

Notwendigkeit und Grenzen der Erfolgskontrolle öffentlicher Interventionsprogramme

von

Hans-Ulrich Derlien

Mit der Zunahme gezielter öffentlicher Interventionen in vormals weitgehend selbstregulative soziale und ökonomische Prozesse ist das System traditioneller, auf Prüfung der Rechtmäßigkeit öffentlichen Handelns ausgerichteter interner und externer Verwaltungskontrollen ergänzt worden um Erfolgskontrollen (Evaluationen), die Wirkungen und Auswirkungen öffentlicher Interventionen prüfen sollen. Diese logische Konsequenz einer stärkeren Betonung politischer Planung läßt sich mehrfach theoretisch begründen (I) und ist faktisch unterschiedlich wirksam geworden; entsprechend fragmentiert stellt sich heute die Organisation der Erfolgskontrolle dar (II). Wenn wir nach nunmehr zwanzigjährigen deutschen und amerikanischen Bemühungen um eine Institutionalisierung der Evaluation als letzter Phase des Planungszyklus heute immer noch kein flächendeckendes System der Erfolgskontrolle auf der Bundesebene, auf die wir unsere Aussagen beschränken werden, feststellen, so ist dieses Defizit auch die Folge methodischer Probleme in der Technik der Erfolgskontrolle und der Eigenart des programmatischen Lernens im politisch-administrativen System, durch die dem, was planungslogisch zwingend erscheint, in der Praxis vielfach Grenzen gesetzt werden (III). Hierbei muß es sich nicht um prinzipielle Grenzen handeln; vielmehr verweisen praktische und theoretische Weiterentwicklungen der letzten Jahre darauf, daß einerseits der theoretische Anspruch der Evaluationsforschung zurückgenommen, andererseits Mutationen des Grundgedankens erfolgversprechend verwirklicht werden (IV).

I. Begründung der Notwendigkeit von Erfolgskontrollen

Die Forderung nach Durchführung und Institutionalisierung von Erfolgskontrollen wird im wesentlichen mit vier, im Einzelfall ineinander übergreifenden Begründungen erhoben (Vgl. Derlien, H.-U. 1976, S. 29-33; Hellstern, G.-M. und Wollmann, H. 1984a, S. 38ff.).

1. Planungslogische Begründung

In kybernetischen Modellen zielorientierten Planens ist die Rückkoppelung der Planungsergebnisse in die laufende Planung, die Analyse der Planabweichung und eine entsprechende Plankorrektur eine Selbstverständlichkeit. Besonders öffentliche Interventionsprogramme, die nicht nur inkrementale Veränderungen am status quo, sondern innovative Reformen anstreben, müssen mit Unsicherheiten der Zielerreichung und unvorhergesehenen Folgen des intentionalen Handelns rechnen, da sozio-ökonomische Theorien kaum treffsichere Prognosen erlauben (Tenbruck, F.H. 1972); "muddling through" (Lindblom, Ch.E. 1969) und "piece-meal engineering" (Popper, K.R. 1961) genügen dem Anspruch innovativer Planung ebensowenig wie die damit verbundene politikwissenschaftliche Konzeption des programmatischen Lernens aus Rückmeldungen des pluralistischen Willensbildungsprozesses (Braybrooke, D. und Lindblom, Ch.E. 1963). Systematisches Lernen aus vorangegangenen Interventionen erfordert vielmehr die wissenschaftliche Analyse von Zielerreichungen, Ursachen mangelnder Zielerreichung und eventueller Nebenfolgen, um eine Präzision zu erreichen, wie sie von politischen Erfolgs- und Mißerfolgsartikulationen nicht zu erwarten ist. In letzter Konsequenz liefe die planungslogische Begründung auf eine Verwissenschaftlichung der Politik zumindest in ihren theoretischen und faktischen Prämissen hinaus.

2. Planungsorganisatorische Begründung

In dem Maße, wie Planungsprozesse systematisiert und von speziellen Planungseinheiten der Exekutive gesteuert werden, sollte, so wird in einer weiteren Begründung gefordert, die Erfolgskontrolle nicht nur verfahrensmäßig institutiona-

lisiert, sondern auch möglichst von Spezialisten in besonderen Evaluationseinheiten wahrgenommen werden; zum einen fehle der Ministerialbürokratie das forschungsmethodische Wissen, um Untersuchungen über Erfolg und Mißerfolg von Programmen beauftragen, auswerten oder gar selbst durchführen zu können; zum anderen rechnet man mit einem Programmkonservatismus der verantwortlichen Linieneinheiten, der eine Umsetzung von Evaluationsergebnissen erschweren würde. Die planungsorganisatorische Begründung schließt zudem eine Kritik der herkömmlichen Kontrollformen ein: weder die Rechtmäßigkeitskontrolle noch die Orientierung an politischen Rückmeldungen oder an verausgabten Haushaltsmitteln können als valide Indikatoren für einen Programmerfolg angesehen werden.

3. Haushaltspolitische Begründung

Schon frühzeitig wurde von Planungspraktikern erkannt, daß Reformpolitiken auf fiskalische Grenzen stoßen würden. In der Evaluation wurde in diesem Zusammenhang auch ein Mechanismus gesehen, ineffektive Programme zu entdecken und zu streichen und damit eine Reallokation von Mitteln zugunsten effektiver, vornehmlich neuer Programme einzuleiten. Insofern wurde vor allem im Ausland die Evaluation als Bestandteil integrierter Aufgaben- und Finanzplanungssysteme betrachtet; dies zeigt der enge historische und systematische Zusammenhang, in dem sowohl in den USA die Evaluation mit dem PPBS als auch in Großbritannien das Programm Analysis and Review (PAR) mit der Finanzplanung stand (Derlien, H.-U 1976, S. 57f.). Selbst wenn nach dem Scheitern dieser integrierten Planungssysteme sowie angesichts des immer noch fehlenden Aufgabenbezugs der Finanzplanung in der Bundesrepublik (Wille, E. 1980) in der Erfolgskontrolle ein Weg zur "Prioritätensetzung durch die Hintertür" (Rürup, B. und Färber, G. 1985) gesehen wird, wird noch deutlich, wie hoch das Reallokationspotential der Evaluation eingeschätzt wird.

4. Planungspolitische Begründung

Schließlich verspricht der systematische, methodisch gesicherte Nachweis von Effektivität oder Ineffektivität eines Programms auch eine politisch legitimierende Wirkung für anstehende Programmkorrekturen, sei es daß Programme finanziell aufgestockt, sei es daß sie beschnitten oder gar eingestellt werden. Der Ver-

weis auf die Ergebnisse einer Evaluation sollte den Programmverteidigern innerhalb und außerhalb der Administration ihr Festhalten an einem ineffektiven und den Programmkritikern eines effektiven Programms ihre Argumentation erschweren, also Programmkorrekturen rational erscheinen lassen und sie damit wissenschaftlich legitimieren.

II. Institutionalisation der Erfolgskontrolle

Wenngleich die Durchführung von Erfolgskontrollen prinzipiell sowohl in Gemeinden und Ländern als auch beim Bund gefordert werden kann, haben sich Praxis und wissenschaftliches Interesse bislang vorwiegend auf die Bundesebene konzentriert. Obwohl auch in der Kommunalpolitik sinnvoll, relativiert sich die Bedeutung der Evaluation hier aufgrund der Geschehensnähe und der Vielfalt relativ unverzerrter traditioneller Rückkoppelungsmechanismen der Kommunalverwaltung (Hellstern, G.-M. und Wollmann, H. 1984 b). Mit Einschränkungen mag dies angesichts der Möglichkeit, eine Vielzahl von programmbezogenen Informationen aus der Vollzugsverwaltung zu gewinnen, auch für die Landesebene gelten, wenngleich hier - wie bei der Stadtentwicklungsplanung auf Kommunalebene - aus planungslogischen Gründen die Erfolgskontrolle in der Regel in der Landesplanung vorgesehen ist (Pfitzer 1985).

1. Historische Entwicklung

Der enge Zusammenhang zwischen staatlicher planerischer Intervention und systematischer Rückmeldung von Planergebnissen wird nicht nur beim Vergleich der Entwicklung in den USA und der Bundesrepublik seit den 60er Jahren deutlich, sondern ist auch an der Geschichte beider Systeme einerseits in der Unterstützung des New Deal durch die in den USA bis heute stärker etablierte angewandte Sozialforschung, andererseits in Deutschland an der gleichgerichteten Entwicklung von merkantilistischer Politik und kameralistischer Methodenentwicklung schon im 18. Jahrhundert abzulesen. Im Zusammenhang mit programmatischen politischen Innovationen nach 1966 schwindet die Planungsabstinenz der Aufbauphase der Bundesrepublik; zunächst werden Planungsmethoden (Nutzen-Kosten-Analysen, Sozialindikatoren, Prognosen) erprobt und planungsorganisatorische Neuerungen eingeführt (Schatz 1973); gleichzeitig wird die amtliche

Statistik in Reaktion auf den gestiegenen planungsinduzierten Informationsbedarf ausgebaut, und das Berichtswesen der Bundesregierung wird verfeinert (Derlien 1975). Ansätze zur Institutionalisierung der Erfolgskontrolle reichen zwar in die späten 60er Jahre zurück, wobei vor allem das parlamentarische Kontrollbedürfnis in den Bereichen der Entwicklungshilfe- und der Forschungs- und Technologiepolitik auslösend war; nennenswerte Erfolge bei der Institutionalisierung und organisatorischen Ausdifferenzierung der Evaluationsfunktion stellen sich jedoch erst mit einem time lag ein, nachdem die Reformpolitiken in die Durchführungsphase gelangt sind (Derlien 1976, S. 82). Dabei sind keineswegs alle Interventionsprogramme mit einer Erfolgskontrolle verbunden worden, sondern - wie in den USA - waren besonders Programme im Bereich der Bildungs- und der Sozialpolitik, kostenintensive "weiche" Interventionspolitiken, einem erhöhten Legitimationszwang durch Erfolgsnachweis ausgesetzt. Diese Einseitigkeit (Derlien 1984 S. 11) hat neben planungspolitischen auch planungslogische Gründe, sind doch aufgrund von Theorie-Schwächen planerische Erfolgsprognosen hier in der Regel unsicherer als in "harten" Politikbereichen, in denen das Planungsziel nicht selten lediglich in der Bereitstellung einer bestimmten Infrastruktur besteht und man sich daher stärker auf die Nutzen-Kosten-Analyse des § 7 Abs. 2 BHO zu verlassen können glaubt (Derlien 1979). Aus heutiger Sicht wird auch deutlich, wie sehr planende Intervention in sozio-ökonomische Zusammenhänge und Strukturen sowie die damit strukturell konsistente Erfolgskontrolle in den Kontext expansiver Budgetierung eingebettet waren: vergleicht man den Ausbaustand der Erfolgskontrolle im Jahre 1974 (Derlien 1976) mit dem Ende der 70er Jahre erreichten Stand (Hellstern/Wollmann 1984) so erkennt man zumindest im Hinblick auf die Institutionalisierung kaum Fortschritte. Zwingend ist dieser Stillstand jedoch auch bei einer fiskalischen Konsolidierungspolitik, die programmatisch neo-konservativ ausgerichtet ist, keineswegs, gelten doch hier ebenfalls die Begründungsstränge, die für die Einführung von Erfolgskontrollen sprechen (Derlien 1985).

2. Programmspezifische Institutionalisierungen

Nicht zuletzt, weil sich die Einsicht in die Notwendigkeit von Programmevaluationen in einer primär juristisch professionalisierten Ministerialverwaltung erst durchsetzen und das anfangs verbreitete Mißverständnis von Erfolgskontrolle als personenbezogener Leistungskontrolle abgebaut werden mußte, sind Untersu-

chungen mit evaluativen Fragestellungen nicht selten als schlichte Ressortforschung unbeachtet geblieben. Erst mit der (gesetzlichen) Verankerung von *Berichtspflichten* der Bundesregierung an den Bundestag wird der Umfang regelmäßig durchgeführter Evaluationen deutlicher erkennbar. Gleichwohl werden darüberhinaus Erfolgskontrollen aber auch ad hoc durchgeführt, sei es daß ein Programm politischer Kritik ausgesetzt ist, sei es daß Mittel der Ressortforschung auch für derartige Fragestellungen eingesetzt werden. Insofern ist es nach wie vor schwierig, präzise empirisch zu bestimmen, in welchem Umfang tatsächlich Erfolgskontrollen durchgeführt werden, welchen Anteil sie an der sozial- und wirtschaftswissenschaftlichen Programmforschung (policy research) einnehmen und welche Mittel sie im Rahmen der Auftragsforschung binden (Derlien 1981, S. 24; Hellstern/Wollmann 1984a, S. 35). Geht man von den Berichtspflichten an den Bundestag aus, die mehr oder weniger regelmäßig von der Bundesregierung zu erfüllen sind und sich lediglich auf einzelne Programme beziehen, so kann man für 1974 etwa 50 (Derlien 1975) und für 1980 etwa 80 (Zeh, W. 1984, S. 40) Erfolgskontrollen unterstellen. Beispiele von Erfolgskontrollen in Berichtform sind Untersuchungen im Zusammenhang mit der Durchführung des ArbeitsförderungsG von 1969 (Erfolg von Umschulungs- und Fortbildungsmaßnahmen), des Lohnfortzahlungsg von 1969 (Auswirkungen auf den Krankenstand), Novelle des BundessozialhilfeG von 1969 (Erfolg der Eingliederung Behinderter), AusbildungsförderungsG von 1969 (Lage der Schüler und Studierenden), Wohnraum-KündigungsschutzG von 1971, Liberalisierung des § 218 Strafgesetzbuch von 1975, Durchführung der Gemeinschaftsaufgabe "Förderung der regionalen Wirtschaftsstruktur" von 1970, KrankenhausfinanzierungG von 1971 oder die Subventionsberichte. Gerade die Subventionsberichterstattung verweist aber darauf, daß diese Evaluationen von unterschiedlicher methodischer Güte und programmatischer Aussagekraft sind.

Auf dem Hintergrund der planungslogischen Funktion der Erfolgskontrolle als Mechanismus systematischen programmatischen Lernens aus der Kausalität von Maßnahmen und Wirkungen mußte sich neben gesetzesförmigen Interventionsprogrammen gerade auch eine *experimentelle Politik* in den Politikfeldern aufdrängen, in denen Erfolge zu ungewiß und Nebenfolgen politisch zu kontrovers waren, als daß man zu generellen Regelungen hätte übergehen können. Wenn gleich die in der Praxis als "Modellversuch" deklarierten Experimentalvorhaben in der Sozial-, Gesundheits- und Bildungspolitik nicht selten von der Bundesre-

gierung initiiert wurden, um verfassungsrechtlichen Bedenken gegen bundesgesetzliche Regelungen zu entsprechen, so sind doch gerade diese Modellversuche evaluiert worden. Zwischen 1971 und 1980 sind 1368 Schulversuche, 166 Modellvorhaben im Hochschulbereich und 225 Modelle der beruflichen Bildung durchgeführt worden (Bundestagsdrucksache 9/699); bekannt geworden ist in der Öffentlichkeit vor allem die - politisch kontroverse - Evaluation der Gesamtschulversuche (vgl. Raschert, J. 1981). Im Geschäftsbereich des BMJFG konnten 1982 allein 56 Modellprojekte registriert werden (Hellstern/Wollmann 1984a, S. 42). Hier wie dort haben diese Experimente jedoch eine ambivalente Funktion erfüllt, indem sie nur z.T. informationelle Unsicherheiten einer generellen Regelung reduzieren, sondern auch der politischen Durchsetzung des Bundes gegenüber den Bundesländern dienen sollten (Hellstern, G.-M. und Wollmann, H. 1983); daß gelegentlich weder das eine noch das andere Motiv vorherrscht, sondern mittels Modellversuchen und zeitaufwändiger Evaluation auch Konflikte vertagt werden sollen, machen die jüngsten Versuche mit Tempobegrenzungen auf Autobahnen unter Umweltgesichtspunkten ebenso deutlich wie ein ähnlicher Versuch in den 70er Jahren, der in erster Linie der Steigerung der Verkehrssicherheit dienen sollte (Derlien, 1976, S. 116).

Abgesehen davon, daß derartige Experimente besondere methodische Voraussetzungen schaffen (können), um die meist schwierigen Fragen der kausalen Zurechnung gemessener Programmeffekte auf spezielle Interventionen beantworten zu können, sind sie dadurch gekennzeichnet, daß ihre Zielgruppe regional oder sozial begrenzt ist und freiwillig an dem betreffenden Modellversuch teilnimmt; ferner sind Experimente ex definitione auch zeitlich befristet. In der Praxis sind seit einigen Jahren nun Tendenzen zu beobachten, auch generelle Regelungen und *Gesetze zeitlich zu befristen* (Hugger, W. 1979; Böhret, C. 1985, S. 148-151). Trotz einiger rechtspolitischer Bedenken scheint die Tendenz zum Gesetz auf Zeit, für das es wiederum in den USA die Parallele der "sunset legislation" gibt (Adams, B. and Sherman, B. 1978; Bothun, D. and Comer, J. 1979), sich auch in der Bundesrepublik durchzusetzen. Allerdings ist das Anliegen dieser gesetzestechnischen Variante nicht nur darin zu sehen, daß der Evaluationszwang gegenüber der Exekutive gesteigert werden soll; denn die Fortschreibung des Gesetzes wird hier vom Nachweis der Effektivität abhängig gemacht; diese Experimentierhaltung mag noch bei der Terminierung der Regelungen des 2. WohnraumkündigungsschutzG von 1974 und ihrer Fortschreibung nach Vorlage des Untersu-

chungsberichts (Bundestagsdrucksache 8/2610) dominiert haben. Vielmehr wird die Befristung gelegentlich auch genutzt, um Umverteilungspolitik politisch zu entschärfen, etwa wenn die Beteiligung an Kosten der Krankenversorgung 1982 befristet wurde. Schließlich wird in der Terminierung auch ein Mittel der Deregulierung gesehen; besonders bei untergesetzlichem Recht sollen auf diese Weise Vorschriften nach einer gewissen Zeit automatisch auslaufen, wenn sie nicht zuvor ausdrücklich erneut beschlossen worden sind.

3. Einrichtung spezieller Evaluationseinheiten

In dem Maße, wie die Häufigkeit zunahm, mit der Erfolgskontrollen oder andere analytische Studien im Regierungsapparat anfielen, sind auch spezielle Referate eingerichtet worden, die mit dieser Aufgabe betraut worden sind (vgl. Derlien, 1976, S. 76-80; Hellstern, G.-M. und Wollmann, H. 1984). Zu nennen sind hier vor allem das Bundesministerium für wirtschaftliche Zusammenarbeit, in dem 1970 ein Referat "Inspektion" eingeführt worden ist, das - entsprechend der in diesem Politikbereich international älteren Tradition - die geförderten Entwicklungsprojekte planmäßig einer Erfolgskontrolle unterzieht; das Landwirtschaftsministerium, in dem in der Planungsabteilung ein Referat nicht nur für die Kosten-Nutzen-Untersuchungen nach § 7 Abs. 2 BHO, sondern seit der Ausdehnung der Verwaltungsvorschriften hierzu auf die ex-post-Analyse auch für Evaluationen zuständig ist; drittens das Verkehrsministerium, in dem auch häufiger "Maßnahmen von erheblicher finanzieller Bedeutung" und somit Studien nach § 7 Abs. 2 BHO anfallen; hier erhielt das ebenfalls in der Grundsatzabteilung angesiedelte Referat 1978 die ergänzende Bezeichnung "Erfolgskontrolle". Andere Ministerien begnügen sich mit weniger deutlich sichtbaren Forschungsreferaten oder lassen Evaluationsstudien von Forschungsanstalten im nachgeordneten Bereich des Ministeriums durchführen. In Überblicken wird auch immer wieder das Presse- und Informationsamt der Bundesregierung aufgeführt, das seine Informationspolitik seit 1968 evaluiert, wie dies auch die Bundeszentrale für gesundheitliche Aufklärung seit 1970 für ihre pädagogischen Informationsprogramme unternimmt, schließlich wurde die Forschungsmethodik der Evaluation zuerst im Bildungsbereich entwickelt und erprobt.

Das *Bundeskanzleramt* hat seit eh und je eine sehr zurückhaltende Rolle gegen-

über diesen Evaluationsbemühungen der einzelnen Ministerien gespielt, wenn gleich man seit dem Regierungswechsel von 1982 ähnlich wie das amerikanische Office of Management and Budget verstärkt den methodischen Informationsaustausch zwischen den Ressorts (König, K. 1986, S. 33) fördern will. Der *Bundesrechnungshof*, geprägt von der traditionellen, auf die Richtigkeit und Rechtmäßigkeit ausgerichteten, in der Sache primär als Sparsamkeitskontrolle durchgeführten Finanzkontrolle, scheint neuerdings ebenfalls seine Prüfungspraxis aus der Sicht der Programmevaluation zu überdenken (vgl. Wittrock, K. 1984; Zavelberg, H.G. 1986). Hierzu mag beigetragen haben, daß der Ausdruck "politische Erfolgskontrolle" (Bohne, E. und König, H. 1976) sich nicht hat durchsetzen können und mithin die von BRH zu recht abgelehnte politische Bewertung von Programmen mit der Evaluation nicht verbunden sein muß, sondern diese wie die "Bemerkungen" ganz allgemein als Feststellung zur Wirtschaftlichkeit (speziell: zur Effektivität) dem Bundestag zugeleitet würde. Es dürfte allerdings fraglich sein, ob der BRH, vorausgesetzt: dies ist überhaupt beabsichtigt, in absehbarer Zeit analytische Kapazitäten wie das amerikanische General Accounting Office (vgl. Marvin, K.E. und Cohen, W.M. 1984) wird aufbauen können; daß ein Evaluationsamt (vgl. Reding, K. 1981, S. 221) als zusätzliche externe Kontrollinstanz eingerichtet werden könnte, erscheint mehr als zweifelhaft.

Das institutionelle Bild der Erfolgskontrolle auf Bundesebene wäre unvollständig, wiese man nicht abschließend darauf hin, daß selbst dort, wo spezielle Einheiten für die Wahrnehmung dieser Aufgabe gebildet worden sind, die Durchführung der Untersuchungen selbst, d.h. i.d.R. die Forschungsarbeit, bei der externen universitären und nicht-staatlichen Forschung liegt; lediglich Beauftragung, administrative Begleitung und Auswertung für Berichte an den Bundestag sind i.d.R. Aufgaben der Spezialeinheiten (Derlien 1978).

III. Methoden und politische Probleme der Erfolgskontrolle

Wenn die Institutionalisierung der Erfolgskontrolle heute immer noch ein fragmentiertes Bild abgibt, so ist dies keineswegs lediglich die Folge organisatorischer Trägheit oder des administrativen Widerstandes gegen Kontrollen schlechthin, sondern läßt sich auch daraus erklären, daß wie bei der Übernahme neuer Planungsverfahren Skepsis hinsichtlich der methodischen Durchführbar-

keit eine Evaluationseuphorie nie hat entstehen lassen. Die Meta-Evaluationsforschung (Cook, Th.D. and Gruder, Ch. 1979) hat inzwischen auf immanente methodische Probleme und schwer einlösbare Prämissen der Verfahrensweisen hingewiesen und aufseiten der Evaluationsforschung zur Entwicklung einer zweiten Generation von Techniken und Selbstverständnissen beigetragen, auf die im Abschnitt IV noch einzugehen sein wird. Aber auch die planungspolitischen Erwartungen sind vielfach unerfüllt geblieben - nicht zuletzt, weil organisatorische, methodische und politische Aspekte der Evaluation interdependent sind (Derlien 1976).

1. Spezifizierung von Effekten

Der Begriff "Erfolgskontrolle" bezeichnet im Sprachgebrauch der Verwaltungspraxis meist lediglich die Überprüfung von Programmen hinsichtlich ihrer Zielerreichung; positive oder negative Nebenfolgen, die gerade im Selbstverständnis der professionellen Evaluationsforschung eine große Rolle spielen, bleiben ausgeblendet, obwohl aus planungslogischen Gründen das Augenmerk auf Nebenfolgen gerichtet sein sollte, wie dies die cost-benefit-analysis schon für die Planungsphase versucht. Im Rahmen der Untersuchung aller möglichen Auswirkungen eines Programms stellt die Zielerreichung mithin nur die Klasse der positiven intendierten Wirkungen dar, während die positiven und negativen unvorhergesehenen Konsequenzen (wir unterstellen: negative intendierte Wirkungen gäbe es nicht) theoretisch abgeleitet werden müssen. Nicht nur bei dieser theoretischen Deduktion steht die Evaluation vor der Schwierigkeit, entweder aus den materiellen Politikprämissen wie schon die Planer keine Implikationen ableiten zu können, weil der Bestätigungsgrad dieser Prämissen zu gering ist, oder aus dem Universum möglicher Folgen einige als untersuchenswert und untersuchbar auswählen zu müssen. Zudem erweist sich auch die Spezifizierung der intendierten positiven Effekte als problematisch, weil Programme gelegentlich gar keine Zielformulierungen aufweisen (und insofern keine Programme, sondern Einzelinterventionen oder Maßnahmenbündel wären) oder weil diese Zielformulierungen zu allgemein oder inhaltsleer sind, als daß man das Programm und seine Maßnahmen als Operationalisierung der Zielerreichung ansehen könnte (ex-ante-Operationalität). Dieser Zusammenhang verweist darauf, daß politische Zielsetzungen nicht selten am Schluß der Programmentwicklung formuliert werden, symbolisch wirken oder ungelöste Zielkonflikte verdecken sollen. Natürlich gibt es auch Zielfor-

mulierungen, die ex post operational sind: das Ausmaß von Arbeitslosigkeit, die ein Programm zu reduzieren versucht, läßt sich messen, selbst wenn das Ziel der Verminderung von Arbeitslosigkeit in dem Sinne ex ante nicht operational ist, daß man genau sagen könnte, mit welchen Mitteln das Ziel zu erreichen wäre. Ferner wird die Spezifizierung und Auswahl der als relevant definierten Wirkungen nicht selten im Hinblick auf die Verfügbarkeit von Daten vorgenommen. Die Selektivitäten, die sich aus Gründen der fehlenden Zieloperationalität, der theoretischen Ableitbarkeit und der Datenbasis ergeben, stellen letztlich Evaluationsergebnisse in ihrer Verbindlichkeit für die Entscheidungsträger infrage, weil möglicherweise politische Erwartungen, die in der Zielformulierung nicht berücksichtigt worden sind, auch vom Forscher vernachlässigt werden.

2. Kausale Zurechnung gemessener Effekte

Um überhaupt Wirkungen eines Programms vermuten zu können, müssen Differenzen der Indikatorwerte für den Zustand vor und nach der Intervention feststellbar sein. Wenn Evaluationen nicht schon bei der Programmverabschiedung geplant sind, ist es häufig jedoch nicht mehr möglich, den Zustand des Interventionsbereichs vor der Durchführung des Programms zu ermitteln oder gar Zeitreihen von Daten zu gewinnen. Selbst wenn Veränderungen einzelner Indikatoren nachweisbar sind, ist zudem fraglich, ob diese Effekte autonom oder durch das Programm verursacht wurden; vielleicht wäre der Zustand der Zielvariablen auch negativ ausgefallen, anstatt unverändert zu bleiben, hätte man gar nicht interveniert? Schließlich: selbst wenn Effekte auf Plausibilitätsbasis kausal auf das Programm zugerechnet werden können, welches seiner Elemente hat sich als besonders (in-)effektiv erwiesen? Ist möglicherweise gar kein Programmfehler für fehlende Effekte, sondern die Art des Vollzuges für den fehlenden Programmserfolg verantwortlich (Derlien 1984a)? Derartige Fragen ließen sich letztlich nur einer Klärung zuführen, wenn das betreffende Programm experimentell angelegt wäre: Messung der Zielindikatoren vor und nach Intervention, zufällig ausgewählte Zielgruppe und nicht tangierte Kontrollgruppe, gezielte Variation einzelner Programmelemente und alternativer Formen der Implementation sind Voraussetzung. Derartige Laborbedingungen lassen sich jedoch selbst dort, wo tatsächlich Sozialexperimente im Feld durchgeführt werden (s.o.II.1) kaum erreichen, sondern diese Modellversuche sind bestenfalls Quasi-Experimente. Schon die geographische Lokalisierung ist nicht selten ein Politikum (welcher Wahl-

kreis?); die Repräsentativität der Ergebnisse leidet ferner nicht selten unter Selbstrekrutierungseffekten der Teilnehmer (Schüler und Lehrer bei Gesamtschul-Versuchen); spill-over-Effekte in der Kontrollgruppe sind schwer methodisch zu kontrollieren. Schließlich ist das Programmdesign während der Laufzeit des Experimentes häufig nicht vor Interventionen (Verbesserungen) vonseiten des Programmpersonals zu "schützen": Lehrer sind schließlich am Wohlergehen ihrer Schüler interessiert und nicht an der Zuverlässigkeit von Forschungsergebnissen.

Aus diesen und einer Reihe anderer Gründe gilt für die Evaluationsforschung das "eherne Gesetz der methodischen Fehler".

3. Politische Umsetzung der Evaluationsergebnisse

Angewandte Forschung allgemein und die Evaluationsforschung im besonderen ist über die erhoffte Rationalisierungswirkung, die sie auf politische Entscheidungen ausüben will, in den letzten Jahren gerade auch in den USA mit ihrem höher entwickelten Forschungsstand enttäuscht worden (Wildavsky, A. 1979; Weiss, C. and Bucuvalas, M.J. 1980). Politische Entscheidungsträger und administrative Programmverantwortliche folgen nicht der Popperschen Falsifikationslogik, korrigieren nicht immer ein Programm, dessen mangelnde Effektivität wissenschaftlich erwiesen schien, oder streichen Programme, an deren Effektivität kein Zweifel entstehen konnte. Im politischen Entscheidungsprozeß über Fortführung, Verbesserung oder Beschneidung eines Programms spielen die Ergebnisse von Erfolgskontrollen als Argumente für die Beibehaltung oder Veränderung des status quo zwar eine gewisse Rolle, zumal wenn sie der Öffentlichkeit zugänglich werden. Allerdings werden dann nicht selten die Untersuchungen selbst politisiert und in ihrer methodischen Qualität und Legitimation angezweifelt. Dies gilt selbst für amerikanische Experimentalprogramme, deren methodisches Raffinement in der Bundesrepublik noch gar nicht erreicht worden ist, wie z.B. das kostspielige, zeitaufwendige Experiment mit der negativen Einkommensteuer (als Alternative zur Zahlung von einzelnen Transfers oder Bereitstellung öffentlicher Hilfen); hier wie auch bei den vier anderen Experimentalprogrammen in den USA mußten zudem Null-Effekte festgestellt werden, durch die - unter konservativer Regierung - ein Teil der amerikanischen Wohlfahrtspolitik des "Kampfes gegen die Armut" aus den 60er Jahren politisch infrage gestellt

wurde, obwohl nicht auszuschließen war, daß die Forschungsergebnisse methodische Artefakte darstellten (vgl. Rossi, P. and Wright, S.R. 1977).

IV. Evaluationsforschung der zweiten Generation

Wie man bei den Programmbudgets inzwischen mehrere Generationen unterscheiden kann (vgl. Rürup, B. und Färber, G. 1980), so hat auch die Evaluationsforschung aus den Erfahrungen gelernt. Zum einen versucht die Evaluationsforschung stärker schon auf die Programmgestaltung einzuwirken, um schon in dieser Phase des Politik-Zyklus für die methodischen Voraussetzungen der Erfolgskontrolle zu sorgen. Offenbar steigen die Chancen hierfür, wenn die Evaluationsfunktion organisatorisch in Spezialeinheiten ausdifferenziert ist. Zum anderen hat man den Schluß gezogen, daß bei Programmen, die nicht auf diese Weise methodisch vorbereitet worden sind, in einer Vorstudie zunächst die Evaluierbarkeit geprüft werden sollte (Wholey, J. 1976). Methodische Probleme selbst mit Experimentalprogrammen führten zu einer Rückbesinnung auf qualitative Verfahren der Datenerhebung. Zudem wird seit einigen Jahren dem Implementationsprozeß verstärkte Aufmerksamkeit geschenkt, weil Störungen in diesem Bereich viel früher erkennbar sind als bei erst langfristig auftretenden Effekten in der sozio-ökonomischen Umwelt und somit den Programmverantwortlichen rechtzeitige Korrekturmöglichkeiten zudem in Bereichen gegeben werden, die unmittelbar zu steuern sind. Allerdings droht die Implementationsforschung in der Bundesrepublik sich zu verselbständigen und Wirkungsfragen auszublenden; auf die Effektivierung des Verwaltungsvollzuges von theoretisch fehlerhaften Programmen kann es aber wohl kaum ankommen. Der Tendenz zur frühzeitigen Einschaltung der Politikberatung und Programmforschung entspricht es auch, wenn Gesetze zunächst einem *Praxistest* unterzogen werden (Böhret, C. 1985), bevor sie verbindlich gemacht werden. Diese methodischen Umorientierungen sind begleitet von einem Wandel des Rollenverständnisses und des Verbindlichkeitsanspruches der Evaluationsforschung. *Planungspolitisch* wird eingesehen, daß programmatisches Lernen im politisch-administrativen System nicht der Planungslogik folgt, sondern eben auch von Machtgewinn und Machterhaltung motiviert wird. Die *planungslogische* Begründung der Erfolgskontrolle wird insofern etwas zurückgenommen, als die Evaluationsforschung ihrem eigenen Exaktheitsanspruch nicht genügen konnte und daher keinen Rationalitätsvorsprung behaupten konnte.

ten kann; sie wird damit zu nur noch einem, wenngleich wohl nicht unwichtigen, Informationslieferanten des Entscheidungssystems. *Planungsorganisatorisch* impliziert diese Reduktion des hohen Anspruches bei allem Bemühen um den weiteren Ausbau analytischer Kapazitäten die Besinnung auf die Qualität traditioneller Rückkoppelungsmechanismen wie insbesondere der Erfahrungen aus dem Verwaltungsvollzug. *Haushaltspolitisch* schließlich ist eine Rationalisierung der Mittelallokation unter den gegebenen Bedingungen der Haushaltsplanung, d.h. der fehlenden systematischen Verzahnung mit der programmatischen Seite des Budgets, auch "durch die Hintertür" kaum zu erwarten.

Literaturverzeichnis

- (1) Adams, Bruce and Sherman, Betsy (1978),
Sunset Implementation: A Positive Partnership to Make Government Work, in: Public Administration Review, Vol. 38, S. 78-81
- (2) Böhret, Carl (1985),
Konsolidierung durch effektive Normsetzung: Gesetz auf Zeit und Gesetzestest, in: Derlien, Hans-Ulrich (Hrsg.), Programmforschung unter den Bedingungen einer Konsolidierungspolitik, München, S. 139-170
- (3) Bohne, Eberhard und König, Herbert (1976),
Probleme der politischen Erfolgskontrolle, in: Die Verwaltung, Bd. 9, S. 19-38
- (4) Bothun, Douglas and Comer, John C. (1979),
The Politics of Termination: Concepts and Process, in: Policy Studies Journal, Vol. 7, S. 540-553
- (5) Braybrooke, David and Lindblom, Charles E. (1963),
A Strategy of Decision. Policy Evaluation as a Social Process, New York
- (6) Cook, Thomas D. and Gruder, Charles (1979),
Metaevaluation Research, in: Evaluation Studies Review Annual, Vol. 4, S. 469-515
- (7) Derlien, Hans-Ulrich (1975),
Das Berichtswesen der Bundesregierung - ein Mittel der Kontrolle und Planung, in: Zeitschrift für Parlamentsfragen, Bd. 8, S. 42-47
- (8) Derlien, Hans-Ulrich (1976),
Die Erfolgskontrolle staatlicher Planung. Eine empirische Untersuchung über Organisation, Methode und Politik der Programmevaluation, Baden-Baden

- (9) Derlien, Hans-Ulrich (1978),
Organisatorische Aspekte der Programmevaluation, in: Verwaltung und Fortbildung, Bd. 6, S. 51-61
- (10) Derlien, Hans-Ulrich (1979),
Die Effizienz von Entscheidungsinstrumenten für die staatliche Ressourcenallokation - Versuch einer Evaluation von Entscheidungstechniken, in: Pfohl, Hans-Christian und Bert Rürup (Hrsg.), Anwendungsprobleme moderner Planungs- und Entscheidungstechniken, Königstein S. 311-326
- (11) Derlien, Hans-Ulrich (1981),
Stand und Entwicklung der Programmforschung in der öffentlichen Verwaltung, in: Derlien, Hans-Ulrich (Hrsg.), Programmforschung in der öffentlichen Verwaltung, München, S. 5-45
- (12) Derlien, Hans-Ulrich (1984),
Programme Evaluation in the Federal Republic of Germany, European Group of Public Administration, Occasional Paper, Brüssel.
- (13) Derlien, Hans-Ulrich (1984a),
Implementationsprobleme: Bürokratische Ineffizienz oder politische Programmfehler?, in: Verwaltungsarchiv, Bd. 75, S. 256-271
- (14) Derlien, Hans-Ulrich (1985),
Planungsinstrumentarium und Planungsorganisation im Kontext expansiver und kontraktiver Haushalte, in: Derlien, Hans-Ulrich (Hrsg.), Programmforschung unter den Bedingungen einer Konsolidierungspolitik, München, S. 5-16
- (15) Hellstern, Gerd-Michael und Wollmann, Helmut (Hrsg.) (1983),
Experimentelle Politik - Reformstrohfeuer oder Lernstrategie. Bestandsaufnahme und Evaluierung, Opladen
- (16) Hellstern, Gerd-Michael und Wollmann, Helmut (Hrsg.) (1984),
Handbuch zur Evaluierungsforschung. Band 1, Opladen
- (17) Hellstern, Gerd-Michael und Wollmann, Helmut (1984a),
Evaluierung und Evaluierungsforschung - ein Entwicklungsbericht, in: Hellstern, G.-M. und Wollmann, H., S. 17-93
- (18) Hellstern, Gerd-Michael und Wollmann, Helmut (1984b),
Evaluierung auf der kommunalen Ebene, in: Hellstern, G.-M. und Wollmann, H., S. 491-523
- (19) Hugger, Werner (1979),
Legislative Effektivitätssteigerung: Von den Grenzen der Gesetzesevaluierbarkeit zum Gesetz auf Zeit, in: Politische Vierteljahresschrift, Bd. 20, S. 202-220
- (20) König, Klaus (1986),
Zur Evaluation staatlicher Programme, in: Eichhorn, Peter und v. Kortzfleisch, Gert (Hrsg.), Erfolgskontrolle bei der Verausgabung öffentlicher Mittel, Baden-Baden, S. 19-34

- (21) Lindblom, Charles E. (1969),
The Science of "Muddling Through", in: Etzioni, Amitai (ed.), Readings on Modern Organization, Englewood Cliffs, S. 154-166
- (22) Marvin, Keith M. und Cohen, Wallace M. (1984),
Rechenschaftsprüfung und Evaluierung im General Accounting Office, in: Hellstern, G.-M. und Wollmann, H., S. 270-485
- (23) Pfitzer, Christine (1985),
Integrierte Entwicklungsplanung als Allokationsinstrument auf Landesebene, Bern
- (24) Popper, Karl R. (1961),
The Poverty of Historicism, London 1961
- (25) Raschert, Jürgen (1981),
Probleme der Evaluation von Modellversuchen am Beispiel des Gesamtschulmodells in Nordrhein-Westfalen in: Derlien, Hans-Ulrich (Hrsg.), Probleme der Methodik und der Rezeption von Programmforschung, München, S. 55-72
- (26) Reding, Kurt (1981),
Die Effizienz staatlicher Aktivitäten. Probleme ihrer Messung und Kontrolle, Baden-Baden
- (27) Rossi, Peter and Wright, Sonja R. (1977),
Evaluation Research. An Assessment of Theory, Practice, and Politics, in: Evaluation Quarterly Vol 1, S. 5-52
- (28) Rürup, Bert und Färber, Gisela (1980),
Programmhaushalte der "zweiten Generation", in: Die öffentliche Verwaltung, Bd. 33, S. 661-672
- (29) Rürup, Bert und Färber, Gisela (1985),
Konzeptioneller Wandel von integrierten Aufgaben- und Finanzplanungssystemen, in: Derlien, Hans-Ulrich (Hrsg.), Programmforschung unter den Bedingungen einer Konsolidierungspolitik, München, S. 17-42
- (30) Schatz, Heribert (1973),
Auf der Suche nach neuen Problemlösungsstrategien: die Entwicklung der politischen Planung auf Bundesebene, in: Mayntz, Renate und Scharpf, Fritz W. (Hrsg.), Planungsorganisation, München, S. 9-67
- (31) Tenbruck, Friedrich (1972),
Zur Kritik der planenden Vernunft, Freiburg u. München
- (32) Weiss, Carol, H. and Bucuvalas, M.J. (1980),
Truth Tests and Utility Tests: decision-makers' frame of reference for social science research, in: American Sociological Review, Vol. 45, S. 302-315

- (33) Wholey, Joseph S. (1976),
The role of Evaluation and the Evaluator in Improving Public Programs:
The Bad News, the Goods News, and a Bicentennial Challenge, in: Public
Administration Review Vol. 36, S. 679-683
- (34) Wildavsky, Aaron (1979),
Speaking Truth to Power. The Art and Craft of Policy Analysis, Boston,
Toronto
- (35) Wille, Eberhard (1980),
Finanz- und Aufgabenplanung, in: HdWW Bd. 5, S. 591-620
- (36) Wittrock, Karl (1984),
Im Gespräch, in: Hellstern, G.-M. und Wollmann H., S. 486-490
- (37) Zavelberg, Heinz-Günter (1986),
Staatliche Rechnungsprüfung und Erfolgskontrolle. Möglichkeiten und
Grenzen, in: Eichhorn, Peter und v. Kortzfleisch, Gert (Hrsg.), Erfolgskon-
trolle bei der Verausgabung öffentlicher Mittel, Baden-Baden 1986, S. 103-
120
- (38) Zeh, Wolfgang (1984),
Wille und Wirkung der Gesetze, Heidelberg

Summary

This contribution gives an overview on the state of the art of programme evaluation in the Federal Republic of Germany. After having explained the reasons why evaluation as a new technique for feeding back information into policy development was claimed, since policy planning became institutionalized, different forms of organizational arrangements for evaluation in federal government are described. Methodological and political obstacles, which are dealt with, can also account for the fragmented institutionalization and have led to lowering the level of aspiration with respect to the political impact of evaluation research as well as to modifying some methodological devices.