Internationale Hochschulschriften

Susanne Giel

Theoriebasierte Evaluation

Konzepte und methodische Umsetzungen

WAXMANN

Susanne Giel

Theoriebasierte Evaluation

Konzepte und methodische Umsetzungen

Waxmann 2013
Münster / New York / München / Berlin

Bibliografische Informationen der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.d-nb.de abrufbar.

D 188

Internationale Hochschulschriften, Bd. 584
Die Reihe für Habilitationen und sehr gute und ausgezeichnete Dissertationen.

ISSN 0932-4763
ISBN 978-3-8309-7855-8

Postfach 8603, 48046 Münster

www.waxmann.com
info@waxmann.com

Umschlaggestaltung: Christian Averbeck, Münster
Umschlagbild: Mathias Waller, Berlin
Satz: Stoddart Satz- und Layoutservice, Münster
Druck: Hubert & Co., Göttingen

Gedruckt auf alterungsbeständigem Papier,
säurefrei gemäß ISO 9706

Printed in Germany

Inhalt

1.	Die Suche nach alternativen Wegen – Einleitung	9
2.	Evaluation – Was ist das eigentlich?	15
2.1	Entstehung einer neuen Profession	17
2.2	Irgendetwas – Gegenstände und typische Fragestellungen	21
2.3	Irgendwie – Das Vorgehen von Evaluationen	28
2.4	Irgendjemand – Wer bewertet nach welchen Kriterien?	31
2.5	Rollen und Aufgaben von Evaluatorinnen und Evaluatoren	37
2.6	Evaluation – Zwischen allen Stühlen	40
2.7	Das Praxisbeispiel: Die Rahmenbedingungen	45
2.7.1	Der Evaluationsgegenstand	45
2.7.2	Optionale Evaluationsfunktionen und Rollenverteilungen	51
3.	Klassische Konzepte zur Durchführung von Evaluationen	55
3.1	Zielorientierte Evaluation (Ralph W. Tyler)	56
3.1.1	Grundzüge des Konzepts	56
3.1.2	Rolle der Evaluatorinnen und Evaluatoren	60
3.1.3	Kritische Würdigung	61
3.2	Experimental- und Quasiexperimentaldesigns (Donald T. Campbell)	63
3.2.1	Grundzüge	63
3.2.2	Die Rolle von Evaluatorinnen und Evaluatoren	69
3.2.3	Kritische Würdigung	70
3.3	Nutzungsfokussierung – Pragmatistisches Paradigma (Patton)	72
3.3.1	Grundzüge	73
3.3.2	Die Rollen von Evaluatorinnen und Evaluatoren	76
3.3.3	Kritische Würdigung	77
3.4	„Evaluation der vierten Generation“ (4th Generation Evaluation – Guba/Lincoln)	78
3.4.1	Grundzüge	79
3.4.2	Die Rolle von Evaluatorinnen und Evaluatoren	83
3.4.3	Kritische Würdigung	84
3.5	Überwindung des Paradigmenstreits: Methodenpluralismus	86
3.6	Die vorgestellten Konzepte am Beispiel	88
3.6.1	Zugang über die Ziele	88
3.6.2	Zugang über Wirkungen mit (quasi-)experimentellen Designs	93
3.6.3	Zugang über Nutzungsabsichten	95
3.6.4	Zugang über Konstruktionen	97

4. Konzepte theoriebasierter Evaluationen ... 101
4.1 Bezeichnungen und Relevanz theoriebasierter Evaluationskonzepte ... 101
4.2 Kritik an bestehenden Evaluationskonzepten ... 105
4.2.1 Das Kreuz mit den Zielen – Kritik an zielorientierter Evaluation ... 106
4.2.2 Labor oder Praxis – Kritik an experimentellen Designs ... 109
4.2.3 Relativistische Konstruktionen – Kritik an der „Evaluation der vierten Generation" ... 114
4.3 Von Theorien bis zu Programmtheorien ... 116
4.4 Welche und wessen Programmtheorien? ... 129
4.4.1 Chens Integrationskonzept ... 131
4.4.2 Der Prozess der Programmtheorieentwicklung bei Weiss ... 133
4.4.3 Zyklische Programmtheorieentwicklung bei Pawson/Tilley ... 135
4.5 Die Funktion von Programmtheorien für die Evaluation und deren Nutzen fürs Programm ... 137
4.6 Programmtheorien zur internetbasierten Lernumgebung ... 139
4.6.1 Sozialwissenschaftliche Zugänge zu Programmtheorien ... 140
4.6.2 Zugang zu Programmtheorien über die Praxis ... 148
4.7 Was – vorläufig – offen bleibt ... 155

5. Methodische Umsetzung theoriebasierter Evaluationen ... 157
5.1 Probleme und Herausforderungen ... 157
5.2 Naturalistische Forschung und Symbolischem Interaktionismus ... 160
5.2.1 Grundzüge ... 160
5.2.2 Methodisches Vorgehen ... 164
5.2.3 Nutzungspotentiale für theoriebasierte Evaluationen ... 167
5.2.4 Nutzungspotentiale für die Evaluation der internetbasierten Lernumgebung ... 169
5.3 Aktionsforschung ... 172
5.3.1 Grundzüge ... 172
5.3.2 Methodisches Vorgehen ... 177
5.3.3 Nutzungspotentiale für theoriebasierte Evaluationen ... 181
5.3.4 Nutzungspotentiale für die Evaluation der internetbasierten Lernumgebung ... 185
5.4 Dokumentarische Methode der Interpretation ... 187
5.4.1 Grundzüge ... 187
5.4.2 Methodisches Vorgehen ... 191
5.4.3 Nutzungspotentiale für theoriebasierte Evaluationen ... 196
5.4.4 Nutzungspotentiale für die Evaluation der internetbasierten Lernumgebung ... 198
5.5 Grounded-Theory-Methodologie ... 200
5.5.1 Grundzüge ... 201
5.5.2 Methodisches Vorgehen ... 205
5.5.3 Nutzungspotentiale für die Umsetzung theoriebasierter Evaluation ... 212
5.5.4 Nutzungspotentiale für die Evaluation der internetbasierten Lernumgebung ... 215

5.6 Standardisiert verfahrende Forschung 219
5.6.1 Grundzüge 220
5.6.2 Methodische Umsetzungen 221
5.6.3 Kausalmodelle testen und Pattern Matching 223
5.6.4 Nutzungspotentiale für theoriebasierte Evaluationen 229
5.6.5 Nutzungspotentiale für die Evaluation der internetbasierten Lernumgebung 233

6. Die Integration von Methoden zur Umsetzung theoriebasierter Evaluationen 241
6.1 Methodenmix – Triangulation – Methodenintegration 242
6.2 Grundtypen der Integration 246
6.3 Integrationsstrategien entlang des Programms 250
6.4 Integrationsstrategien je nach Funktion der Evaluation 253
6.4.1 Kontrollparadigma 253
6.4.2 Entwicklungsparadigma 254
6.4.3 Forschungsparadigma 255
6.5 Die Integration von Methoden zur Evaluation der internetbasierten Lernumgebung 257

7. Potentiale theoriebasierter Evaluationskonzepte und zu lösende Herausforderungen 272

Mein Dank 280

Literatur 281

Abbildungsverzeichnis 304

Tabellenverzeichnis 305

1. Die Suche nach alternativen Wegen – Einleitung

Zweifellos kann Evaluation als eine empirisch-wissenschaftliche Beschreibung und Bewertung von Programmen, Projekten, Maßnahmen, Organisationen einen steilen Aufschwung verzeichnen. Sie etabliert sich als eine Strategie zur Qualitätssicherung und -entwicklung in fast allen gesellschaftlichen Bereichen: Hochschule und Schule (Bildung), Gesundheitswesen, Kinder- und Jugendhilfe, in den Politikfeldern Arbeitsmarkt, regionale Entwicklung, Umwelt usw., nahezu überall ist eine Zuwendung öffentlicher Mittel an die Durchführung von Evaluationen geknüpft. So lässt sich eine „zunehmende Verankerung dieses Instruments im Rahmen gesellschaftlicher bzw. administrativer Regelungen und Vorschriften" (Brand 2009, S. 11) beobachten.

Aus diesem hohen Verbreitungsgrad folgt jedoch nicht automatisch, dass die Funktionen und die Vorgehensweisen von Evaluation eindeutig profiliert wären. Eher schon zeichnet sich seit einigen Jahren die Tendenz ab, dass der Begriff inflationär benutzt wird, denn in manchen Kontexten wird jegliche Form von Bewertung mit dem Label „Evaluieren" versehen (vgl. Kromrey 2000, S. 19). Tasso Brand (2009, S. 14) sieht gar die Gefahr, dass Evaluation zu einer „rein sprachlichen Leerformel degeneriert, ohne eine angemessene Berücksichtigung der mit diesem Begriff verknüpften theoretischen Konzepte und qualitativen Ansprüche zu erfahren." Folglich ist es kaum verwunderlich, dass sich Evaluationen mit vielfältigen Erwartungen konfrontiert sehen: Das Spektrum reicht von der Evaluation als Rechtfertigung von Entscheidungen bis hin zur Vorbereitung von Entscheidungen, manchmal interessieren sich Auftraggebende für die Akzeptanz von und die Zufriedenheit mit Angeboten, in anderen Fällen erwarten sie die Feststellung von Wirkungen. So vielfältig die Erwartungen und Nutzungsabsichten sind, so breit gefächert sind die Einschätzungen darüber, was Evaluationen eigentlich leisten können, müssen oder auch dürfen.

Gleichwohl hat sich hinsichtlich einer Professionalisierung der Evaluation in den letzten Jahren viel getan: So wurden inzwischen im deutschsprachigen Raum Berufsverbände gegründet (die „Schweizerische Evaluationsgesellschaft" SEVAL, für Österreich und Deutschland die „Gesellschaft für Evaluation" DeGEval), Standards für die Durchführung von Evaluationen sowie Empfehlungen für die Aus- und Weiterbildung von Evaluatorinnen und Evaluatoren verabschiedet und erste Studiengänge installiert. Einen ausführlichen Überblick über den beachtlichen Professionalisierungsgrad ebenso wie die noch einzulösenden Herausforderungen liefert Tasso Brand (2009) in seiner Dissertation. Er macht vor allem Defizite auf der Ebene der theoretischen Reflexion und Profilierung aus. So konstatiert er für den Bereich der Methoden der Evaluation zu Recht eine große Vielfalt, die „durch unterschiedliche disziplinäre Zugänge geprägt ist" (Brand 2009, S. 247). Der Fachaustausch findet vorwiegend feldspezifisch statt, z. B. in disziplinär ausgerichteten Arbeitskreisen oder im Rahmen von Tagungen der DeGEval. Es existieren derzeit lediglich zwei übergreifende Arbeitskreise (Aus- und Weiterbildung, Gender Mainstreaming) sowie der „Arbeitskreis in Gründung Methoden".

Die vorliegende Arbeit geht davon aus, dass nach wie vor ein erheblicher Diskussions- und Entwicklungsbedarf hinsichtlich der Methodologie von Evaluationen besteht. Damit verbindet sich die Hoffnung, dass durch eine methodologische Debatte eine Brücke zwischen Theorie und Methoden sowie eine zwischen den verschiedenen Disziplinen geschlagen werden kann. Auch eine weitere zu diagnostizierende Diskrepanz soll angegangen werden: Auf der einen Seite werden experimentelle Designs als der Goldstandard für die Durchführung von Evaluationen ausgewiesen, auf der anderen Seite sind in der Regel die Voraussetzungen für deren forschungspraktische Umsetzung nicht gegeben (siehe z. B. Kromrey 2001a, S. 121). In der Praxis bleiben zahlreiche Evaluationen bei einfachen Befragungen zur Akzeptanz oder Zufriedenheit stehen, die in der Regel wiederum dem Evaluationsgegenstand nicht gerecht werden (Kromrey 2004, S. 244ff.). Zwischen diesen beiden Polen – Experimentaldesigns hier und Zufriedenheitsmessungen dort – ist nach wie vor die Diskussion von verschiedenen Designs notwendig und essentiell. Vor allem im Bereich von Humandienstleistungen stehen Evaluationen nach wie vor vor der Herausforderung, ihren hoch komplexen und facettenreichen Gegenständen gerecht zu werden. Erfolge von Programmen und Maßnahmen sind eben nicht so ohne Weiteres feststellbar, geschweige denn, dass empirisch fundierte Aussagen darüber getroffen werden können, wodurch solche Erfolge erzielt wurden. Die besonderen Ansprüche an den methodologischen Zuschnitt von Evaluationen formuliert die DeGEval (Gesellschaft für Evaluation, 2010) zu ihrer 12. Jahrestagung in Luxemburg folgendermaßen:

> „Durch die Besonderheiten der Evaluation stellen sich jedoch auch zusätzlich methodische Anforderungen. Eingebettet in politische und praktische Kontexte müssen Evaluationen nicht nur gegenstandsangemessene Methoden einsetzen, sondern auch die zeitlichen und strukturellen Rahmenbedingungen sowie Fragen der Durchführbarkeit und Fairness gegenüber den Betroffenen berücksichtigen."

Auf diese genannten Herausforderungen wollen theoriebasierte Evaluationskonzepte – begrifflich und konzeptionell in den USA entwickelt – eine Antwort liefern.[1] Bereits 2001 schlägt Helmut Kromrey (2001a, S. 123) diese als eine Alternative zu experimentellen Designs vor. Auch 2011 gelten theoriebasierte Ansätze noch als zu ergründende und aussichtsreiche Variante. Eine von proVal und dem Zentrum für interdisziplinäre Forschung vom 16. bis 18. Februar 2011 in Bielefeld durchgeführte Fachtagung unter dem Titel „Evaluation von Programmen und Projekten zur Förderung einer pluralistischen und demokratischen Kultur" räumte einer theoriebasierten Evaluation als alternativem Design einen breiten Raum ein. Im Ankündigungstext heißt es:

1 Um dem Anspruch gerecht zu werden, das Konzept theoriebasierter Evaluationen für den deutschsprachigen Raum zu erschließen, wurde im Fließtext auf englischsprachige Zitate verzichtet und versucht, nur solche englischsprachigen Begriffe zu verwenden, die eindeutig bestimmt sind.

„So ist es für eine Evaluation aus verschiedenen Gründen schwierig, die Wirkung von Programmen und Projekten zu ermitteln. Klassische Ansätze wie experimentelle oder quasiexperimentelle Designs sind oft nicht praktikabel, und deshalb müssen andere Möglichkeiten wie die theoriebasierte Evaluation in diesem Zusammenhang diskutiert werden (….), ein viel versprechender [wahrscheinlich vielversprechender S.G.] Ansatz für die Bewertung der Wirksamkeit von Programmen, die für konventionelle Herangehensweisen entweder zu komplex sind oder die zu wenige Teilnehmer für den Einsatz quantitativer Methoden haben" (http://www.proval-services.net/evalconference/index.html).

Eine Spezifik der theoriebasierten Ansätze liegt darin begründet, dass Programmtheorien – dem Programm zugrunde liegende Annahmen sowie Hypothesen über das Programm – zum Dreh- und Angelpunkt der Durchführung von Evaluationen erklärt werden. Theoriebasierte Evaluation proklamiert für sich nicht nur Auskunft darüber zu geben, ob ein Programm greift, sondern wie und weshalb dies (nicht) gelingt.

Auch wenn die Schlagworte Theoriebasierung oder Programmtheorie im deutschsprachigen Raum mehr und mehr fallen (z. B. Stockmann 2007, S. 50) und unter verschiedenen Aspekten beleuchtet werden (z. B. Haubrich 2009, Hense/Kriz 2006) – eine umfassende, systematische Darstellung der sich hinter der gemeinsamen Klammer „theoriebasiert" verbergenden verschiedenen Vorschläge steht noch aus. Eine noch größere Lücke öffnet sich hinter der Frage, wie sich theoriebasierte Evaluationen forschungspraktisch und methodisch umsetzen lassen. Hier sind gleich zwei Aufgaben zu bewältigen: Zum einen muss ein methodischer Zugang zu den Programmtheorien gefunden werden und zum anderen müssen diese mit dem tatsächlichen Programmgeschehen konfrontiert werden.

Ziel dieser Arbeit soll es insofern sein, Konzepte theoriebasierter Evaluationen in differenzierter und umfassender Weise für den deutschsprachigen Raum nachvollziehbar, nutzbar und anwendbar zu machen. Damit soll für die Praxis von Evaluationen sowie für die Aus- und Fortbildung von Evaluatorinnen und Evaluatoren eine zusätzliche alternative Vorgehensweise erschlossen werden.

Die übergeordnete Fragestellung dieser Arbeit lautet: Wie können Evaluationen gegenstandsangemessen durchgeführt werden, so dass sie für Beteiligte und Betroffene einen Nutzen erzeugen, realistisch umsetzbar sind, zu genauen Ergebnissen führen und fair den Beteiligten gegenüber sind (vgl. Standards für Evaluation DeGEval 2002). Diese Fragestellung gilt es in besonderer Weise an theoriebasierte Evaluationskonzepte heranzutragen: Welche besonderen konzeptionellen Vorschläge unterbreiten Protagonistinnen und Protagonisten theoriebasierter Evaluation? Welchen Nutzen ziehen Evaluationen und Evaluationsgegenstände und -beteiligte aus theoriebasierten Herangehensweisen? Wie lassen sich diese Konzepte forschungspraktisch und methodisch realisieren?

Zum Einstieg und um das Feld abzustecken erfolgt zunächst eine Einschätzung darüber, welchen Ansprüchen Evaluationen im Allgemeinen und damit auch theo-

riebasierte Evaluationskonzepte genügen müssen. Neben einem kurzen Abriss zur Entstehung und Entwicklung sollen vor allem die Einsatzfelder von sowie die Erwartungen an Evaluation aufgezeigt werden. Diese Aspekte werden im folgenden Kapitel „Evaluation – was ist das eigentlich?" (2) behandelt. Darüber hinaus wird ein Blick auf die Aufgabenfelder von sowie Anforderungen an Evaluatorinnen und Evaluatoren geworfen und die typischen Konfliktfelder, in denen sie sich bewegen, beleuchtet.

Zum Verständnis des alternativen Angebots theoriebasierter Evaluationen müssen zunächst klassische Designs für Evaluationen besprochen werden. Vor allem sind solche relevant, die die Praxis und Theorie von Evaluationen bestimmen und grundlegende „Schulen" und Paradigmen von Evaluationen repräsentieren. Zur genaueren Betrachtung sind solche Designs ausgewählt, die ein breites Spektrum abdecken und gleichermaßen für den Ansatzpunkt theoriebasierter Evaluationskonzepte stehen. In Kapitel 3 werden demzufolge die zielorientierte Evaluation nach Ralph Tyler (3.1), experimentelle und quasiexperimentelle Designs in Anlehnung an Donald T. Campbell, Julian Stanley und Thomas D. Cook (3.2), die nutzungsfokussierte Evaluation nach Michael Q. Patton (3.3) sowie die konstruktivistische Evaluation bzw. „Evaluation der vierten Generation" nach Egon Guba und Yvonna Lincoln (3.4) vorgestellt. Diese vier Typen werden jeweils in ihren methodologischen Grundzügen erläutert, die daraus resultierenden Rollen von Evaluatorinnen und Evaluatoren behandelt und abschließend jeweils kritisch gewürdigt.

Die Auseinandersetzung mit diesen Designs ist nicht zuletzt deswegen relevant, weil die Protagonistinnen und Protagonisten verschiedener theoriebasierter Evaluationsansätze in ihrer Argumentation für eine Alternative genau an den konzeptionellen und forschungspraktischen Schwächen dieser Klassiker ansetzen. Die Kritikpunkte an zuvor beschriebenen Designs werden unter 4.2 abgehandelt, nachdem zunächst einmal die Entstehung theoriebasierter Evaluationsansätze historisch verortet sowie das Label „theoriebasiert" begrifflich begründet wird (4.1).

Unmissverständlicher wäre übrigens, das theoriebasierte Vorgehen als „programm-theoriebasierte Evaluation" zu bezeichnen, denn wie der darauf folgende Abschnitt (4.3) zeigt, lässt sich der Begriff „Theorie" präziser als „programmbezogene Theorie" fassen. Statt in allgemeiner Weise auf theoretisches Wissen zurückzugreifen, konzipieren die verschiedenen Autorinnen und Autoren Programmtheorien entlang des Evaluationsgegenstands in einem Spektrum von Theorien über das Programm bis hin zu Theorien *des* Programms, das heißt aus dem Programm heraus. Die verschiedenen Konzepte – von Huey-tsyh Chen und Peter H. Rossi, Rick Pawson und Nick Tilley, Carol H. Weiss, Michael Q. Patton – führen zu unterschiedlichen Modellierungen von Programmtheorien, die ebenfalls unter 4.3 dargestellt sind.

Aus der Erkenntnis, dass es nicht die eine Programmtheorie gibt, resultiert folgerichtig die Frage, welche der zahlreichen Möglichkeiten denn nun die Grundlage für die jeweils durchzuführende Evaluation sein soll. In dieser Arbeit wird die Auffassung vertreten, dass die einzelnen Typen von Programmtheorien je nach Auftrag an die Evaluation und je nach Spezifik des Programms flexibel reagieren können muss.

Ein entsprechend breites Spektrum der hier vorgeschlagenen Varianten und Spielarten wird in Kapitel 4.4 behandelt. Als Zwischenfazit lässt sich daran anschließend feststellen, welche Funktionen Programmtheorien für die Evaluation übernehmen und welchen Nutzen sie für die Programme und Programmbeteiligte haben können (4.5).

Ihre Stärken können theoriebasierte Evaluationsansätze nur dann entfalten, wenn Antworten auf die Frage gefunden werden, wie theoriebasierte Evaluationen methodologisch und forschungspraktisch zu realisieren sind. Theoriebasierte Evaluationen verfügen zwar über einen konzeptionellen Ansatzpunkt, die Programmtheorien, wie diese jedoch methodisch einwandfrei erschlossen und mit der Programmwirklichkeit zu konfrontieren sind, dazu liefert die Literatur nur wenige Vorschläge.[2] Um insofern den spezifischen Herausforderungen theoriebasierter Evaluationen methodisch begegnen zu können, werden im sechsten Kapitel fünf verschiedene methodologische Konzepte auf ihr Potential für die Umsetzung theoriebasierter Evaluationen hin überprüft. Ausgewählt wurden die naturalistische Forschung und der Symbolische Interaktionismus (5.2), die Aktionsforschung (5.3), die Dokumentarische Methode (5.4), die Grounded Theory (5.5) sowie standardisiert verfahrende Forschungsstrategien (5.6). Diese verschiedenen Forschungsrichtungen werden jeweils auf ihre wissenschaftstheoretischen Grundannahmen und auf ihr methodisches Vorgehen hin analysiert, um fundierte Auskunft über deren Nutzungsmöglichkeiten für theoriebasierte Evaluationen treffen zu können. Auch hier gilt dasselbe Prinzip wie in der kritischen Reflexion unterschiedlicher Programmtheorien und -modellierungen: Es kann nicht die eine richtige Methode geben, sondern je nach Rahmenbedingungen und Evaluationszweck kann ein methodischer Zugang angemessener sein als ein anderer.

Demzufolge muss in der Zusammenschau der verschiedenen Methoden geklärt werden, wie sie für unterschiedliche Aufgaben zur Umsetzung theoriebasierter Evaluationen genutzt werden können. Da Programmtheorien erschlossen und überprüft werden müssen, ist es unabdingbar zu erörtern, wie verschiedene Methoden kombiniert werden können. In Kapitel 6 wird – alternativ zu Methodenmix und Triangulation – die integrative Nutzung verschiedener methodischer Zugänge vorgeschlagen. Daraus resultieren unterschiedliche Integrationsstrategien, abhängig vom Auftrag an die Evaluation und dem Reifegrad des Evaluationsgegenstands.

Da sich Evaluationskonzepte in ihrer praktischen Umsetzung beweisen müssen und außerdem die Suche nach der hier vorgeschlagenen Alternative in der Praxis von Evaluationen entstand, sollen die methodologischen Überlegungen und konzeptionellen Vorschläge anhand eines konkreten Fallbeispiels diskutiert werden. Entgegen der üblichen Publikationspraxis, gut gelungene idealtypische Beispiele zu veröffentlichen, wird hier ein Exempel verwendet, dass eher für alltägliche Herausforderungen, Probleme und Missgeschicke steht. Beim Evaluationsgegenstand han-

2 Eine Ausnahme liefert Karin Haubrich (2009), die am Beispiel der Modellförderung in der Kinder- und Jugendhilfe die „Entwicklung der rekonstruktiven Programmtheorie-Evaluation“ ausführlich erörtert.

delte es sich um eine internetbasierte Lernumgebung, die für die Methodengrundausbildung im Rahmen des Soziologiestudiums an der Freien Universität Berlin entwickelt und eingesetzt wurde. Da die Autorin verantwortlich für die Lernumgebung und die Evaluation war, handelte es sich um eine Selbstevaluation. Dieses konkrete Vorhaben soll hier zur Illustration genutzt werden – nicht als Ideal für eine theoriebasierte Evaluation, vielmehr als ein typisches, dorniges Feld, in dem sich die Chancen, wie auch Grenzen theoriebasierter Konzepte zeigen. Das Beispiel wird parallel zu den hier vorgestellten Analyseschritten entfaltet, indem am Ende jeden Hauptkapitels die zentralen Thesen und Ideen anhand des Beispiels rekapituliert werden. Abschließend werden die Stärken und der Nutzen theoriebasierter Evaluationen vor dem Hintergrund der in der vorliegenden Arbeit aufgeworfenen Fragen zusammengefasst ebenso wie die noch zu lösenden Herausforderungen benannt.

2. Evaluation – Was ist das eigentlich?

> „You don't get very far in studying evaluation before realizing that the field is characterized by enormous diversity. From large-scale, long-term, international comparative designs costing millions of dollars to small, short evaluations of a single component in a local agency, the variety is vast. Contrasts include internal versus external evaluation; outcomes versus process evaluation; experimental designs versus case studies; mandated accountability systems versus voluntary management efforts; academic studies versus informal action research by program staff; and published, polished evaluation reports versus oral briefings and discussions where no written report is ever generated. Then there are combinations and permutations of these contrasting approaches." (Patton 1997, S. 64)

Der Begriff der Evaluation hat eine beachtliche Karriere zu verzeichnen. Bestanden vor gut einem Jahrzehnt bei etlichen noch erhebliche Schwierigkeiten, ihn richtig auszusprechen, ist er mittlerweile aus politischen Debatten und wissenschaftlichen Veröffentlichungen kaum noch wegzudenken und wird geradezu inflationär benutzt. Es ist zu vermuten, dass nur selten eine präzise Vorstellung darüber existiert, was sich hinter dieser „Zauberformel" verbirgt:[3]

> „‚Evaluation' ist zu einem äußerst unscharfen Modewort geworden und wird von manchen lediglich als ‚wohlklingendes' Fremdwort für jede Form von Bewertung oder Beurteilung verwendet." (Kromrey 2003, S. 106)

Zwar gibt es kaum noch einen gesellschaftlichen Bereich, der ohne Evaluation auszukommen scheint, jedoch entsteht zuweilen der Eindruck, dass der Begriff bei Auftraggebenden wie auch Auftragnehmenden wenig transparent ist und Entscheidungen über Methoden und Design willkürlich getroffen werden. Unzweifelhaft lässt sich der Begriff „Evaluation" immer noch als Nebelmaschine einsetzen: Er suggeriert Seriosität, Wissenschaftlichkeit und Wichtigkeit. Jede und jeder mag sich erhoffen (oder auch befürchten), was sie oder er will, der Fantasie sind kaum Grenzen gesetzt.

Gerade wenn im Deutschen von Evaluation die Rede ist – was ja zunächst nichts anderes als Bewertung bedeutet –, wird automatisch unterstellt, dass von einer spezifischen Form des Bewertens gesprochen wird, die nicht willkürlich und intuitiv erfolgt, sondern einer gewissen Systematik unterliegt.[4] Um die Unterscheidung zwischen Alltagshandeln und systematischem Vorgehen zu verdeutlichen, verwenden

3 Die Beliebigkeit in der Bedeutungszuweisung beschreibt auch die Aussage von Glass/Ellett (zit. nach Shadish u. a. 1991, S. 30): „Evaluation – more than any science – is what people say it is; and people currently are saying it is many different things."

4 Abgesehen von der Unterscheidung zwischen einem wissenschaftlichen und alltäglichen Bewertungsbegriff ist außerdem zwischen Evaluation als Ergebnis (z. B. in Form eines abschließenden Berichts bzw. eines Urteils) und Evaluation als Prozess/Tätigkeit zu trennen (Kromrey 2001a, 2005b).

einige Autorinnen und Autoren den Begriff „Evaluationsforschung", so z. B. Reinhard Stockmann 2004a und Uwe Flick 2006a. Auch Christian Lüders (2006, S. 49) trennt „zwischen Evaluation im Sinne einer professionellen Praxis einerseits und Evaluationsforschung im Sinne eines besonderen Typs sozialwissenschaftlicher Forschung andererseits". Hier wird im Folgenden der Begriff Evaluation vor allem verwendet, um bereits im Vorfeld eine Engführung auf ausschließlich forschendes Handeln zu vermeiden, denn Evaluation umfasst – wie noch zu zeigen ist – mehr als nur Datenerhebung und -analyse. Außerdem ist der Begriff „Evaluationsforschung" auch in semantischer Hinsicht irreführend, denn üblicherweise suggeriert er „Forschung über Evaluation", wie z. B. im Fall von Klimaforschung, Bildungsforschung oder Lebenslaufforschung (Hense 2006, S. 26f.).

> „Auf der Basis dieser Überlegungen scheint es angemessener, unter Evaluationsforschung die theoretische und empirische Forschung über Bedingungen, Praxis und Wirkungen von Evaluation zu verstehen." (Hense 2006, S. 27)

Die Antwort auf die Frage, wie denn nun Evaluation als systematisches Bewerten zu definieren sei, fällt kompliziert aus. Zu berücksichtigen ist, dass Evaluation unterschiedliche Gegenstände unter die Lupe nimmt, dass ihr variierende Aufgaben zugewiesen werden, sie verschiedene Zwecke verfolgt und dass ihr eine breite Palette an Vorgehensweisen zur Verfügung steht. Die Suche nach Definitionen macht deutlich, dass eine eindeutige Begriffszuweisung bereits eine beachtliche Hürde darstellt. Z. B. beginnen Egon G. Guba und Yvonna S. Lincoln (1989, S. 21) ihr Buch „Fourth Generation Evaluation" mit der Feststellung, dass Versuche zur einheitlichen Definition grundsätzlich überflüssig seien, und Barbara Lee (2004, S. 127) stellt ihrem Aufsatz eine Anekdote voran, die mit der Feststellung endet, dass niemand wirklich wisse, was Evaluation sei.

Betrachtet man gängige Definitionen von Evaluation, so wird deutlich, dass der Begriff multidimensional gefüllt wird. In einigen Fällen, so z. B. bei Peter H. Rossi und Howard E. Freeman,[5] liegt der Fokus der Begriffsbestimmung auf der Verwendung sozialwissenschaftlicher Verfahren, bei Michael Q. Patton hingegen auf Informationsbeschaffung,[6] Peter H. Rossi und James D.Wright bestimmen als Gegenstand Effekte und Wirkungen,[7] andere, wie z. B. Michael Scriven (1991, S. 1), zielen hingegen auf die Untersuchung von Güte und Wert („merit and worth"), wiederum andere, wie Daniel L. Stufflebeam (2000, S. 35) oder auch Donna M.

5 Für Rossi/Freeman (1993, S. 5) ist Evaluation „the systematic application of social research procedures in assessing social intervention programs".

6 Patton (1997, S. 23) benutzt einen breiten Begriff: „Program evaluation is the systematic collection of information about the activities, characteristics, and outcomes of programs to make judgments about the program, improve program effectiveness, and/or inform decisions about future programming."

7 Rossi/Wright (1986, S. 48): „(…) evaluation research, whose intent is the estimation of the net impacts or effects of social programs."

Mertens (2004, S. 40)[8], heben die Unterstützung von Entscheidungen als einen Zweck von Evaluationen hervor.

Der Vergleich der gängigen Bedeutungszuweisungen soll zeigen, dass mit dem Begriff Evaluation viele verschiedene Bedeutungen transportiert und damit auch zahlreiche Erwartungen geschürt werden können. Die Herausforderung besteht nun darin, eine ausreichend breite Begriffsbestimmung vorzunehmen, ohne dabei in Beliebigkeit zu verfallen. Das Ziel des folgenden Einführungskapitels besteht darin, die unterschiedlichen Erscheinungsformen von Evaluationen zu verdeutlichen. Darüber hinaus wird auf die neuralgischen Aspekte in Theorie und Praxis hingewiesen, die wichtigsten Positionen gegeneinander abgewogen und damit das Feld abgesteckt. Ausgehend von einem Alltagsverständnis von Evaluation – nämlich „irgend etwas wird von irgend jemanden nach irgendwelchen Kriterien in irgend einer Weise bewertet" (Kromrey 2000, S. 19) – werden die einzelnen Dimensionen genauer betrachtet: Gegenstände von Evaluation und deren typische Fragestellungen (2.2), die Durchführung von Evaluationen (2.3), verschiedene Konzepte von Bewertung (2.4) und die sich daraus ergebenden unterschiedlichen Rollen- und Aufgabenzuweisungen für Evaluatorinnen und Evaluatoren (2.5). Abschließend wird zusammenfassend das Spannungsfeld, in dem sich Evaluation bewegt, dargestellt.

Lohnend erscheint zuvor ein kurzer Blick auf die Entstehungsgeschichte der Profession „Evaluation" in den USA und Deutschland (2.1). Dieser soll verdeutlichen, dass Evaluation einerseits nicht „brandneu" ist, also durchaus auf theoretischen, methodischen und praktischen Erfahrungen und Reflexionen aus mehreren Jahrzehnten aufbauen kann. Andererseits stecken evaluationsspezifische professionelle Strukturen noch in ihren Kinderschuhen. Außerdem zeigt ein Blick auf die Entstehungsgeschichte, dass Evaluationen erst in ihrem jeweiligen Kontext verständlich werden. Sie fungieren als Bindeglied zwischen Wissenschaft, Politik und Gesellschaft, und ihre Aufgaben sowie Herausforderungen ergeben sich aus den jeweiligen politischen und gesellschaftlichen Bedingungen. Dabei haben sie regionale, nationalstaatliche und ebenso internationale Entwicklungen zu berücksichtigen.[9]

2.1 Entstehung einer neuen Profession

Die Karriere des Begriffs findet ihre Entsprechung in dem Boom eines relativ neuen Phänomens: Weder Bildung, Arbeitsmarktpolitik, Gesundheitswesen und Entwicklungszusammenarbeit noch Kinder- und Jugendhilfe, Regional- und Stadtentwicklung, Ökologie oder Forschungspolitik scheinen aktuell ohne Evaluation auszukommen.

---

8 Mertens (2004, S. 40) benennt als gängige Definition: „Evaluation is the systematic investigation of the merit and worth of an object (program) for the purpose of reducing uncertainty in decisionmaking."

9 „Though evaluation practice has to respect local context and so be tailored to some national realities, there are many other things about evaluation writ large that are supranational." (Cook 1997, S. 30)

Evaluation – im Sinne eines Alltagshandelns – ist selbstverständlich nicht neu.[10] Als systematische Beschreibung und Bewertung aber nahm sie ihren ersten zaghaften Ausgangspunkt in den USA der 1930/40er Jahre im Rahmen einer Reform des Bildungswesens, deren Erfolg mithilfe einer wissenschaftlichen Begleitung durch Ralph Tyler und sein Team überprüft werden sollte (vgl. Kapitel 3.1). Auslöser für das Entstehen einer neuen Profession ca. dreißig Jahre später war die von Präsident John F. Kennedy initiierte und unter Präsident Lyndon B. Johnson ausgeweitete „Great Society", die sich einem „Krieg gegen Armut" (War against Poverty) verpflichtet sah. Der Staat legte verschiedene Programme auf – von Bildung über Gesundheit bis hin zur Kriminalitätsprävention –, um Benachteiligungen und negative Folgen von Armut zu bekämpfen. Diese Maßnahmen waren mit erheblichen finanziellen Ressourcen ausgestattet.[11] So lag es nahe, dass die politisch Verantwortlichen (und vermutlich auch deren Gegnerinnen und Gegner) überprüft wissen wollten, ob sich die Investitionen lohnten und ob die Programme erfolgreich waren. Regierungseigene Gremien und Abteilungen (vor allem Ökonomen und Budget-Spezialisten) waren nicht in der Lage, den Verantwortlichen in der Politik brauchbare Antworten zu liefern – und so wurden Aufträge nach außen delegiert (Shadish u. a. 1991, S. 23f.). Auftragnehmende fanden sich in Universitäten und privaten Instituten.[12] 1967 erschien die erste von Edward Suchman veröffentlichte Monographie zu dem Thema mit dem Titel „Evaluative Research", und spätestens seit Beginn der 1970er Jahre folgten zahlreiche Publikationen, die sich vorwiegend mit methodologischen Fragen oder politischen Herausforderungen – wie beispielsweise der Nutzung von Evaluationen – beschäftigten.

Der neu entstandene Berufszweig erhielt einen ersten Dämpfer Anfang der 1980er Jahre, in denen im Zuge der „Reagonomics" nicht nur die Ausgaben für den Wohlfahrtsstaat und seine Programme, sondern dementsprechend auch für Evaluationen drastisch heruntergefahren wurden. Diese Entwicklung traf die Evaluatorinnen und Evaluatoren gerade in einer Zeit, in der sie heftige Auseinandersetzungen um die „richtigen" wissenschaftstheoretischen Paradigmen austrugen. Der enger werdende Markt sowie die konzeptionellen Debatten führten 1986 zum Zusammenschluss des Evaluation Network und der Evaluation Research Society zur American Evaluation Association (AEA) und schließlich zur Verabschiedung von Standards für die Durchführung von Evaluationen (Joint Committee 1999).[13]

In Deutschland waren die 1970er Jahre ebenfalls durch zahlreiche Reformanstrengungen in fast allen Politikbereichen geprägt. Auch hier stand zur Debatte, ob diese Veränderungen von Erfolg gekrönt waren und zu Verbesserungen führten. Zur

10 Treffend ist die Formulierung Scrivens (1987, S. 95): „Evaluation, like technology or psychology or philosophy, is a subject with a very long history and a much shorter conciousness."

11 Zahlen zum Umfang der staatlichen Aufwendungen für diese Programme finden sich in Shadish u. a. (1991, S. 22).

12 „Professional evaluation became a viable careeer alternative to academic employment. Thus evaluation met a need of the day, and a supply of labor existed to conduct its tasks, which led to a profession of evaluation." (Shadish u. a. 1991, S. 25)

13 Einen knappen Überblick über die Geschichte der US-amerikanischen Evaluation gibt Mertens (2004).

Beantwortung dieser und ähnlicher Fragen wurde die Wissenschaft herangezogen, auch wenn man weniger von Evaluation als von Erfolgsmessung und Monitoring (Eekhoff u. a. 1977), von Akzeptanz- und Wirkungsforschung (Kromrey 1988), von Implementationsforschung (Mayntz 1980) oder auch allgemein von wissenschaftlicher Begleitung sprach.[14] Die Bedeutung (sozialwissenschaftlicher) Forschung für die Planung sowie die Beurteilung von Politik wurde in den 1970er Jahren außerdem durch die Indikatorenbewegung umfassend diskutiert und bearbeitet (Zapf 1974-76).

Wissenschaftliche Begleitungen von Politik und Modellprojekten, vor allem im Bildungsbereich (vgl. z. B. Holz/Schlemme 2005), in der Stadt- und Regionalplanung (Eekhoff u. a. 1977; Kromrey 1976, 1981) sowie in der Entwicklungszusammenarbeit (vgl. Stockmann 2004b) blicken in Deutschland auf eine lange Tradition zurück. Der von Christoph Wulf 1972 herausgegebene Sammelband mit übersetzten US-amerikanischen Evaluationskonzepten sowie das 1974 in deutscher Sprache erschienene Handbuch von Carol Weiss (1972a) transportierten den Begriff „Evaluation" in die deutsche Debatte. Bezeichnend für diese sprachliche Wandlung ist beispielsweise, dass das noch 1978 unter dem Titel „Sanierungsmassnahmen. Städtebauliche und stadtstrukturelle Wirkungen" veröffentlichte Werk von Gerd-Michael Hellstern und Hellmut Wollmann 1983 in einer unveränderten Neuauflage in „Evaluationsforschung. Ansätze und Methoden – dargestellt am Beispiel des Städtebaus" umbenannt wurde.

Der entscheidende Durchbruch für die Benutzung des Begriffs „Evaluation" gelang jedoch erst in der zweiten Hälfte der 1990er Jahre. Dies war eine Phase, in der zum einen die Ausgaben der öffentlichen Hand für Bildung, Soziales und Gesundheit heruntergefahren wurden und zum anderen eine damit eng verwobene Qualitätsdebatte einsetzte. Die Botschaft lautete: Öffentliche Gelder müssen gezielter und effizienter eingesetzt werden, Zuwendungsempfänger sollen für Qualitätsentwicklung und -sicherung sorgen. Neben Qualitätsmanagementverfahren wie TQM, Zertifizierung nach ISO 9000ff. und Neuen Steuerungsmodellen (New Public Management) in der öffentlichen Verwaltung wird seither auch Evaluation als Verfahren vermehrt eingefordert und finanziert.[15]

Mit Evaluationen verknüpft sich die Hoffnung, das gängige „Rasenmäherprinzip" (gleichmäßige Kürzungen über alle Angebote und Einrichtungen hinweg) durch gezielte Einsparungen ersetzen zu können. So ist es kaum verwunderlich, dass in allen Bereichen, die nun mit Evaluationen konfrontiert sind – ob Hochschule, Jugendarbeit, Sozial- oder Gesundheitswesen, seit Ende der 1990er Jahre auch Schulen –, die Begeisterung nicht gerade überschwappt: Für die Praktikerinnen und Praktiker steht die Qualität der von ihnen geleisteten Arbeit infrage, neue Aufgaben kommen auf die Beteiligten zu, und gleichzeitig sehen sie ihre Arbeitsplätze durch Kürzungen bedroht. Allerdings erkennen vor allem Organisationsleitungen vermehrt, dass alleine

14 Haubrich (2006, S. 103) spricht im Fall der gängigen Praxis, Modellvorhaben wissenschaftliche Begleitungen zur Seite zu stellen, von einer „deutschen Besonderheit, die keine Entsprechung in anderen Ländern findet."

15 Zu den Gemeinsamkeiten und Unterschieden zwischen Qualitätsmanagementverfahren und Evaluation siehe Stockmann (2006).

die Tatsache, dass Evaluationen durchgeführt werden – unabhängig von deren Seriosität – bereits ein Wettbewerbsvorteil sein kann. Stockmann (2004a, S. 15f.) spricht in diesem Zusammenhang z. B. von einer „taktischen Funktion" oder von „dekorativen Symbolen für eine moderne Politik", die die eher „pathologische Seite" von Evaluation darstellen.

Weitere Impulse – einerseits zusätzliche Mittel, andererseits spezifische Anforderungen – bekommen Evaluationen durch die Stärkung der Europäischen Union.[16] Diese ist zu einer der wichtigsten Finanzierungsquellen arbeitsmarkt- und strukturpolitischer Maßnahmen geworden. Die von der EU geförderten Programme, Projekte und Maßnahmen sind fast durchweg zur Durchführung von Evaluationen verpflichtet.

Parallel zur gestiegenen Nachfrage entsteht folgerichtig ein neues Tätigkeitsfeld für die Sozialwissenschaften: Lehrstuhlinhaberinnen und Lehrstuhlinhaber an (Fach-)Hochschulen, Personen, die bisher als Praxisforschende tätig waren, Beschäftigte aus Programmen und Verwaltung, private Forschungs- und Beratungsinstitute oder Einzelunternehmen sowie sozialwissenschaftliche Hochschulabsolvierende werden mit der „neuen" Aufgabe betraut. Ähnlich wie in den USA der 1960er Jahre steigt die Nachfrage nach Evaluationen rasant, jedoch fehlen adäquate Durchführungsstrukturen und qualifiziertes Personal. Auf der europäischen Ebene wurde die steigende Bedeutung von Evaluation früh erkannt: Bereits Ende der 1980er Jahre gründete sich die European Evaluation Society (nahezu zeitgleich mit der britischen UK Evaluation Society).

In Deutschland begannen Ende der 1990er Jahre die ersten groß angelegten Professionalisierungsoffensiven. 1998 fand die Gründungsveranstaltung der Deutschen Gesellschaft für Evaluation (DeGEval)[17] statt, die jährlich eine Tagung durchführt, an der mittlerweile mehrere hundert Personen teilnehmen. 1999 erschienen die von Wolfgang Beywl und Thomas Widmer übersetzten Standards der American Evaluation Association (Joint Committee 1999), womit die Verabschiedung der deutschsprachigen Standards (DeGEval 2002, SEVAL 2001) forciert wurde. Seit 2002 werden pro Jahr zwei Ausgaben der „Zeitschrift für Evaluation" (ZfEv) veröffentlicht, die neben methodologischen und theoretischen Beiträgen auch Praxisbeispiele enthalten. Ab Mitte der 1990er Jahre werden vereinzelt berufsqualifizierende Qualifikationen angeboten. Seit 2002 kann man in Bern den Nachdiplomstudiengang „Evaluation" absolvieren. Ein Jahr später entstand der erste deutsche Masterstudiengang in Saarbrücken, der seit dem Sommersemester 2008 auch an der Universität Bonn angeboten wird.[18]

---

16 Eine Einführung zur europäischen Evaluationspolitik findet sich in Leeuw (2004). Auf einige der spezifischen Herausforderungen geht z. B. Stame (2004) ein.

17 Während der IX. Jahrestagung 2006 wurde die „Deutsche Gesellschaft für Evaluation" in „Gesellschaft für Evaluation (DeGEval)" umbenannt, weil vor allem für österreichische Durchführende und Auftraggebende von Evaluationen die Türen offen stehen sollen.

18 Ausführlich dargestellt sind die Entwicklung der Evaluation in Deutschland sowie der Professionalisierungsstand in Brand (2009).

Als Zwischenfazit lässt sich feststellen: Zweifellos ist mit Evaluation ein vielfältiges, ertragreiches und gleichermaßen anspruchsvolles Beschäftigungsfeld für Sozialwissenschaftlerinnen und Sozialwissenschaftler entstanden. Jedoch stehen nicht nur Einsteigerinnen und Einsteiger vor grundlegenden praktischen und methodischen Herausforderungen. Aufgrund der Unschärfe des Begriffs besteht die Gefahr, dass Evaluation als Verfahren diskreditiert wird. Vor allem die enge Anbindung an Qualitätsmanagementverfahren steigert das Risiko, Evaluationen auf die schlichte Feststellung von „Kundenzufriedenheit" zu reduzieren.[19] Neben den „Standards für Evaluation" bietet ein kaum zu überschauender und zum Teil nicht leicht zugänglicher Berg an Literatur (von grundlegenden theoretischen und methodologischen Texten bis hin zu Praxisberichten) hilfreiche Orientierungen. Zu berücksichtigen sind neben den US-amerikanischen und europäischen Quellen auch die reichhaltigen Erfahrungsschätze aus wissenschaftlichen Begleitungen in den 1970/80er Jahren in Deutschland.

2.2 Irgendetwas – Gegenstände und typische Fragestellungen

Grundsätzlich kann mehr oder weniger alles evaluiert werden, sogar Evaluationen selbst (Shadish u. a. 1991, S. 19). In Anlehnung an John M. Owen und Patricia J. Rogers (1999, S. 22-38) lässt sich eine erste Grobeinteilung möglicher Evaluationsgegenstände in Programme, Organisationen, Produkte und Personen vornehmen.[20] Diese Gegenstandsbereiche sind allerdings nicht trennscharf voneinander abgrenzbar: Organisationen führen von der Politik verabschiedete Programme durch, die Durchführung erfolgt durch Personen, die sich innerhalb von Organisationen bewegen, und dabei entstehen Produkte.

Zwar gibt es zahlreiche Bewertungsverfahren von Personen, angefangen bei Schulnoten bis hin zu Personalauswahlverfahren (z. B. durch Assessment-Center), jedoch werden diese im deutschsprachigen Raum (noch) nicht unter dem Label „Evaluation" entwickelt und angewandt. Sie werden dementsprechend im Folgenden ausgespart.[21] Auch die Spezifik von Organisationen als Evaluationsgegenstand wird hier außer Acht gelassen, obwohl in Deutschland Organisationsevaluationen weit verbreitet und aktuell sind; vor allem ist hier die Hochschulevaluation zu nennen. Auch wenn sich zahlreiche Parallelen zur Programmevaluation ziehen lassen, ergeben sich besondere Schwierigkeiten. Helmut Kromrey (2004, S. 237) bemerkt zum Evaluationsgegenstand Hochschule: „Es existiert weder ein präzise beschreibbares ‚Programm' mit klar definierten Zielen und ihnen zugeordneten Maßnahmen sowie ein-

19 Ein typisches Beispiel sind Lehrevaluationen an Hochschulen, die einzig auf Zufriedenheitsmessungen bei Studierenden abstellen. Zur deren Kritik siehe auch Kromrey (z. B. 1994, 1999, 2004).

20 Owen/Rogers (1999, S. 31ff.) erwähnen außerdem „Politik" als einen möglichen Evaluationsgegenstand. Da in der Praxis die Evaluation eines solchermaßen allgemein gehaltenen Gegenstands nicht in Auftrag gegeben wird, wird er im Folgenden ausgeblendet.

21 In den USA hingegen existieren Standards für Personenevaluationen (Joint Committee on Standards for Educational Evaluation 2011).

deutig festgelegte Zielerreichungskriterien noch ein konkretes Produkt (…)."[22] In der Praxis wird dieses Dilemma dadurch aufgelöst, dass sich Hochschulen auf Verfahren einigen, die das Vorgehen sowie die zu beleuchtenden Aspekte festschreiben.[23] Aktuell spielen Evaluationen eine wichtige Rolle im Vorlauf auf anstehende Zertifizierungen, beispielsweise von Studiengängen oder medizinischen Einrichtungen.[24]

Produkte entstehen im Rahmen von Programmen, ihre Evaluation kann also im Rahmen der Bewertung von Programmen erfolgen. Typische Produkte sind Lernmaterialien, Konzepte, Ablaufpläne, Software etc. Speziell im Rahmen des Einsatzes so genannter „Neuer Medien" nimmt die Evaluation von Produkten eine herausragende Rolle ein. Für die Einschätzung von Learning Management Systemen und Lern-CDs bestehen bereits umfassende Verfahrensvorschläge und Leitfäden (z. B. Baumgartner u. a. 2002a), um beispielsweise eine Entscheidung zu treffen, was angeschafft werden soll.

Im Mittelpunkt dieser Arbeit stehen *Programm*evaluationen, da sie die größte Verbreitung vorzuweisen haben, weil sie umfassend funktionieren und sie letztlich implizit Produkte und Organisationsstrukturen mit erfassen können müssen.[25] Sie stehen damit vor spezifischen Herausforderungen und sind zudem weniger auf konkrete Felder zugeschnitten.

> „Als *Programme* sind komplexe Aktionsmodelle zu bezeichnen, die auf die Erreichung bestimmter Politikziele gerichtet sind, auf bestimmten Handlungsstrategien beruhen und für deren Abwicklung bestimmte finanzielle, personelle und sonstige administrative Ressourcen (Richtlinien, ‚flankierende Maßnahmen' usw.) bereit gestellt werden." (Hellstern/Wollmann 1983, S. 7, Herv. im Orig.).

Diese Definition verweist auf die kaum zu überschauende Komplexität von Programmen und unterstreicht, dass es sich nicht um konkret fassbare und abgegrenzte Gegenstände handelt, sondern vielmehr um Konstruktionen, die im Zuge der Untersuchung von Forschenden geleistet werden (vgl. Mayntz 1980). Und doch weisen Programme strukturelle Gemeinsamkeiten auf. Unter anderem funktionieren sie dadurch, dass Ideen und Pläne umgesetzt und von Zielgruppen mehr oder weniger angenommen werden können. Die Interventionen verfolgen die Absicht, bei den erreichten Zielgruppen Veränderungen (oder auch Stabilisierungen) herbeizuführen, und die erreichten Personen sowie die Interventionen sind wiederum vielfältigen externen Einflüssen ausgesetzt.

---

22 Ganz ähnlich benennen Owen/Rogers (1999, S. 34) als Problem für Organisationsevaluationen im Allgemeinen: „(…) it is more difficult to describe the ends-means and cause-effect relationships by which the outcomes are delivered – in other words, it is more difficult to describe the underlying program logic."

23 Einen Überblick über die wichtigsten Verfahrenswege geben Mittag u. a. (2003).

24 Zur Abgrenzung und Überschneidung zwischen Zertifizierung und Evaluation sei auf Kromrey (2005a) verwiesen.

25 Scriven (1987, S. 99) hält die Fokussierung auf Programmevaluation für einen grundsätzlichen Fehler: „If social scientists had spent more time thinking about (…) product evaluation (…) we would have detected immediately the anti-consumer and indeed anti-scientific bias in the approach."

Im Zuge des Imports von Qualitätsmanagementverfahren und Evaluationskonzepten aus dem Amerikanischen verwendet man in Deutschland zahlreiche Begriffe, die in der Literatur und den Fachdebatten nicht unbedingt stringent verwendet werden. Deswegen werden zunächst die für diese Arbeit gewählten Bedeutungszuweisungen erläutert (vgl. Tabelle 1):

Tabelle 1: Die wichtigsten Begriffe im Überblick[26]

Inputs	Ressourcen, die in ein Programm einfließen, wie z. B.: finanzielle Zuwendungen, Qualifikationen und Kompetenzen von Durchführenden
Incomes	Voraussetzungen und der Bedarf von Zielgruppen, deren Einstellungen, Wissen und Fertigkeiten zum Programmbeginn
Outputs	Produkte und Leistungen des Programms, wie z. B.: durchgeführte Beratungen, Seminarkonzepte, Absolventenzahlen
Outcomes	Veränderungen und Folgen für Zielgruppen durch das Programm in Wissen, Einstellungen, Verhalten, wie z. B. Kompetenzzuwachs
Impacts	(Strukturelle) Veränderungen und Folgen über die eigentlichen Zielgruppen hinaus, z. B. neue Gesetze, Gremien, Entscheidungswege, Verfahren
Effekt	Gesamtwirkung eines Programms bzw. Gesamtheit der Veränderungen in Outcomes und Impacts inclusive der nicht-intendierten Wirkungen
Effektivität	Wirksamkeit relativ zum Input und Income oder auch zu alternativen Strategien
Effizienz	Verhältnis von Kosten und Nutzen, das Verhältnis von Input zu allen angestrebten und erreichten Resultaten.

Eine relevante Dimension eines Programms kann z. B. dessen Output sein, d. h. ein Produkt, das Ergebnis von Aktivitäten ist. So können Outputs einer Bildungsmaßnahme beispielsweise Absolventenzahlen sein, für ein Arbeitsvermittlungsprojekt vermittelte Praktikums- oder Arbeitsplätze, für eine Gesundheitspräventionsinitiative die erreichte Anzahl von Teilnehmenden einer Informationsveranstaltung. Outcomes eines Programms bezeichnen hingegen Veränderungen und Folgen, die sich durch Aktivitäten und Produkte bei den Zielgruppen ergeben. Für die Bildungsmaßnahme kann sich der Outcome in erworbenem Wissen oder erlernten Kompetenzen zeigen, langfristig evtl. in einem verbesserten Zugang zum Arbeitsmarkt. Die Vermittlung von Praktika oder Beschäftigung kann zu einer mittel- oder langfristigen Integration in den Arbeitsmarkt führen, sie kann ein gesteigertes Selbstbewusstsein sowie eine Verbesserung der materiellen und immateriellen Lebenssituation zur Folge haben. Die gesundheitspräventive Informationsveranstaltung kann zu neuem Wissen, zu neuen Einstellungen und evtl. sogar zu gesünderem Verhalten führen. Darüber hinaus können Programme auch einen Impact erzielen, nämlich Veränderungen und Folgen, die über die eigentliche Zielgruppe hinausgehen. Die Bildungsmaßnah-

26 Die Begriffserläuterungen orientieren sich am Glossar von Univation (2004).

me könnte von anderen Organisationen aufgegriffen und übernommen werden. Aus dem Vermittlungsprojekt könnte sich ein Netzwerk von Betrieben entwickeln, das sich z. B. für Verbundausbildungen stark macht; die Informationsveranstaltung zur Gesundheitsprävention könnte für alle Eltern von Kita-Kindern verbindlich angeboten werden.

Vor allem für Outcomes und Impacts muss davon ausgegangen werden, dass sie erst mittel- oder gar langfristig zustande kommen. Besonders augenfällig ist dieser Umstand für Präventionsprogramme, bei denen es zumeist darum geht, Verschlechterungen in der Zukunft zu verhindern. Auch bei der Vermittlung von Beschäftigung wird wohl relevant sein, ob die Beschäftigung eine kurzfristige oder langfristige Integration in den Arbeitsmarkt bedeutet. Für erworbenes Wissen ist es wichtig, wie lange das Wissen präsent bleibt und welche Anwendungsmöglichkeiten es für erworbene Kompetenzen auf längere Sicht gibt.

Outputs, Outcomes und Impacts können geplant und absichtsvoll verfolgt werden, sie können jedoch auch nicht-intendiert und unbeabsichtigt zustande kommen. Das Konzept der Bildungsmaßnahme könnte beispielsweise zu erhöhten Abbruchquoten führen; die in Arbeit Vermittelten könnten weit unter ihrer Qualifikation arbeiten und deswegen unzufrieden sein; die Eltern könnten die Verantwortung für die Gesundheitsprävention an Kitas übertragen.

Zusätzlich komplizierter werden Programme dadurch, dass sie nicht abgeschottet von ihrer Umwelt stattfinden. Vielmehr sind sie in einen konkreten Kontext eingebettet und müssen vor dem Hintergrund gegebener Bedingungen betrachtet werden. Berücksichtigt werden müssen beispielsweise die vorhandenen bzw. zur Verfügung gestellten Ressourcen (Input) und die Merkmale der Zielgruppen. Soll eine Bildungsmaßnahme beurteilt werden, so ist ausschlaggebend, ob sie sich an Schulverweigerer oder an intrinsisch motivierte Schülerinnen und Schüler richtet. Ebenso beeinflussen auch regionale Bedingungen ein Programm: Vermittlungserfolge von Jobcentern in Bayern und Mecklenburg-Vorpommern werden unterschiedlich ausfallen.

Programme werden gänzlich unübersichtlich, wenn man berücksichtigt, dass Bedingungen selbstverständlich nicht stabil bleiben, sondern sich permanent verändern (können). Die Entwicklung des Programms und seiner Bestandteile ist sich wandelnden externen Einflüssen ausgesetzt. Das kann soweit gehen, dass Zielvorgaben modifiziert sowie Pläne angepasst werden und sich dementsprechend auch der Ertrag verändert.

Neben diesen verschiedenen Dimensionen von Programmen bieten sie als Gegenstand

> „ein kaum vollständig aufzählbares Spektrum an möglichen Variationen (…): Der zu evaluierende Sachverhalt kann schon lange bestehen, sich gerade im Prozess der Realisierung befinden oder gar erst als Planungs- und Entwicklungsabsicht existieren; er kann sehr umfassend und abstrakt oder aber eng umgrenzt und konkret sein; er kann (im Sinne ‚experimenteller Politik') ein Pilotvorhaben sein, das in einem abgegrenzten (zumindest prinzipiell abgrenzbaren) Feld durchge-

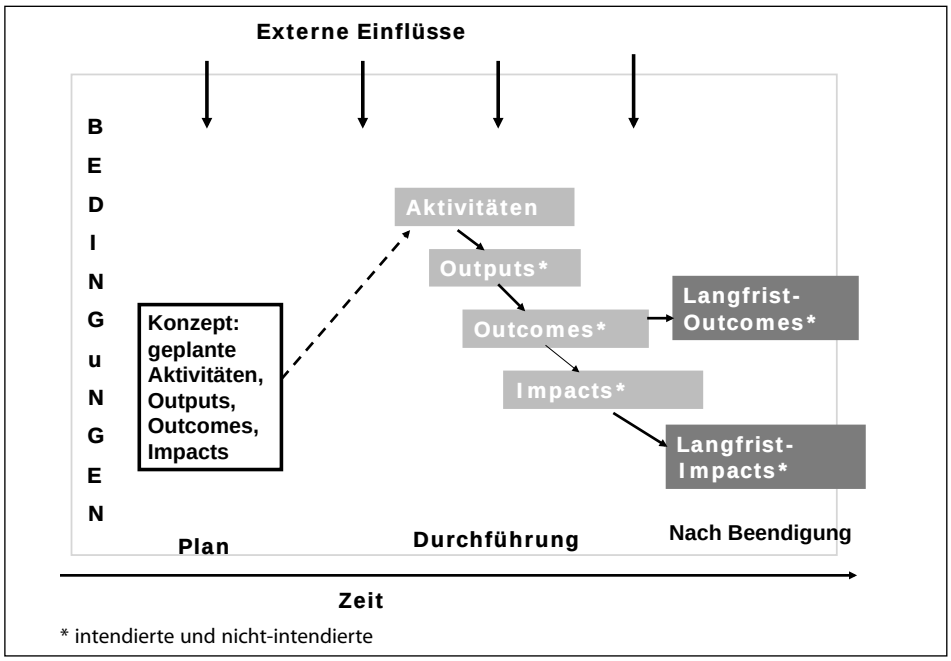

Abbildung 1: Das Programm im Überblick

führt wird, oder aber eine Innovation, die sich in Konkurrenz zu bestehenden Angebotsalternativen behaupten soll." (Kromrey 2001a, S. 109)

Zu der bislang ausgeführten Vielfalt von Programmen und deren Dimensionen ist weiterhin zu ergänzen, dass vor allem dann Evaluationen angestrengt werden, wenn es sich um Humandienstleistungen handelt, die nicht am Profit orientiert sind. Gerade im Non-Profit-Bereich spielen Evaluationen eine herausragende Rolle, da ein wichtiges und zentrales Erfolgskriterium, nämlich monetärer Gewinn, gerade nicht zur Bewertung und Beurteilung herangezogen werden kann. Und doch besteht ein Bedarf, den „Mehrwert" oder Erfolg der zumeist durch Steuergelder finanzierten Aktivitäten zu bemessen.[27] Außerdem zeichnen sich Humandienstleistungen dadurch aus, dass sie nicht darauf ausgerichtet sind, Gebrauchsgüter zu produzieren. Ihr zentraler Kern besteht vielmehr in sozialer Interaktion. Sie basieren darauf, „dass eine Austauschbeziehung zwischen Menschen im Kontext ihrer Lebenswelt entsteht" (Kromrey 2000, S. 24). Welche tiefgründigen und kompliziert zu überprüfenden Mechanismen in sozialen Interaktionen zur Geltung kommen, zeigt sich anschaulich bei doch relativ leicht standardisierbaren Medikamententests, in denen sogenannte Placeboeffekte zum Tragen kommen (siehe z. B. Suchman 1967, S. 96ff.). Auch die

27 Berk/Rossi (1990, S. 7) formulieren das so: „Program evaluation derives from the commonsense idea that social programs should have demonstrable benefits." Chen (2005, S. 3) schreibt: „… they all share the common feature of being organized efforts to enhance human well-being – wether by preventing desease, reducing poverty, or teaching skills."

Feststellung von Wirkungen und Nebenwirkungen bereitet selbst bei dieser klar eingrenzbaren Intervention erhebliche Probleme.

Damit steht die Evaluation von sozialen Dienstleistungen vor besonderen Herausforderungen. Statt an relativ unproblematisch festzustellenden Outputs, besteht das Interesse vor allem an Outcomes und Impacts (Kromrey 2000, S. 24). Nicht nur, dass die Folge von Programmen zumeist nur zeitverzögert beobachtbar ist, hinzukommt, dass Veränderungen nicht ohne Weiteres auf Programme zurückgeführt werden können.

Aus diesen vielfältigen Komponenten und unzählbaren Varianten von Programmen lassen sich unterschiedliche Fragestellungen für Evaluationen ableiten: Werden Konzepte wie geplant umgesetzt? Welches sind die Bedingungen, unter denen das Programm umgesetzt wird, verändern sie sich, bleiben sie stabil? Werden die angestrebten Outputs, Outcomes und Impacts kurz-, mittel- oder langfristig erreicht? Sind diese Veränderungen und Ergebnisse dem Programm oder anderen Einflüssen zuzuschreiben? Welche externen Einflüsse behindern oder fördern das Erreichen der Erfolge? Wie ist das Verhältnis zwischen Input zu Output und Outcome? Treten negative „Nebenwirkungen" oder auch positive Effekte auf, mit denen im Vorfeld nicht gerechnet wurde? Dies ist nur ein kleiner Ausschnitt generischer Fragestellungen an Evaluationen (siehe auch Kromrey 2000, S. 21).

Um Ordnung in die Vielfalt von möglichen Evaluationen zu bekommen, bietet sich eine Klassifizierung nach verschiedenen Zwecken an. Eine erste Unterscheidung führte Scriven (1972a, 1972b) zwischen summativer und formativer Evaluation ein, die auch heute noch Beachtung findet. Erstere hat die Aufgabe, bilanzierend darüber Auskunft zu geben, welche Ergebnisse ein Programm erzielen konnte. Letzterer geht es hingegen darum, von der Entwicklung des Konzepts über die Durchführung bis hin zu den Ergebnissen ein Programm zu begleiten und solche Informationen beizusteuern, die helfen, das Programm mitzugestalten.[28] Angesprochen ist in dieser Unterscheidung auch die zeitliche Dimension: Summative Evaluationen gehen zumeist rückblickend, ex-post vor, formative finden parallel zum Programm statt.

Eine weitere nützliche Unterteilung nach Funktionen ist die von Eleonor Chelmisky (1997) eingeführte Differenzierung zwischen drei Perspektiven, die von Kromrey (2000, 2001a, 2004, 2005b) mit der Einteilung in Forschungs-, Kontroll- und Entwicklungsparadigmen für die deutsche Debatte aufgegriffen wurde.[29] Diese jeweiligen Perspektiven verfolgen unterschiedliche Verwertungsinteressen der Evaluationsergebnisse, haben Konsequenzen für die Akzentsetzung und werden vorzugsweise von bestimmten Auftraggebenden von Evaluationen eingenommen: Programmfinanzierenden wird daran gelegen sein, Informationen darüber zu erhalten, ob die von ihnen investierten Mittel ihren Absichten gemäß eingesetzt wurden, ob

28 Zur Veranschaulichung des Unterschieds zwischen formativ und summativ wird immer wieder das folgende anschauliche Bild von Stake angeführt: „When the cook tastes the soup, that's formative; when the guests taste the soup, that's summative." (zit. nach Patton 1997, S. 69)

29 Chelimsky (1997, S. 10ff.) spricht von „knowledge", „accountability" und „development" Perspektiven.

die vorgegebenen Ziele erreicht wurden oder ob die Erfolge mit geringeren Ressourcen zu erzielen sind, um eine Entscheidungsgrundlage über Fortführung, Ausweitung oder Beendigung von Programmen zu erlangen (Kontrolle). Damit verfolgt diese Art von Evaluation einen „Beitrag zur Planungsrationalität durch Erfolgskontrolle des Programmhandelns" (Kromrey 2001a, S. 114).

Programmdurchführenden wird eher daran gelegen sein, Informationen darüber zu erhalten, welche Strategie gut funktioniert und welche Verbesserungsmöglichkeiten bestehen, um das eigene Vorgehen weiterzuentwickeln. Damit wird Evaluation zu einem Teil und Gestaltungselement von Programmen (Entwicklung). Im Fall des Forschungsparadigmas sind Evaluatorinnen und Evaluatoren interessiert an der „Verbreiterung der Wissensbasis" (Kromrey 2000, S. 25), insbesondere an Handlungswissen über die Wirksamkeit von Interventionen. Solche Evaluationen sind vorwiegend als Wirkungsforschung angelegt.

Selbstverständlich lassen sich Schnittmengen bezüglich der Zwecke und Perspektiven von Evaluationen ausmachen: Eine programmdurchführende Organisation könnte z. B. an dem Beleg interessiert sein, erfolgreich gearbeitet zu haben, um weiterhin finanziert zu werden; staatliche Institutionen greifen evtl. auf Wirkungsstudien zurück, um künftige Programmausschreibungen zu gestalten. Trotz dieser Schnittmengen ist eine Unterscheidung nach Funktionen von zentraler Bedeutung, denn aus dem grundlegenden Zweck und der leitenden Perspektive ergeben sich Konsequenzen für unumgängliche Priorisierungen. Programme sind so vielfältig, dass es in der Regel nicht möglich ist, sie in all ihren Facetten zu untersuchen und zu bewerten.[30] So stellt Kromrey (2001a, S. 109) fest:

> „Selbst wenn eine ‚umfassende Evaluation' (…) angestrebt würde, wäre doch noch (stark selektiv) zu entscheiden, welche Teilaspekte denn tatsächlich im Detail einer systematischen Beurteilung unterzogen werden sollen und welche allenfalls als Randbedingungen berücksichtigt werden könnten. Jede Evaluation wäre überfordert, wollte sie ein Programm, eine Einrichtung o. ä. quasi ‚ganzheitlich' *zu ihrem Gegenstand machen*." (Herv. im Orig.)

Auch für die Aufgaben, Kompetenzen und Rollen, die Evaluatorinnen und Evaluatoren zukommen, bedeuten die dargestellten Paradigmen zentrale Wegweiser. Das Gelingen einer Evaluation wird entscheidend davon abhängen, dass der Zweck frühzeitig und eindeutig geklärt sowie transparent gemacht wird.

---

30 Chelimsky (1997, S. 22): „(...) the perspectives represent different ways to think about evaluation, and each tends to solve particular evaluative problems while complicating others. None of them, however, can solve all of the problems or answer all of the questions posed to evaluation; (...) by treating evaluation as a single entity, we allow a certain blurring of the perspectives to occur that further raises tensions, as, for example, when claims are made about the universal applicability of a particular evaluation approach that in fact adresses the questions of one perspective much better than it does those of the others."

2.3 Irgendwie – Das Vorgehen von Evaluationen

Aufgrund des kaum zu überblickenden Felds, der Komplexität der möglichen Gegenstände, der Verschiedenartigkeit der Zwecke von Evaluationen und der Multiplizität von Fragestellungen kann es auch keine einfache und eindeutige Antwort auf die Frage nach dem „Wie" geben. Zunächst einmal besteht Einigkeit darüber, dass Evaluationen von intuitiven, willkürlichen Bewertungen abzugrenzen sind, stattdessen systematisch vorzugehen haben. Diese Systematik wird durch die Einbeziehung vorliegender und/oder neu erhobener Daten erreicht. Auf welche Art und Weise Daten erhoben, analysiert und interpretiert werden, darüber gibt es unterschiedliche Ansichten.

Die in der Evaluation angewandten Methoden speisen sich aus verschiedenen Quellen. Zuallererst ist festzustellen, dass Evaluatorinnen und Evaluatoren in der Regel nicht als solche das Feld betreten, sondern dass sie in unterschiedlichen wissenschaftlichen Disziplinen sozialisiert wurden. Zu nennen sind vor allem die Psychologie, die Erziehungs- und Politikwissenschaften, die Soziologie, aber auch Medizin, Wirtschafts- und Rechtswissenschaften ebenso wie Technik und Ingenieurswesen. Jede Disziplin steuert ein mehr oder weniger eigenständiges Methodenrepertoire und spezifische theoretische Konzepte bei.[31] Zusätzlich – auch innerhalb der einzelnen Disziplinen – sind verschiedene wissenschaftstheoretische Paradigmen vertreten: vom Positivismus über den Realismus hin zum Konstruktivismus, von kritisch-rationaler Sozialforschung über das interpretative Paradigma hin zur Phänomenologie.

Grundsätzlich – hierin besteht weitestgehend Konsens in der Literatur – kann sich Evaluation aller Verfahren empirischer Forschung bedienen. Im Vorhinein sind weder bestimmte Designtypen noch Erhebungsverfahren gänzlich auszuschließen. Das „Geheimnis" besteht jedoch darin, dass die für den jeweiligen Informationsbedarf angemessenen Verfahren eingesetzt werden.[32] Deswegen lassen sich – unabhängig vom Design – für alle Evaluationen folgende Arbeitsschritte vorgeben (vgl. Abbildung 2).

Unter methodischen und methodologischen Gesichtspunkten besteht das Ziel also darin, auf die jeweiligen Zwecke eng abgestimmte oder wie Rossi/Freeman (1993) es nennen, „maßgeschneiderte" (tailored) Designs zu entwickeln.

Neben der wissenschaftlichen Basis schöpfen Konzepte zur Durchführung auch aus der Auseinandersetzung mit den praktischen Erfahrungen in Evaluationsprojekten und Programmen.[33] Neben der akademischen Ausbildung bringen viele Evaluatorinnen und Evaluatoren Kenntnisse aus der Programmadministration, der -durchführung und -begleitung und damit auch Feldkompetenzen ein. Vor der bzw. parallel zur Evaluation sind oder waren sie Fachkräfte im Bildungsbereich, im Gesundheits-

31 Weiss (1972a, S. XII) spricht diesbezüglich pointiert von einer „balkanization of the territory".

32 In der US-amerikanischen Boomphase der Evaluation in den 1970er Jahren wird der Löwenanteil der Evaluationen von Privatinstitutionen durchgeführt (Shadish u. a. 1991, S. 24).

33 In der US-amerikanischen Boomphase der Evaluation in den 1970er Jahren wird der Löwenanteil der Evaluationen von Privatinstitutionen durchgeführt (Shadish u. a. 1991, S. 24).

Abbildung 2: Grobverlauf von Evaluationen

wesen, in Erziehung und in Sozialer Arbeit, in arbeitsmarktpolitischen Feldern, in der Gewaltprävention bis hin zu Umweltschutz und Stadtentwicklung, um nur einige zu nennen. Zumindest in den USA gab es gar eine Debatte, ob Feldkompetenz eine zwingende Voraussetzung zur Durchführung von Evaluationen sei (Lee 2004, S. 128).

Diese breit gestreuten Quellen von Wissen und Kompetenzen bedeuten natürlich einen reichen Erfahrungsschatz für die Evaluations-Community.[34] Gleichzeitig birgt diese Vielfalt jedoch auch Gefahren. Zum einen trägt sie nicht unbedingt zur Profilierung der Profession und damit zu einer Abgrenzung bei, was Evaluation leisten kann und was nicht. Außerdem ist speziell im deutschsprachigen Raum zu beobachten – beispielsweise an der Organisation der DeGEval[35] in feldbezogene Arbeitskreise, an den Veröffentlichungen in der „Zeitschrift für Evaluation" (ZfEv) und auch in anderen Publikationen –, dass das „Wie" von Evaluationen sehr feldspezifisch behandelt wird, was wiederum zu einer Vernachlässigung von feldübergreifendem Austausch und Kommunikation führen kann.

Auch wenn es durchaus Vorhaben gibt, die nach forschungsparadigmatischer Logik vorgehen, ist Evaluation jedoch eindeutig nicht der Grundlagen-, sondern der

34 Lee (2004, S. 127) kann der Unbestimmtheit und dem breiten Erscheinungsbild von Evaluation so auch einen gewissen Charme abgewinnen: „On the positive side, this has resulted in a wonderful level of creativity and diversity in evaluation methods and thinking."

35 Über die Organisationsstruktur der DeGEval finden sich Informationen unter www.degeval.de, zur Schweizerischen Evaluationsgesellschaft unter http://www.seval.ch.

angewandten Sozialforschung zuzurechnen.[36] Laut Evert Vedung (2004, S. 113) besteht der zentrale Unterschied zwischen angewandter Forschung und Grundlagenforschung darin, dass Ergebnisse der ersteren zur Nutzung vorgesehen sind. Er unterscheidet insgesamt fünf Nutzungsarten: instrumentelle, konzeptionelle (enlightenment), taktische, diskursive und legitimierende.

Bis in die 1980er Jahre hinein gingen viele Evaluationsansätze wie selbstverständlich von einem rationalen „Problem-Lösungs-Konzept" aus, bei dem erstens das Problem klar definiert ist, zweitens potentielle Lösungen vorgeschlagen, drittens ausgewählte Lösungsstrategien ausgeführt, die dann viertens evaluiert werden. Im Anschluss daran verbreitet sich fünftens das Wissen um erfolgreiche Lösungen und sechstens nutzen die politisch Verantwortlichen dieses Wissen für anstehende Entscheidungen (Cook/Shadish 1987, S. 34). Auch die Indikatorenbewegung legt ihren Überlegungen ein rationales Planungsmodell zugrunde, „beginnend mit der grundlegenden Zielbestimmung über die empirische Feststellung von Handlungsbedarf, die Entwicklung von gedanklichen Handlungskonzepten bis zu deren Implementation und nachträglicher Erfolgskontrolle" (Kromrey 2003, S. 101).

Diese Modelle halten der Wirklichkeit nicht stand. Einerseits können Wissenschaftlerinnen und Wissenschaftler ihre Versprechen nicht einlösen; sie scheitern an der Wandelbarkeit und Komplexität der gesellschaftlichen Realität. Andererseits sind auch Politik und Praxis nicht bereit, sich ausschließlich auf die Wissenschaft als Informationsbasis zu verlassen.[37] Vielmehr ist davon auszugehen, dass Evaluationsergebnisse in der Vorbereitung auf Entscheidungen eine Informationsquelle unter vielen anderen darstellen (z. B. Weiss 1977). Seit den 1970er Jahren sind demzufolge zahlreiche Schriften rund um das Thema Nutzen und Nutzung von Evaluationen erschienen und zahlreiche Konzepte zur Erhöhung der Nützlichkeit vorgeschlagen worden (siehe Kapitel 4.3).

Wenn man zustimmt, dass es sich bei Evaluationen um angewandte Forschung handelt, deren Ergebnisse entweder für die Administration oder aber für die Praxis von Programmen genutzt werden sollen, dann hat das vielfältige Konsequenzen: In der Planung sowie der Durchführung sind vorhandene zeitliche und materielle Ressourcen, beispielsweise termingerechte Fertigstellungen, unbedingt zu berücksichtigen. Darüber hinaus gilt, dass Evaluation eine Dienstleistung für das Programm darstellt, denn sie soll für Programmverantwortliche und Durchführende nützliche Informationen bereitstellen. Daraus folgt ein „Primat der Praxis vor der Wissenschaft":

---

36 Dass Grundlagen- und angewandte Forschung nicht nur voneinander abzugrenzen sind, sondern auch zusammengehören, das stellen Rossi/Wright (1986, S. 67) heraus: „A discipline that does not have an applied side loses a certain richness of theory and method. An applied field that loses touch with its basic discipline also runs a risk of parochialism and overly narrow attention to policymakers' definitions of social problems and their most feasible solutions."

37 Ausführlicher sind die Gründe für das „Versagen" der Wissenschaft von Kromrey (2003, S. 103ff.) dargestellt.

„Geraten wissenschaftlich-methodische Ansprüche einer möglichst objektiven Erkenntnisgewinnung (etwa methodische Kontrolle ‚störender' Umgebungseinflüsse) mit den Funktionsansprüchen des zu evaluierenden Gegenstands in Konflikt, haben die wissenschaftlichen Ansprüche zurückzutreten und ist nach – aus wissenschaftlicher Perspektive – suboptimalen Lösungen zu suchen, nach Lösungen jedenfalls, die das Funktionsgefüge im sozialen Feld nicht beeinträchtigen." (Kromrey 2001a, S. 113)

Auf jeden Fall ist Evaluation darauf angewiesen, das eigene Handeln mit Politik sowie Praxis abzustimmen, und deswegen benötigen Evaluatorinnen und Evaluatoren weitaus mehr als nur Methodenwissen. Notwendig sind kommunikative Kompetenzen, um das Vorgehen abzustimmen sowie (Zwischen-)Ergebnisse verständlich vermitteln zu können. Ebenso sind Projektmanagement-Fähigkeiten gefragt, damit die geforderten Informationen bedarfs- und fristgerecht zur Verfügung stehen und die vorhandenen Ressourcen angemessen verwendet werden.

Im Zusammenspiel der vielfältigen Herausforderungen, der Ressourcen aus zahlreichen Disziplinen und der unterschiedlichsten Erfahrungsquellen verwundert es nicht, dass neben- und miteinander eine kaum noch zu überblickende Anzahl an Modellen und Konzepten zur Durchführung von Evaluationen entstanden sind. Stufflebeam u. a. (2000) tragen beispielsweise 22 verschiedene Modelle zusammen,[38] Patton (1997, S. 192ff.) präsentiert gar eine Übersicht von 58(!) Designtypen. Diese sind überwiegend auf spezifische Fragestellungen und Settings ausgerichtet und spielen dort jeweils ihre Stärken aus. Alle denkbaren Varianten zu präsentieren ist wenig erhellend. Jedoch werden im folgenden Kapitel (3) vier prominente und unterschiedliche Grundmodelle vorgestellt und jeweils auf ihre Stärken und Schwächen hin beleuchtet.

2.4 Irgendjemand – Wer bewertet nach welchen Kriterien?

Ohne das Thema „Bewertung" anzusprechen, gibt es keinen Anlass, von Evaluation als einer spezifischen Form angewandter Sozialwissenschaft zu reden, denn das Ziel von Evaluation kann „präzisiert werden als: empirisch gestützte Gewinnung von Bewertungen mit intersubjektivem Geltungsanspruch" (Kromrey 2007, S. 113). In der Literatur zu Theorie, Methodologie und Praxis wird dieses Thema jedoch in der Regel ausgespart.[39]

Das Wertfreiheitspostulat des kritischen Rationalismus verpflichtet Forschung traditionell zu objektivem Vorgehen; Werte können zwar Untersuchungsgegenstand

38 Eine knappe und vergleichende Zusammenschau dieser 22 Konzepte findet sich bei Stufflebeam (2001).

39 Ausnahmen bestätigen die Regel: So setzen sich z. B. Beywl (2006) und Lüders (2006) mit Bewertungen durch Evaluation auseinander. Ein weiterer Beitrag stammt von Kromrey (2007), der folgenden Untertitel für seinen Beitrag in der Zeitschrift für Evaluation wählt: „Oder: Wie sich die Evaluationsforschung um das Evaluieren drückt".

sein, Bewertungen sind jedoch nicht als Aufgabe für Forschung vorgesehen (vgl. z. B. Shadish/Reichardt 1987, S. 17f.). Sie bleiben ebenso wie Zwecksetzung und Nutzung dem außerwissenschaftlichen Entdeckungs- und Verwertungszusammenhang zugewiesen. In der Auseinandersetzung mit Stakeholdern von Evaluationen zeigt sich jedoch, dass Evaluation auch in einem politischen Rahmen stattfindet, letztlich selbst auch politisches Handeln ist:

> „Wer ausdrücklich Evaluationsforschung betreibt muss damit rechnen und suggeriert seinen Partnern gegenüber, dass er oder sie nicht nur empirisch valide Sachverhaltsdarstellungen liefert, sondern darüber hinaus auch eine *qualitativ andere, eben wissenschaftliche Bewertung* dieser Sachverhalte zu offerieren vermag." (Lüders 2006, S. 51, Herv. im Orig.).

Michael Scriven (z. B. 1987, 1991, 1997) gehört zu den beständigen Mahnern, der Evaluatorinnen und Evaluatoren in zahlreichen Veröffentlichungen immer wieder daran erinnert, dass Bewertungen unbedingt in deren Tätigkeitsspektrum inbegriffen sind.[40] Auch in der praktischen Umsetzung wird deutlich, dass Evaluationen auf vielfältigen Ebenen mit Werten konfrontiert sind: Jedem Programm liegt eine politische, wertgebundene Entscheidung zu Grunde.[41] Welche Aspekte eines Programms untersucht werden, ist ebenfalls immer durch Werte bestimmt. Je nach Design werden Werte auf verschiedene Weise berücksichtigt (Beywl 2006), und natürlich ist die Festlegung von Güte- oder Qualitätskriterien wertegeladen. Wenn also Evaluation auf Programme trifft, muss sie sich zwangsläufig mit Werten auseinandersetzen und Wege im Umgang mit Werten finden.[42] Zumindest müssen unbedingt folgende Fragen aufgeworfen werden:

- Um bewerten zu können, braucht es Kriterien, anhand derer die Bewertung vorgenommen werden kann. *Welches* sind also die Kriterien, nach denen Evaluation bewerten kann?
- Wer legt die Kriterien fest?
- Wer führt die Bewertung durch?

Insgesamt hält die Literatur drei Strategien zum Umgang mit Werten bereit: eine metatheoretische, die Aussagen zur Rechtfertigung von Werten vorgibt (Scriven), eine präskriptive, die bestimmten Werten Priorität über andere einräumt (z. B. House) und zuletzt eine deskriptive, die Werte, so wie sie vorkommen beschreibt, ohne irgendwelchen Werten den Vorzug zu geben (vgl. Shadish u. a. 1991, S. 48). Zunächst ist Scrivens (1991, 1997) Position zu nennen, der die Ansicht vertritt, dass Evalua-

40 „The correct formulation of the role of evaluation research is to say that the evaluator *must* draw evaluative conclusions (otherwise they are doing less than their job) (...)." (Scriven 1987, S. 109)

41 Dies illustriert Gueron (1997) prägnant am Beispiel des Wechsels von welfare- zu workfare-Politik in den USA.

42 Damit sind Evaluatorinnen und Evaluatoren vor große Herausforderungen gestellt: „While evaluators have increasingly come to acknowledge that values deserve more attention, they have not known how to proceed in this delicate task, for most evaluators were trained to believe that values are not part of ‚science'." (Cook/Shadish 1987, S. 46)