, 152

Datenauswertung 230-233, 243-253

Datenerhebung 192-193, 223-230, 233-242

Gruppendiskussion 228-230

Indizes 239-40

Expertengespräch 227

Demokratie 10, 13, 33, 43

Deregulierung 8, 125

Deutschland 4, 6-7, 12, 15, 19, 21, 60, 106, 136, 163, 170-171

direktdemokratisch s. *Demokratie*

direkte Demokratie s. *Demokratie*

Dokumentenanalyse 24, 223-224

Effekte s. *Wirkungen*

Effektivität 37, 69, 73, 89, 93, 99, 104-109, 112, 119, 131, 148, 150, 168

Effizienz s. *Wirtschaftlichkeit*

Evaluation

als Bewertung 45-48

Begriff 39-40

und Implementationsforschung 24-25, 134-135

und Mediation 156-157, 163-168, 172

- und Monitoring 150
- und Politikanalyse 134-135
- und staatliches Handeln 40-44
- von Arbeitsprogrammen 67
- von Gesetzen 67
- von Massnahmen 66-67
- von öffentlicher Politik 67-68
- von Projekten 67
- wissenschaftliche Untersuchungsanlage 48-54
 - formativ 47, 52, 68, 268
 - interaktiv 24, 157, 218-220, 223, 233, 264-265, 267, 270
 - lernorientiert 218, 257, 264-265, 268-270, 279
 - partizipativ 144, 163, 166, 178, 226, 264, 266, 269, 279
 - prospektiv 26, 39, 68, 74-76, 81-82, 92, 118, 120
 - retrospektiv 21, 68, 74-76, 82, 92, 99, 101, 118, 121
 - summativ 47, 269
 - symbolisch 127, 269
 - von Mediationsverfahren 167-168
 - zielfrei 47, 178
 - zielorientiert 47, 178
- Evaluationsdesign 182, 185-198
- Evaluationsforschung 2-3, 17-18, 23-24, 39-40, 47-48, 54-57, 134-135, 219-221, 236-238, 253-254
- Evaluationsklausel 27-28
- Evaluationsmethode *s. Evaluation*
- Evaluationsstandard 55-57, 253-255
- Evaluationssynthese 117-118, 120, 280
- Evalulierbarkeitsprüfung 68, 179
- Experiment 15-16, 32, 58, 121, 174, 195, 198, 202, 206, 273
- Experimentelle Designs 194-95, 215
- Fallstudie 24, 174, 187-189, 199-204, 212, 215
- finanzielle Anreize 42, 44, 82, 273
- Finanzkontrolle 10, 30, 45, 126
- Föderalismus 10, 13, 17, 25, 43, 59, 60, 62
- Forschungsparadigma 156, 214-216
 - analytisch 49, 54, 174, 199, 203, 214-216
 - hermeneutisch 199, 214-217
 - ideologisch 262
 - interaktiv 218-220, 264-265

- interpretativ 214-217, 220-221, 223, 226, 257, 270
- konstruktivistisch 157, 233, 261
- partizipativ 264-266
- qualitativ 186, 200, 202, 216-233
- quantitativ 186, 233-255, 265
- Forschungstechnik (*s. auch Untersuchungsmethodik*) 175, 182, 214-255
- Frankreich 5, 8, 17, 60, 67, 106, 136
- Gebote 41, 63, 80
- Generalisierung
 - von Evaluationsresultaten 128-129, 201, 222-223
- Gesetz 15, 66-67, 95, 162, 190
- Gesetzgebung
 - experimentell 28-29, 273
- Glaubwürdigkeit 56, 117, 222, 279, 280
- Grossbritannien 4-12, 163, 259, 262
- Handlungshypothese 81
- Ideologie 128
- ideologisch 20, 78, 120, 131, 262, 270
- Impact 68, 73, 100, 102-105, 107, 135, 148
- Implementationsforschung 2, 16, 19-21, 24, 136
- Indikatoren 22, 86, 96, 105, 149, 152, 178, 185, 190, 200, 239
- Information 39, 42, 44, 57, 65, 115, 137, 143, 154, 226-227, 275
- Inhaltsanalyse
 - qualitativ 224, 231-232, 238
 - quantitativ 224, 236, 238-239
- Institution 29-31, 58-62, 125, 163, 228, 240
- Instrumente 21, 41-43, 87-88, 101, 106, 115, 119; 129-130, 133, 148, 152, 170-171, 195, 202, 206, 219, 236, 257, 271-273, 277
- Interventionen *s. Instrumente*
- Interventionshypothese 80-81
- Interviews 24, 200, 203, 206, 221, 225-228, 236-238
- Japan 162
- Kanada 4-12, 162
- Kausalhypothese 79, 109, 112, 180
- Kausalität 75, 121, 176, 187, 205, 245
- Kohärenz 78-83, 91-95, 98, 176
- Kontrolle 4, 10, 21, 23, 29-31, 102, 107, 119, 133, 147-148, 177-179, 184, 190, 254
- Kontrollgruppen 15, 104, 188-189, 194, 206-207, 250

- Kontrollorgane 25, 126
Korrelation, Korrelationsanalyse 246
Kosten-Effektivitäts-Analyse 2, 114
Kosten-Nutzen-Analyse 2, 114
Kreuztabelle, Kreuztabellenanalyse 244-246
Kunden *s. Politikadressaten*
Längsschnittanalyse 182, 187-189, 190-191, 207-210, 248-252
Legalität *s. Rechtmässigkeit*
Leistungsindikatoren 21-22, 39, 149, 151
Lernen
 in öffentlichen Politiken 26, 110, 157, 177, 223, 257-80
 personenorientiert 107, 262, 265
Machbarkeitsprüfung 179-180
Massnahme
 staatliche 43-44, 49, 66, 80
Mediation 62, 126, 133, 138, 156-172, 268
Metaanalyse 116-117
Metaevaluation 11, 26, 68, 74, 115-117
Methode *s. Untersuchungsmethode*
Monitoring 21, 31, 39, 45, 101, 120-121, 133, 147-155, 165
Nationales Forschungsprogramm "Wirksamkeit staatlicher Massnahmen"
 20, 26, 265, 270
Netzwerk 34, 80, 139-146, 268, 271-272, 274-276, 279
New Public Management 22-23, 39, 58, 61, 88, 96, 131, 138, 151
Niederlande 4-12, 131, 163, 259
NPM *s. New Public Management*
Nutzung 56, 119-132, 183-184, 257-280
 von Evaluationen 13, 31, 117, 119-121, 127-128, 130, 235, 254
OECD 125-126, 131, 162, 259
öffentliche Politik *s. Politik, öffentliche*
Outcome (*s. auch Wirkungen*) 23, 37, 39, 73, 75-77, 82, 103, 107, 109-111, 113-114, 116, 135, 148, 152, 154, 177, 180-181
Output 5, 23, 72, 76, 100-103, 105, 107-108, 114, 135, 148, 153-154, 176, 272
Paradigma *s. Forschungsparadigma*
Planung 6, 19, 21, 74-77, 92, 98-99, 146, 179

Politik

- öffentliche Politik 19, 21, 41, 49, 58-77, 82-83, 86-91, 93-95, 102-104, 106-111, 113-116, 119-132, 134-138
- substantielle Politik 58-62
- Politikadressaten 42, 63-64, 71-73, 79, 83, 92, 99, 103-104, 108, 112, 136, 180
- Politikanalyse 2, 25, 63-64, 66, 76, 129, 133, 134-138, 141, 156, 158, 257, 261
- Politikbeobachtung 22, 150
- Politikberatung *s. Beratung*
- Politikbetroffene 63, 73, 86, 110, 137, 171, 178, 265
- Politikformulierung 10, 37, 48, 64
- Politikkonzept 71, 78- 82, 116
- pooled time series 251
- Primärerhebung 234, 236-242
- Programm (*s. auch Verwaltungsprogramm*) 7, 10, 15, 21, 43-44, 47, 49-53, 67, 86, 98-99
- Programmformulierung 76, 93, 95
- Projekt 60, 67, 123, 166, 177, 267, 275
- Qualität, Qualitätskontrolle
 - von Evaluationen 3, 10, 55, 115-118, 167, 178, 201
- quasi-experimentell 195
 - Ansatz 182
 - Design 186, 194-196, 204, 215
 - Untersuchungsanlagen 198
 - Vergleich 198, 204-207
- Querschnittanalyse 189-190, 202, 243-244, 248, 251-252
- Rechnungshof *s. Finanzkontrolle*
- Rechtmässigkeit 30, 91, 94-95, 106
- Regressionsanalyse 244, 246-247, 251
- Regressionsansatz 250
- Reliabilität 56-57, 217, 220-221, 242, 250
- Ressourcen 7, 41, 46, 48, 57, 62-65, 72, 85, 88, 92-93, 99, 108-109, 113-115, 136-138, 142-143, 182, 191, 230, 242, 268, 275
- Ressourcentausch 275, 279
- Schweden 4, 6-8, 10, 131
- Schweiz 1, 5, 12-35, 42-43, 46, 57-60, 67, 91, 94, 96, 106, 113, 116-117, 123, 126, 131, 136, 142, 150, 152, 162, 180, 210, 212, 235-236, 239, 250-251, 259

- Sekundärdaten 117
- Sekundärerhebung 234-236, 242
- Signifikanz 245
- Skalen 236, 239-240, 249
- Stichprobe 192-193, 241
- Triangulation 174, 210-213, 230
- Übertragbarkeit von Evaluationsresultaten *s. Generalisierung*
- Überzeugung, Überzeugungspolitiken 42-44
- Umsetzung
 - von Evaluationen *s. Lernen sowie Nutzung*
 - von öffentlichen Politiken *s. Vollzug*
- Umverteilung 42, 88
- Untersuchungsmethodik
 - qualitativ 186, 200, 202, 216-223
 - quantitativ 186, 233-255
- USA 1, 2, 4, 6, 8-9, 12, 15, 17-19, 21, 24, 55, 57, 67, 76, 80, 106, 118, 130-131, 136, 162, 167, 170, 259
- Validität 56-57, 189, 217, 221-222, 242, 250
- Variable
 - situativ 138
- Vereinigte Staaten *s. USA*
- Versuchserlass 28-29
- Verwaltungsprogramm (*s. auch Programm*) 62-63, 65, 67, 71-73, 76, 83-85; 87-95, 135-136
- Verwendung von Evaluationen *s. Nutzung*
- Verwendungsforschung 117, 257-258, 278
- Vollzug 16, 19-22, 27, 48, 52, 64, 68, 72, 76, 93-98, 101, 104, 107, 115, 120, 133, 143, 147, 279
- Vollzugsdefizit 20-21, 52, 73, 92, 100, 105, 170
- Vollzugsforschung *s. Implementationsforschung*
- Vollzugsgrad 105, 107
- Wirksamkeit 30, 111-113
- Wirksamkeitsanalyse 30, 32, 48
- Wirkungen (*s. auch Outcomes*) 19-23, 45-53, 68
 - langfristig 53, 62, 259
 - unbeabsichtigt 51
 - unerwünscht 51, 82, 112
- Wirkungsanalyse 29-30, 32, 165

Wirkungsmodell

der Evaluation 54, 78-83, 134, 175, 179, 180-181, 215

der Politik 71, 79-81, 134, 180

Wirtschaftlichkeit 20, 25, 37, 48, 73, 96, 109, 113-115, 121

Wirtschaftlichkeitsanalyse 48, 114, 151

Zeitreihe, Zeitreihenanalyse s. *Längsschnittanalyse*

Ziele 47-48, 83, 86-87, 93, 96, 98, 109, 111, 119, 179

AUTORENVERZEICHNIS

Andreas Balthasar, Dr. rer.-pol., ist seit 1991 Leiter des Instituts für Politikstudien in Luzern und Lehrbeauftragter am Institut für Politikwissenschaft der Universität Bern und am Institut de recherche sur l'environnement construit (IREC) der Lausanne. Er ist Verfasser und Mitverfasser zahlreicher Evaluationsstudien. Zu seinen jüngeren Veröffentlichungen zählen *Technik-Standort Schweiz, Von der Forschungs- zur Technologiepolitik* (zusammen mit Willi Zimmermann und Carlo Knöpfel unter der Leitung von Dieter Freiburghaus, Bern, Stuttgart 1991), *Umweltschutz und technische Entwicklung* (mit Carlo Knöpfel, Basel 1994) und *Umsetzung und wirtschaftliche Wirkungen des Impulsprogrammes RAVEL, Eine Evaluation im Auftrage der Programmleitung* (erscheint demnächst).

Adresse: INTERFACE Institut für Politikstudien, Kapellgasse 1, CH-6004 Luzern, Tel: 41 41 410 07 12, Fax: 41 41 412 51 82, E-Mail: Interface@Access.

Hans-Martin Binder, lic.phil.I, war 1984-1996 Forscher und Berater bei der SYNTROPIE-Stiftung für Zukunftsgestaltung. Die Schwerpunkte lagen im Bereich von lernorientierten Evaluationen politischer Programme (Energie- und Technologiepolitik, Wirtschaftsförderung). Neuere Publikationen sind *Unternehmerische Innovationsprozesse; Die Wirksamkeit staatlicher Förderungsmassnahmen - Eine Evaluation* (zusammen mit Willy Bierter, Chur/ Zürich 1993), *Lernprozess als Ziel einer interaktiven Evaluationsstrategie*, (in: Werner Bussmann (Hrsg.): Lernen in Verwaltungen und Policy-Netzwerken; Chur/Zürich 1994) und *Begleitende Wirkungsanalyse der MICROSWISS-Zentren*, (Begleitforschung MICROSWISS Hauptphase 1/1994; Zürich 1995).

Adresse: Wildbachstrasse 74, CH-8008 Zürich, Tel./Fax: 41 13817471.

Werner Bussmann, Dr. oec, war 1989-1996 Programmleiter des Nationalen Forschungsprogramms "Wirksamkeit staatlicher Massnahmen" und ist Mitarbeiter des Bundesamts für Justiz, Bern. Seine Forschungsschwerpunkte sind Föderalismus, Vollzug und Evaluation öffentlicher Politiken, Verwaltungsführung und organisationales Lernen. Veröffentlichungen im Zusammenhang mit dem Natio-

nalen Forschungsprogramm sind *Lernen in Verwaltungen und Policy-Netzwerken* (Hrsg.; Chur/Zürich 1994), *Evaluationen staatlicher Massnahmen erfolgreich begleiten und nutzen: Ein Leitfaden* (Chur/Zürich 1995) und *Evaluations and grassroots politics: The case of Switzerland* (in Knowledge and Policy 1995, Vol. 8, Number 3).

Adresse: Bundesamt für Justiz, CH-3003 Bern,

Tel: 41 31 322 47 98, Fax: 41 31 322 84 01,

E-mail: Werner.Bussmann@mbox.gsejpd.admin.ch

Hans-Ulrich Derlien ist Professor an der Universität Bamberg. Seine Forschungsschwerpunkte sind der öffentliche Dienst und die Evaluation öffentlicher Programme. Zu seinen jüngeren Veröffentlichungen zählen *Compétence Bureaucratique et Allégances Politiques* (in Ezra Suleiman und Henry Madras (éd.), *Le recrutement des élites en Europe*. Paris 1995), *Public Administration in Germany: Political and Societal Relations* (in Ton Pierre (ed.), *Bureaucracy in the Modern State. An Introduction to Comparative Public Administration*. Aldershot 1994) und *Evaluation zwischen Programm und Budget* (in Albert Hofmeister (Hrsg.), *Möglichkeiten und Grenzen der Programmsteuerung: Controlling und Evaluation*. Schriftenreihe der Schweizerischen Gesellschaft für Verwaltungswissenschaften Bd.21. Bern 1994)

Adresse: Universität Bamberg, Lehrstuhl für Verwaltungswissenschaft, Feldkirchenstrasse 21, D-96045 Bamberg,

Tel: 49 951 863-2625/2626, Fax: 49 951 863-2627,

E-Mail: Hans-Ulrich.Derlien@sowi.uni-Bamberg.de.

Ingrid Kissling-Näf, ist Assistenzprofessorin für Forstliche Ressourcenökonomie an der Universität Zürich. Ihre Arbeitsschwerpunkte sind die Evaluations- und Vollzugsforschung unter besonderer Berücksichtigung des Umweltbereichs, umwelt- und ressourcenökonomische Fragestellungen sowie verwaltungswissenschaftliche Forschung. Ihre Dissertation befasste sich mit dem Thema *Lernprozesse und Umweltverträglichkeitsprüfung - Staatliche Steuerung über Verfahren und Netzwerkbildung in der Abfallpolitik* (Universität St. Gallen).

Adresse: -Zentrum HG G 12, Rämistr. 101, CH-8092 Zürich,

Tel: 41 1 632 32 25, Fax: 41 1 632 11 10,

E-Mail: kissling@waho.ethz.ch.

Ulrich Klöti ist Professor und Leiter des Instituts für Politikwissenschaft der Universität Zürich. Zu seinen Forschungsschwerpunkten zählen das Regierungs- und Verwaltungssystem und dessen Reform sowie Föderalismusfragen. Neue Veröffentlichungen sind *Die Stadt im Bundesstaat - Alleingang oder Zusammenarbeit?* (mit Theo Haldemann und Walter Schenkel, Chur/Zürich 1993), *Switzerland: Serving the State and Maximizing Income* (in Christopher Hood and B. Guy Peters (ed.): *Rewards at the Top*, London 1994) und *Regierungsreform in der Schweiz: ein langwieriger Prozess* (in Kleinstaat und Menschenrechte, Festgabe für Gerald Batliner zum 65. Geburtstag, hrsg. von Alois Riklin, Luzius Wildhaber und Herbert Wille, Basel 1993).

Adresse: Institut für Politikwissenschaft, Universität Zürich,
Karl Schmid Strasse 4, CH-8006 Zürich, Tel: 41 1 257 38 41,
Fax: 41 1 252 54 11, E-Mail: ukloeti@pwi.unizh.ch.

Peter Knoepfel ist Professor und Direktor des Institut de hautes études en administration publique (IDHEAP), Lausanne. Er befasst sich schwergewichtig mit Politikanalysen aller Art, Umweltpolitik und Verwaltungsreorganisation. Neue Veröffentlichungen sind: *Total quality management et fédéralisme: le point de vue de l'analyse des politiques publiques I Total Quality Management und Föderalismus - Betrachtungen aus der Sicht der Politikanalyse*, Cahier de l'EDHEAP no 159, Chavannes-près-Renens: IDHEAP 1996; *Approaches to an Effective Framework for Environmental Management*, (in: Carius, A., Höttler, L. und Mercker, H. (Hrsg.), *Environmental Management in Kenya, Tanzania, Uganda and Zimbabwe*, 1995) und *Von der konstitutionellen Konkordanz über administrative Konsenslösungen zum demokratischen Dezisionismus - zur Vielfalt von Verhandlungsarrangements in Konfliktlösungsverfahren in der Schweiz*, (in: Knoepfel, Peter (Hrsg.), *Die Lösung von Umweltkonflikten durch Verhandlung / La solution de conflits environnementaux par la négociation*, Basel 1995).

Adresse: IDHEAP, Route de la Maladière 22,
CH-1022 Chavannes-près-Renens, Tel: 41 21 694 06 40/44,
Fax: 41 21 694 06 09, E-Mail: Peter.Knoepfel@idheap.unil.ch.

Luzius Mader ist Vizedirektor im Bundesamt für Justiz, Bern, und Professor am Institut de hautes études en administration publique

(IDHEAP), Lausanne. Diverse Publikationen in den Bereichen Implementations- und Evaluationsforschung, Gesetzgebungsmethodik sowie Staats- und Verwaltungsrecht (in neuerer Zeit z.B. zum Gleichstellungsgesetz und zum Verfahren der Totalrevision der Bundesverfassung).

Adresse: Bundesamt für Justiz, CH-3003 Bern,
Tel: 41 31 322 41 02, Fax: 41 31 322 78 37.

Frédéric Varone, lic.rer.pol. und MPA, ist Doktorand an der Universität Bern und wissenschaftlicher Mitarbeiter am IDHEAP. Seine Forschungsschwerpunkte bilden das Policy Design, die Politikevaluation sowie die Technologiefolgen-Abschätzung. Seit 1991 hat er sich an verschiedenen interdisziplinären Evaluationsaufträgen beteiligt: z.B. *Evaluation der Bewilligungspflicht für elektrische Widerstandsheizungen* (hrsg. vom Bundesamt für Energiewirtschaft 1993), *Energetische Bedeutung der LESIT-Technologien* (hrsg. vom Schweizerischen Wissenschaftsrat 1996).

Adresse: IDHEAP, Route de la Maladière 22,
CH-1022 Chavannes-près-Renens, Tel: 41 21 694 06 00,
Fax: 41 21 694 06 09, E- Mail: IDHEAP@idheap.unil.ch.

Helmut Weidner, Dr. phil., ist seit 1978 wissenschaftlicher Mitarbeiter in der Abteilung Normbildung und Umwelt des Wissenschaftszentrum Berlin für Sozialforschung. Seine Hauptforschungsgebiete sind die Umweltpolitik im internationalen Vergleich sowie alternative Konfliktregelungsverfahren (insbesondere Mediation). Zu seinen jüngsten Veröffentlichungen zählen *Umwelt-Information. Berichterstattung und Informationssysteme in zwölf Ländern* (hrsg. mit Peter Knoepfel und Roland Zieschank, Berlin 1992), *Successful Environmental Policy. A Critical Evaluation of 24 Cases* (hrsg. mit Martin Jänicke, Berlin 1995; Basiselemente einer erfolgreichen Umweltpolitik, Berlin 1996) und *Alternative Dispute Resolution in Environmental Conflicts. Experiences in 12 Countries* (Hrsg.; im Erscheinen, Berlin 1996.).

Adresse: Wissenschaftszentrum Berlin für Sozialforschung (WZB), Reichpietschufer 50, D-10785 Berlin,
Tel.: 49 30 25 491-269/-284, Fax: 49 30 25 491-254.

Thomas Widmer, Dr. phil. I, ist wissenschaftlicher Mitarbeiter und Lehrbeauftragter am Institut für Politikwissenschaft der Universität Zürich. Seine Forschungsschwerpunkte sind die Methodik der Sozialforschung und die Policy-Analyse. Für das NFP 27 führte er die Studie *Meta-Evaluation. Kriterien zur Bewertung von Evaluationen* (Bern 1996) durch. Weitere Veröffentlichungen sind *Evaluation von Massnahmen zur Luftreinhaltepolitik in der Schweiz. Eine quasi-experimentelle Interventionsanalyse nach dem Ansatz von Box/Tiao* (Chur/Zürich 1991) und *Kurz und gut? Qualität und Effizienz von Kurzevaluationen*. (in Zusammenarbeit mit Christine Rothmayr und Uwe Serdült) (Chur/Zürich 1996).
Adresse: Institut für Politikwissenschaft, Universität Zürich,
Karl Schmid Strasse 4, CH-8006 Zürich, Tel: 41 1 257 38 41,
Fax: 41 1 252 54 11, E-Mail: thow@pwi.unizh.ch.

**ANHANG: VERZEICHNIS DER VERÖFFENT-
LICHUNGEN AUS DEM NATIONALEN
FORSCHUNGSPROGRAMM
"WIRKSAMKEIT STAATLICHER
MASSNAHMEN"**

Beim Schweizerischen Nationalfonds (Abt. IV/NFP 27, Postfach 2338, 3001 Bern, Tel 031 308 23 49) erhältliche Veröffentlichungen:

SCHLUSSBERICHTE

Blanc, Jean-Marc, Boyer, François und Piotet, Daniel 1995. *Evaluation comparative du processus d'apprentissage de la gestion de crise au sein de l'administration suite à la catastrophe de Tchernobyl.* Berne, 20 francs

Evaluationen in der Agrarpolitik I, Bern 1994, 192 Seiten, Fr. 20.-
Felder, Stefan und Rieder, Peter. *Analyse des Angebotsverhaltens und der Produktionsstruktur im Agrarsektor mit Hilfe eines LP Strukturmodells*

Evaluationen in der Agrarpolitik II, Bern 1994, 154 Seiten, Fr. 10.-
Baur, Priska, Meyer, Leo und Rieder, Peter. *Politökonomische Analyse der Getreidepolitik* Stefan Felder,
Cavegn, Gion, Meyer, Leo und Peter, Rieder. *Analyse des Angebotsverhaltens der Getreideproduzenten*
Meyer, Leo und Rieder, Peter. *Vorzeitige Stilllegung von Höchsttierbeständen in der Fleisch- und Eierproduktion 1980-1984;*

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

Lüscher, Claude und Wyss, Christian 1994. *Auswirkungen des UVP-Verfahrens in einem Schweizer Kanton.* Bern, 85 Seiten, Fr. 20.—

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

- Stalder, Peter, Frick, Andres, Gaillard, Serge und Salzgeber, Renate 1994. *Die Auswirkungen der Ausländerpolitik auf den Strukturwandel der Schweiz*. Bern, 97 Seiten, Fr. 20.--
- Urio, Paolo, Bari, Jean-Paul et Glauser, Christoph 1996. *L'évaluation comme instrument d'apprentissage et de changement organisationnels: l'exemple des programmes cantonaux de promotion de santé*. Berne, 20 francs
- Widmer, Thomas 1996. *Fallstudien zur Meta-Evaluation*, 7 Teile, Bern, 52 bis 78 Seiten, je Fr. 5.--

KURZBERICHTE

- "Avenir du tourisme de ski", Communauté de travail (éd.) 1995. *Tourisme de ski: de l'analyse à la stratégie*, Berne
- Balthasar, Andreas und Knöpfel, Carlo 1994. *Umweltvorschriften und technische Entwicklung*. Bern
- Bierter, Willy und Binder, Hans-Martin 1994. *Unternehmerische Innovationsprozesse, Die Wirksamkeit staatlicher Förderungsmassnahmen - eine Evaluation*. Bern
- Boyer, François et Piotet, Daniel 1995. *Leçons tirées de Tchernobyl par les administrations suisse et allemande*. Berne
- Bussmann, Werner 1996. *Evaluationen: Mehr Transparenz über die Wirkungen staatlichen Handelns*. Bern
- Bussmann, Werner 1996. *Les évaluations: plus de transparence sur les effets de l'action étatique*. Berne
- Delley, Jean-Daniel 1994. *Impact et limites des mesures de protection de l'air: Le cas de l'industrie et de l'artisanat*. Berne
- Gerheuser, Frohmüt W. und Schmid, André 1993. *Bedingt wirksam: Das Lohnungleichheitspostulat des Heimarbeitsgesetzes*. Bern
- Hanser, Christian, Küster, Jürg und Cavelti, Guido 1994. *Auswirkungen der staatlichen Hotellerieförderung*. Bern
- Lüscher, Claude und Wyss, Christian 1994. *Geringere Umweltbelastung dank Umweltverträglichkeitsprüfung?* Bern

- Meyer, Leo und Rieder, Peter 1994. *Evaluationen in der Agrarpolitik - Fallbeispiele im Getreidebau und der Tierhaltung*. Bern
- Rey, Michel et Vodoz, Luc 1993. *La promotion économique valaisanne*. Berne
- Schulz, Hans-Rudolf, Muggli, Christoph und Hübschle, Jörg 1993. *Die Wirksamkeit der staatlichen Wohneigentumsförderung*. Bern
- Stalder, Peter, Frick, Andres, Gaillard, Serge und Salzgeber, Renate 1994. *Die Auswirkungen der Ausländerpolitik auf den Strukturwandel der Schweiz*. Bern
- Widmer, Thomas 1996. *Evaluationen und ihre Qualität*. Bern
- Widmer, Thomas 1996. *Kurzevaluationen. Eine Alternative zu umfangreicheren Evaluationsstudien?* Bern

Im Buchhandel erhältliche Veröffentlichungen:

SYNTHESEN

- Bussmann, Werner 1995. *Evaluationen staatlicher Massnahmen erfolgreich begleiten und nutzen, Ein Leitfaden*. Chur/Zürich: Verlag Rüegger
- Bussmann, Werner 1995. *Accompagner et mettre à profit avec succès les évaluations des mesures étatiques, guide de réflexion*. Genève: Edition Georg
- Bussmann, Werner (Hrsg.) 1994. *Lernen in staatlichen Verwaltungen und Policy-Netzwerken*. Chur/Zürich: Rüegger, 144 Seiten
- Morand, Charles-Albert (éd.) 1993. *Evaluation législative et lois expérimentales.*: Aix-en-Provence: Presses universitaires d'Aix-Marseille
- Verwaltung und Management*, Mai/Juni 1996: S. 129-141 (Beiträge von Bundesrat Koller, ., Buschor, E., Bussmann, W., Klöti, U., Widmer T. und Mader, L.)

PROJEKTE

- Balthasar, Andreas und Knöpfel, Carlo 1994. *Umweltschutz und technische Entwicklung*. Basel: Helbing & Lichtenhahn
- Bierter, Willy und Binder, Hans-Martin 1993. *Unternehmerische Innovationsprozesse, Die Wirksamkeit staatlicher Förderungsmassnahmen - eine Evaluation*. Chur/Zürich: Rüegger
- Brandner, Balthasar et al. 1995. *Skitounsmus: Von der Vergangenheit zum Potential der Zukunft*. Chur/Zürich: Rüegger
- Delley, Jean-Daniel 1994. *Impact et limites des mesures de protection de l'air: Le cas des installations industrielles et artisanales*. Genève: CETEL (Faculté de droit, Uni Mail, CH-1211 Genève 4)
- Hanser, Christian, Küster, Jürg und Cavelti, Guido 1994. *Hotellerieförderung durch Bund und Kantone, Evaluation der Auswirkungen in der Hotellerie*. Schriftenreihe BIGA, Beiträge zur Tourismuspolitik Nr. 3, Bern: EDMZ
- Hohermuth, Susanne und Meier, Hans-Peter 1996. *Lokale Chancen für Nachhaltigkeit: Wintertourismus aus der Sicht der Bevölkerung*. Zürich: edition cultur prospektiv (Bezugsquelle: cultur prospektiv, Gasometerstrasse 28, 8005 Zürich)
- Kissling-Näf, Ingrid 1996. *Lernprozesse und Umweltverträglichkeitsprüfung - Staatliche Steuerung über Verfahren und Netzwerkbildung in der Abfallpolitik*. Dissertation Universität St. Gallen Nr. 1744 (erscheint 1997 im Verlag Helbing & Lichtenhahn, Reihe Ökologie & Gesellschaft, Band 12, Basel)
- Knoepfel, Peter, Kissling-Näf, Ingrid und Marek, Daniel 1997. *Lernen in öffentlichen Politiken*. Basel: Helbing & Lichtenhahn
- Sangra, Emmanuel 1992. *Evaluation des actions d'information de la protection civile sur le comportement de la population en cas d'alarme*. Le Mont-sur-Lausanne: Editions J-M Blanc SA
- Sauvain, Paul et al. 1995. *Zukunft des Skitounsmus / Avenir du tourisme de ski*. Schriftenreihe BIGA/OFIAMT: Beiträge zur Tourismuspolitik Nr. 7, Bern: EDMZ
- Schulz, Hans-Rudolf, Muggli, Christoph und Hübschle, Jörg 1993. *Die Wirksamkeit der staatlichen Wohneigentumsförderung*. Schriften-

reihe Wohnungswesen (Hrsg. Bundesamt für Wohnungswesen),
Band 55, Bern: EDMZ

Widmer, Thomas 1996. *Meta-Evaluation. Kriterien zur Bewertung von Evaluationen*. Bern: Haupt

Widmer, Thomas, Rothmayr, Christine und Serdült, Uwe 1996. *Kurz und gut? Qualität und Effizienz von Kurzevaluationen*. Zürcher Beiträge zur Politikwissenschaft, Band 19, Chur/Zürich: Rüegger.

KURZEVALUATIONEN

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

ECOPLAN 1993. *Schneeräumung von Urner Passstrassen, Evaluation im Auftrag der Baudirektion des Kantons Uri*, (erhältlich beim Bauamt des Kantons Uri, 6460 Altdorf, Tel. 044 4 26 11)

Forster, Sophie und Jeanrenaud, Claude 1994. *La promotion économique neuchâteloise, bilan et perspectives*. Neuchâtel (peut être obtenu auprès de l'Institut de recherches économiques et régionales, Tel. 038 21 10 85)