



Reinhard Stockmann (Hrsg.)

Evaluationsforschung

Grundlagen und ausgewählte Forschungsfelder

3. Auflage

WAXMANN

Sozialwissenschaftliche Evaluationsforschung

Herausgegeben von
Reinhard Stockmann

Band 1



Waxmann Münster / New York
München / Berlin

Reinhard Stockmann (Hrsg.)

Evaluationsforschung

Grundlagen und
ausgewählte Forschungsfelder

3. Auflage



Waxmann Münster / New York
München / Berlin

Bibliografische Informationen Der Deutschen Bibliothek

Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.ddb.de> abrufbar.

Sozialwissenschaftliche Evaluationsforschung, Band 1

ISSN 1861-244X

ISBN-10 3-8309-1734-1

ISBN-13 978-3-8309-1734-2

© Waxmann Verlag GmbH, Münster 2006

www.waxmann.com

info@waxmann.com

Umschlaggestaltung: Pleßmann Kommunikationsdesign, Ascheberg

Umschlagbild: Ulrich Thul, Ludwigsburg

Satz: Stoddart Satz- und Layoutservice, Münster

Druck: Hubert & Co., Göttingen

Gedruckt auf alterungsbeständigem Papier, DIN 6738

Alle Rechte vorbehalten

Printed in Germany

Inhalt

Vorwort zur 3. Auflage	7
Vorwort zur 2. Auflage	9
Vorwort zur 1. Auflage	11

I. Grundlagen der Evaluationsforschung

<i>Reinhard Stockmann</i> Evaluation in Deutschland	15
--	----

<i>Donna M. Mertens</i> Institutionalizing Evaluation in the United States of America	47
--	----

<i>Frans L. Leeuw</i> Evaluation in Europe	64
---	----

<i>Thomas Widmer</i> Qualität der Evaluation – Wenn Wissenschaft zur praktischen Kunst wird	85
---	----

<i>Evert Vedung</i> Evaluation Research and Fundamental Research	113
---	-----

<i>Barbara Lee</i> Theories of Evaluation	137
--	-----

<i>Valerie J. Caracelli</i> Methodology: Building Bridges to Knowledge	177
---	-----

II. Ausgewählte Felder der Evaluationsforschung

Hellmut Wollmann

Evaluation und Verwaltungspolitik
Konzepte und Praxis in Deutschland und
im internationalen Kontext.....207

Helmut Kromrey

Qualität und Evaluation im System Hochschule234

Xaver Büeler

Qualitätsevaluation und Schulentwicklung.....260

Stefan Kuhlmann

Evaluation in der Forschungs- und Innovationspolitik.....289

Bettina Bangel, Christian Brinkmann & Axel Deeke

Evaluation von Arbeitsmarktpolitik.....311

Joseph Huber und Axel Müller

Zur Evaluation von Umweltschutz-Maßnahmen in
Staat und Unternehmen345

Reinhard Stockmann

Evaluation staatlicher Entwicklungspolitik378

Autorenverzeichnis415

Vorwort zur 3. Auflage

Seit dem Erscheinen der ersten Auflage dieses Buches im Jahr 2000 hat die Professionalisierung der Evaluation in Deutschland enorme Fortschritte gemacht. Die Gesellschaft für Evaluation (DeGEval) ist zur Heimat vieler geworden, die sich als Auftraggeber oder als Evaluatoren mit Evaluation beschäftigen (www.degeval.de).

Die Zeitschrift für Evaluation (www.zfev.de) ist fest etabliert und erfreut sich weiter ansteigender Abonentenzahlen. Das Centrum für Evaluation (www.ceval.de) ist gerade einer externen Evaluation unterzogen worden, in der seine besondere Bedeutung für die deutsche Evaluationslandschaft betont wurde. Der berufsbegleitende Studiengang „Master of Evaluation“ an der Universität des Saarlandes (www.master-evaluation.de) wird international nachgefragt und scheint sich zu einer Keimzelle einer professionalisierteren Evaluation zu entwickeln. In Kürze dürften auch an anderen Hochschulen in Deutschland konkurrierende Angebote entstehen. Die Angebotsseite der Evaluation hat in den letzten Jahren deutlich an Qualität gewonnen.

Die Nachfrage nach Evaluation ist ebenfalls weiterhin ungebrochen und bietet für Evaluatoren gute Forschungs- und Beschäftigungsmöglichkeiten. Allerdings sind auch problematische Entwicklungen festzustellen, die sich insbesondere auf die Qualität von Evaluation auswirken können.

Während international und auch auf Deutschland bezogen von einer dritten Welle der Evaluation in der Nachkriegszeit gesprochen wird, bleibt Deutschland in einigen Politikfeldern deutlich zurück, nicht unbedingt, was die Quantität an Evaluation angeht, aber ihre Qualität.

In einem Sammelwerk von Sabine Kuhlmann, Jörg Bogumil und Hellmut Wollmann (2004) zur Leistungsmessung in Politik und Verwaltung wird konstatiert, dass die dritte Welle der Evaluation die deutsche Politik- und Verwaltungswelt nur in vergleichbar geringem Umfang erfasst hat (S. 41). Im Umweltsektor spielen Evaluationen kaum (mehr) eine Rolle, nachdem Ministerien und die Stiftungen hier nahezu jedes Interesse verloren zu haben scheinen. Im Arbeitsmarktbereich wird zwar heftig evaluiert, doch bei genauerer Betrachtung erkennt man, dass es sich dabei hauptsächlich um ökonometrische Analysen handelt.

In der deutschen Entwicklungszusammenarbeit ist die Evaluation seit über drei Jahrzehnten ein fester Bestandteil der Projekt- und Programmsteuerung. Doch konzeptionell-methodisch sind im Vergleich zur internationalen Entwicklung nur wenig Fortschritte zu erkennen. Aufwändige methodische Studien, z.B. für Länderprogramm- oder Sektorevaluationen, zur Messung der Zielerreichung bei den Millennium Goals etc. sucht man in Deutschland vergebens. Intensive Versuche, die methodische Qualität zu verbessern und den Standards für Evaluation zur Geltung zu verhelfen, treffen allzu oft auf Unverständnis.

Im Schul- und Hochschulbereich ist der Evaluationsboom ungebrochen. Doch auch in diesem Feld darf über die Qualität, z.B. der im Schulbereich etablierten Evaluationssysteme, zumindest gestritten werden. Wenn dann zudem wichtige Auftraggeber, die ihre Evaluationen immerhin ausschreiben, erklären, dass mehr als 50% der für die Auftragsvergabe zählenden Kriterienpunkte vom Angebotspreis bestimmt werden, dann fördert dies sicherlich nicht die Qualität der Studien. Manchmal scheint die Devise vorzuherrschen, eine konzeptionell und methodisch schwache Studie mit guten Evaluationsergebnissen ist allemal besser als eine gute Studie mit schlechten Ergebnissen. Hier ist noch viel Aufklärungsarbeit nötig. Es ist zu wünschen, dass dieses Buch, das sich aus Sicht des Herausgebers und der Autoren ungebrochener Beliebtheit erfreut, einen Beitrag dazu leistet.

Bürrstadt und Saarbrücken, 7. Juli 2006

Reinhard Stockmann

Vorwort zur 2. Auflage

Die Professionalisierung der Evaluation in Deutschland hat in den letzten Jahren einen enormen Entwicklungsschub erfahren. Die Mitgliederzahl der Deutschen Gesellschaft für Evaluation (DeGEval) ist auf über 300 (davon fast 50 institutionelle Mitglieder) angewachsen. Die von der Gesellschaft veranstalteten Jahrestagungen verzeichnen von Jahr zu Jahr steigende Teilnehmerzahlen. Mit der Entwicklung von Evaluationsstandards gibt es erstmals auch in Deutschland einen Satz von Qualitätskriterien, an denen sich Auftraggeber, Evaluatoren und die Öffentlichkeit orientieren können. Mittlerweile ist auch eine deutschsprachige ‚Zeitschrift für Evaluation‘ (ZfEv) gegründet worden, die die DeGEval in dem Bemühen unterstützen möchte, den Austausch über Theorien, Methoden und Kontextbedingungen von Evaluation zwischen Fächern und zwischen Wissenschaft und Praxis zu fördern. Interessant ist in diesem Zusammenhang die Gründung eines ‚Centrums für Evaluation (CEval)‘ an der Universität des Saarlandes, das Theorien und Methoden der Evaluation weiter entwickelt, bei der Planung und Durchführung von Evaluationen berät, selbst Evaluationen durchführt und Aus- und Weiterbildung anbietet.

Gerade im Bereich der Aus- und Weiterbildung hat es bisher kaum umfassende Angebote gegeben. An der Universität Bern wurde erstmals 2002/2003 ein postgraduales ‚Ergänzungsstudium Evaluation‘ angeboten. In Deutschland führten 2003 die ‚Arbeitsgemeinschaft Entwicklungspolitischer Gutachter‘ (AGEG) und das CEval als wissenschaftliche Evaluationseinrichtung gemeinsam ein einjähriges Fortbildungsprogramm für Evaluatoren in der Entwicklungszusammenarbeit (FEEZ) durch. Beide Programme sollen auf Grund der hohen Nachfrage erneut angeboten werden. An der Universität des Saarlandes, die sich zunehmend zu einem Zentrum der Evaluationsforschung in Deutschland entwickelt, wird darüber hinaus ein Masterstudiengang in Evaluation eingerichtet. In diesem ersten Evaluationsstudiengang Europas wird es ab 2004 möglich sein, nach einem viersemestrigen Vollzeitstudium den akademischen Grad eines ‚Master of Evaluation‘ zu erwerben.

Diese Entwicklung der letzten Jahre deutet darauf hin, daß die Evaluation in Deutschland auf die gestiegene Nachfrage und die erhöhten Qualitätsansprüche mit einer Qualifizierungs- und Professionalisierungsoffensive reagiert. Doch

bis zu einer eigenständigen sozialwissen-schaftlichen Evaluationsdisziplin ist der Weg noch weit. Auch von der Existenz einer Evaluationskultur kann in Deutschland nicht gesprochen werden. Evaluation ist noch längst nicht zu einem Regelinstrument der politischen Steuerung, der Qualitätssicherung oder gar der politischen Aufklärung geworden. Ziel dieses Buches ist es deshalb, zu einem wachsenden Verständnis für die Ziele, Theorien, Methoden, spezifischen Handlungsbedingungen und Anwendungsfelder der Evaluation beizutragen. Es richtet sich insbesondere an Auftraggeber und Nutzer von Evaluation sowie an Evaluatoren und solche, die es werden wollen.

Zu danken habe ich allen Autoren und Autorinnen dieses Bandes, die ihre Beiträge überarbeitet und aktualisiert haben sowie Angelika Nentwig und Nicolà Reade für Ihre Unterstützung bei der Fertigstellung des Buches.

Bürstadt und Saarbrücken, den 7. Juli 2003

Reinhard Stockmann

Vorwort zur 1. Auflage

Evaluation in Deutschland bietet ein sehr vielfältiges und buntes Bild. In vielen Politikfeldern wird aus den verschiedensten Motiven und Gründen und mit den unterschiedlichsten Ansätzen und Methoden evaluiert. Dennoch wird der Erfolg und die Wirksamkeit der weit überwiegenden Mehrzahl von politischen und sozialen Interventionsprogrammen überhaupt nicht eruiert. Wiederholt hat der Bundesrechnungshof festgestellt, daß in Deutschland nur selten Erfolgskontrollen durchgeführt werden, daß in fast allen Ressorts dafür die Voraussetzungen fehlen und daß in den wenigen durchgeführten Evaluationen oft die vorhandenen methodischen Möglichkeiten zur Ermittlung von Erfolg und Wirksamkeit nicht ausgeschöpft werden.

So bunt wie das Bild der Evaluation ist auch die Gilde derer, die Evaluationen durchführen. Im Unterschied zur USA konnte sich in Deutschland bisher keine sozialwissenschaftliche Evaluationsdisziplin durchsetzen. Die Evaluationsforschung in Deutschland ist fachlich zersplittert, weist nur eine geringe interdisziplinäre Kommunikation und Zusammenarbeit auf und ist noch immer mit dem Makel anwendungsbezogener Auftragsforschung behaftet, deren Wissenschaftlichkeit teilweise bestritten wird.

Allerdings lassen sich erste Professionalisierungserfolge erkennen. Im September 1997 wurde die Deutsche Gesellschaft für Evaluation gegründet, die es sich zur Aufgabe gemacht hat, das Verständnis, die Akzeptanz und die Nutzbarmachung von Evaluation in der Gesellschaft zu fördern, Prinzipien und Qualitätsstandards für Evaluationen zu entwickeln und den interdisziplinären Austausch zu unterstützen. Auch ist eine vermehrte Zahl von Fachpublikationen und ein verstärktes Interesse an Aus- und Weiterbildungsangeboten zu beobachten.

Doch nach wie vor fehlen Lehrstühle, Sonderforschungsbereiche oder wissenschaftliche Institute, die sich mit grundlegenden theoretischen und methodischen Fragen der Evaluationsforschung beschäftigen. Studiengänge, in denen das „Handwerk“ des Evaluierens erlernt werden könnte, gibt es nicht. Auch ein außeruniversitärer Aus- und Weiterbildungsmarkt für Evaluatoren ist kaum existent. Bisher gibt es keine deutschsprachige Evaluationszeitschrift und es fehlt an einschlägigen Sammelwerken oder Lehrbüchern zur Evalua-

tionsforschung, die die fachliche Zersplitterung überwinden und das bisher zusammengetragene Wissen bündeln.

Mit dieser im Leske & Budrich Verlag aufgelegten Reihe zur sozialwissenschaftlichen Evaluationsforschung, in der das vorliegende Buch den ersten Band stellt, soll ein Beitrag dazu geleistet werden, diese Lücke etwas zu füllen.

Der Themenband ist so aufgebaut, daß zuerst die Evaluationsentwicklung in Deutschland, in den USA und in Europa charakterisiert wird, bevor grundsätzliche Evaluationsthemen wie die Etablierung gemeinsamer Evaluationsstandards, die Differenz zwischen sozialwissenschaftlicher Grundlagenforschung und anwendungsorientierter Evaluationsforschung sowie Theorie- und Methodenentwicklung in der Evaluationsforschung behandelt werden. Im zweiten Teil des Bandes werden ausgewählte Felder der Evaluationsforschung untersucht.

Als Herausgeber dieses Bandes freut es mich außerordentlich, daß es gelungen ist, auf internationaler, europäischer und nationaler Ebene führende Evaluationsforscher als Autoren und Autorinnen zu gewinnen. Ihnen gilt mein besonderer Dank!

Herzlich zu danken habe ich auch Angelika Nentwig, die sehr freundlich und mit unerschöpflicher Geduld den Kontakt zu den Autoren und Autorinnen gehalten und dafür gesorgt hat, daß alle Beiträge rechtzeitig eingegangen sind.

Ein Dankeschön sage ich auch Nicolà Reade und Anette Becker, die die verschiedenen Texte in eine einheitliche Form brachten.

Bürrstadt und Saarbrücken, den 7. Juli 2000

Reinhard Stockmann

I.
Grundlagen der
Evaluationsforschung

Evaluation in Deutschland

Ziel dieses ersten Beitrags ist es, eine kurze Einführung in den Forschungsbereich der Evaluation zu geben. Dabei steht die Entwicklung in Deutschland im Kontext internationaler und europäischer Strömungen sowie im Rahmen nationaler Besonderheiten im Vordergrund. Zudem werden die Themen abgesteckt, die in den weiteren Beiträgen dieses Sammelbandes behandelt werden. Abschließend werden einige Professionalisierungsdefizite in Deutschland aufgezeigt und Vorschläge zu deren Beseitigung unterbreitet.

1. Evaluation, was ist das?

Auch wenn den meisten Menschen der Begriff „Evaluierung“ oder „Evaluation“ wenig vertraut sein dürfte¹, ist die Tätigkeit, die damit gemeint ist, so alt wie die Menschheit selbst. Wenn jemand z.B. überprüft, ob gebratenes Fleisch besser schmeckt und bekömmlicher ist als rohes oder wenn jemand ausprobiert, ob bestimmte Pilze oder Pflanzen essbar sind oder nicht, dann führt er eine Evaluation durch. Evaluationen stellen ein wichtiges Instrument zur Generierung von Erfahrungswissen dar. Sie werden durchgeführt, indem Informationen gesammelt und anschließend bewertet werden, um letztlich Entscheidungen zu treffen. Die dabei verwendeten Bewertungskriterien können sehr verschieden sein, orientieren sich jedoch sehr oft am Nutzen eines Gegenstandes, Sachverhaltes oder Entwicklungsprozesses für bestimmte Personen oder Gruppen. Je nach Kriterienauswahl kann die Nutzenbewertung durch einzelne Personen oder Gruppen dementsprechend sehr unterschiedlich ausfallen.

Deshalb kommt es sehr darauf an, wer eine Evaluation durchführt, welche Ziele von wem damit verfolgt werden, welche Vorgehensweisen gewählt und welche Methoden angewendet werden. Damit sind schon einige wichtige

1 Die Begriffe „Evaluierung“, „Evaluation“ und „Evaluationsforschung“ werden hier synonym verwendet. Zu definitorischen Unterschieden vgl. Wottawa & Thierau (1998: siehe Literaturliste)

Fragen umrissen, mit denen sich Evaluationen auseinandersetzen müssen und die in diesem Band behandelt werden.

Wissenschaftliche Evaluationen unterscheiden sich von Alltagsevaluationen unter anderem dadurch, daß Bewertungen nicht willkürlich, sondern anhand explizit auf den zu evaluierenden Sachverhalt und anhand präzise festgelegter Kriterien vorgenommen werden. Ein weiteres Unterscheidungsmerkmal besteht darin, daß wissenschaftliche Evaluationen von ‚Experten‘ durchgeführt werden, die für die Datenerhebung systematisch objektivierende Verfahren anwenden. Hierfür kann die gesamte Bandbreite sozialwissenschaftlicher Forschungsmethoden genutzt werden und es gelten die in der Wissenschaft grundlegenden Regeln für die Sammlung valider und reliabler Daten (vgl. Rossi u.a. 1988: 1ff.; Wottawa/Thierau 1998: 9f.; Kromrey 1995: 314f.).

Evaluationsforschung ist ein Teilgebiet der angewandten Sozialforschung und will deshalb wie diese zur Lösung praktischer, gesellschaftspolitischer Probleme beitragen, in dem sie gezielt und systematisch Grundlagen für außerwissenschaftliche Entscheidungsprozesse bereitzustellen versucht (vgl. Clemens 2000: 215). Von der Grundlagenforschung unterscheidet sich die angewandte Sozialforschung und damit auch die Evaluationsforschung in mehrfacher Hinsicht. *Evert Vedung*, der in dem vorliegenden Band diesen Unterschieden nachgeht, hebt hervor: „the basic difference between evaluation research and fundamental research is that the former is intended for use“. Während Grundlagenforschung relativ zweckfrei nach Erkenntnissen streben kann, hat Evaluationsforschung zumeist einen Auftraggeber, der damit bestimmte Absichten verfolgt². Evaluation hat sich deshalb an den extern gesetzten Zielbestimmungen zu orientieren. Die Definition und Präzisierung ihrer Fragestellungen richtet sich an außerwissenschaftlichen Erkenntnisinteressen und Verwertungskontexten aus (vgl. Kromrey 2003: 96f.).

Ein weiterer wesentlicher Unterschied zwischen Evaluations- und Grundlagenforschung besteht darin, daß Evaluationen immer mit Bewertungen verbunden sind. Hierfür werden die Beurteilungskriterien zumeist aus dem zu evaluierenden Programm selbst abgeleitet. In diesem Fall werden die Implementation des Programms sowie seine Wirkungen im Lichte seiner eigenen Ziele bewertet. Kromrey (1995: 320) weist darauf hin, daß es sich dabei jedoch nicht um subjektive Werturteile des Evaluationsforschers handelt, sondern um ‚technologische Einschätzungen‘, die intersubjektiv nachprüfbar sein müssen. Dabei geht der Forscher in der Regel so vor, daß er empirisch die Zustände in den Zielbereichen des Programms vor und nach den ergriffenen Maßnahmen feststellt und untersucht, welche Veränderungen auf welche Elemente des Programms zurückzuführen sind. Durch Vergleich der empirischen Fakten (Ist-Zustand) mit den im Programm formulierten Zielniveaus (Soll-Zustand) können anschließend deduktiv Aussagen darüber abgeleitet werden, ob und in

2 Natürlich kann Evaluationsforschung auch als akademische Forschung mit eigener Zwecksetzung betrieben werden. Doch dies ist eher die Ausnahme.

welchen Teilen das durchgeführte Programm erfolgreich bzw. nicht erfolgreich war. Um zu einer abschließenden Bewertung zu kommen, sind allerdings auch nicht-intendierte Wirkungen zu berücksichtigen. Darüber hinaus können zudem andere oder weitere Bewertungskriterien ergänzend herangezogen werden, die z.B. von Auftraggebern oder den betroffenen Zielgruppen formuliert werden (z.B. Politikrelevanz, Nutzen für die Zielgruppen).

Keine prinzipiellen Unterschiede sind hingegen zwischen Evaluations- und Grundlagenforschung im Hinblick auf die Auswahl des Untersuchungsgegenstandes sowie die Verwendung von Datenerhebungs- und Analysemethoden zur Identifizierung von Wirkungen und der Bearbeitung der Kausalitätsfrage (Ursache-Wirkungszusammenhänge) zu erkennen.

Gegenstände einer Evaluation können z.B. Personen, Organisationen, Produkte, Reformen, Gesetze, Maßnahmen, Projekte, Programme oder gar Evaluationen selbst sein. Da mit Evaluationen sehr unterschiedliche Ziele, Erkenntnis- und Verwertungsinteressen verfolgt werden, werden auch sehr unterschiedliche Ansätze, Modelle und Evaluationsdesigns verwendet. Dementsprechend gibt es auch eine Vielfalt unterschiedlicher Definitionen.

Eine Definition, die sowohl auf die verschiedenen Evaluationsgegenstände paßt, als auch auf die besondere Aufgaben- und Zielbestimmung der Evaluations- im Unterschied zur Grundlagenforschung hinweist, schlägt Donna Mertens vor:

“Evaluation is the systematic investigation of the merit or worth of an object (program) for the purpose of reducing uncertainty in decision making.” (Mertens 1998: 219)

Eine ebenfalls weithin akzeptierte Begriffsbestimmung ist die von Peter Rossi/Howard Freeman, die Evaluationsforschung definieren

„als systematische Anwendung sozialwissenschaftlicher Forschungsmethoden zur Beurteilung der Konzeption, Ausgestaltung, Umsetzung und des Nutzens sozialer Interventionsprogramme. Evaluationsforschung bezeichnet den gezielten Einsatz sozialwissenschaftlicher Forschungsmethoden zur Verbesserung der Planung und laufenden Überwachung sowie zur Bestimmung der Effektivität und Effizienz von (...) sozialen Interventionsmaßnahmen.“ (Rossi u.a. 1988: 3)³

Die Evaluationsforschung kann sich mit drei Phasen des politischen Prozesses beschäftigen und dabei drei verschiedene Analyseperspektiven einnehmen (vgl. Tabelle 1):

1. Die Evaluationsforschung kann die Phase der Programmentwicklung behandeln, die die Konzeptualisierung und Ausarbeitung einer geplanten Intervention mit einschließt. Dabei hat die Evaluationsforschung die Aufgabe, „die materiellen, personellen, institutionellen, finanziellen, theoretischen

3 Sinngemäße Definitionen finden sich z.B. bei Wittmann (1985:17); Wottawa/Thierau (1998: 9).

schen Rahmen- bzw. Eingangsbedingungen eines Programms“ zu untersuchen, um zur Erstellung eines Programmdesigns beizutragen (vgl. Brandtstädter 1990: 217). Dabei sollen möglichst schon frühzeitig negative Effekte eines Programms oder einer Maßnahme abgeschätzt werden (vgl. Hellstern/Wollmann 1980: 13, 1984: 24; Staudt u.a. 1988: 40f.). Solche Untersuchungen werden „ex-ante evaluations“, „input-evaluations“ oder „preformative evaluations“ (Scriven 1991: 169) genannt.

2. Während der Implementationsphase übernimmt die Evaluationsforschung vor allem Kontroll- und Beratungsfunktionen. Indem Informationen über den Programmverlauf und die Programmresultate gesammelt und bewertet werden, sollen Entscheidungshilfen für die Steuerung der Durchführung des Programms gegeben und durch eine frühzeitige Korrekturmöglichkeit Veränderungen des Programmdesigns ermöglicht werden (vgl. Staudt u.a. 1988: 27; Rossi u.a. 1988: 12, 31 u. 63; Wottawa/Thierau 1998: 54). Solche Evaluationen zur Überwachung der Umsetzung und Ausführung von laufenden Programmen („on-going“) werden als „Begleitforschung“ (Rossi u.a. 1988: 11) oder als „formative Evaluationen“ bezeichnet (vgl. Scriven 1980, 1991). Sie beschäftigen sich mit derselben Phase des politischen Prozesses wie die Implementationsforschung⁴ und verfolgen dabei ähnliche Zielsetzungen. Begleitforschung und formative Evaluation müssen nicht identisch sein. Auch summativ Evaluation kann ihre Daten im Rahmen des „Monitoring“ per Begleitforschung sammeln. Formative Evaluation und Implementationsforschung müssen zudem nicht ähnliche Zielsetzungen verfolgen, tun es aber oft. So kann Implementationsforschung auch reinen Kontrollzwecken dienen.
3. Nach Abschluß der Implementation eines Programms kommt der Evaluationsforschung die Aufgabe zu, möglichst umfassend die Effekte, die durch ein Programm ausgelöst wurden, zu erfassen und zu bewerten sowie Zusammenhänge aufzudecken (vgl. Wottawa/Thierau 1998: 55; Scriven 1991: 340). Solche Untersuchungen werden als „summative Evaluationen“ bezeichnet, und in der Regel ex-post durchgeführt.⁵

4 Zur Implementationsforschung vgl. u.a. Mayntz (1977, 1980, 1983); Windhoff-Héritier (1980, 1983, 1987); Wollmann (1980, 1994).

5 Die terminologische Unterscheidung zwischen formativer und summativer Evaluation geht auf Scriven (1967, 1972a u. b, 1980, 1983, 1991) zurück. Eine aufschlußreiche Gegenüberstellung der Merkmale von formativer und summativer Evaluation geben Wottawa/Thierau (1998: 56). Vgl. auch Rossi u.a. (1988), Will u.a. (1987: 20ff.).

Tabelle 1: Dimensionen der Evaluationsforschung

Phasen des politischen Prozesses	Analyseperspektive	Erkenntnisinteresse	Evaluationskonzepte
Programmformulierung/ Planungsphase	ex-ante	„analysis for policy“ „science for action“	preformativ/ formativ: aktiv gestaltend, prozeßorientiert, konstruktiv
Implementationsphase	on-going	beides möglich	formativ/summativ: beides möglich
Wirkungsphase	ex-post	„analysis of policy“ „science for knowledge“	summativ: zusammenfassend, bilanzierend, ergebnisorientiert

Demnach kann das Erkenntnisinteresse in der Evaluationsforschung (1.) mehr darauf gerichtet sein, Programme und den Prozeß ihrer Durchführung und Wirkungsweise zu verbessern, oder (2.) mehr darauf, die Resultate vollzogener „policies“ zu analysieren. Folgende von Jann (1994: 311) für die Politikfeldanalyse getroffene Unterscheidung läßt sich deshalb auch auf die Evaluationsforschung übertragen:

„Während auf der einen Seite ‚analysis for policy‘ im Vordergrund steht, d.h. die Verbesserung zukünftiger *policies* durch die Anwendung wissenschaftlicher Erkenntnisse und Methoden (*science for action*), ist die zweite Perspektive eher ‚analysis of policy‘, mit Erklärung und Verallgemeinerung als vorrangigen Zielsetzungen (*science for knowledge*).“⁶

Dementsprechend können Evaluationen mehr *formativ*, d.h. aktiv-gestaltend, prozeßorientiert, konstruktiv und kommunikationsfördernd angelegt sein, oder mehr *summativ*, d.h. zusammenfassend, bilanzierend und ergebnisorientiert. Prinzipiell können beide Evaluations-Perspektiven bei allen *Phasen des politischen Prozesses* eingenommen werden. Da es in der Planungs- und Designphase eines Programms jedoch kaum Ansatzpunkte für eine summative Evaluation gibt, kann sie in der Durchführung nur formativen Charakter haben. Während der Durchführungsphase sind sowohl formative als auch summative Evaluationen möglich. Ex-post Analysen sind in der Regel summative Evaluationen, da der Gestaltungsaspekt entfällt. Durch entsprechende informationelle Rückkopplungsschleifen für Folgeprojekte können sie jedoch auch formative Bedeutung gewinnen.

Generell können Evaluationen vier Funktionen zugeschrieben werden (vgl. Schaubild 1). Diese sind natürlich nicht trennscharf, sondern miteinander verbunden. Dennoch macht es für die Ausrichtung einer Evaluation, ihr Design und die angewendeten Methoden sowie für die an Evaluation gerichteten Erwartungen einen gravierenen Unterschied, welche Funktion im Vordergrund steht:

⁶ Hervorhebungen durch Jann (1994: 311).

1. Erkenntnisfunktion

Mit Hilfe von Evaluationen werden entscheidungsrelevante Daten gesammelt. Es soll u.a. festgestellt werden, ob die Maßnahmen die Zielgruppe erreichen, welche Bedarfe die Zielgruppe hat, wie es mit der Akzeptanz des Programmes bestellt ist, ob die Träger in der Lage sind, das Programm effektiv und effizient umzusetzen, wie sich die Rahmenbedingungen verändert haben, welche Kausalbeziehungen bestehen etc. Die durch Evaluationen gesammelten Informationen sollen dann für Steuerungsentscheidungen genutzt werden.

2. Kontrollfunktion

Bei der Beobachtung eines Programms und seiner Wirkungen steht zwar in erster Linie das Interesse im Vordergrund, Defizite zu erkennen, um möglichst rasch steuernd eingreifen zu können. Gleichzeitig gewinnt man natürlich aber auch Informationen, die erkennen lassen, ob alle Beteiligten ihre Aufgaben erfüllen, den eingegangenen Verpflichtungen nachkommen etc., so daß direkt oder indirekt auch eine Form von Kontrolle mit Evaluationen verbunden ist.

3. Dialog-/Lernfunktion

Durch Evaluationen werden Informationen bereitgestellt die den Dialog zwischen verschiedenen ‚Stakeholdern‘ (Mittelgeber, Durchführungsorganisation, Zielgruppen, sonstige Beteiligte und Betroffene) auf eine solide Grundlage stellen. Auf der Basis der ermittelten Ergebnisse kann gemeinsam und für alle transparent bilanziert werden, wie erfolgreich die Zusammenarbeit verlief und wo Defizite auftraten, um daraus Konsequenzen für die Gestaltung der weiteren Zusammenarbeit zu ziehen. Dadurch werden wichtige Voraussetzungen für gemeinsames Lernen geschaffen.

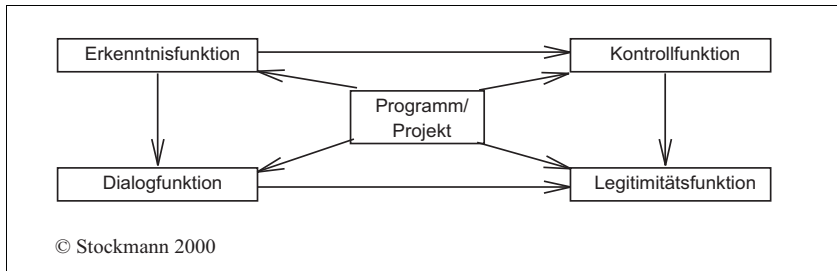
4. Legitimierungsfunktion

Die gewonnene Datenbasis bietet die Möglichkeit, nachprüfbar nachzuweisen, mit welchem Input, welcher Output und welche Wirkungen über die Zeit hinweg (Prozeßperspektive) erzielt wurden. Bei Ex-post Evaluationen läßt sich zudem die Nachhaltigkeit der Programmwirkungen prüfen. Dadurch können Mittelgeber und Durchführungsorganisationen belegen, wie effizient sie mit Finanzmitteln umgegangen sind und welchen Wirkungsgrad ihre Projekte und Programme erreicht haben.

Sehr oft werden Evaluationen auch ‚taktische‘ Funktionen zugeschrieben. Davon wird dann gesprochen, wenn die Ergebnisse von Evaluationen nur dazu verwendet werden sollen, um lediglich bestimmte politische Entscheidungen (manchmal sogar nachträglich) zu legitimieren, z.B. weil ein Programm weitergeführt oder im Gegenteil eingestellt werden soll. Mittlerweile ist es für Politiker auch ‚schick‘ geworden „to use evaluations as baubles or as bolsters“ (Pollitt 1998: 223), als dekorative Symbole für eine ‚moderne‘ Politik, ohne die Ergebnisse von Evaluationen ernsthaft nutzen zu wollen. Diese Art von ‚takti-

scher‘ Funktion läßt sich jedoch kaum mit dem eigentlichen Zweck von Evaluationen vereinbaren und stellt eher ihre pathologische Seite dar. Deshalb wurde sie hier auch nicht in das Aufgabenfeld von Evaluationen mit aufgenommen.

Schaubild 1: Funktionen von Evaluation



Da Evaluationen unterschiedliche Analyseperspektiven einnehmen und unterschiedliche Erkenntnisinteressen verfolgen können, werden auch verschiedene theoretische Ansätze, methodologische Paradigmen und Erhebungsmethoden verwendet. Diese unterscheiden sich prinzipiell nicht von den in der sozialwissenschaftlichen Forschung üblichen Ansätzen und Verfahren. Allerdings gelten *spezielle Anwendungsbedingungen*, da die Evaluationsforschung in einem Kontext konkreten Handelns agiert und ihrer Intention nach nutzungs- und handlungsorientiert ausgerichtet ist (vgl. Weiss 1974: 7 u. 25; Rossi u.a. 1988: 3; Vedung 1999: 12). Wenn sie darauf abzielt, die Entwicklung von Programminitiativen zu unterstützen, die Planung und Umsetzung sozialer Interventionen zu beeinflussen und ihre Durchführung zu verbessern, ist sie „in einem weiten Sinn auch eine politische Tätigkeit“ (Rossi u.a. 1988: 183). Indem sie Gestaltungs-, Kontroll-, Steuerungs- und Bewertungsfunktionen innerhalb von Handlungsfeldern der Politik wahrnimmt und ihre Ergebnisse in den Verwaltungs- und Politikprozeß zurückmeldet, wird sie selbst zu einem Politikum und bewegt „sich notwendig in einem Minenfeld politischer, administrativer und gesellschaftlicher Interessen“ (Hellstern/Wollmann 1980: 61). Evaluationsforscher müssen deshalb die soziale Ökologie ihres Arbeitsumfeldes berücksichtigen. Verschiedene Interessengruppen sind direkt oder indirekt an den Evaluierungen beteiligt und können die Durchführung behindern oder fördern. Solche ‚Stakeholders‘ können politische Entscheidungsträger, die Auftraggeber der Evaluation, Durchführungsorganisationen, Implementationsträger, Programmteilnehmer, Zielgruppen, Projektmitarbeiter, Programmkonkurrenten und andere sein.

Aufgrund dieser ‚Dualität‘ der Evaluationsforschung, die sich darin ausdrückt, daß sie einerseits Teil der empirischen Sozialforschung ist und sich ihrer Theorien und Methoden bedient, aber andererseits auch Teil des politi-

schen Prozesses ist, den sie selbst mit ihren Ergebnissen beeinflußt und umgekehrt als Instrument zur Entscheidungsfindung für die politische Steuerung wissenschaftsfremden Anforderungen ausgesetzt ist, haben sich im Laufe der Entwicklung der Evaluationsforschung unterschiedliche methodische Ansätze herausgebildet. Diese orientieren sich entweder stärker an wissenschaftlichen Standards oder stärker an den Informationsbedürfnissen der Auftraggeber bzw. der Zielgruppen.

2. Theorien und Methoden der Evaluationsforschung

Da sich die Evaluationsforschung nicht losgelöst von der allgemeinen wissenschaftlichen Forschung entwickelt, fanden und finden in ihr auch die dort geführten Debatten mit nicht geringerer Heftigkeit ihre Fortsetzung.

Vor allem in den Anfangsjahren der Evaluationsforschung herrschte ein *methodologischer Rigorismus* vor. Experimentellen Designs wurde der Vorzug gegeben um „die wahren Zusammenhänge zwischen kausalen Kräften“ (Cook/Matt 1990: 20) aufzudecken, damit politische Entscheidungen über Interventionsprogramme nicht auf falschen Behauptungen hinsichtlich dessen beruhen, was ein Programm leistet und was nicht. Donald Campbell (1969: 409ff.), mißt deshalb der ‚internen Validität‘ die größte Bedeutung bei Evaluationen zu, d.h. der Sicherheit, daß die Verbindung zwischen zwei Variablen eine kausale Beziehung einer antezedenten Variable und ihrer Konsequenz reflektiert (vgl. Cook/Campbell 1979; Scriven 1972, 1980, 1991).

Nicht zuletzt als Reaktion auf die Erfolglosigkeit der Versuche, das Experimentaldesign zu realisieren, und in dem Bestreben, die ‚externe Validität‘ zu stärken, um die Verwendungsfähigkeit von Evaluationsergebnissen zu erhöhen, formierte sich ein Lager von Evaluatoren, das vor allem die instrumentelle Funktion der Evaluationsforschung für die Entscheidungsfindung im politischen Prozeß in den Vordergrund stellte: Evaluation wird eher als politischer, denn als wissenschaftlicher Akt gesehen. Aus Gründen der Ethik und wegen der politischen Nützlichkeit sollten Evaluationen vor allem den Interessen der Betroffenen dienen. Die Evaluatoren sollen dabei die Rolle von Moderatoren übernehmen und allen beteiligten Parteien bei der Evaluation beratend zur Seite stehen (vgl. Cronbach u.a. 1981; Cronbach 1982; Wottawa/Thierau 1998: 33).

Die Handlungs-Forschung führt diese Gedanken weiter und postuliert, daß Evaluationen nicht nur die Kontrolle der Qualität von Innovationen beinhalten darf „sondern gleichzeitig die Konstruktion, Optimierung und die Legitimierung der Modellmaßnahmen zu beinhalten habe“ (Lange 1983: 256). Daraus ergeben sich folgende, dem konventionellen sozialwissenschaftlichen Forschungsparadigma zuwiderlaufende, methodische Konsequenzen:

- Nicht die Falsifikation von Theorien oder Hypothesen ist primäres Erkenntnisziel, sondern die Angabe von Handlungsalternativen zur Lösung auftretender Probleme.
- Die Trennung zwischen Evaluatoren und den Evaluationsobjekten wird aufgehoben. Die Wissenschaftler geben ihre distanzierte Position zum Untersuchungsgegenstand auf und werden zu gleichberechtigten Partnern von den unmittelbar an der Evaluation Beteiligten und den davon Betroffenen (Evaluationsforschung mit stark partizipativen Komponenten, im Extremfall als Aktionsforschung).
- Nicht die Forschungsfragen des Evaluators stehen im Mittelpunkt des Interesses, sondern die Informationsbedarfe der Zielgruppen.
- Nicht Wertneutralität in den Aussagen wird angestrebt, sondern es werden im Gegenteil stellungsbeziehende Wertungen verlangt.
- Gütekriterien der Evaluation sind nicht mehr primär Validität, Reliabilität und Objektivität, sondern Kommunikation, Intervention, Transparenz und Relevanz. (Vgl. Gruschka 1976: 142-151; Weiss 1972: 6f.; Rein 1984: 179; Lachenmann 1987: 320; Staudt et al. 1988: 27f.; Gagel 1990: 45ff.; Schneider-Barthold 1992: 379ff.)

Der methodologische Rigorismus wurde zudem vom *konstruktivistischen Paradigma* in Frage gestellt. Das Vorhandensein einer einzigen ‚wahren‘ Realität wird bestritten. Statt dessen wird angenommen, daß Realität aus verschiedenen Perspektiven sozial konstruiert ist, die in Konflikt zueinander stehen können. Deshalb fordern die Anhänger dieses Ansatzes ein ‚qualitatives‘ Denken, um die verschiedenen Sichtweisen und Interpretationen der ‚Realität‘ erfassen zu können. Nach der erkenntnistheoretischen Position des Konstruktivismus sind grundsätzlich keine Aussagen über die tatsächliche Beschaffenheit der Welt möglich; „sie zeigen nur, ob eine Erkenntnis mit der Beschaffenheit der Welt vereinbar ist, ob sie ‚passt‘ – nicht aber, daß sie ‚wahr‘ (im Sinne eines ‚einzig richtig‘) ist“ (Meinefeld 1995: 100). Diese Auffassung erfordert deshalb auch bei Evaluationen eine andere wissenschaftliche Vorgehensweise als im Rahmen einer analytisch-nomologisch orientierten Erfahrungswissenschaft. (Vgl. vor allem Guba/Lincoln 1989; Patton 1987; Stake 1983)

Als eine Weiterführung des konstruktivistischen Paradigmas kann das *transformative/emanzipatorische Paradigma* betrachtet werden, das die unterschiedliche Machtstellung einzelner Stakeholder-Gruppen hervorhebt, die zur Überbetonung bestimmter Interessen führen kann (vgl. Mertens 1998 und in diesem Band).

Mittlerweile gilt der ‚kalte Krieg‘ der Paradigmen zwar keineswegs als endgültig beendet. Unterschiedliche Sichtweisen bleiben weiterhin bestehen. Doch es herrscht weitgehend Konsens darüber, daß Evaluationen die Perspektiven und Bedürfnisse der Stakeholder zu berücksichtigen haben, und daß quanti-

tative und qualitative Methoden oft gut miteinander kombiniert werden können (Multimethodenansatz).

Von vielen wird mittlerweile die Auffassung geteilt, „daß eine Untersuchung gleichzeitig strengen wissenschaftlichen Anforderungen genügen und für den Auftraggeber und andere Interessengruppen von maximalem Nutzen sein kann“ (Rossi u.a. 1988: 10). Sie folgen dabei der kritisch-rationalen Forschungslogik und halten prinzipiell alle bekannten empirischen Forschungstechniken für einsetzbar.

Chelimsky (1995: 6) beschreibt die neuere Entwicklung so:

“We think less today about the absolute merits of one method versus another, and more about whether and how using them in concert could result in more conclusive findings.”

Die eigentlichen Evaluationsfragen rücken dabei mehr in den Vordergrund:

“We have learned that the choice of methods (and measures and instruments and data) depends much more on the type of question being asked than on the qualities of any particular method.” (ebenda)

Über alle theoretischen und methodischen Fragen hinweg besteht zudem Einigkeit darin, daß es vor allem darauf ankommt, daß Evaluationsergebnisse im politischen Prozeß genutzt werden (vgl. Chelimsky 1995: 8). Damit rückt das Kriterium der Nützlichkeit von Evaluationsergebnissen für die Auftraggeber und die unterschiedlichen Stakeholder-Gruppen in den Vordergrund. Seit Mitte der 80er Jahre ist dies ein Schwerpunkt in der Evaluationsdiskussion (vgl. vor allem Patton 1997, Alkin 1990).

Die Qualität von Evaluationen kann demnach nicht allein an einem Gütekriterium gemessen werden. Neben der Wissenschaftlichkeit der Ergebnisse, kommt es eben auch darauf an, daß Evaluationen nützlich sind, denn nur dann werden sie auf politische und soziale Veränderungsprozesse einwirken können.

Im Zuge der Professionalisierung der Evaluationsforschung haben Ende der 70er Jahre verschiedene Organisationen in den USA eine Reihe von Kriterien entwickelt mit denen die Qualität von Evaluationen erfaßt werden soll. Am weitesten verbreitet haben sich die ursprünglich vom „Joint Committee on Standards for Educational Evaluation“ vorgelegten „Standards for Evaluations of Educational Evaluation“, die zahlreichen nationalen Evaluationsverbänden als Grundlage für die Erarbeitung eigener Standards dienten, so auch der Deutschen Gesellschaft für Evaluation (DeGEval). Nach diesen, 2001 entwickelten und sich jetzt in der Diskussion befindlichen Standards, sollen Evaluationen vier grundlegende Eigenschaften aufweisen: Nützlichkeit, Durchführbarkeit, Fairness und Genauigkeit (vgl. DeGEval 2002: 8f.) Im einzelnen sollen

- die Nützlichkeitsstandards sicherstellen, daß die Evaluation sich an den geklärten Evaluationszwecken sowie am Informationsbedarf der vorgesehenen Nutzer und Nutzerinnen ausrichtet,

- die Durchführbarkeitsstandards gewährleisten, daß eine Evaluation realistisch, gut durchdacht, diplomatisch und kostenbewußt geplant und ausgeführt wird,
- die Fairnessstandards regeln, daß in einer Evaluation respektvoll und fair mit den betroffenen Personen und Gruppen umgegangen wird und
- die Genauigkeitsstandards dafür sorgen, daß eine Evaluation gültige Informationen und Ergebnisse zu dem jeweiligen Evaluationsgegenstand und den Evaluationsfragestellungen hervorbringt und vermittelt.

Mit der *Qualität von Evaluationen* befaßt sich in diesem Band der Beitrag von *Thomas Widmer*, der auch auf die Entstehung, Struktur und den Inhalt der Standards sowie ihre Anwendung in der Evaluationspraxis eingeht.

Die Entwicklung der *Theorien und Methoden* in der Evaluationsforschung, die hier thematisch nur kurz angerissen werden konnte, wird in diesem Band vornehmlich in den Beiträgen von *Barbara Lee* und *Valerie Caracelli* behandelt. Aber auch in dem Beitrag von *Donna Mertens* zur *Institutionalisierung der Evaluation in den USA* finden sich hierzu einige Ausführungen, denn die Geschichte der Evaluationsforschung kann nicht ohne die historische Entwicklung von Paradigmen und den mit ihnen verbundenen Methoden beschrieben werden.

3. Entwicklung der Evaluationsforschung

Im folgenden soll auf die Evaluationsforschung in Deutschland eingegangen werden. Diese wurde vor allem von den Entwicklungen in den USA, Europa und in Deutschland selbst geprägt.

3.1 Entwicklung in den USA

Evaluationen, die auf systematisch gesammelten Daten basieren, sind ein relativ modernes Phänomen. Sie stehen in Zusammenhang mit den Demokratisierungsprozessen dieses Jahrhunderts, und sie wurden ermöglicht durch die Entwicklung sozialwissenschaftlicher Methoden.

In den USA wurden bereits in den *30er und 40er Jahren* Reformprogramme zur Verminderung der Arbeitslosigkeit und der Verbesserung der sozialen Sicherheit im Rahmen des „New Deal“ von Evaluationsstudien begleitet (vgl. Deutscher/Ostrander 1985: 17f.).

Als wissenschaftliche Pionierarbeiten der Evaluationsforschung gelten Lewins (1951) Feldstudien, die Arbeit von Lippitt und White über demokratische und autoritäre Führungsstile (Lippitt 1940; White/Lippitt 1953) so-

wie die Western-Electric-Studie (Hawthorne-Werke) über die psychischen und sozialen Folgen technologischer Innovationen (Roethlisberger/Dickson 1934).

Während des 2. Weltkriegs versuchte sich die U.S. Army die angewandte Sozialforschung zu Nutze zu machen. In ihrem Auftrag entwickelten Stouffer und seine Mitarbeiter (Stouffer u.a. 1949) Instrumente zur kontinuierlichen Messung der Stimmung in der Truppe und zur Evaluierung bestimmter Maßnahmen im Bereich des Personalwesens und der Propaganda (vgl. Rossi u.a. 1988: 5).

Der eigentliche Boom der Evaluationsforschung setzte in den USA zu Beginn der 60er Jahre im Gefolge umfangreicher Sozial-, Bildungs-, Gesundheits-, Ernährungs- und Infrastrukturprogramme ein. Von Anfang an waren diese mit dem Auftrag verknüpft, die Wirkungen dieser Programme zu überprüfen. In einem Großteil der Fälle wurden die Evaluationen sogar gesetzlich vorgeschrieben und eigene Haushaltsmittel dafür bereitgestellt.

Am bekanntesten sind die Wohlfahrts- und Reformprogramme der demokratischen Präsidenten John F. Kennedy (1961-1963) und Lyndon B. Johnson (1963-1969). Mit Hilfe des „War on poverty“ sollte die „Great Society“ verwirklicht werden. Die Evaluationsforschung sollte dabei nach Hellstern und Wollmann (1984: 27) Zeuge, Rechtfertigungsinstrument und Stimulator für diese Politik werden.

Mit dem „Economic Opportunity Act“ wurde 1964 eine von den Regierungsministerien unabhängige „Behörde für wirtschaftliche Chancengleichheit“ (Office of Economic Opportunity) für die Verwaltung und Finanzierung von Programmen gegründet, wie Headstart (Vorschulprogramm für Unterschichtkinder)⁷, Job Corps (Arbeitsbeschaffungsprogramm für arbeitslose Jugendliche), Manpower Training (Berufsbildung), Gesundheitsdienste und Legal Services (Rechtsberatung). Große Bedeutung errang vor allem das „Community Action Programm“, das nicht nur die materiellen Nöte der Bevölkerung abbauen, sondern auch neue Formen der Demokratie und Mitbestimmung auf kommunaler Ebene einführen sollte (vgl. Lachenmann 1977: 31).

Ein ganz wesentlicher Impuls für die Evaluationsforschung ging von der Einrichtung eines Planungs-, Programmgestaltungs- und Haushaltsplanungssystems (Planning, Programming and Budgeting System, PPBS) aus, das im Verteidigungsministerium durch Robert McNamara eingeführt und unter Präsident Johnson 1965 auf den gesamten Regierungsapparat und alle Ministe-

7 Head Start wurde zu einem der bekanntesten, intensivst evaluierten und umstrittensten sozialpolitischen Programme der USA. An dem zwischen 1965 und 1980 durchgeführten Programm nahmen 7,5 Mill Kinder teil. Es wurden mehr als 50 größere Evaluierungen durchgeführt und etwa 800 Zeitschriftenartikel publiziert. Bis erste Ergebnisse des Head Start Programms vorlagen, war bereits Präsident Richard M. Nixon (1969-1974) im Amt, der die Mittel für die Reformprogramme stark einschränken wollte. Die Evaluierungsergebnisse, die kaum einen Effekt des Vorschulprogramms nachweisen konnten, kamen ihm deshalb gerade recht. Die negativen Testergebnisse lösten eine Welle weiterer Evaluationen und eine umfassende Methodendiskussion aus (vgl. Hellstern/Wollmann 1984: 29ff.).

rien ausgedehnt wurde.⁸ Unter Präsident Nixon wurde ein Großteil dieses Systems allerdings schon 1971 wieder aufgegeben.

Durch die Reformprogramme entstand ein großer Auftragsmarkt, von dem die Evaluationsforschung stark profitierte. Es wird geschätzt, daß 1976 bereits 600 Mill. \$ für die Evaluation von sozialen Dienstleistungsprogrammen aufgewendet wurden (vgl. Wittmann 1985: 9). Das „Evaluationsbusiness“ wurde zu einer Wachstumsindustrie.

Als in den *70er* und vor allem *80er Jahren* die Kritik an solchen Programmen wuchs, gingen zwar insgesamt die Mittel für innovative Modellvorhaben zurück, doch die Bedeutung der Evaluationsforschung wurde dadurch nicht geschmälert. Lediglich ihr Schwerpunkt verlagerte sich. Zunehmend rückte das Verhältnis von Kosten und Nutzen, die Effizienz des Managements und die Rechenschaftsberichterstattung in den Mittelpunkt des Interesses. So wurden z.B. sogenannte „Sunset-Gesetze“ erlassen, die eine automatische Beendigung der Programme vorsahen, wenn nicht innerhalb einer festgelegten Zeitspanne der Nachweis ihrer Wirksamkeit erbracht werden konnte (vgl. Rossi u.a. 1988: 6). Mit dem „Government Performance and Result Act“ wurde die Rolle der Evaluationsforschung noch einmal deutlich gestärkt. Dieses 1993 vom US-Kongreß verabschiedete Gesetz, schreibt Mertens in diesem Band: „shifted the focus of federal management and decision making away from preoccupation with the activities that are undertaken under the auspices of federal funding to a focus on the results of those activities“. Mittlerweile ist Evaluation in den USA zu einem integrierten Bestandteil von Programmen geworden, um deren Implementation und Wirksamkeit zu überprüfen.⁹

Die *Institutionalisierung der Evaluation in den USA* sowie neuere Entwicklungen, insbesondere auch die Auswirkungen legislativer Einflüsse, werden hier im Beitrag von *Donna Mertens* behandelt.

3.2 Entwicklung in Europa

In Europa setzte die „moderne“ Evaluationsforschung im Vergleich zu den USA mit einer zehnjährigen Verspätung Ende der *60er Jahre* ein. Wie dort gewann sie mit dem Aufkommen umfassender politischer Reformprogramme an Bedeutung und Profil. Schweden, Großbritannien und auch Deutschland werden zu den „frontrunners“ (vgl. Leeuw in diesem Beitrag)¹⁰ gezählt. Während diese erste Evaluationswelle getragen wurde von einer neo-keynesiani-

8 Deutscher & Ostrander (1985:18) halten die Einführung des PPBS für einen Markstein in der Evaluationsgeschichte: „This concept and McNamara’s influence are crucial to the history of evaluation research.“ Dieser Auffassung ist u.a. auch Wittmann (1985: 5).

9 Besonders interessant sind an dieser Stelle die Überlegungen von Derlien (1990: 9f.) über die Auswirkungen unterschiedlicher Verfassungskontexte auf die Bedeutung von Rechnungshöfen und die Funktion von Evaluation.

10 Vgl. zur europäischen Entwicklung auch Derlien (1990); Rist (1990); Wollmann (1997); Pollit (1998); Wollmann (2000).

schen Politik, dem Ausbau des Wohlfahrtsstaates und der Verbesserung staatlicher Infrastruktureinrichtungen und dabei noch aus „vollen“ Kassen schöpfen konnte, setzten in Folge der Erdölpreiserhöhung von 1973 und der dadurch ausgelösten weltweiten Wirtschaftsrezession staatliche Finanzierungsengpässe ein. Die Diskussion zur Modernisierung von Staat und Verwaltung drehte sich in den westlichen Industrieländern zunehmend um das Problem der Haushaltskonsolidierung. Im Rahmen neo-liberaler und neo-konservativer Strömungen wurde der Rückbau des Sozial- und Wohlfahrtsstaates propagiert.

Bis Mitte der 70er Jahre wurde die Evaluationsforschung – wie zuvor in den USA – vor allem dazu genutzt, die Effektivität von Programmen nachzuweisen, um die Durchsetzungschancen innovativer Maßnahmen zu verbessern. Evaluationen sollten nicht nur die Wirksamkeit von Programmen belegen, sondern auch zur Korrektur und Steuerung laufender Maßnahmen beitragen. Dabei wurden analytische Interessen und legitimatorische Absichten häufig miteinander vermengt.

Ab Mitte der 70er Jahre gewannen – wie in den USA – Kosten-Nutzen-Erwägungen an Bedeutung und die Administratoren erhofften sich von Evaluationen rationale Entscheidungs- und Argumentationshilfen für die Priorisierung und Selektion von Programmen. In die Evaluierungsforschung wurden große Hoffnungen gesetzt:

„*Öffentlichkeit* und *Parlament* versprachen sich eine verbesserte Kontrolle staatlicher Maßnahmen, die Berücksichtigung von Neben- und Folgewirkungen; der *Verwaltung* eröffnete Evaluierung Eingriffschancen, um eine verbesserte Zielgenauigkeit der Programme, verbesserte Wirtschaftlichkeit der Maßnahmen und Senkung der Kosten zu erreichen; der beteiligten *Wissenschaft* stellte sie nicht nur den Gewinn zusätzlicher Ressourcen, sondern auch die Möglichkeit eines experimentellen Tests ihrer Theorien in Aussicht.“ (Hellstern/Wollmann 1984: 23)¹¹

In den späten 80er und 90er Jahren bekam Evaluation einen neuen Aufschwung. Die Reform- und Modernisierungsdebatte gewann unter dem Stichwort „New Public Management“ neue Konturen, zuerst in Holland, Großbritannien und in den skandinavischen Ländern (vgl. Wollmann 1994: 99). Dabei handelt es sich nicht etwa um ein kohärentes Konzept, sondern eher um ein Bündel von Organisations- und Verfahrensprinzipien, mit denen vor allem eine Reduzierung der Staatsaufgaben (insbesondere durch Privatisierung) ein Abbau staatlicher Regelungsdichte (Deregulierung) und die Erhöhung administrativer Effizienz (interne Ökonomisierung, value for money) durch binnensstrukturelle Managementreformen sowie durch die Einführung von Wettbewerb erreicht werden soll. Im Prinzip geht es darum, das an ökonomischer Rationalität orientierte privatwirtschaftliche Unternehmens- und Marktmodell auf den öffentlichen Sektor zu übertragen (vgl. Schröter/Wollmann 1998:

¹¹ Hervorhebungen durch Hellstern und Wollmann.

59ff.). Der Evaluation kommt in diesem Kontext vor allem die Rolle zu, die Effizienz staatlicher Maßnahmen zu überprüfen.

Ein weiterer maßgeblicher Faktor, der auf die Entwicklung der Evaluationsforschung Einfluß nahm, ist in der fortschreitenden europäischen Integration zu sehen, die nicht nur zu einer Ausdehnung eines europäisch bestimmten Verwaltungsregimes führte (vgl. Kohler-Koch 1991: 47ff.), sondern auch zur Initiierung einer Vielzahl unterschiedlichster Programme. Dies stellt die Evaluationsforschung, die vor allem national organisiert ist, vor große Herausforderungen. Zu beobachten ist mittlerweile: „In a variety of sectors – most notably for the EC Structural Funds – evaluation has been made legally or conventionally mandatory.“ (Pollitt 1998: 214)

Insgesamt ist sowohl für die EU-Institutionen als auch für die meisten europäischen Länder im letzten Jahrzehnt eine deutliche Zunahme an Politik- und Programmevaluationen zu konstatieren, was Christopher Pollitt (1998: 214) zu der Aussage veranlaßt: „These are grand days for European evaluators.“

Mit der *Entwicklung der Evaluation in Europa* setzt sich in diesem Band der Beitrag von *Franz Leeuw* auseinander.

3.3 Entwicklung in Deutschland

Die Entwicklung der Evaluation in Deutschland wurde von den genannten internationalen und europäischen Strömungen erfaßt und mitgeprägt. Allerdings haben auch eine Reihe *nationaler Rahmenbedingungen* die Evaluationsentwicklung beeinflußt.

Nach einer Phase der „institutionellen Restauration“ (Alber 1989: 60) nahm der Problemdruck in der BRD deutlich zu. Die sogenannte „Bildungskatastrophe“, ein allgemein konstatiertes ‚Reformstau‘ und die Sorge um die internationale wirtschaftliche Wettbewerbsfähigkeit führten zu einem breiten gesellschaftlichen Konsens, daß weitreichende Reformen notwendig seien. Bereits die große Koalition aus CDU/CSU und SPD (1966-1969) brachte eine Reihe von Vorhaben auf den Weg, die dann in der sozial-liberalen Koalition unter Kanzler Willy Brandt (1969-1974) als Bestandteil einer Politik der inneren Reform zum Regierungsprogramm erhoben wurde. Es wurde eine umfassende Staats- und Verwaltungsmodernisierung angestrebt, die auf der Vorstellung beruhte, daß „der Staat als zentrale gesellschaftliche Steuerungsinstanz fungieren, „aktive Politik“ betreiben und die Gesellschaft langfristig planend gestalten sollte“ (Mayntz 1997: 68). Die Erweiterung der Handlungs- und Gestaltungsfähigkeit des Staates wurde insbesondere von der Einführung neuer Planungsverfahren erwartet, wobei Evaluation als ein wichtiges analytisches Instrument verwendet wurde. Evaluation erlebte in Deutschland eine erste Blüte. Nicht nur die nationalstaatlichen Einrichtungen, sondern auch die Länder und Gemeinden wurden vom Reformfeifer erfaßt und bedienten sich der

Evaluation als Steuerungs- und Kontrollinstrument (vgl. Wollmann 1994, 1998, 1999; Derlien 1976, 1990, 1994). Vor allem in den Bereichen Bildung und Erziehung, Stadterneuerung und Infrastruktur wurden umfangreiche und aufwendige Evaluationsstudien durchgeführt.

Dabei griffen die Ministerien weitgehend auf externe Evaluationen zurück, die zumeist öffentlich ausgeschrieben wurden. Vom Aufbau hauseigener Evaluationskapazitäten wurde weitgehend abgesehen. Eine der wenigen Ausnahmen war das noch junge Bundesministerium für wirtschaftliche Zusammenarbeit, das bereits 1972 ein hausinternes Evaluationsreferat einrichtete (zur *Entwicklung der Evaluation in der staatlichen Entwicklungszusammenarbeit* siehe den Beitrag von Reinhard Stockmann in diesem Band).

Das Kanzleramt scheiterte bei dem Versuch, „to use evaluation as its analytical muscle for co-ordinating (and possibly controlling) the ministries „sectoral policies“ (Wollmann 1997: 4). So blieb es dabei, daß die einzelnen Fachministerien relativ unabhängig voneinander ihre Evaluationen sektorbezogen und speziell an ihren eigenen spezifischen Bedürfnissen und Vorstellungen ausgerichtet durchführten.

Unterstützt wurde der Evaluationsboom durch die Haushaltsreform von 1970. In den Vorläufigen Verwaltungsvorschriften (VV) zu §7 der Bundeshaushaltsordnung (BHO) wurden ausdrücklich Erfolgskontrollen bei ganz oder teilweise abgeschlossenen Maßnahmen vorgeschrieben. Nr. 1.3 der VV-BHO bestimmt:

- „1.3 Im Wege der Erfolgskontrolle (Ergebnisprüfung) soll insbesondere untersucht werden
- 1.3.1 während der Durchführung von mehrjährigen Maßnahmen mindestens jährlich, ob die Zwischenergebnisse im Rahmen der Planung liegen, die Planung anzupassen ist und die Maßnahmen weiterzuführen oder einzustellen sind,
- 1.3.2 nach der Durchführung von Maßnahmen, ob das erreichte Ergebnis der ursprünglichen oder angepaßten Planung entspricht, die Maßnahmen zu revidieren sind und Erfahrungswerte gesichert werden können.“ (Bundesrechnungshof 1989: 13 und Anlage 1: 49ff.)

Innerhalb weniger Jahre entstand ein Evaluationsmarkt, an dem zwar auch die Universitäten partizipierten, der aber vor allem von einer rasch expandierenden Consultingwirtschaft beherrscht wurde. Hellmut Wollmann kann hierzu nicht nur als Wissenschaftler sondern auch als Zeitzeuge zitiert werden: „Commercial research and consultancy firms mushroomed and succeeded to produce the lion's share of the evaluation research funding.“ (Wollmann 1997: 4)

Doch dem Aufblühen der Evaluation in Deutschland wurde, wie in vielen westeuropäischen Ländern, durch die weltweiten ökonomischen und fiskalischen Auswirkungen der Erdölpreiserhöhung von 1973 ein abruptes Ende be-