

Thomas Widmer
Wolfgang Beywl
Carlo Fabian (Hrsg.)

Evaluation

Ein systematisches Handbuch



BUNDESTAG GRUNDGESETZ POLITISCHES SYSTEM EUROPÄISCHE UNION
WAHLEN VERFASSUNG INTERNATIONALE BEZIEHUNGEN POLITISCHE THEO
RIE PARTEIEN INSTITUTIONEN POLITISCHE KULTUR POLITISCHE ELITEN
PARLAMENTARISMUS DEMOKRATIE MACHT REGIERUNG VERWALTUNG FÖDER
ALISMUS POLITISCHE SOZIOLOGIE GLOBALISIERUNG POLITISCHE KOMMU



VS VERLAG FÜR SOZIALWISSENSCHAFTEN

Thomas Widmer · Wolfgang Beywl · Carlo Fabian (Hrsg.)

Evaluation

Thomas Widmer
Wolfgang Beywl
Carlo Fabian (Hrsg.)

Evaluation

Ein systematisches Handbuch



VS VERLAG FÜR SOZIALWISSENSCHAFTEN

Bibliografische Information 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.

1. Auflage 2009

Alle Rechte vorbehalten

© VS Verlag für Sozialwissenschaften | GWV Fachverlage GmbH, Wiesbaden 2009

Lektorat: Frank Schindler

VS Verlag für Sozialwissenschaften ist Teil der Fachverlagsgruppe Springer Science+Business Media.
www.vs-verlag.de



Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Verlags unzulässig und strafbar. Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in diesem Werk berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, dass solche Namen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürften.

Umschlaggestaltung: KünkelLopka Medienentwicklung, Heidelberg

Druck und buchbinderische Verarbeitung: Krips b.v., Meppel

Gedruckt auf säurefreiem und chlorfrei gebleichtem Papier

Printed in the Netherlands

ISBN 978-3-531-15741-2

Inhalt

| | |
|---------|----|
| Vorwort | 11 |
|---------|----|

Einleitung

| | |
|---|----|
| <i>Wolfgang Beywl und Thomas Widmer</i> Evaluation in Expansion: Ausgangslage für den intersektoralen Dreiländer-Vergleich | 13 |
|---|----|

Institutionelle Einbettung der Evaluationsfunktion

| | |
|---|----|
| <i>Götz Konzendorf</i> Institutionelle Einbettung der Evaluationsfunktion in Politik und Verwaltung in Deutschland | 27 |
|---|----|

| | |
|--|----|
| <i>Rupert Pichler</i> Institutionelle Dimensionen von Evaluierung in Österreich | 40 |
|--|----|

| | |
|--|----|
| <i>Luzius Mader</i> Die institutionelle Einbettung der Evaluationsfunktion in der Schweiz | 52 |
|--|----|

| | |
|---|----|
| <i>Thomas Widmer und Frans L. Leeuw</i> Die institutionelle Einbettung der Evaluationsfunktion: Deutschland, Österreich und die Schweiz im Vergleich | 64 |
|---|----|

Agrarpolitik

| | |
|--|----|
| <i>Regina Grajewski</i> Evaluation in der Agrarpolitik in Deutschland | 75 |
|--|----|

| | |
|--|----|
| <i>Karl M. Ortner</i> Evaluation in der österreichischen Agrarpolitik | 87 |
|--|----|

| | |
|---|----|
| <i>Vinzenz Jung</i> Schweizer Agrarpolitik: Stand der Evaluation | 99 |
|---|----|

| | |
|--|-----|
| <i>Günter Tissen</i> Evaluation von Agrarpolitik im Ländervergleich | 109 |
|--|-----|

Arbeitsmarktpolitik

Gerald Wagner

Evaluation in der Arbeitsmarktpolitik in Deutschland 117

Angela Wroblewski

Evaluation von arbeitsmarktpolitischen Programmen und Maßnahmen in Österreich 129

Thomas Ragni und Oliver Bieri

Evaluation der Arbeitsmarktpolitik in der Schweiz 137

Wolfgang Helmstädter

Ländervergleich – Evaluation Arbeitsmarktpolitik 148

Bildung

Katharina Maag Merki

Evaluation im Bildungsbereich Schule in Deutschland 157

Uwe Schmidt

Evaluation an deutschen Hochschulen – Entwicklung, Stand und Perspektiven 163

Werner Specht

Österreich: Evaluation im Bereich der Schule 170

Alexander Kohler

Evaluation im österreichischen Hochschulsystem 177

Heinz Rhyn

Evaluation im Bildungsbereich in der Schweiz 182

Lars Balzer

Evaluation im Bildungsbereich: Ein Ländervergleich zwischen Deutschland, Österreich und der Schweiz 193

Energie- und Umweltpolitik

Markus Duscha, Herbert Klemisch und Wolfgang Meyer

Umweltevaluation in Deutschland – Entwicklungstrends mit Fokus auf dem Energiesektor 203

André Martinuzzi und Melanie Leschnik

Evaluation im Themenfeld Energie und Umwelt – Länderstudie Österreich 213

| | |
|---|-----|
| <i>Kurt Bisang</i> Energie und Umwelt: Stand der Evaluationstätigkeiten in der Schweiz | 226 |
|---|-----|

| | |
|--|-----|
| <i>Christian Hirschi</i> Ländervergleich Energie und Umwelt | 236 |
|--|-----|

Entwicklungszusammenarbeit

| | |
|--|-----|
| <i>Michaela Zintl</i> Evaluierung in der deutschen Entwicklungszusammenarbeit | 245 |
|--|-----|

| | |
|--|-----|
| <i>Anton Mair</i> Evaluierung in der Österreichischen Ost- und Entwicklungszusammenarbeit | 255 |
|--|-----|

| | |
|---|-----|
| <i>Dieter Zürcher</i> Evaluation in der schweizerischen Entwicklungszusammenarbeit | 263 |
|---|-----|

| | |
|---|-----|
| <i>Nicolà Reade</i> Ländervergleich: Evaluierung in der Entwicklungszusammenarbeit | 273 |
|---|-----|

Forschung und Technologie

| | |
|---|-----|
| <i>Stefan Kuhlmann</i> Evaluation von Forschungs- und Innovationspolitik in Deutschland – Stand und Perspektiven | 283 |
|---|-----|

| | |
|---|-----|
| <i>Klaus Zinöcker und Michael Dinges</i> Evaluation von Forschungs- und Technologiepolitik in Österreich | 295 |
|---|-----|

| | |
|--|-----|
| <i>Peter Farago</i> Evaluation von Forschung und Technologie in der Schweiz | 305 |
|--|-----|

| | |
|--|-----|
| <i>Wolfgang Polt</i> Evaluation in den Bereichen Forschung, Technologie und Innovation in Deutschland, Österreich und der Schweiz | 313 |
|--|-----|

Gesundheit

| | |
|---|-----|
| <i>Philipp Mayring</i> Evaluation im Bereich Gesundheit – Beispiel Deutschland | 321 |
|---|-----|

| | |
|---|-----|
| <i>Georg Spiel und Joachim Petscharnig</i> Evaluation im österreichischen Gesundheitswesen – ein Überblick | 331 |
|---|-----|

Ambros Uchtenhagen

Evaluation im Gesundheitsbereich: Schweiz 341

Stefan Spycher

Mehr Hetero- als Homogenität. Zum Stand der Evaluation im Bereich Gesundheit in Deutschland, Österreich und der Schweiz 353

Institutionelle Politik

Fritz Sager

Die Evaluation institutioneller Politik in der Schweiz 361

Sabine Kuhlmann

Die Evaluation von Institutionenpolitik in Deutschland: Verwaltungsmodernisierung und Wirkungsanalyse im föderalen System 371

Werner Pleschberger

Evaluation institutioneller Politik in Österreich 381

Werner Bussmann

Evaluation institutioneller Politik in der Schweiz, Deutschland und Österreich im Vergleich 392

Raumentwicklungspolitik

Oliver Schwab

Evaluierung von Raumentwicklungspolitik in Deutschland 403

Richard Hummelbrunner und Andreas Maier

Evaluation von Raumentwicklungspolitik in Österreich 413

Alain Thierstein

Evaluation von Raumentwicklung, Regional- & Strukturpolitik in der Schweiz 422

Frohmut Gerheuser

Die Evaluation räumlicher Entwicklung im Vergleich der drei Länder 433

Soziale Arbeit

Karin Haubrich

Evaluation in der Sozialen Arbeit in Deutschland. Entwicklungslinien und Besonderheiten der Evaluationsdebatte am Beispiel der Kinder-, Jugend- und Familienhilfe 441

Rainer Loidl-Keil

Evaluation in der Sozialarbeit in Österreich – auf dem Weg zu einem neuen
Verständnis und aktiver Auseinandersetzung 450

Matthias Drilling

Evaluationen in der Sozialen Arbeit in der Schweiz: Eine Einschätzung der Situation
aus professionsorientierter Sicht 458

Karin Haubrich, Rainer Loidl-Keil und Matthias Drilling

Evaluation in der Sozialen Arbeit im Ländervergleich 469

Nutzung der Evaluationsfunktion

Christiane Spiel und Evelyn Bergsmann

Zur Nutzung der Evaluationsfunktion: Partielle Bestandsaufnahme und Pilotstudie
in österreichischen und deutschen Bundesministerien 477

Andreas Balthasar

Evaluationen in der Schweiz: Verbreitung und Verwendung 486

Carlo Fabian

Nutzung der Evaluationsfunktion im Vergleich 498

Sektorale und nationale Trends

Thomas Widmer und Wolfgang Beywl

Divergente Entwicklungspfade der Evaluation 509

Herausgeber 529

Autorinnen und Autoren 530

Personenverzeichnis 535

Sachverzeichnis 549

Literatur 575

Vorwort

Vor zehn Jahren war die Evaluation im deutschsprachigen Raum auf einige wenige Gesellschaftsbereiche beschränkt. In der Zwischenzeit sind die Akzeptanz und damit auch die Anwendungsbreite grösser und die methodischen Ansätze vielfältiger geworden. Zum Teil wird kritisch sogar von einem Evaluationsboom gesprochen. Parallel zur steigenden Bedeutung der Evaluierung für Politik und Gesellschaft haben sich breit abgestützte Fachgesellschaften entwickelt und etabliert: Die DeGEval – Gesellschaft für Evaluation in Deutschland und Österreich sowie die Schweizerische Evaluationsgesellschaft SEVAL in der Schweiz. Beide Gesellschaften haben in ihren Einzugsgebieten viel zu Kompetenzaufbau und Professionalisierung auf Seiten der Anbietenden und der Nachfragenden, zur Qualitätssteigerung und zur Stärkung der politischen Relevanz von Evaluation beigetragen. Die Verankerung der Verpflichtung des Parlaments zur Wirkungsprüfung in Artikel 170 der Bundesverfassung der Schweizerischen Eidgenossenschaft sowie die Qualitätsstandards der SEVAL und der DeGEval zeugen davon. Weitere Beispiele dafür sind die Empfehlungen der DeGEval zur Aus- und Weiterbildung in der Evaluation und deren Empfehlungen für Auftraggebende von Evaluationen.

Trotz zahlreicher individueller Kontakte auf internationalen Tagungen und Kongressen blieben die Aktivitäten der Evaluationsgesellschaften vorerst mehrheitlich auf die jeweiligen Länder ausgerichtet. Dies war in der Phase des Aufbaus durchaus zweckmässig. Die Herausforderungen an Evaluationen werden jedoch sowohl von ihren inhaltlichen Fragestellungen als auch von den methodischen Anforderungen immer komplexer und anspruchsvoller. Damit ist für die Weiterentwicklung von Evaluation der systematische Blick über die Grenzen unabdingbar. Die Akzeptanz von Evaluationen wird zunehmend von deren Orientierung an internationalen Qualitätsstandards abhängen. Von besonderer Bedeutung ist daher – neben den persönlichen Kontakten und Kooperationen mit Evaluierenden im Ausland – die regelmäßige wissenschaftliche Auseinandersetzung auf Tagungen sowie in Form von Publikationen.

Hier setzte die im Frühjahr 2007 durchgeführte Fachtagung an, welche von der SEVAL und der DeGEval gemeinsam mit der Fachhochschule Nordwestschweiz in Basel durchgeführt wurde. Erstmals haben sich Evaluationsexpertinnen und -experten aus den drei Ländern versammelt, um sich über ihre Praxis, Methodik und Erfahrung in zahlreichen Themenfeldern auszutauschen. Auf der Basis dieser Tagung entstand die vorliegende Publikation. Sie bietet eine breite Bestandsaufnahme zum Stand der Evaluation in den drei Ländern. Gleichzeitig zeigt sie Unterschiede und Gemeinsamkeiten bezüglich Einsatz und Methodik von Evaluation in den verschiedenen Politikfeldern auf.

Den Organisatoren der Tagung und gleichzeitig Herausgebern dieses Bandes, Thomas Widmer (Universität Zürich), Wolfgang Beywl (Univation, Köln und Universität Bern) sowie Carlo Fabian (Fachhochschule Nordwestschweiz, Basel) sind die DeGEval und die SEVAL für diese Initiativen äusserst dankbar. Sie haben es ermöglicht, den fachlichen Dis-

kurs zu fördern, der zu oft durch thematische Limitierungen und Landesgrenzen eingeschränkt ist. Es ist nicht nur gelungen, eine erfolgreiche Tagung zu realisieren, sondern mit der vorliegenden Publikation auch einen Meilenstein in der deutschsprachigen Evaluationsliteratur zu setzen.

Wir möchten allen, die zum Gelingen der Tagung und der vorliegenden Publikation beigetragen haben, herzlich danken. Neben den Organisatoren der Tagung gehören dazu folgende schweizerische Stellen: das Bundesamt für Energie BFE, das Bundesamt für Gesundheit BAG, das Bundesamt für Justiz BJ, das Bundesamt für Umwelt BAFU, die Direktion für Entwicklung und Zusammenarbeit DEZA, die Eidgenössische Finanzkontrolle EFK, die Parlamentarische Verwaltungskontrolle PVK, das Staatssekretariat für Wirtschaft seco, der Regierungsrat des Kantons Basel-Stadt, die Hochschule für Soziale Arbeit der Fachhochschule Nordwestschweiz FHNW, das Hotel Bildungszentrum 21 (Basel), der Schweizerische Nationalfonds sowie – aus Deutschland – Unipark. Ein ganz besonderer Dank gebührt auch all jenen, die inhaltliche Beiträge zur Tagung und zu diesem Band geleistet haben.

Wir hoffen sehr, dass der im Zusammenhang mit der Tagung lancierte wissenschaftliche Diskurs sowie die geknüpften Kontakte gepflegt und weiterentwickelt werden. Ebenso wie Evaluationen für Auftraggebende und Stakeholder ein Lernpotential bereit stellen, können und sollen auch Evaluierende voneinander lernen. Evaluation ist eine internationale Herausforderung. Den Diskurs wollen wir im deutschsprachigen Raum und darüber hinaus weiterhin pflegen.

Andreas Balthasar

Präsident (bis 2007)

Schweizerische Evaluationsgesellschaft SEVAL

Christiane Spiel

Vorstandsvorsitzende

DeGEval – Gesellschaft für Evaluation

Evaluation in Expansion: Ausgangslage für den intersektoralen Dreiländer-Vergleich

1 Evaluation – Inflation eines Wortes oder Expansion einer wissenschaftlichen Praxis?

In der deutschen Sprache breitet sich das Wort „Evaluation“ – besonders im Österreichischen auch in der Variante „Evaluierung“ – in den letzten zwanzig Jahren aus. Auch das Verb „evaluieren“ kann nicht mehr – wie vielleicht in den 1980er Jahren – nahezu exklusiv dem Wortschatz eines kleinen Kreises von Fachpersonen zugerechnet werden. Diese und davon abgeleitete Terme haben sich im Sprachgebrauch weit verbreitet und werden in Parlamentsdebatten, Tageszeitungen, Talkshows mit politischen oder gesellschaftlichen Inhalten, Radio- und Fernsehnachrichtensendungen genutzt. Der Erfolg der Bezeichnung „Evaluation“ bietet Anlass, den Prozessen ihrer Verbreitung nachzugehen.

Vermutlich werden „Evaluation“ und „Evaluierung“, wird das Verb „evaluieren“ bald in der Belletristik und anderen Zeugnissen des kulturellen Lebens auftauchen. Auf einen Evaluationsroman, wie den von Ernest House hingegen, wird die deutschsprachige Lesegemeinschaft wohl warten müssen (House 2007).

Schülerleistungsmessungen und darauf aufbauende Schulvergleiche sowie Lehrveranstaltungsfeedbacks an Hochschulen – ob zu recht oder unrecht mit dem Terminus versehen – haben sicher zur Verbreitung des Wortes Evaluation massgeblich beigetragen. Hier wird er mit dem Bildungsbereich in einer mit der Geschichte der Evaluation (Alkin 2004; Beywl 1988) eng verbundenen Gegenstandsdomäne genutzt. Doch kann das Wort „Evaluation“ auf nahezu alle Gegenstände angewandt werden und vieles kann damit gemeint sein. Sein Bedeutungsraum weitet sich aus und verliert an Trennschärfe. Es ist oft synonym mit Bewerten durch Personen oder Gremien, die sich qua Amt, Status, Expertise, Profession, Erfahrung oder andere Ausweise dazu legitimiert sehen. Nicht notwendig ist es verbunden mit einer systematischen, transparenten und datenbasierten Vorgehensweise. Mit letzterem ist Evaluation in dem Sinne angesprochen, wie sie in diesem Buch behandelt wird.

Es kann nur spekuliert werden, ob das Wort Evaluation deshalb eine solch starke Verbreitung gefunden hat, da die kommunikationstechnologische Revolution des Internet mit dem rasanten Preisverfall von quantitativen Datenerhebungen, den Möglichkeiten schneller und massenhafter Erhebung und Datenfeedbacks mittels Online-Tools oder SMS, eine neue Kultur des quantifizierenden Bewertens durch bis zu hunderttausende von Personen hervorgebracht hat. Sichtbarer Ausdruck dieses Trends sind die Bewertungsportale, zunächst für Produkte und sachbezogene Dienstleistungen, zwischenzeitlich zunehmend in typischen Gegenstandsfeldern von systematischer und transparenter Evaluation: Im Falle des schlag-

zeilenträchtigen, ursprünglich rein deutschen Portals www.meinprof.de, das heute auch Studierendenurteile für Hochschullehrende in Österreich (www.meinprof.at) und in der Schweiz (www.meinprof.ch) sammelt, ist dies der Hochschulbereich; bei www.spickmich.de und seinem Ableger www.schulradar.de sind es Lehrpersonen und die Schule. Im Gesundheitsbereich existieren mehrere miteinander konkurrierende Bewertungsportale für Krankenhäuser und Ärzte und Ärztinnen. Mit www.wohin-im-alter.de startet ein Bewertungsportal für Alten- und Pflegeheime. Diese Portale sind ein ungewisser, nicht immer stabiler, teils mit starken Interessenkonflikten und wirtschaftlichen Verwertungsabsichten einher gehender Ausdruck (quasi-)demokratischer, populärer Bewertung. In einer zunehmenden Anzahl von Themenfeldern ergänzen oder konkurrieren sie die Evaluation, stellen ihr Image in Frage. Jedenfalls verändern die Bewertungsportale – wie vordem die „Happy-Sheets“, Kurzfragebögen zur Kunden-Zufriedenheitsmessung in Hotels, Gastronomie und anderen Dienstleistungsbetrieben – den Kontext der Evaluationstätigkeit, sowohl was die ständige Präsenz von Erhebungsinstrumenten als auch von Bewertungsergebnissen im Alltag betrifft. Dies gilt länder- und themenfeldübergreifend, wobei die mittel- und längerfristigen Auswirkungen nicht absehbar sind. Sie können für eine Evaluationskultur sowohl positiv als auch negativ ausfallen – letzteres zum Beispiel in Form höheren Widerstandes gegen Evaluation (Taut/Brauns 2003).

Dabei wird es im Deutschen wohl dabei bleiben, dass „evaluieren“ nicht wirklich in die Alltagssprache eingeht und synonym mit „bewerten“, „überprüfen“ oder „beurteilen“ von „irgend etwas, von irgend jemandem, nach irgend welchen Kriterien, in irgend welcher Weise“ verwendet wird (vgl. zu den unterschiedlichen alltäglichen, politischen und fachlichen Sprachverwendungen Kromrey 2001). Im Französischen – woher das Wort über den Umweg aus dem amerikanischen Englisch in die deutsche Sprache gefunden hat – ist die Lage anders: „évaluer“ steht hier auch für die alltägliche Tätigkeit der „Schätzung“, „Ermittlung“ oder „Wertbestimmung“ in allen möglichen Lebenszusammenhängen.¹ Es mag deshalb kein Zufall sein, dass in der Schweizerischen Bundesverfassung (dem schweizerischen Grundgesetz; Artikel 170) in der deutschen Fassung von „Überprüfung der Wirksamkeit“ gesprochen wird, während es in der französischen Version „[é]valuation de l'efficacité“ heisst (vgl. dazu ausführlicher Widmer 2007).

Die Ausbreitung des Terms wird zum Problem, wenn zentrale Definitionsbestandteile von systematischer und transparenter Evaluation in der Domäne der Wortverwendung (also z. B. Politik- oder Organisationsanalyse und -beratung, Reform des Bildungs- oder Gesundheitssystems) nicht mehr thematisiert, abgekoppelt oder stillschweigend suspendiert werden. Es würde zu weit führen, diese konstitutiven Merkmale von Evaluation in einem engeren Sinn im Einzelnen aufzuführen, zu begründen und zu erläutern. Sie sind in einem verbreiteten Korpus von Lehr- und Fachbüchern zur Evaluation kodifiziert, unterliegen einer fortlaufenden Diskussion und Reflexion, sind in Standards für Evaluationen, Leitlinien für Evaluatorinnen und Evaluatoren oder Richtlinien für deren Ausbildung niedergelegt (siehe beispielhaft die im Handbuch der Evaluationsstandards [Joint Committee on Standards for Educational Evaluation 2006] aufgeführten Regelwerke; vgl. auch Weiss 1998a; Rossi et al. 2004; Patton 2008; Russ-Eft et al. 2008). Neben der Abstützung auf wissenschaft-

¹ Vgl. die etymologische Herleitung von Manfred Karbach (<http://evaluelab.de/extra/evalety.html> [29.5.2008])

liche Methodik, empirische Daten, Klärung und Transparenz von Bewertungskriterien, sind dies zahlreiche weitere evaluationsfachliche Anforderungen. Nicht selten ist festzustellen, dass als „Evaluationsberichte“ gekennzeichnete Publikationen kein einziges solcher zentralen Bezugsdokumente zitieren; wobei ihr Auflisten im Literaturverzeichnis keine Gewähr dafür bietet, dass es sich um Evaluationen im Sinne der dargelegten Anforderungen handelt.

Die Lage wird kompliziert, wenn Kritiker und Kritikerinnen der Evaluation die inflationäre Wortverwendung von Evaluation übernehmen und damit zu volatilen Ziel ihrer oft polemischen Kritik machen. Bei sich vielfach ‚überevaluieren‘ sehenden Professionellen – seien dies Ärzte oder Ärztinnen, Pflegepersonen, Sozialarbeitende oder Hochschullehrende – lösen Schlagwörter wie „Evaluitis“ (Simon 2000; später auch Frey 2008; vgl. Widmer 2008a: 276) oder „Ritual“ (Schwarz 2006) oder die augenzwinkernd vorgebrachte Parallele zwischen Inquisition und Evaluation (Hornbostel 2008) einerseits Entlastung und andererseits Selbstvergewisserung in der Abwehr gegen Evaluation – gleich welcher Qualität – aus. Diese In-Eins-Setzung von Evaluation mit zahlreichen anderen Verfahren, die sich in einem oder mehreren Merkmalen substantiell von Evaluation unterscheiden, droht auf lange Sicht auch systematische und transparente Evaluation mit Wirkungspotential zu diskreditieren. Differenzierungen gehen dabei verloren. Als eines für viele Beispiele, wie Kritik an grundsätzlich von Evaluation verschiedenen Ansätzen der Kontrolle (vgl. dazu auch Daintith/Page 1999; Hood 2004) auf diese übertragen wird, sei die Argumentation von Michael Power (1997: 2-7) nachfolgend kurz skizziert.

Der Professor für Rechnungswesen an der London School of Economics beschäftigt sich in seiner stark rezipierten Studie zur „Audit Society“ im Kern mit der Ausdehnung von Kontroll-Instrumenten auf nicht-finanzielle Tatbestände, etwa im Hochschulbereich. In diesem Vordringen standardisierter, flächendeckender und hochpräziser Audit-Routinen, die einseitig managerialistischen Werten verpflichtet seien, in die nicht-fiskalischen Domänen etwa des Bildungs- und Gesundheitsbereichs sieht er „autodestruktive“ Gefahren. Ohne dass er die evaluationstheoretische Literatur im Detail kennen würde, schreibt er der Evaluation eine korrektive Funktion zu, insofern sie „other cultures of performance evaluation“ zur Geltung bringen könne – „In short, auditing needs to be evaluated.“ (Power 1997: 13).

Trotz dieser Differenzierung, die er mit der – aus seiner Sicht durch „Auditisierung“ bedrohten (Power 2008) – Evaluation teilt, wird Power gerne als eine Art Kronzeuge gegen Evaluation ins Feld geführt (Schwarz 2006: 10-11; Frey 2008: 125).

Abseits der inflationären Ausbreitung des Wortes stellt sich die Frage, ob Evaluation in einem engeren Sinne tatsächlich – wie vielfach behauptet – in den vergangenen Jahren so stark expandiert hat. Hans-Ulrich Derlien schreibt der Bundesrepublik Deutschland schon zum Ende der 1960er Jahre zu, dass sie als eines der ersten Länder im Zuge der sozial-liberalen Reformprogrammatik Evaluation (allerdings nur teilweise unter dieser Bezeichnung) als Instrument zur Steuerung nationaler Regierungspolitik institutionalisiert habe, besonders in den Themenfeldern Entwicklungshilfe, Verkehr, Landwirtschaft und Technologie (Derlien 1990a: 38-42; vgl. dazu auch Hellstern/Wollmann 1984b; Stockmann 2006a: 29-30); für frühe Ansätze in westdeutschen Grossstädten, siehe Hellstern/Wollmann 1984c). Aus Derliens Sicht hat sich das Evaluationsrepertoire in Deutschland seit dessen Etablierung

im Jahre 1975 kaum verändert, auch wenn er zugesteht, dass ein gewisses Wachstum zu verzeichnen sei (Derlien 2002: 88-90).

Die Schweiz hingegen sah sich anfangs der 1990er Jahre noch als Stiefkind in Evaluationsfragen (Horber-Papazian/Thévoz 1990: 142). Auch am Ende der 1990er Jahre vermitteln Darstellungen – vielleicht beeinflusst durch Thesen, dass die Schweiz aufgrund ihrer spezifischen Merkmale (Föderalismus, Dezentralisierung, direkte Demokratie; vgl. dazu Widmer/Neuenschwander 2004: 389-391) für Evaluation weniger empfänglich sei – noch einen erheblichen Rückstand gegenüber dem vermeintlichen „Frühstarter“ (Derlien 1997: 6) im Norden (Spinatsch 2002, bes. 375). Ein Vergleich der in verschiedenen Ländern etablierten Evaluationskultur unterstellt ebenfalls einen erheblichen Entwicklungsvorsprung Deutschlands gegenüber der Schweiz (Furubo/Sandahl 2002: 6-12). Andere Autoren verweisen bereits für die Zeit um die Jahrtausendwende (und danach) auf eine hohe Dynamik der Evaluationsentwicklung in der Schweiz (Fornerod 2001; Widmer et al. 2001; Balthasar 2002; Jacob/Varone 2003; Bussmann 2004; Widmer/Neuenschwander 2004: 389-391; Kettiger 2005).

Im Falle von Österreich sind die Kenntnisse zum Stand der Evaluation erheblich spärlicher. In den genannten international vergleichend ausgerichteten Publikationen ist Österreich nicht abgedeckt. Umfassende Publikationen zur Evaluation in Österreich liegen nicht vor. Veröffentlichungen betreffen spezifische Themenfelder oder Instrumente (vgl. dazu etwa die Beiträge in Schäffer 2005b), was eine fundierte Würdigung von Entwicklung und aktuellem Stand verunmöglicht.

2 Zur Abgrenzung und Abdeckung des Gegenstandsbereiches dieses Buches

Angeichts der im vorangegangenen Abschnitt beschriebenen Risiken einer Inflation des Wortgebrauchs wäre es denkbar gewesen, dem Buch eine Definition von Evaluation voranzustellen, die in etwa gelautet hätte:

Wissenschaftliche Dienstleistung, die namentlich öffentlich verantwortete und/oder finanzierte ‚Gegenstände‘ (Politiken, Programme, Projekte, Massnahmen ...) in verschiedensten Themenfeldern systematisch, transparent und datengestützt beschreibt und ausgewogen bewertet, so dass Beteiligte & Betroffene (Auftraggebende und andere Stakeholder) die erzeugten Evaluationsergebnisse für vorgesehene Zwecke wie Rechenschaftslegung, Entscheidungsfindung oder Optimierung nutzen.

Ein solches stark einengendes Vorgehen schien jedoch nicht geeignet, die Pluralität in der Breite der vertretenen Evaluationsansätze angemessen zu repräsentieren. So mag den einen die partizipative Komponente durch die Öffnung gegenüber den Stakeholdern zu stark, den anderen durch die Erstnennung der Auftraggebenden zu schwach formuliert sein. Einigen mag die Nutzenorientierung in obiger Definition zu instrumentell sein; Evaluation zum Zweck der wissenschaftlichen Erkenntnis zu wenig angesprochen. Ein Teil der Beitragenden würde sich statt einer lediglich „ausgewogenen“ – Stärken wie Schwächen ansprechenden –

immer eine „unabhängige“ Bewertung wünschen, was viele intern tätige Evaluatorinnen und Evaluatoren schnell in unlösbare Loyalitätskonflikte stellen könnte et cetera.

Eine Eingrenzung des Evaluationsverständnisses ist über den Auswahlprozess der Autoren und Autorinnen erfolgt. In erster Linie verlief dieser mit Unterstützung der beiden Evaluationsgesellschaften, die sich beide weitgehend übereinstimmende Standards der Evaluation gegeben haben (Widmer 2004; 2006a) und damit ein sehr ähnliches Grundverständnis von Evaluation vertreten. Zwei der Herausgeber verfügten durch ihre langjährige Mitgliedschaft im Vorstand der jeweiligen Gesellschaft über zahlreiche einschlägige Kontakte in die verschiedenen Themenfelder. Die aktuellen Vorstände der beiden Gesellschaften unterstützten die Suche nach geeigneten Autorinnen und Autoren aktiv. Bei einem Grossteil der Beitragenden handelt es sich um Mitglieder der Evaluationsgesellschaften – im Falle der DeGEval teils um Sprecher oder Sprecherinnen bzw. Aktive der im Themenfeld tätigen Arbeitskreise; die SEVAL kennt keine solche Arbeitskreisstruktur. Besonders für Deutschland und Österreich erwies sich die Suche nach Beitragenden in den Feldern, in denen es keine aktiven Arbeitskreise (mehr) gibt, als aufwendiger. Sie konnte schliesslich durch intensive Recherche erfolgreich abgeschlossen werden.

In jedem Falle war das Anliegen der Herausgeber, diejenigen als Autoren oder Autorinnen zu gewinnen, die im jeweiligen Land über den besten Überblick im jeweiligen Themenfeld verfügen, und zwar aus einer evaluationsfachlichen Perspektive (vgl. zum Spannungsfeld zwischen Evaluationsfachlichkeit und Fachlichkeit der Themenfelder den Schlussbeitrag von Widmer/Beywl in diesem Band).

Dabei arbeiten die Personen, welche die Erstellung der Themenbeiträge übernommen haben, zum kleineren Teil schon länger an solchen Überblicksaufgaben. Für viele war es eine neue Aufgabe. Einige verfügten aufgrund längerer Betätigung im entsprechenden Evaluationsfeld über eine gute empirische Basis oder zumindest gute Informationszugänge und Kontakte zu Auskunftspersonen und Quellen. Einzelne haben diese durch schriftliche Umfragen und Auswertung von Datenbanken gezielt erweitert und abgesichert.

In der Schweiz stellt der Schweizerische Informations- und Datenarchivdienst für die Sozialwissenschaften (SIDOS) ein Inventar sozialwissenschaftlicher Forschung zur Verfügung (www.sidos.ch). In Österreich führt das Wiener Institut für Sozialwissenschaftliche Dokumentation und Methodik (WISDOM) vergleichbare Erhebungen durch (www.wisdom.at) und dokumentiert diese in der Forschungsdokumentation Sozialwissenschaften (FODOS). In Deutschland führt die Gesellschaft sozialwissenschaftlicher Infrastruktureinrichtungen (GESIS) Erhebungen zur Identifikation sozialwissenschaftlicher Forschungsarbeiten durch und macht die Projektinformationen im sozialwissenschaftlichen Forschungsinformationssystem SOFIS zugänglich (www.gesis.org/Information/SOFIS). Sein Ziel ist, möglichst hohe Transparenz zu Forschungsaktivitäten herzustellen und Kommunikation im Bereich Wissenschaft und Forschung zu fördern. In dieser Datenbank werden die Informationen aus allen drei Ländern zusammengeführt. Je nach Enge der Suche – Methoden-Schlagwort oder Freitext – ergeben sich für die drei Länder zusammen seit dem Jahr 2000 zwischen rund 2.370 bis zu fast 5.000 gemeldete Studien, die Evaluationen sein können oder evaluative Elemente

aufweisen könnten.² Auch wenn eine detaillierte Nachrecherche sicher zu einer Veränderung der Angaben führt, können die genannten Zahlen als Anhaltspunkte beigezogen werden. Weiter ist zu beachten, dass die tatsächliche Zahl an realisierten Evaluationen in den drei Ländern um ein Mehrfaches höher liegt, und zwar aus folgenden Gründen:

- Es befinden sich in den Forschungsdatenbanken zahlreiche empirische Untersuchungen, bei denen es sich um Evaluationen handelt, die diese Bezeichnung aber weder im Titel noch im Volltext ihrer Beschreibung führen. Sie laufen zum Beispiel unter Bezeichnungen wie „wissenschaftliche Begleitung“, „Praxisforschung“, „Handlungs“- oder „Aktionsforschung“, „angewandte Forschung“, „Erfolgskontrolle“, „Wirkungsforschung“ oder „Vergleichgruppenuntersuchung“. Zu vermuten ist, dass die Bezeichnung „Evaluation“/„Evaluierung“ in den vergangenen Jahren relativ häufiger gewählt wird.
- Die Forschungsdatenbanken konzentrieren sich auf den Nachweis von Studien aus Hochschulen und (überwiegend) öffentlich finanzierten Forschungseinrichtungen (siehe etwa die Beschreibung zu SOFIS). Der Grossteil von Evaluationen wird in vielen Themenfeldern jedoch durch private Anbieter durchgeführt, für die eher geringe Anreize der Meldung an die GESIS-Datenbanken bestehen, die den Hochschulen Recherchemöglichkeiten zu besonders günstigen Konditionen bieten. Interne Evaluationen, auch grössere und längerfristig angelegte, dürften in aller Regel nicht gemeldet werden.
- Kleine und kleinste Kurzevaluationen dürften selten an die Forschungsdokumentationssysteme gemeldet werden, da der Meldeaufwand in Relation zum Projektbudget hoch und der Nutzen für die überwiegend ausserhalb des Wissenschaftssystems tätigen Evaluationsdurchführenden unklar ist.

Andererseits dürften in den Forschungsdatenbanken „Evaluationen“ enthalten sein, die Merkmale einer Evaluation vermissen lassen, zum Beispiel eine Bewertung des Evaluationsgegenstandes.

Die – inzwischen elektronisch ausfüllbaren – Erhebungsformulare zum Beispiel von SOFIS sind auf Forschungsprojekte ausgerichtet und fragen konsequenterweise zwar nach dem Forschungsparadigma, nicht aber nach einem Evaluationsmodell (Stufflebeam et al. 2000; Stufflebeam 2001). Die für die Einordnung einer Evaluation ebenfalls zentrale Frage nach deren Zweck (Rechenschaftslegung, Entscheidung, Optimierung) beziehungsweise danach, ob vorrangig eine summative oder eine formative Funktion erfüllt werden soll, fehlt. Dies verlangt von Evaluationsinteressierten zusätzliche Nachforschungen.

In der Konsequenz bestehen für die Abbildung des Evaluationsgeschehens in den drei Ländern bislang erhebliche Lücken. So gibt es für den besonders wichtigen Bereich Schule, namentlich berufliches Schulwesen, keinerlei bundesweiten Überblick, weder für Deutschland noch für Österreich, über Evaluationen auf Bundes- und Länderebene (geschweige denn kommunaler Ebene).

² Die Zahl von 2.370 ergibt sich durch Suche nach dem methodischen Schlagwort bzw. Deskriptor „Evaluation“ in einem speziellen Suchfeld zu methodischen Vorgehensweisen und Verfahren der Datengewinnung; bei über 80% von diesen ist ein Land zugeordnet: 1.658 mal Bundesrepublik Deutschland, 217 mal Schweiz, 64 mal Österreich. Insgesamt 4.768 (Stand: 2. Juni 2008) werden ausgewiesen nach Eingabe von „Evalu*“ in die Volltextsuche.

Wie weit der Abdeckungsgrad bei der Einschätzung des Standes der Evaluation in den Beiträgen dieses Bandes in den drei Ländern für die verschiedenen Themenfelder jeweils ausfällt, kann nicht genau ausgesagt werden. Der Schwierigkeitsgrad, sich in einem bestimmten Feld und in einem bestimmten Land einen Überblick zu verschaffen, hängt namentlich von den folgenden Faktoren ab:

- Grösse des Landes,
- Grad der Vernetzung der Evaluierenden im jeweiligen Themenfeld,
- Organisationsgrad der jeweiligen Evaluationsgesellschaft,
- Bestand an Parallelstrukturen wissenschaftlicher Fachgesellschaften, die sich mit evaluationsähnlichen Fragen beschäftigen, und
- Forschungsstand über Evaluation, besonders der öffentlich geförderten Forschung durch nationale Fördereinrichtungen.

Es wird nicht der Anspruch erhoben, dass für die drei Länder sämtliche Themenfelder der Evaluation lückenlos abgedeckt sind. Beispiele sind die Themenfelder Gleichstellung oder der Katastrophenschutz, die innere Sicherheit oder die Aussenpolitik, die entweder gar nicht oder bestenfalls peripher in einem angrenzenden Themenfeld angesprochen sind.

Andere Themenfelder sind nicht umfassend, sondern selektiv behandelt; herausragend gilt dies für die Bildung, wo die (allgemeinbildenden) Schulen sowie die Hochschulen im Zentrum stehen. Die berufliche Bildung, die Weiterbildung oder die frühe Bildung werden hingegen lediglich cursorisch angesprochen, obwohl sie relevante Evaluations-Teilfelder darstellen. Gerade im Bildungsbereich fehlt eine alle Bildungsphasen übergreifende Bestandsaufnahme und Forschung zu Evaluationsaktivitäten. Eine vollständige Abdeckung durch mehrere Teilbeiträge hätte vom Platzbedarf und in zeitlicher Hinsicht das Vorhaben gesprengt. Wünschenswert für die nähere Zukunft wäre die Erarbeitung eines systematischen Überblicks über die Evaluationsaktivitäten in einem breiten Bereich von Bildung, wobei ein Ländervergleich weitere Aufschlüsse verspricht.

Die Themenfelder enthalten im Wesentlichen Evaluationsgegenstände, die zumindest zu einem grossen Teil direkt oder indirekt durch öffentliche Mittel finanziert sind. Es sind entweder Programme, die von der öffentlichen Hand selbst geplant und umgesetzt werden oder die Dritte im Auftrag der öffentlichen Hand durchführen (öffentliche Förderung). Teilweise handelt es sich um Programme oder auch Organisationen, die von Stiftungen oder Hilfsorganisationen, Wohlfahrtsverbänden und anderen Organisationen realisiert werden oder errichtet sind, die aufgrund ihres gemeinnützigen Zweckes steuerbefreit oder -begünstigt sind, respektive wegen ihrer Finanzierung durch Spenden gegenüber der öffentlichen Hand oder der Öffentlichkeit in einer besonderen Rechenschaftspflicht stehen. Evaluationen im Bereich von Unternehmen (besonders bei der Weiterbildungsfunktion oder auch im Bereich von Kundengewinnung oder Kundenbindung) werden in diesem Buch nicht respektive randständig behandelt. Dies auch deshalb, da sich hier der Zugang zu Evaluationsberichten in aller Regel schwierig gestaltet, insofern diese schützenswerte Betriebsgeheimnisse enthalten.

Ein systematisches Beobachtungssystem für Evaluationen ist in allen drei Ländern sehr schwach ausgeprägt. Weder gibt es Forschungsprojekte noch Institutionen innerhalb oder

ausserhalb von Hochschulen, welche diese Aufgabe themenfeldübergreifend wahrnehmen würden. Die Evaluationsgesellschaften vermögen bei knappen Ressourcen auf ihren Webseiten oder in gelegentlichen Publikationen die Evaluationsaktivitäten lediglich ausschnittsweise abzubilden (vgl. Deutsche Gesellschaft für Evaluation 2001). Innerhalb von Themenfeldern gibt es vereinzelte Ausnahmen, wie etwa in der Sozialen Arbeit in Deutschland (Projekt eXe; siehe den Beitrag von Haubrich in diesem Band) oder bei der Technologie in Österreich (fteval; siehe den Beitrag von Zinöcker/Dinges in diesem Band).

Dieser Sammelband will einen Beitrag zur Verbesserung der Informationsbasis leisten. Zentrale Ressource sind die beteiligten Autorinnen und Autoren, meist aktive Evaluationsfachleute auf einem bestimmten Gebiet, die in Evaluationsnetzwerken, in Austausch- und Kooperationsbeziehungen tätig sind, oft im Rahmen der Evaluationsgesellschaften einen professionellen Diskurs pflegen und aus dieser Position heraus einen Einblick in die Vielfalt der Evaluationsaktivitäten im jeweiligen Evaluationsfeld haben.

Das Vorhaben wird insofern erleichtert, als es sich bei den drei Ländern um solche handelt, in denen die deutsche Sprache gesprochen wird. Dies erleichtert den Austausch und die Erschliessung und Zugänglichkeit von Literatur und Evaluationsberichten.

Dabei ist zu beachten, dass die Schweizer Länderberichte die Evaluationsaktivitäten in den Landessprachen abdecken und sich nicht auf die deutschsprachigen Evaluationsaktivitäten beschränken.

Mit der Möglichkeit, sowohl länder- als auch themenfeldübergreifende Vergleiche aufgrund der Darstellungen vorzunehmen, werden neue Möglichkeiten des Lernens voneinander geschaffen. Sowohl was Theorien, Methoden und Instrumente der Evaluation in verschiedenen Ländern, als auch was deren praktische Anwendung in verschiedenen Themenfeldern betrifft: Von den Verfahren der Ausschreibung und Beauftragung über die Durchführung seitens verschiedener Arten von Anbietern und Kooperationen bis hin zur Berichterstattung, deren Veröffentlichung, Zugänglichkeit in Evaluationsdatenbanken und schliesslich Meta-Evaluationen und Evaluationssynthesen. Hier sehen wir die besondere Leistungsfähigkeit dieser Publikation – länder- und themenfeldübergreifend einen Diskurs zu ermöglichen über Möglichkeiten der Weiterentwicklung guter Methodik und Praxis der Evaluation.

Dieser Band erfasst erstmals systematisch den Entwicklungsstand der Evaluation in der Breite von zehn Themenfeldern in den drei Ländern Deutschland, Österreich und Schweiz. Ausserdem wird zu den drei Ländern ein aktueller Überblick gegeben, wie der Stand der institutionellen Verankerung der Evaluationsfunktion im jeweiligen Regierungssystem auf den verschiedenen föderalen Ebenen ist. Schliesslich wird der Forschungsstand dazu referiert, in welchem Masse Evaluation durch die Auftraggebenden und darüber hinaus durch weitere Berechtigte für ausgewiesene Zwecke genutzt wird.

3 Aufbau des Buches und Gliederungssystematik der Teilkapitel

Das Ziel der Publikation besteht darin, einen systematischen Überblick über die Evaluation in den drei Ländern in breiten Themenfeldern zu geben und – gerade auch im Ländervergleich – den Lesenden zu ermöglichen, den Entwicklungsstand der Evaluation zu bewerten und sich ein Bild von den Perspektiven der weiteren Entwicklung zu machen.

Dementsprechend ist das Buch in folgende drei Hauptteile gegliedert.

- Institutionelle Einbettung der Evaluationsfunktion
- Themenfelder der Evaluation
- Nutzung der Evaluationsfunktion sowie sektorale und nationale Trends

Diese drei Teile sind nachfolgend kurz skizziert.

Der **erste Teil** enthält neben dem vorliegenden Beitrag das Kapitel zur institutionellen Einbettung der Evaluationsfunktion. Darin eingeschlossen ist eine kurze Darstellung der institutionellen, namentlich verfassungsrechtlichen und weiteren relevanten gesetzlichen Grundlagen, die für die Evaluationstätigkeit im öffentlichen Bereich relevant sind. Die drei Länderbeiträge klären, wie die Evaluationsfunktion in den Organen von Parlament, Regierung und Verwaltung institutionell verankert ist, welche Konzeptionen umgesetzt sind und welche Ressourcen eingesetzt werden. Diese „institutionelle Einbettung“ stellt einen relevanten Ausschnitt der Kontextbedingungen dar, welche das Evaluationshandeln in den drei Ländern prägen. Diese Ausführungen werden in den Themenfeldern des nachfolgenden Hauptteils vertieft, ergänzt und ausdifferenziert. Nach den Beiträgen zu den drei Ländern erfolgt in einem vierten Unterkapitel ein resümierender Ländervergleich.

Der **zweite Teil** mit den zehn Themenbereichen stellt den umfangreichsten und zentralen Bereich des Buches dar. Es enthält zehn Hauptkapitel mit in der Regel vier Unterkapiteln – jeweils drei Länderbeiträge und einen ländervergleichenden Beitrag.

Die Themenfelder in alphabetischer Reihenfolge sind:

- Agrarpolitik
- Arbeitsmarktpolitik
- Bildung
- Energie- und Umweltpolitik
- Entwicklungszusammenarbeit
- Forschung und Technologie
- Gesundheit
- Institutionelle Politik
- Raumentwicklungspolitik
- Soziale Arbeit

In einigen Themenfeldern führten die Länderberichterstattenden vorab einen intensiven Austausch, um eine Arbeitsdefinition ihres Gegenstands zu erarbeiten. Beispielfhaft verwiesen sei auf die beiden Themenfelder „institutionelle Politik“ sowie „Raumentwicklungspolitik“. In der Regel wird der vierte, ländervergleichende Beitrag von einer Person geschrieben, die nicht mit einem Länderbeitrag befasst ist. Eine Ausnahme bildet hier das Themenfeld „Soziale Arbeit“, in dem die drei Länderberichterstattenden für den Ländervergleich verantwortlich zeichnen.

Für die Unterkapitel haben die Herausgeber im Vorfeld einen Rahmen gesetzt. Die damit vorgeschlagene inhaltliche Ausrichtung und Strukturierung soll sowohl den intersektoralen wie den internationalen Vergleich erleichtern und Anschlusspunkte bieten:

- Die Darstellung soll sich auf die neuere Entwicklung, etwa ab dem Jahr 2000, konzentrieren.
- Sowohl die Länderberichte wie der Ländervergleich sollen folgende fünf Aspekte behandeln:
 - a. Kontext der Evaluationstätigkeit: Skizze von Struktur und Entwicklung des Themenfeldes, und zwar so, dass Personen, die mit den länderspezifischen Besonderheiten nicht vertraut sind, die wesentlichen Merkmale erfassen können (z. B. bedeutende rechtliche und institutionelle Eigenschaften und Veränderungen).
 - b. Evaluationsaktivitäten im jeweiligen Themenfeld: Beschreibung auf allen Staatsebenen (Bund, Bundesländer/Kantone, kommunale Ebene). Berücksichtigung verschiedener Arten von Auftraggebern (staatliche und nicht-staatliche, auch Evaluationsaktivitäten ohne Auftraggebende), wie von Evaluationsanbietenden (private Unternehmen, Hochschulen etc.); sowohl von intern wie von extern durchgeführten Evaluationen. Falls sich eine spezifische Evaluation dazu eignet, können an diesem Beispiel typische Merkmale der Evaluationstätigkeit im Themenfeld herausgearbeitet werden.
 - c. Verbindungen mit anderen Evaluationsfeldern und internationale Entwicklungen: Herausragende Errungenschaften zum Beispiel konzeptioneller oder methodischer Art, die auch in anderen Themenfeldern beachtet werden (sollen); Einflüsse anderer Evaluationsfelder und aus anderen Ländern; Stellenwert internationaler Entwicklungen (besonders europäische Ebene).
 - d. Stand der Evaluation: Gegenüberstellung der Situation um 2000 mit der aktuellen; Institutionalisierung der Evaluation, Organisation der Evaluierenden; Stellenwert von Evaluationsstandards im jeweiligen Themenfeld. Wichtige begünstigende und hemmende Faktoren für eine qualitativ und quantitativ befriedigende Evaluationstätigkeit.
 - e. Bilanzierung: Diskussion besonderer Stärken und Schwächen; Entwicklungsperspektiven, wobei eine Stellungnahme der jeweiligen Beitragenden eingehen kann.
- Für jedes Themenfeld soll länderspezifisch und -übergreifend die relevante aktuelle Literatur eingearbeitet werden, sodass die Beiträge als Ausgangspunkt für eine systematische Beschäftigung mit Evaluation in bestimmten Evaluationsfeldern, zum Beispiel im Rahmen von Forschungsprojekten oder Lehrveranstaltungen, dienen können.

Der **dritte Teil** des Buches enthält zwei Kapitel mit resümierendem Charakter.

Das erste behandelt die „Nutzung der Evaluationsfunktion“ und damit die Frage, in welchem Masse der Prozess, insbesondere aber die Ergebnisse – Daten, zur Verfügung gestellte verdichtete Informationen, Schlussfolgerungen, Empfehlungen – aber auch die „Produkte“, welche diese Ergebnisarten enthalten (Präsentationen, Kurzfassungen, gedruckte und im Internet zur Verfügung gestellte Berichte usw.) – durch die vorgesehenen Adressa-

ten und Adressatinnen sowie durch die Öffentlichkeit genutzt werden. Abweichend vom Schema der vorangegangenen Hauptkapitel wird hier der Bericht zu Deutschland und Österreich zusammengefasst und jenem zur Schweiz gegenübergestellt. Die Grenzen dieser beiden empirisch basierten Studien sind in diesem Fall durch die beiden für die jeweiligen Länder zuständigen Evaluationsgesellschaften, die Gesellschaft für Evaluation (DeGEval) – für Deutschland und Österreich –, und die Schweizerische Evaluationsgesellschaft (SEVAL) definiert. Im Anschluss diskutiert ein weiterer Beitrag die vorgetragenen Befunde und stellt sie in einen weiteren Kontext.

Das abschliessende Kapitel zu sektoralen und nationalen Trends stellt Bezüge zwischen den Beiträgen her und formuliert zentrale Erkenntnisse, die sich aus Sicht der Autoren in der Auseinandersetzung mit den insgesamt fünfzig vorangehenden Beiträgen ergeben.

4 Anliegen dieses Buches – und Dank

Die Idee dieses Buches ist keinesfalls, die ‚besten‘ Themenfelder oder die ‚besten‘ Länder der Evaluation zu küren. Es geht darum transparent zu machen, wie Evaluation zur Gestaltung einer effizienten, sozial verantwortlichen, gerechten und nachhaltigen Politik beitragen kann, welche alternativen Möglichkeiten der institutionellen Einbindung von Evaluation in Parlament, Regierung und Verwaltung bestehen, und was nicht nur Evaluatoren und Evaluatorinnen, sondern auch Verwaltungsfachleute, Politikerinnen und Politiker und die interessierte Öffentlichkeit aus einer guten Evaluationspraxis lernen können. Dabei ist der Blick über das eigene Politikfeld hinaus und in andere Länder oft lohnenswert, um Anregungen für Variationen zu erhalten und so auf neue erfolgversprechende Lösungen zu treffen. Die Breite der Themenfelder und die Unterschiedlichkeit der Erfahrungen in verschiedenen Ländern in den Blick zu nehmen, dem liegt die Überzeugung zugrunde, dass die Evaluation nicht nur als interdisziplinäre wissenschaftliche Vorgehensweise angelegt werden muss. Darüber hinaus muss sie auch offen sein für Wissen und für Werte aus verschiedenen Themenfeldern, sie muss verschiedenste wissenschaftliche Vorgehensweisen mobilisieren, damit sie ihren besonderen Nutzen in einer Gesellschaft entfalten kann, die sich in immer kürzeren Abständen auf schnelle Veränderungen der politischen, sozialen, ökonomischen und natürlichen Bedingungen einstellen muss.

Die Herausgeber danken den Autoren und Autorinnen für die Mitwirkung bei diesem Vorhaben. Sie haben sich mit grossem Engagement und mit grosser Flexibilität der durch die Herausgeber gestellten Aufgabe angenommen und den definierten Bedingungen unterzogen. Den Autorinnen und Autoren sind wir besonders auch deshalb dankbar, weil sie in ihren Beiträgen ihre fundierten Kenntnisse und reichhaltigen Erfahrungen einfließen liessen, welche den Wert dieser Publikation ausmachen.

Es sei hervorgehoben, dass die Beitragenden ihre persönlichen Sichtweisen, Erkenntnisse und Überzeugungen vertreten. Sie sprechen nicht für die Institutionen, für die sie tätig sind. Die entsprechenden Institutionen können von den vorliegenden Äusserungen abweichende Auffassungen vertreten.

Unser besonderer Dank gilt aber auch Céline Widmer, Nicole Staubli und Marianne Müller für ihre verlässliche Unterstützung bei der redaktionellen Aufarbeitung des Textes.

Institutionelle Einbettung der Evaluationsfunktion

Institutionelle Einbettung der Evaluationsfunktion in Politik und Verwaltung in Deutschland

Die institutionelle Einbettung der Evaluationsfunktion in Politik und Verwaltung in der Bundesrepublik Deutschland im Detail nachzuzeichnen, bedürfte einer eigenständigen, umfassenden Monographie. In nahezu allen Politikfeldern werden Evaluationen mit je eigenen Fragestellungen, Schwerpunkten und methodischen Besonderheiten durchgeführt. Zudem haben sich das Ausmaß der Evaluation und die Art ihrer Institutionalisierung in der Systemgeschichte der BRD verändert. In diesem Beitrag können deshalb nur grob Entwicklungslinien der Institutionalisierung der Evaluationsfunktion nachgezeichnet werden (1). Darüber hinaus wird die Institutionalisierung der Evaluierung im Bereich der querschnittlichen Rechtsoptimierung fokussiert. Da die Gesetzesfolgenabschätzung (2) und das Standardkosten-Modell (3) als herausgehobene (Evaluierungs-)Verfahren in diesem Bereich angesehen werden können, werden ihre Verankerungen beachtet. Abschliessend werden zentrale Erkenntnisse des Artikels zusammengefasst (4).

1 Organisatorische und rechtliche Institutionalisierung der Evaluationsfunktion

Eine zentrale Stelle, die mit den Evaluationsaktivitäten in Politik und Verwaltung befasst ist, gibt es in der Ministerialverwaltung Deutschlands nicht. Auch ein gemeinsames Archiv oder eine Internet-Seite, auf der die von den Ministerien selbst durchgeführten oder von ihnen in Auftrag gegebenen Evaluationen¹ dokumentiert werden, wurde bislang nicht eingerichtet. Dies soll sich jedoch nun ändern: In der Umsetzung des von der Bundesregierung am 13. September 2006 beschlossenen Programms „Zukunftsorientierte Verwaltung durch Innovationen“ bestehen Bestrebungen, innerhalb der Bundesverwaltung eine Internet-Plattform mit Informationen zum Thema Evaluation einzurichten. Angesichts einer Entwicklung, nach der in allen Politikbereichen eine steigende Nachfrage nach Evaluationen festzustellen ist und in einigen Politikfeldern – wie zum Beispiel der Entwicklungspolitik – eine längere Evaluationstradition vorherrscht (Toepel/Tissen 2000), wird dabei keine vollständige Dokumentation der Evaluationsaktivitäten anzustreben sein. Ziel kann es jedoch sein, das Wissen um Evaluation und den Informationsaustausch – innerhalb der Verwaltung und mit externen Experten – zu verbessern.

¹ In der Praxis werden umfassende Evaluationen begrifflich vielfach nicht von anderen prüfenden (Forschungs-) Aktivitäten (z. B. Benchmarking) unterschieden. Vor diesem Hintergrund wird in diesem Beitrag ein breiter Evaluationsbegriff verwendet.

Evaluationen haben in Deutschland eine längere Tradition. Hellmut Wollmann spricht aktuell von einer dritten Welle der Evaluierung (Wollmann 2005b: 509). Während – vereinfacht gesprochen – die erste Welle der Evaluierung in den 1960er und 1970er Jahren im Wesentlichen auf die Verbesserung der Zielerreichung von Politik durch externe Evaluation abgezielt habe, hätten die Evaluierungsbestrebungen der zweiten Welle (späte 1970er und 1980er Jahre) primär eine Reduktion der Kosten von Politik unterstützen sollen. Heute in der dritten Welle würden – mit den Bestrebungen der Verwaltungsreform – interne Evaluationsverfahren (Kosten-Leistungsrechnung, Controlling usw.) und deren systematische Einbindung in die Verwaltungsführung im Mittelpunkt der Evaluationen stehen. Dieses Wellenmodell der Evaluation korrespondiert mit paradigmatischen Phasenmodellen der Systemgeschichte der Bundesrepublik Deutschland (z. B. Böhret/Konzendorf 1997), in denen nach einer Aufbauphase grob vereinfacht reformorientierte von am Status quo orientierten Phasen unterschieden werden können. Die unterschiedliche Intensität der Evaluationsaktivitäten sowie die politische Nutzung der Evaluationsstudien scheinen also mit politischen Grundausrichtungen zu korrelieren. Es kann vermutet werden, dass in reformorientierten Zeiten die Evaluationsaktivitäten zunehmen, da die Legitimierung einer Reformpolitik im Allgemeinen mit höherem Aufwand verbunden ist als die einer am Status quo orientierten Politik.

Frägt man vor diesem Hintergrund nach der Institutionalisierung der Evaluationsfunktion in Deutschland, so fällt auf, dass sich in der Phase der Reformpolitik der späten 1960er und der 1970er Jahre etliche Institute etablierten, die Evaluationsaufgaben übernommen haben.² So wurde zum Beispiel im Jahre 1967 das Institut für Arbeitsmarkt- und Berufsforschung (IAB) als Forschungsinstitut der Bundesanstalt für Arbeit³ gegründet.

„Ausschlaggebend war die Erkenntnis, dass eine aktive Arbeitsmarktpolitik auf die wissenschaftlich fundierte Analyse und Vorausschau angewiesen ist. Das IAB erhielt vom Gesetzgeber den Auftrag, Arbeitsmarktforschung aus der Perspektive verschiedener Disziplinen zu betreiben, um so den Arbeitsmarkt besser verstehen und Problemlösungen entwickeln zu können. (...) Erarbeitet werden Projektionen, Politiksimulationen, internationale und regionale Vergleiche, Gesamtrechnungen, sektorale und betriebliche Analysen, Evaluationen arbeitsmarktpolitischer Programme oder simultane Analysen von Angebot und Nachfrage.“⁴

Auch das Bundesinstitut für Berufsbildung (BIBB), das 1970 als Kompetenzzentrum zur Erforschung und Weiterentwicklung der beruflichen Aus- und Weiterbildung gegründet wurde, übernimmt Evaluationsfunktionen in seinem Politikbereich.

Aber auch schon vor diesen Institutsgründungen während der Reformphase um 1970 gab es Institutionalisierungen der Evaluationsfunktion. So hatte bereits das 1952 als Nachfolgeorganisation des Reichsgesundheitsamts gegründete Bundesgesundheitsamt (BGA) den Auftrag, Risiken für die Gesundheit von Mensch und Tier früh zu erkennen, diese zu

² Die im Weiteren genannten Institutionen wie das Institut für Arbeitsmarkt und Berufsforschung und das Bundesinstitut für Berufsbildung haben – neben der Evaluationsfunktion – weitere (Forschungs-)Aufgaben, so z. B. im Bereich der Grundlagenforschung. Es kann im Rahmen dieses Beitrags nicht ermittelt werden, wie sich das Verhältnis dieser unterschiedlichen Aktivitäten innerhalb der Institutionen entwickelt hat. Es wird vermutet, dass die Evaluationstätigkeit vielfach einen verhältnismäßig geringen, aber steigenden Umfang einnimmt.

³ Heute: Bundesagentur für Arbeit.

⁴ Internet-Seite des IAB. <http://www.iab.de/iab/ueberblick/ueberblick.htm> [27.3.2008].

bewerten und im Rahmen seiner gesetzlichen Kompetenzen einzudämmen. Nach einer Umstrukturierung des Gesundheitssystems im Jahre 1994 wurden die Aufgaben neuen Institutionen wie dem Bundesinstitut für Arzneimittel und Medizinprodukte und dem Bundesinstitut für gesundheitlichen Verbraucherschutz und Veterinärmedizin übertragen. Evaluationsfunktionen übernehmen aber auch die Landesrechnungshöfe sowie der in der Tradition der Preußischen Oberrechnungskammer 1950 in Frankfurt am Main errichtete Bundesrechnungshof (BRH). Dieser prüft nach Art. 114 Abs. 2 Satz 1 und 2 des Grundgesetzes für die Bundesrepublik Deutschland die Rechnung sowie die Wirtschaftlichkeit und Ordnungsmäßigkeit der Haushalts- und Wirtschaftsführung des Bundes. Weitere wichtige Rechtsgrundlagen seiner Prüfungen sind das Haushaltsgrundsätzegesetz, die Bundeshaushaltsordnung, das BRH-Gesetz, die Geschäftsordnung und die Prüfungsordnung des BRH.

Aber auch später in der eher am Status quo orientierten Phase um 1990 wurde die Evaluierungsfunktion organisatorisch verankert. So wurde 1990 das Büro für Technikfolgen-Abschätzung (TAB) beim Deutschen Bundestag mit dem Ziel eingerichtet, Beiträge zur Verbesserung der Informationsgrundlagen für forschungs- und technologiebezogene Beratungs- und Entscheidungsprozesse im Deutschen Bundestag zu leisten. Allerdings belegt der langwierige Prozess der organisatorischen Institutionalisierung, wie schwierig diese zwischenzeitlich geworden war.

Im Unterschied zu diesen organisatorischen Verankerungen etablieren sich in der derzeitigen Reformphase andere Formen der Institutionalisierung und ergänzen die älteren. Diese Formen haben einen geringeren Organisationsgrad und sind mithin flexibler. So werden vielfach wissenschaftliche Institute und private Auftragnehmer mit der Durchführung einer Evaluation beauftragt; die Evaluation wird als Projekt definiert und oftmals in Netzwerken von unterschiedlichen Partnern durchgeführt.

Beispielhaft für diese Netzbildung kann auf das Modellprogramm zur „Förderung der medizinischen Qualitätssicherung“ des Bundesministeriums für Gesundheit (BMG) hingewiesen werden. Innerhalb dieses Programms werden seit dem Jahr 2003 zehn Modellprojekte im Rahmen des Förderschwerpunktes „Benchmarking in der Patientenversorgung“ gefördert. Ziel des Förderschwerpunktes ist es, innovative Maßnahmen zur Verbesserung der Versorgungsqualität modellhaft zu entwickeln und zu erproben. Hierzu haben sich Einrichtungen aus dem ambulanten und stationären Bereich zusammengeschlossen. Der so ermöglichte kollegiale Erfahrungsaustausch zu medizinischen oder pflegerischen Themen soll einen systematischen Verbesserungsprozess durch Lernen von anderen (Benchmarking) einleiten.⁵

Handelt es sich bei diesem Beispiel um ein verwaltungsexternes Netzwerk, so können die in der deutschen Verwaltung vor allem seit den 1990er Jahren durchgeführten Vergleichsringe als Beispiele für verwaltungsinterne Evaluationen in Netzwerken angeführt werden.

Gemeinsam ist diesen Formen der Verankerung von Evaluation, dass sie nicht wie in den 1970er Jahren in Neugründungen von dauerhaften Organisationen münden, sondern dass in ihnen Wissen und Erfahrungen unterschiedlicher Institutionen befristet zur Durchführung von Evaluationen genutzt werden.

⁵ Siehe Internet-Seite des BMG: http://www.bmg.bund.de/nn_605054/DE/Themenschwerpunkte/Gesundheit/Krankenhaeuser/Voneinander-lernen-Benchmarki-6781.html [27.3.2008].

Einen Indikator für die Zunahme der Bedeutung von Evaluation als Instrument der Planung und Steuerung sieht die DeGEval, Gesellschaft für Evaluation, in der deutlichen Zunahme von Fundstellen zum Thema Evaluation im Koalitionsvertrag der großen Koalition vom 11.11.2005 mit 28 Nennungen im Vergleich zum Koalitionsvertrag der Vorgängerregierung mit vier Fundstellen.⁶

Zugenommen hat in der jüngeren Vergangenheit wohl auch die Verankerung von Evaluations- und Berichtspflichten in Gesetzen. Beispielfhaft kann hier auf das SGB XI (Soziale Pflegeversicherung) hingewiesen werden, in dem in § 10 Abs. 4 eine Berichts- und damit eine Evaluationspflicht verankert ist. Dort heißt es:

„Das Bundesministerium für Gesundheit berichtet den gesetzgebenden Körperschaften des Bundes im Abstand von drei Jahren, erstmals im Jahre 1997, über die Entwicklung der Pflegeversicherung, den Stand der pflegerischen Versorgung in der Bundesrepublik Deutschland und die Umsetzung der Empfehlungen und Vorschläge des Ausschusses für Fragen der Pflegeversicherung.“

Noch deutlicher ist die Pflicht zur Evaluation in § 55 des SGB II verankert. Der Paragraph lautet:

„§ 55 Wirkungsforschung

Die Wirkungen der Leistungen zur Eingliederung und der Leistungen zur Sicherung des Lebensunterhalts sind regelmäßig und zeitnah zu untersuchen und in die Arbeitsmarkt- und Berufsforschung nach § 282 des Dritten Buches (SGB III; G.K.) einzubeziehen. Das Bundesministerium für Arbeit und Soziales und die Bundesagentur können in Vereinbarungen Einzelheiten der Wirkungsforschung festlegen. Soweit zweckmäßig, können Dritte mit der Wirkungsforschung beauftragt werden.“

Über diese Verankerung der Evaluation in einzelnen Gesetzen hinaus findet in einigen Ländern Deutschlands (z. B. Hessen, Nordrhein-Westfalen) auch die generelle Befristung von Gesetzen (Sunset Legislation) Anwendung. So wird auf der Internet-Seite des Landes Nordrhein-Westfalen dargelegt, dass die Landesregierung dem Landtag bei allen zukünftigen Gesetzen eine Befristung vorschlagen werde. Danach solle das Gesetz an einem festgesetzten Datum außer Kraft treten. Ein in Kraft bleiben komme nur dann in Betracht, wenn im Rahmen einer „Beweislastumkehr“ die Fortdauer des Gesetzes als unverzichtbar notwendig begründet werden könne. „Alternativ kann im Gesetz gegenüber dem Landtag eine Berichtspflicht vorgesehen werden. In diesem Fall muss die Landesregierung zu dem festgesetzten Zeitpunkt dem Landtag die Wirkungen dieses Gesetzes, seine Notwendigkeit und Zweckmäßigkeit darlegen. Die Befristung gilt in gleicher Weise für Verordnungen“ und wurde auch – soweit es sich nicht um „Fundamentalrecht“ (z. B. Staatsverträge) handelt – auf bereits verabschiedete Rechtsnormen angewandt.

Alle diese Regeln und Vorschriften zur Evaluation von rechtlichen Normierungen lassen eine Betrachtung der querschnittigen Verankerungen von Evaluationen im Rechtsetzungsbereich als lohnend erscheinen. Deshalb sollen im Folgenden die institutionelle Ein-

⁶ Internet-Seite der Gesellschaft für Evaluation: http://www.degeval.de/index.php?class=Calimero_Article&id=329 [27.3.2008].

bettung der Gesetzesfolgenabschätzung sowie des Standardkosten-Modells als herausgehobene Evaluationsverfahren im Bereich der Rechtsoptimierung in den Mittelpunkt der Betrachtung gestellt werden.

2 Gesetzesfolgenabschätzung (GFA): Verfahren und Institutionalisierung

Gesetzesfolgenabschätzungen sind spezifische Evaluationsverfahren, mit deren Hilfe den politisch Verantwortlichen in allen Phasen des Rechtsetzungsprozesses rechtzeitig adäquate Informationen über zu erwartende oder eingetretene Wirkungen (z. B. Kosten, Nutzen, Praktikabilität) einer Normierungsabsicht oder einer bestehenden Regelung zur Verfügung gestellt werden sollen. Das heißt, Gesetzesfolgenabschätzungen sollen a) helfen, die beste programmatische Regelungsoption zu ermitteln, b) die Erstellung und Ausarbeitung des (vorläufigen) Gesetzestextes (z. B. einen Referentenentwurf) begleitend unterstützen, c) sie sind Verfahren zur Evaluierung von in Kraft getretenen Rechtsvorschriften (Gesetze, Verordnungen usw.) Aus diesen Anforderungen lassen sich drei Module der Gesetzesfolgenabschätzung ableiten (Böhret/Konzendorf 2001):

- *Die prospektive GFA (pGFA):* Als prälegislative Analytik liefert sie in einem noch „offenen Regelungsfeld“ alternative Regelungsmodelle (programmatische Leitvorstellungen). Darüber hinaus werden mit ihrer Hilfe Folgen der Regelungsalternativen abgeschätzt und bewertet. Die pGFA unterstützt damit den Rechtsetzungsprozess durch Systematisierung bei der Entwicklung von Regelungsalternativen und durch Abschätzungen in einer Phase, in der noch überlegt wird, ob man überhaupt und in welcher „Denk“richtung man eine rechtsförmige Regelung erwägen könnte.
- *Die begleitende GFA (bGFA):* Bei den Analysen der bGFA werden vorformulierte Rechtssätze (z. B. von Referentenentwürfen) auf Vollzugspraktikabilität, Befolgbarkeit, Verständlichkeit, Kosten-Nutzen-Relationen sowie institutionelle und rechtliche Funktionalitäten geprüft. Dabei wird unterstellt, die Rechtsvorschrift wäre schon in Kraft. Testmethoden wie Praxistests und Planspiele kommen hier ebenso zum Einsatz wie Prüfverfahren (Nutzwertanalyse, Schnittstellenanalyse, Funktionendiagramm, Standardkosten-Modell etc.).
- *Die retrospektive GFA (rGFA):* Diese ex-post Evaluierung (die ggf. periodisch wiederholt wird) soll die Bewährung von Rechtsvorschriften in ihrer Praxis nachprüfen. Die zentralen Fragen lauten: Werden die politischen Absichten mit den Regelungen erreicht und welche Nebenfolgen treten auf? Daraus ergeben sich Folgerungen für Novellierung, Aussetzung oder Neugestaltung.

Die Module greifen zu unterschiedlichen Zeitpunkten in den Rechtsetzungsprozeß ein und unterscheiden sich hinsichtlich ihres konkreten Forschungsziels und der angewandten Verfahren (Methode). Die nachfolgende Tabelle (Tabelle 1) gibt hierzu eine Übersicht.

Tabelle 1: Module der Gesetzesfolgenabschätzung nach Phasen, Ausrichtungen, Ergebnissen

| Phase der Rechtsetzung | Ausrichtung | | | Ergebnis („Produkt“) | GFA-Modul |
|------------------------------|--|--|---|---|------------------------------------|
| | sachlich | zeitlich | methodologisch | | |
| Regelungs- intention | Regelungsnot- wendigkeit; Alternativen- entwicklung | vor der Recht- setzung (ex ante) | Optimierung Intention/ Folgen | kein Regelungs- bedarf oder optimale Regelungsalternative | prospektive GFA (pGFA) |
| Regelungs- entwurf | Rechtsförmige Ausgestaltung | während/ begleitend | Folgenorientier- te Analyse, Formulierungs- hilfen | Diskussionsentwurf Referentenentwurf | begleitende GFA (bGFA) |
| Geltende Rechtsvorschrift | Bewährungs- prüfung (laufend, perio- disch, final) | nach der Rechtsetzung (ex post) | Realisierung/ Intention | überprüfte Normen (Bestätigung; Ände- rungsempfehlung) | retrospektive GFA (rGFA) |

Laut Handbuch für Gesetzesfolgenabschätzung sind für Gesetzesfolgenabschätzungen die vergleichende Vorgehensweise in allen Modulen, die Einbeziehung von Normadressaten und Experten in den Rechtsetzungsprozess (Konsultation) sowie die Beachtung unterschiedlicher Zeitdimensionen (kurz- und langfristige Folgen) konstitutiv. Zusätzlich sollte über die angewandten Verfahren und die Ergebnisse Transparenz hergestellt werden. Das heißt, eine hinreichende Dokumentation der Folgenabschätzung ist notwendig, um die Nachvollziehbarkeit der angewandten Verfahren und ihrer Ergebnisse zu ermöglichen.

Gesetzesfolgenabschätzungen sind komplexe Verfahren, die den Entstehungsprozess eines Rechtsentwurfs aktiv mitgestalten. Hierzu werden zum Beispiel in der prospektiven GFA Regelungsalternativen entwickelt und empirisch vergleichend unter Einbeziehung von Normadressaten und Experten im Hinblick auf Prüfkriterien wie Zielerreichung, Praktikabilität oder Kostenwirkungen abgeschätzt. Für die Methodik ist prägend, dass sie aktiv im Entstehungsprozess eines Rechtsentwurfs eingesetzt wird und den Erkenntnisstand in einem Regelungsfeld praxisrelevant verdichtet.

Eine OECD-Studie (2004) bescheinigt Fortschritte der Gesetzesfolgenabschätzung in Deutschland. Allerdings wird zugleich betont, dass noch erhebliche Defizite bei ihrer Anwendung bestehen. Nach einer kurzen Darstellung der vorherrschenden Institutionalisierungsformen der Gesetzesfolgenabschätzung soll im Folgenden nach den Ursachen für diese Anwendungsdefizite gefragt werden.

Institutionelle Einbettung der Gesetzesfolgenabschätzung

Die Gesetzesfolgenabschätzung ist auf Bundesebene in der Gemeinsamen Geschäftsordnung der Bundesministerien (GGO) rechtsförmig verankert. Auch in verschiedenen Ländern (z. B. Rheinland-Pfalz, Thüringen, Sachsen-Anhalt, Sachsen) finden sich Verankerun-

gen in der GGO oder in eigens geschaffenen Richtlinien und Verwaltungsvorschriften. Beispielfhaft soll hier auf die GGO des Bundes Bezug genommen werden.

Nach der am 1. September 2000 in Kraft getretenen überarbeiteten Gemeinsamen Geschäftsordnung der Bundesministerien des Bundes ist bei jeder Vorlage von Gesetzen und Rechtsverordnungen eine Gesetzesfolgenabschätzung durchzuführen. Deren Ergebnisse sind in der Begründung und auch im Vorblatt darzustellen. Gemäss § 44 GGO soll die Gesetzesfolgenabschätzung insbesondere folgende Inhalte einschließen:

- Beabsichtigte Wirkungen, unbeabsichtigte Nebenwirkungen (wobei die Darstellung im Einvernehmen mit den jeweils fachlich zuständigen Bundesministerien erfolgen und hinsichtlich der finanziellen Auswirkungen erkennen lassen muss, worauf die Berechnungen oder die Annahmen beruhen)
- Haushaltsauswirkungen beim Bund, den Ländern und Kommunen
- Vollzugsaufwand
- Kosten für die Wirtschaft, insbesondere auch für mittelständische Unternehmen
- Auswirkungen auf Einzelpreise und das Preisniveau
- Auswirkungen auf die Verbraucherinnen und Verbraucher
- Bürokratiekosten.

Zudem ist in der GGO vorgesehen, dass vom federführenden Ressort festzulegen ist, „ob und nach welchem Zeitraum zu prüfen ist, ob die beabsichtigten Wirkungen erreicht worden sind, ob die entstandenen Kosten in einem angemessenen Verhältnis zu den Ergebnissen stehen und welche Nebenwirkungen eingetreten sind.“ (GGO § 44 Abs. 7).

Trotz dieser verpflichtenden rechtlichen Verankerung kann die Anwendung der Gesetzesfolgenabschätzung in der Regel – wie bereits gesagt – bislang nicht überzeugen. Fragt man die Mitarbeiterinnen und Mitarbeiter der Ministerialverwaltung nach den Ursachen für diese Anwendungsdefizite, so werden zum Beispiel der mangelnde politische Wille zur Transparenz, zeitliche Restriktionen und nicht ausreichende Methodenkenntnisse als Gründe angeführt. Es wird hier jedoch vermutet, dass ein weiterer Grund in der Institutionalisierung der Gesetzesfolgenabschätzung selbst zu suchen ist. Zwar wurden Referate oder Normprüfungsstellen eingerichtet, die sich mit dem Thema Gesetzesfolgenabschätzung befassen, doch können diese aus Kompetenz- und Kapazitätsmangel nur in Ausnahmefällen eine aktive Begleitung eines Rechtssetzungsprozesses durchführen. In der Regel werden sie vielmehr nach Erstellung eines Referentenentwurfes um Stellungnahme gebeten, sie werden also im arbeitsteiligen Prozess der Ministerialverwaltung am Ende des Produktionsprozesses des rechtsförmigen Entwurfes um Mitzeichnung betreffend der Ausführungen zu den Gesetzesfolgen im Vorblatt und in der Begründung des Gesetzes gebeten. Da jedoch in den rechtsförmigen Bestimmungen zur Gesetzesfolgenabschätzung kaum verfahrensmässige Vorgaben (z. B. keine Notwendigkeit zur Darstellung von Regelungsalternativen sowie zur Konsultation) und Mindeststandards für die Gesetzesfolgenabschätzung vorgegeben sind, ist eine hinreichende Qualitätssicherung auf diese Weise kaum zu gewährleisten. So ist aufgrund dieser mangelnden Vorgaben in der Regel nicht zu erkennen, auf welchen Grundlagen die Angaben zu den Folgen des Gesetzes beruhen.

Diese in Deutschland vorherrschende Form der Institutionalisierung der Gesetzesfolgenabschätzung ist unzureichend und dürfte ein gewichtiger Erklärungsfaktor für das eklatante Anwendungsdefizit der Gesetzesfolgenabschätzung in Deutschland sein.

Um dieses zu überwinden, ist es nützlich, sich wesentliche Eckpunkte der Institutionalisierung der Folgenabschätzung (Impact Assessment) in der Europäischen Kommission zu vergegenwärtigen, denn dort werden Folgenabschätzungen seit dem Jahr 2005 für alle politikdefinierenden Entwürfe des Arbeitsprogramms der Kommission standardmäßig angewandt.⁷

Die Folgenabschätzung der EU, die methodisch weitgehende Überschneidungen mit der prospektiven Gesetzesfolgenabschätzung aufweist, wurde Anfang des Jahres 2003 im Bereich der Kommission eingeführt und ersetzte die bis dahin vorherrschenden sektoralen Abschätzungsverfahren.

Für jede Initiative des Arbeitsprogramms der Kommission ist ein Ablaufplan für eine Folgenabschätzung zu erstellen, der von der federführenden Institution (Generaldirektion) möglichst frühzeitig an alle anderen Generaldirektionen übermittelt werden soll. Für jede Institution und die interessierte Öffentlichkeit ist anhand des Ablaufplans ersichtlich, zu welchen Initiativen (Themen) eine Folgenabschätzung durchgeführt wird. Zudem können sich die in der Folgenabschätzung einzubeziehenden Stellen frühzeitig für ihre Mitwirkung einrichten.

Für die Durchführung der Folgenabschätzung bestehen verfahrensmäßige Mindeststandards. So besteht die Pflicht, ein Konsultationsverfahren durchzuführen, für das ebenfalls Mindeststandards verabschiedet wurden. Zudem muss die Folgenabschätzung dokumentiert werden, wobei die Dokumentation wie folgt aufgebaut sein sollte:

Zusammenfassung

Teil 1: Verfahrensfragen und Konsultation der betroffenen Parteien

Teil 2: Problemdefinition

Teil 3: Ziele

Teil 4: Politikoptionen

Teil 5: Analyse der Auswirkungen

Teil 6: Vergleich der Optionen

Teil 7: Monitoring und Bewertung.

Dieser Aufbau folgt den im Wesentlichen vorgegebenen analytischen Schritten der Folgenabschätzung. Verpflichtend sind die Analyse des Problems, die Formulierung von Zielvorgaben und von Regelungsalternativen (Politikoptionen), die vergleichende Analyse der Auswirkungen im Hinblick auf Soziales, Wirtschaft und Umwelt sowie eine abschließende Bewertung der Regelungsalternativen.

Verantwortlich für die Durchführung zeichnet die jeweils federführende Institution, die jedoch frühzeitig eine dienststellenübergreifende Steuerungsgruppe unter Einbeziehung des Referates „Strategische Planung und Koordinierung“ des Generalsekretariats einberufen

⁷ Siehe http://ec.europa.eu/governance/impact/practice_en.htm [27.3.2008].

muss. Nur in begründeten Ausnahmefällen kann von der Einrichtung dieser Steuerungsgruppe abgesehen werden.

Für die Qualitätssicherung der Folgenabschätzung ist es entscheidend, dass das Generalsekretariat im Rahmen der dienststellenübergreifenden Konsultation die Qualität des Berichts prüft. Dabei kommen ihm die verfahrensmäßigen Mindeststandards zugute.

Vergleicht man diese Form der Institutionalisierung mit der in Deutschland vorherrschenden Form, so ist leicht erkennbar, dass sie der aktiv gestaltenden Methodik besser gerecht wird. Zudem wird durch die verfahrensmäßigen Mindeststandards eine Qualitätssicherung ermöglicht. Aufgrund dieser Erkenntnisse ist es empfehlenswert, ein ähnliches Verfahren auch in deutschen Verwaltungen zu implementieren.

3 Standardkosten-Modell (SKM)⁸

Im Bereich der querschnittigen Rechtsoptimierung hat das in den Niederlanden entwickelte Standardkosten-Modell auch in Deutschland vor dem Hintergrund des von der Bundesregierung beabsichtigten Abbaus überflüssiger Bürokratie besondere Bedeutung gewonnen.

3.1 Methode

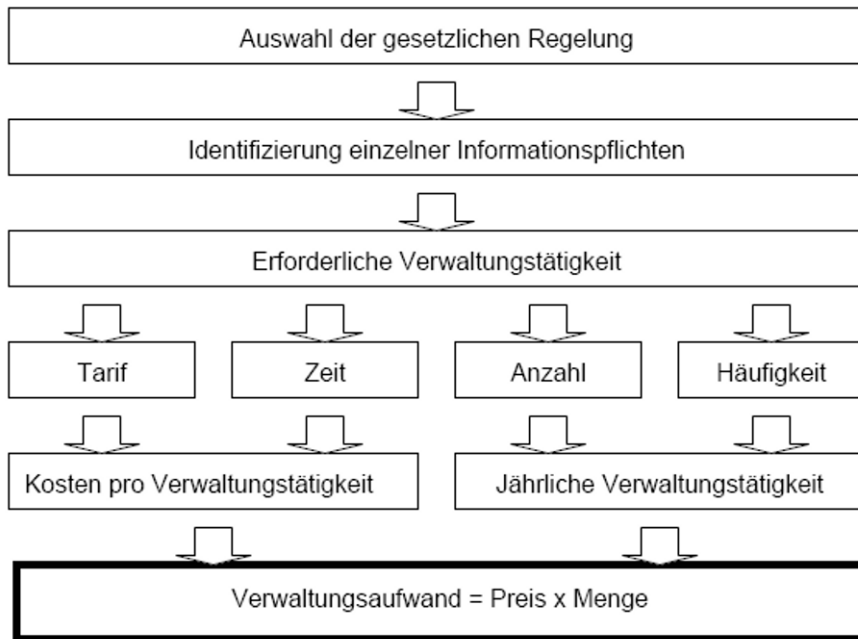
Das Standardkosten-Modell ist ein Verfahren zur Quantifizierung von Belastungen durch gesetzlich veranlasste Informationspflichten. Diese werden im Methodenhandbuch der Bundesregierung definiert „als aufgrund von Gesetz, Rechtsverordnung, Satzung oder Verwaltungsvorschrift bestehende Verpflichtungen, Daten und sonstige Informationen für Behörden oder Dritte zu beschaffen, verfügbar zu halten oder zu übermitteln.“ (Bundesregierung 2006: 7) In der Diskussion um das Standardkosten-Modell wird immer betont, die Reduktion dieser Informationspflichten sei ein unpolitischer Akt, da die Informationspflichten unabhängig von materiellen Standards seien.

Die Methode macht es zunächst notwendig, die Rechtsgebiete auf Informationspflichten zu untersuchen. Werden Pflichten gefunden, können Mittel – wie die Verlängerung von Berichtsintervallen oder die Absenkung von Informationsstandards – zu ihrer Reduktion genutzt werden. Die Höhe der Bürokratiekosten hängt von der Zahl der betroffenen Unternehmen, der Häufigkeit der Informationsübermittlung und dem Zeitaufwand für diese ab.

⁸ Das SKM hat einen eingeschränkten Fokus: Es stellt keine umfassende Evaluation eines Gesetzes (Programms) dar, zumal der zentrale Evaluationsgegenstand (Informationskosten) von vorneherein feststeht. Trotz dieser Einschränkung soll die Institutionalisierung des SKM hier analysiert werden, um Voraussetzungen für einen hohen Institutionalisierungsgrad in Bund und Ländern herauszuarbeiten. So wird argumentiert, dass die Reduktion von Komplexität die Institutionalisierung erleichtert. Gleichzeitig wird durch diese Reduktion auf eine umfassende, unterschiedliche Folgen (Ziele, Kosten, Nebenwirkungen etc.) beachtende Evaluation verzichtet. Methodisch kann argumentiert werden, dass die multidimensionale Fragestellung nach „besserer Rechtsetzung“ auf eine Dimension reduziert und unterkomplex behandelt wird.

Zur Ermittlung der Informationspflichten werden deshalb methodisch im Wesentlichen die in der nachfolgenden Abbildung dargestellten Schritte vollzogen. Eine vertiefende Darstellung der Methode findet sich im Methodenhandbuch der Bundesregierung.

Abbildung 1: Ablauf Standardkosten-Modell (Quelle: Bundesregierung 2006: 26)



3.2 Institutionalisierung

In den letzten Jahren haben bereits etliche Landesregierungen unter der Projektleitung der Bertelsmann Stiftung in Pilotprojekten Anwendungserfahrungen im Bereich des Standardkosten-Modells gewonnen.⁹ Neben einem SKM-Scan Projekt, mit dem die wichtigsten Kostentreiber im Bereich des Landesrechts identifiziert wurden, wurden SKM-Projekte in den Bereichen Landesbauordnung und Weiterbildungsrecht durchgeführt.

Die Bundesregierung hat im April 2006 das Programm „Bürokratieabbau und bessere Rechtsetzung“ beschlossen, um gezielt Informationspflichten bei Wirtschaft, Bürgerinnen und Bürgern sowie Verwaltung abzubauen bzw. zu vereinfachen und so mehr Raum für privates Engagement, Innovationen und Investitionen zu eröffnen. Zum Abbau dieser Informationspflichten greift die Bundesregierung ebenfalls methodisch auf das Standardkosten-Modell zurück. Ziel der Bundesregierung ist es, mit dieser Methode bis zum Jahre 2011 25 Prozent der Kosten einzusparen, die durch staatliche Informationspflichten hervorgerufen werden. Da die Bundesregierung dem Bürokratieabbau einen hohen Stellenwert bei-

⁹ Siehe http://www.bertelsmann-stiftung.de/cps/rde/xchg/bst/hs.xsl/prj_5036.htm [27.3.2008].

misst und das Standardkosten-Modell auf das gesamte Bundesrecht anwendet, wurde ein hoher Grad der Institutionalisierung gewählt.

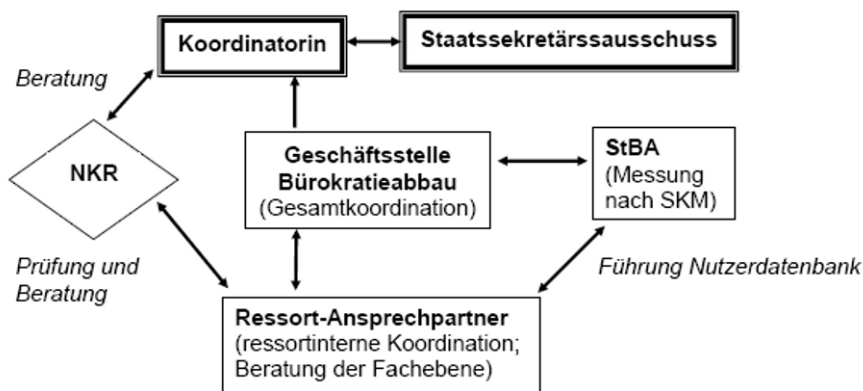
In rechtlicher Hinsicht wurde das Standardkosten-Modell im Gesetz zur Einsetzung eines Nationalen Normenkontrollrates verankert. In § 1 Abs. 2 weist das Gesetz dem Normenkontrollrat die Aufgabe zu, „die Bundesregierung dabei zu unterstützen, die durch Gesetze verursachten Bürokratiekosten durch Anwendung, Beobachtung und Fortentwicklung einer standardisierten Bürokratiekostenmessung auf Grundlage des Standardkosten-Modells zu reduzieren“. Für die Bundesministerien wurde die Befassung mit den Bürokratiekosten in § 44 Abs. 5 GGO verankert. Demnach müssen die Bundesministerien „die Bürokratiekosten im Sinne des § 2 Abs. 1 des Gesetzes zur Einsetzung eines Nationalen Normenkontrollrats ermitteln und darstellen.“

In organisatorischer Hinsicht wurde die unten skizzierte Form der Institutionalisierung gewählt (Abbildung 2).

Koordiniert wird der Bürokratieabbau von einem mit besonderen Aufgaben betrauten Staatssekretär des Bundeskanzleramtes (Koordination), der einem Staatssekretärsausschuss vorsitzt. Unterstützt wird er bei seiner Tätigkeit von einer ebenfalls im Kanzleramt angesiedelten Geschäftsstelle. Diese berät die Bundesministerien bei der Anwendung des Standardkosten-Modells. Für die Reduzierung der Informationspflichten in ihrem Geschäftsbereich sind die Ressorts jedoch selbst verantwortlich. Die Datenbasis liefert das Statistische Bundesamt (StBA). Es führt die Aufwandsmessung und Berechnung der Bürokratiekosten für die Ressorts durch.

Im September 2006 konstituierte sich der aus acht Mitgliedern bestehende unabhängige Nationale Normenkontrollrat (NKR). Er hat die Aufgabe, die Gesetzesentwürfe der Bundesregierung auf Bürokratiebelastungen durch Informationspflichten zu prüfen. Durch eine Änderung der Geschäftsordnung der Regierung wurde dem Normenkontrollrat die Möglichkeit zur Stellungnahme bei Rechtsetzungsvorhaben ermöglicht. Über diese organisatorische Struktur hinaus werden Wirtschafts- und Kommunalverbände in den Bürokratieabbau einbezogen.

Abbildung 2: Standardkostenmodell – Organisationsstruktur



Angesichts dieser organisatorischen Ausdifferenzierung kann von einem beachtlichen Grad der Institutionalisierung gesprochen werden, insbesondere wenn man den im Vergleich niedrigen Institutionalisierungsgrad der Gesetzesfolgenabschätzung betrachtet. Thematisiert man die Ursachen für diese Unterschiede im Ausmaß der Institutionalisierung, so ist zunächst der deutliche politische Wille zum Abbau bürokratischer Lasten für die Wirtschaft anzuführen. Er wird unterstrichen durch die Verankerung des Standardkosten-Modells an prominenter Stelle im Bundeskanzleramt. Zusätzlich dürfte aber auch die methodische Beschränkung des Standardkosten-Modells auf das Prüfkriterium „Informationspflichten“ eine Erleichterung für die Institutionalisierung darstellen. Im Gegensatz zur Gesetzesfolgenabschätzung, die ein multidimensionales Prüfraster nutzt, um optimale Regelungsalternativen zu finden, findet im Standardkosten-Modell angesichts vielfältiger Anforderungen an ein gutes Gesetz eine erhebliche methodische Reduktion statt. Diese hat jedoch auch Vorteile. Denn die Auswertung dieses einen Prüfkriteriums kann von den im Rechtsetzungsprozess beteiligten Stellen methodisch leichter durchgeführt werden als eine multidimensionale Prüfung. Durch diese Praktikabilität und durch die Darstellung der Informationspflichten als unpolitische Größen kann die Akzeptanz gegenüber der Methodik und ihrer breiten Anwendung erhöht werden.

4 Resümee

- Verschiedene Indikatoren weisen darauf hin, dass das Thema Evaluation in Deutschland derzeit Konjunktur hat. So gibt es Bestrebungen zur Einrichtung einer Website zum Thema Evaluation in der Bundesverwaltung; zudem wird das Thema vergleichsweise häufig im Koalitionsvertrag der Großen Koalition genannt. Auch wird die Evaluationsfunktion vermehrt in Gesetzen verabschiedet und in einigen Ländern als Sunset Legislation implementiert. Vieles spricht dafür, dass die Häufigkeit von Evaluationsaktivitäten in Politik und Verwaltung mit der paradigmatischen Ausrichtung von Politik korreliert. In Reformphasen scheinen die Evaluationsaktivitäten gegenüber am Status quo orientierten Phasen gesteigert zu werden.
- Die Art der institutionellen Einbettung der Evaluationsfunktion hat sich in der Geschichte der Bundesrepublik Deutschland verändert – flexible, temporäre Institutionalisierungen ergänzen oder ersetzen derzeit dauerhafte Organisationsformen.
- Dass noch Defizite in der Form der Institutionalisierung bestehen, belegt die institutionelle Einbettung der Gesetzesfolgenabschätzung. Hier können Politik und Verwaltung in Deutschland von der Europäischen Kommission lernen.
- Am hohen Grad der Institutionalisierung des Standardkosten-Modells wird deutlich, welche Bedeutung dem politischen Willen bei der Institutionalisierung der Evaluationsfunktion zukommt. Zudem kann vermutet werden, dass die Reduktion von methodischer Komplexität sowie die Entpolitisierung des Evaluationsverfahrens die Akzeptanz erhöhen und die Institutionalisierung in Politik und Verwaltung erleichtern.
- Mit dem beabsichtigten Aufbau einer Internet-Plattform zum Thema Evaluation in der Bundesverwaltung bestehen erste Bestrebungen, um einen interministeriellen Austausch zum Themenkomplex zu etablieren. Es bleibt abzuwarten, ob dies der Nukleus

eines zukünftigen Kompetenzzentrums „Evaluation“ der obersten Bundesbehörden ist. Freilich müssten dafür notwendige Voraussetzungen erst geschaffen werden. Zu denken ist beispielsweise an eine bessere zeitliche Verflechtung von Evaluationsaktivitäten mit politischen Entscheidungsprozessen.

Institutionelle Dimensionen von Evaluierung in Österreich

Die Notwendigkeit, staatliche Interventionen zu rechtfertigen, verschafft deren Analyse und Bewertung eine wichtige Funktion im politisch-institutionellen Kontext. Evaluierung wird damit integraler Bestandteil des Politikzyklus und des institutionellen Regelwerks des politischen Systems. Positionierung und Bedeutung von Evaluierung in politischen Prozessen und Institutionen in Österreich sind jedoch überwiegend heterogen und von unterschiedlichen Rahmenbedingungen, Traditionen und Funktionen geprägt.

1 Rechtliche Rahmenbedingungen

Eine generelle Verpflichtung, die Wirkungen des Handelns staatlicher Organe regelmäßig und systematisch zu überprüfen, kennt die österreichische Bundesverfassung nicht, wohl aber Vorschriften, welche die effiziente und effektive Erreichung eines Ziels – was allenfalls durch eine Evaluierung zu überprüfen wäre – sicherstellen sollen. So normiert das Bundes-Verfassungsgesetz, dass bei der Haushaltsführung die „Grundsätze der Sparsamkeit, Wirtschaftlichkeit und Zweckmäßigkeit“ zu beachten sind (Art. 51a Abs. 1 B-VG).

Dem steht – enger als es ein heutiger Evaluierungsbegriff nahe legt – die klassische Rechnungs- und Gebarungskontrolle gegenüber. Diese erfolgt durch den Rechnungshof, der unmittelbar dem Nationalrat¹ untersteht und als dessen Organ tätig wird. Der Rechnungshof kann auch Länder und Gemeinden prüfen und wird in diesem Fall als Organ des jeweiligen Landtages tätig.² Der Rechnungshof ist von der Bundes- und den Landesregierungen unabhängig, bestimmt seine Prüfungsgegenstände selbst (oder auf Verlangen des Nationalrates) und genießt daher im politischen Institutionengefüge Österreichs eine bedeutende Position. Neben der unmittelbaren Rechnungskontrolle soll der Rechnungshof auch die Einhaltung der in Art. 51a festgelegten Grundsätze überprüfen, was den Ausgangspunkt auch für inhaltlich tief gehende Prüfungen bietet – zumal auf Bundesebene nicht nur die Bundesorgane selbst, sondern auch vom Bund kontrollierte Unternehmen und eine Reihe weiterer gesetzlich bestimmter Organisationen, wie zum Beispiel Selbstverwaltungskörper, von der Prüfung durch den Rechnungshof umfasst sind (Art. 121-128 B-VG).

¹ Das österreichische Parlament besteht aus zwei Kammern, dem Nationalrat und dem Bundesrat (der Länderkammer).

² Zusätzlich verfügen jedoch auch die Bundesländer außer Wien über eigene Landesrechnungshöfe, Wien über das Kontrollamt.

Die Aufgaben des Rechnungshofes werden durch das Rechnungshofgesetz 1948 (RHG) konkretisiert. Dabei wird auch der Gesichtspunkt der Überprüfung der inhaltlichen Angemessenheit geschärft. In § 2 Abs. 1 RHG wird ausdrücklich festgehalten: „Keinesfalls darf er [der Rechnungshof, Anm. R.P.] sich auf die bloß ziffernmäßige Nachprüfung beschränken.“ Nicht zuletzt gestützt auf diese Bestimmungen versteht sich der Rechnungshof daher immer wieder nicht nur als Berater und Evaluator der geprüften Stellen, sondern führt auch als Vergleichsbasis für spätere Gebarungüberprüfungen Begutachtungen von Gesetzes- und Verordnungsentwürfen durch. Über diese Normenbegutachtungen wird dem Nationalrat vom Rechnungshof jährlich berichtet. Der Rechnungshof ist somit das wesentlichste Instrument des Parlaments zur Kontrolle, Analyse und Bewertung der Bundesverwaltung. Darüber hinaus sehen zahlreiche Materiengesetze regelmäßige Berichterstattungspflichten der Bundesregierung an den Nationalrat vor, die Elemente einer ex-post-Evaluierung enthalten können (Jetzinger 2007).

Aus nahe liegenden Gründen sollte, noch bevor es zu einer Prüfung durch den Rechnungshof kommt, Sorge getragen werden, dass die Angemessenheit einer geplanten Maßnahme im Sinn der mit ihr verbundenen Ziele gewährleistet ist. Indem die Schaffung gesetzlicher Normen das zentrale staatliche Instrumentarium darstellt, ist die Prüfung und Bewertung von Gesetzesfolgen und -wirkungen der erste Schritt zu einer institutionell breit abgestützten Evaluierungsfunktion im politischen Entscheidungsprozess. Dass Letzteres bis heute ein Desiderat bleibt, wird noch zu sehen sein.

Auf Bundesebene besteht die Verpflichtung der Bundesregierung, die Auswirkungen – wenn auch nur primär hinsichtlich der finanziellen Belastung des Bundes – von geplanten Rechtsvorschriften³ möglichst genau ex ante abzuschätzen und darzustellen, explizit seit 1979 (Irresberger 2005). Eine für die gesamte Bundesverwaltung übergreifend gültige gesetzliche Regelung fand diese Verpflichtung im Bundeshaushaltsgesetz 1986 (BHG). Dieses bestimmt, dass neben den erwarteten Kosten und Erlösen auch der genaue Nutzen darzustellen ist, der die erwarteten Kosten rechtfertigt (§ 14 Abs. 1 BHG; vgl. auch Simma 2007: 48).

Höhere Bedeutung kommt Evaluierungsfunktionen in den Rechtsgrundlagen für staatliche Interventionen zu, wenn diese durch finanzielle Anreize und nicht durch hoheitliche Anordnungen erfolgen. Wenn eine bestimmte, erwünschte Verhaltensweise nicht angeordnet werden kann oder soll, sondern die Adressaten der Regelung einen finanziellen Vorteil für dieses Verhalten erhalten, scheint zumindest aus der Sicht des Gesetzgebers der Wirkungszusammenhang zwischen Norm und Verhalten lockerer und das Risiko für finanzielle Schäden höher zu sein. Folglich findet sich für die Durchführung von Vorhaben oder Programmen – also sich finanziell auswirkenden Maßnahmen auf bestehender gesetzlicher Grundlage – zusätzlich zu den oben genannten Vorschriften die Bestimmung, dass diese dahin gehend zu prüfen sind, „ob der mit der Durchführung solcher Vorhaben und Programme angestrebte Erfolg und die zu seiner Erreichung vorgesehenen Maßnahmen noch mit den ... Zielen im Einklang stehen“; erforderlichenfalls wären die Maßnahmen einzustel-

³ Das betrifft Regierungsvorlagen zu Bundesgesetzen und Verordnungsentwürfe, nicht aber selbständige Gesetzesanträge („Initiativanträge“) im Parlament.

len oder abzuändern (§ 47 BHG). Der Begriff „Evaluierung“ (bzw. Evaluation) kommt im Haushaltsgesetz dennoch kein einziges Mal vor.

Eine weitere Konkretisierung erfährt diese implizite Evaluierungsverpflichtung im Bereich der Förderung. Während die zuvor genannte haushaltsgesetzliche Bestimmung auf alle konkreten finanzwirksamen Maßnahmen zutrifft, bestehen zusätzliche Richtlinien für den Bereich der privatrechtlichen Förderungen durch Zuschüsse, zinsbegünstigte Darlehen und dergleichen. Dazu ist zu bemerken, dass Förderungen im Rahmen der so genannten Privatwirtschaftsverwaltung des Bundes (Art. 17 B-VG) sehr breite Anwendung finden. Da hier somit keine gesetzlichen Ansprüche und daher ein größerer Spielraum der Verwaltung bei der Gestaltung entsprechender Förderungsmaßnahmen gegeben sind⁴, ist die Verwaltung gehalten, die Wirkung der Maßnahmen besonders sorgfältig zu überprüfen. Die Allgemeinen Rahmenrichtlinien zur Gewährung von Förderungen aus Bundesmitteln 2004 (ARR) verpflichten daher die zuständigen Organe, „nach Abschluss der geförderten Leistung eine Kontrolle und Evaluierung durchzuführen, ob und inwieweit der mit der Förderungsgewährung angestrebte Erfolg erreicht wurde.“ Darüber hinaus sind Zwischenevaluierungen durchzuführen. Diese Bestimmungen gelten subsidiär auch für alle anderen Förderungsrichtlinien des Bundes (§ 13 ARR).

Dieser in den Allgemeinen Rahmenrichtlinien verwendete Begriff der Evaluierung bezieht sich nicht nur auf die der Verwaltung selbst entstehenden Kosten, sondern auch auf die Wirkungen, welche die betreffenden Maßnahmen auf Dritte haben. Eine solche Einbeziehung des nicht-staatlichen Bereichs in die Gesetzesfolgenabschätzung ist gegenüber der Ermittlung der unmittelbaren administrativen Kosten des Bundes in Österreich noch vergleichsweise neu. Aufgrund eines Ministerratsbeschlusses vom 22. Jänner 1999 sollen seither im Zuge von Regierungsvorlagen für Bundesgesetze nicht nur die budgetären Auswirkungen, sondern auch die Auswirkungen auf Beschäftigung und Wettbewerbsfähigkeit dargestellt werden (Irresberger 2005), denn: „Vor der Verfassung schuldet die Politik nur das einwandfreie Gesetz. Aber vor der Bevölkerung ... das wohlerwogene ... Gesetz.“ (Schäffer 2005a: 11)

Die Tendenz zu einer gesamthafteren Betrachtung von Gesetzesfolgen, die auch nicht-finanzielle Dimensionen und die rechtsunterworfenen Personen einbezieht, hat sich seit einigen Jahren – auch im internationalen Kontext⁵ – verstärkt (vgl. Binder et al. 1999). Da aber eine komplexe ex-ante-Bewertung kaum zu bewerkstelligen ist, enthält eine zunehmende Anzahl österreichischer Bundesgesetze „echte“ Evaluierungsbestimmungen. Beispiele dafür sind etwa folgende Bestimmungen: § 113 Abs. 6 des Telekommunikationsgesetzes 2003 (TKG), der vorschreibt, „regelmäßig eine Evaluierung der gesetzlichen Bestimmungen vorzunehmen und im Abstand von zwei Jahren dem Nationalrat einen diesbezüglichen Bericht vorzulegen“; § 124b⁶ Abs. 6 des Universitätsgesetzes 2002 (UG, in der Fassung des BGBl. I Nr. 74/2006), der vorschreibt, „die Auswirkungen ... in Zusammenarbeit mit den Universitäten zu evaluieren und dem Nationalrat ... einen Bericht vorzulegen“; oder § 9c

⁴ Man spricht hier deshalb auch von „Ermessensausgaben“.

⁵ Die maßgebliche Rolle der EU als Mitinitiatorin dieser Entwicklung muss hier aus Platzgründen ausgeblendet bleiben.

⁶ In dieser Bestimmung geht es um die Zulassungsbestimmungen für vom deutschen Numerus Clausus betroffene Studien und die so genannte Quotenregelung für die Medizin- und Zahnmedizin-Studienplätze.

Abs. 7 des Immissionsschutzgesetzes-Luft (IG-L, in der Fassung des BGBl. I Nr. 34/2006), der die Evaluierung der Maßnahmen zur Einhaltung von Grenz- oder Zielwerten vorschreibt. Diese Aufzählung ließe sich weiter fortsetzen. Darüber hinaus gibt es, oft im EU-Kontext, gesetzliche Regelungen, die selbst eine Prüfung (und daher in gewissem Sinn auch Evaluierung) Dritter begründen. Ein klassisches Anwendungsgebiet dafür ist die Umweltpolitik mit dem Umweltkontrollgesetz 1998 (UKG), mit dem auch das Umweltbundesamt in seiner derzeitigen Form eingerichtet wurde, und weiterführenden, spezielleren Gesetzen, wie beispielsweise dem Umweltverträglichkeitsprüfungsgesetz 2000 (UVP-G).

Auf materienübergreifender Ebene wird ein anderer Teilaspekt von Evaluierung im Rahmen der amtlichen Statistik abgedeckt. Das Bundesstatistikgesetz 2000 (BStatG) setzt als Ziel „den Bundesorganen zur Planung, Entscheidungsvorbereitung und Kontrolle von Maßnahmen sowie der Wissenschaft, der Wirtschaft und der Öffentlichkeit“ Daten bereit zu stellen (§ 1). Mit dem BStatG wurde auch die „Statistik Österreich“ als Anstalt des öffentlichen Rechts in ihrer derzeitigen Form eingerichtet. Durch den gesamthaften Ansatz der rechtlichen Grundlagen der Bundesstatistik wurde die Möglichkeit geschaffen, auf statistische Grundlagen auch im Zusammenhang mit den oben genannten Berichts- und Evaluierungspflichten zurückzugreifen.

Schließlich hinterlässt auch die „administrative burdens“-Welle in Österreich ihre Spuren. Mit Hilfe des Standardkostenmodells (vgl. Frick et al. 2007) werden alle Rechtsnormen, die Informationsverpflichtungen für Unternehmen enthalten, bewertet. Es handelt sich somit nur um eine Evaluierung der Auswirkungen dieser Verpflichtungen auf Dritte in der (nicht evaluierten) Annahme, dass eine Reduktion der gegenständlichen Informationspflichten positive Nettoeffekte zeitigen würde. Dieses Vorhaben soll bis 2010 abgeschlossen sein (Ministerratsbeschlüsse vom 27.4.2006 und 14.2.2007). Zusätzlich wurde die Anwendung des Standardkostenmodells bei der Schaffung neuer Bundesgesetze, Verordnungen und Richtlinien durch den 2007 neu in das BHG eingefügten § 14a festgelegt. Diese Maßnahmen stellen jedoch einen eher isolierten, nicht auf eine Evaluierung der Gesamtwirkungen staatlicher Intervention bedachten Ansatz dar, der nicht zu einer systematischeren Integration von Evaluierung in den Politikzyklus beiträgt.

Insgesamt ist auf Bundesebene Vorsorge getroffen, dass von Organen des Bundes Evaluierungen nachgefragt werden, insbesondere wenn Maßnahmen des Bundes wirtschaftliche Auswirkungen auf diesen selbst oder aber auch auf Dritte haben. Die betreffenden Vorschriften sind allerdings häufig so allgemein gehalten, dass bestimmte Ausprägungen von Evaluierung in formeller oder materieller Hinsicht nicht determiniert werden. Eine Institution im engeren Sinn, die als Organisation für Evaluierung auf Bundesebene zuständig wäre, gibt es nicht. Allerdings weist die Tätigkeit des Rechnungshofes Elemente einer Evaluierungsfunktion auf.

Exkurs: Der Österreich-Konvent

Von Juni 2003 bis Jänner 2005 trat der „Österreich-Konvent“ zur Beratung einer Staats- und Verfassungsreform zusammen. Insofern stellte der Konvent selbst gewissermaßen ein Evaluierungsprojekt dar. Zusammengesetzt war der Konvent aus 71 VertreterInnen der maß-

geblichen Organe auf Bundes-, Landes- und Gemeindeebene, der politischen Parteien, der Sozialpartner sowie aus VirilistInnen. Den Vorsitz führte Franz Fiedler, der während des ersten Tagungsjahres noch Präsident des Rechnungshofes war. Auch die Rolle von Evaluierung als Instrument in Gesetzgebung und Verwaltung wurde thematisiert. Aber obwohl im Mandat der Ausschüsse enthalten war, Evaluierung als Teil der Wirkungsorientierten Verwaltungsführung zu beraten, ergibt sich kein Bild einer konsistenten Einbettung von Evaluierung in den Berichten und Vorschlägen des Konvents. Vor allem im Zusammenhang mit der Finanzverfassung wird „regelmäßige Evaluierung“ gefordert (Bericht des Ausschusses 10 vom 15.7.2004; Österreich-Konvent Endbericht, Teil 3, vom 31.1.2005), während dies in anderen Themenbereichen nur ganz am Rand berührt wird.

Freilich war das Hauptthema des Österreich-Konvents eine Neuordnung der Kompetenzverteilung zwischen Bund, Ländern und Gemeinden, die häufig als kompliziert und ineffizient kritisiert wird. Gemäß der geltenden Bundesverfassung gibt es mehrere Ebenen des Zusammenwirkens von Bund und Ländern, vom Tätigwerden der Länder als Bundesbehörden (mittelbare Bundesverwaltung) über die selbstständige Vollziehung von Bundesgesetzen durch die Länder bis hin zur Gesetzgebung der Länder in Ausführung von Bundesgesetzen oder in ihren selbstständigen Wirkungsbereichen. Art. 15 Abs. 1 B-VG normiert nämlich, dass alle nicht ausdrücklich dem Bund übertragenen Zuständigkeiten in Gesetzgebung und Vollziehung Landessache sind. Tatsächlich sind aber die staatlichen Kernaufgaben dem Bund zugewiesen, sodass der gestalterische Spielraum der Landesgesetzgebung – insbesondere im Vergleich mit Deutschland und der Schweiz – eher gering ist. Auf der Landesgesetzgebung lastet daher ein hoher Rechtfertigungsdruck, der unter Umständen die Bereitschaft, Evaluierungselemente in die Gesetzgebung einfließen zu lassen, erhöht (Raich 2005; Steiner 2005; Simma 2007: 52–59).

2 Evaluierungsfunktion in drei ausgewählten Materien

Die allgemeinen Rechtsvorschriften, die von der Bundesebene ausgehen, stellen – ergänzt durch spezielle materienrechtliche Bestimmungen – nur einen Teil der Randbedingungen für die institutionelle Einbettung von Evaluierung dar. Anhand von drei ausgewählten Beispielen sollen zunächst die jeweiligen Rechtsgrundlagen im Zusammenhang mit der Evaluierungspraxis dargestellt werden.

2.1 Forschung und Technologie

Maßgebliche Rechtsgrundlagen sind das Forschungs- und Technologieförderungsgesetz (FTF-G), das Gesetz zur Errichtung der Österreichischen Forschungsförderungsgesellschaft 2004 (FFG-G), das Forschungsorganisationsgesetz (FOG; Berichtswesen: §§ 6–9) sowie die auf diesen Gesetzen basierenden Förderungsrichtlinien. Das FTF-G (§ 15 Abs. 2) normiert erstmals auf gesetzlicher Ebene die Evaluierungsgrundsätze als Mindestanforderungen für die Richtlinien. Diese verlangen ein schriftliches Evaluierungskonzept für jedes Programm mit genauen Angaben über die anzuwendenden Evaluierungsverfahren und -indikatoren

(Punkt 2.2 FTE-Richtlinien). Evaluierungsfunktionen sind darüber hinaus in den durch die genannten Gesetze eingerichteten Förderungsinstitutionen FFG und FWF verankert, die dabei weitgehend unabhängig vorgehen können.

Insbesondere ex-ante-Projektevaluierungen gehören zum Standardrepertoire sämtlicher Forschungs- und Technologieförderungsinstitutionen und -programme, für welche die entsprechenden Verfahren – meist in Form von Peer Reviews und Expertenpanels – ebenfalls auf Richtlinienenebene kodifiziert sind. Demgegenüber schienen interim- und ex-post-Evaluierungen für die Funktionstüchtigkeit des Förderungssystems lange von untergeordneter Bedeutung zu sein und werden massiv erst seit etwa Mitte der 1990er Jahre eingefordert. Erst die Zunahme an forschungs- und technologiepolitischen Maßnahmen seit diesem Zeitraum, gepaart mit den Erfahrungen aus der vollen Teilnahme Österreichs an den EU-Rahmenprogrammen, verstärkte den Bedarf an Evaluierungen zur Rechtfertigung dieser Initiativen. Seither ist eine große Anzahl von Evaluierungen durchgeführt worden (Platform/Austrian Council 2007), die fallweise auch im Gesetzgebungsprozess Bedeutung erlangen, wie dies bei der Evaluierung des Fonds zur Förderung der wissenschaftlichen Forschung – FWF – und des Forschungsförderfonds für die gewerbliche Wirtschaft – FFF – (Arnold 2004) im Zuge der Forschungsförderungs-Strukturreform 2004 der Fall war.⁷

2.2 Tertiäre Bildung

Maßgebliche Rechtsgrundlagen für das öffentliche Hochschulwesen sind das Universitätsgesetz 2002 (UG) und das Fachhochschul-Studiengesetz 1993 (FHStG). Das Universitätsgesetz bestimmt, dass die Universitäten ihr gesamtes Leistungsspektrum nach internationalen Standards regelmäßig zu evaluieren haben, auf Veranlassung der Universitätsleitung oder des Ministeriums extern (§ 14 UG).⁸ Die entsprechende Praxis ist erst in Entwicklung und jedenfalls mit internationalen Beispielen, wie der britischen Research Assessment Exercise, nicht vergleichbar (Pechar 2006; Campbell 2006).

Im Fachhochschulsektor ist die Evaluierungsfunktion im Akkreditierungsverfahren eingebettet. Die Verlängerung der Akkreditierung eines Studienganges ist dabei an das Vorliegen einer positiven Evaluierung geknüpft (§ 13 Abs. 2 und 2a FHStG). Die Ziele, Methoden, Verfahren, Gegenstände und Umsetzung dieser Evaluierungen wurden detailliert 2004 in einer Verordnung des Fachhochschulrates festgelegt. Diese Evaluierungsverfahren kombinieren interne und externe Elemente und gehören zur regelmäßigen Praxis im Fachhochschulsektor (EvalVO FHR).

⁷ Diese Evaluierung wurde ergänzt durch eine Prüfung durch den Rechnungshof. Zwischen beiden Untersuchungen bestand ein gewisses Konkurrenzverhältnis, das sich auch in unterschiedlichen Ergebnissen und Empfehlungen niederschlug. So empfahl der Rechnungshof die Zusammenlegung von FWF und FFF, während die internationale Evaluierung davon ausdrücklich abriet; die Politik orientierte sich dabei an der Evaluierung.

⁸ Diese Bestimmungen lösten die detaillierte Evaluierungsverordnung 1997 aufgrund des früheren Universitätsorganisationsgesetzes 1993 (UOG) ab.

2.3 Regionalentwicklung

Die institutionelle Einbettung der Evaluierung regionalpolitischer Maßnahmen ist geprägt von der Dominanz der EU-Strukturfonds in diesem Bereich. Die entsprechenden EU-Vorschriften werden hier in den EU-Mitgliedstaaten direkt wirksam. Die EU-Ratsverordnung über den europäischen Regional-, Sozial- und Kohäsionsfonds sieht einen Katalog von Maßnahmen zur Evaluierung der Regionalförderung aus Strukturfondsmitteln vor, wobei für die ex-ante-Evaluierung die Mitgliedstaaten, für die ex-post-Evaluierung diese gemeinsam mit der Kommission zuständig sind (Art. 47–49 Verordnung (EG) Nr. 1083/2006 des Rates).

Diese Bestimmungen werden in mehreren Leitlinien der Kommission zu Methoden, Verfahren und Indikatoren spezifiziert. Die Leitlinien legen Bestandteile und Vorgangsweisen fest und bieten einen umfangreichen methodischen Rahmen (Europäische Kommission 2006a-d; 2007a, b). Evaluierung ist in diesem Kontext also sehr weit gehend normiert, die Evaluierungsfunktion ist fest verankert, da gar keine andere Wahl bleibt, als die vorgeschriebenen Evaluierungen durchzuführen. Dies galt selbstverständlich auch für die Strukturfondsperiode 2000 bis 2006, im Zuge derer ein entsprechender Bestand an Evaluierungen hervorgebracht wurde.

3 Evaluierungsfunktion, Akteurskonstellationen und Institutionalisierungsformen

Die für die drei beispielhaften Sektoren beschriebene Evaluierungspraxis auf Basis spezifischer Rechtsgrundlagen erfüllt jeweils bestimmte Funktionen, die im Kontext der betreffenden Akteurskonstellationen zur Ausformung unterschiedlicher institutioneller Merkmale führt. Diese umfassen über die formellen Rechtsvorschriften hinaus auch informelle Regeln, Handlungsweisen und Anreize, die die institutionelle Dimension von Evaluierung prägen.

3.1 Forschung und Technologie

Im Rahmen forschungs- und technologiepolitischer Entscheidungen besteht eine wesentliche Funktion von Evaluierung darin, sachgerechte Entscheidungsgrundlagen bereit zu stellen. In der Regel verfügen Entscheidungsträger in Politik und Verwaltung nicht über ausreichendes Wissen, um die Qualität von Forschungs- und Technologieentwicklungsvorhaben vollständig beurteilen zu können und eine genaue Vorstellung vom Ergebnis dieser Vorhaben zu besitzen. Mit anderen Worten: Forschungs- und Technologiefinanzierung durch die öffentliche Hand ist kein Beschaffungsvorgang. Der Einsatz von Evaluierungsverfahren ermöglicht daher Qualitätssicherung und legitimiert den Einsatz öffentlicher Mittel, da andere Beurteilungskriterien wie etwa der Markterfolg nicht anwendbar sind.

Daher haben die Politik als Geldgeber, intermediäre Einrichtungen als Geldverteiler und Forschungseinrichtungen als Geldempfänger gleichermaßen Interesse an funktionie-

renden Evaluierungsstrukturen: Die Politik, um sicher zu gehen, dass Steuergelder sinnvoll verwendet werden; die Intermediäre, um den politischen Auftrag effizient umsetzen zu können; schließlich die Forschung, um ihre Tätigkeit zu rechtfertigen. Diese Interessen treffen zunächst bei Projektentscheidungen aufeinander, weshalb ex-ante-Projektevaluierungen in den intermediären Förderungseinrichtungen fest institutionalisiert sind.

Die anderen Abschnitte des Politikzyklus erfordern auch Evaluierungen auf einem projektübergreifenden, programm- oder institutionenbezogenen Niveau, wenn Entscheidungen über neue Maßnahmen getroffen und diese umgesetzt werden sollen. Die Einbettung solcher, meist formativer oder summativer Evaluierungen in der österreichischen Forschungs- und Technologiepolitik ist weitgehend von allen beteiligten Akteuren getragen, da aus den vorher dargestellten Gründen jeder Akteur für sich einen spezifischen Nutzen aus der Evaluierungsfunktion ziehen kann.

Dieses Bewusstsein, in einem gemeinsamen Boot zu sitzen, findet seinen institutionellen Niederschlag in der Plattform Forschungs- und Technologieevaluierung. Die Einrichtung der Plattform ist aber auch im Kontext der seit dem EU-Beitritt neu hinzugekommenen, auch die nationale Ebene massiv beeinflussenden europäischen Aufgaben der Ministerien zu sehen (Gottweis/Latzer 2006: 722), deren Sensibilität für die Notwendigkeit umfassender Evaluierungsfunktionen dadurch wesentlich verstärkt wurde. Die Plattform ist ein 2006 aus einer 1996 gegründeten Arbeitsgemeinschaft hervorgegangener Verein, dessen Mitglieder vorwiegend die zuständigen Bundesministerien, Förderungsorganisationen und Forschungsinstitute sind. Die Mitglieder haben sich zur Einhaltung der von der Plattform beschlossenen „Standards der Evaluierung in der Forschungs- und Technologiepolitik“ verpflichtet (Plattform Forschungs- und Technologieevaluierung 2005).

Die Evaluierungsplattform ermöglicht im Interesse einer stärkeren Evaluierungskultur den Austausch zwischen allen betroffenen Akteuren im forschungs- und technologiepolitischen Prozess und stellt somit im Dreieck Politik – Intermediäre – Forschung einen wichtigen Balancefaktor dar. Gleichzeitig führt diese Konstellation aber dazu, dass sich keine dieser Akteursebenen die Evaluierungsfunktion als Governance-Instrument überwiegend aneignen kann, womit aber auch die Akzeptanz von Evaluierung sichergestellt bzw. erhöht werden kann. Umgekehrt schließen die damit gegebenen Interessenkonflikte der Plattform-Mitglieder eine direkte Mitwirkung der Plattform bei konkreten Evaluierungen aus. Programm- oder Institutionenevaluierungen werden daher immer ausgeschrieben.

Evaluierung in der österreichischen Forschungs- und Technologiepolitik besitzt also eine stark konsensorientierte institutionelle Dimension, die aber im Wesentlichen auf dieses Politikfeld beschränkt bleibt, sehr stark von informellen Übereinkommen der Beteiligten untereinander gekennzeichnet ist und wenig Berührungspunkte zu Evaluierungsfunktionen anderer Politikfelder aufweist. Daneben hat der 2000 eingerichtete Rat für Forschung und Technologieentwicklung versucht, sich an einer bestimmten Stelle des Politikzyklus zu positionieren und systematische und regelmäßige Programm- und Institutionenevaluierungen einzufordern (Empfehlung des Rates für Forschung und Technologieentwicklung vom 12.4.2005). Der Rat ist mit der Beratung der Bundesregierung in Fragen von Forschung, Technologie und Innovation betraut und soll Vorschläge für ein entsprechendes „Monito-

ring“⁹ erstatten (§ 17b Abs. 1 FTF-G). Der Rat ist gleichzeitig Mitglied der Plattform Forschungs- und Technologieevaluierung.

3.2 Tertiäre Bildung

Im Rahmen der Bereitstellung des tertiären Bildungsangebotes ergibt sich eine wesentliche Evaluierungsfunktion aus dem Monopolcharakter des öffentlich finanzierten Angebotes. Im Gegensatz zu Forschung und Technologieentwicklung kann die öffentliche Hand den Gegenstand ihrer finanziellen Zuwendung – Universitäten und Fachhochschulen – aufgrund eigener Kompetenz definieren und beurteilen. Die Quasi-Monopolstellung des öffentlich finanzierten Angebots schließt den Marktmechanismus der Qualitätssicherung durch Wettbewerb weitgehend aus.¹⁰

Interesse an Evaluierung hauptsächlich zur Qualitätssicherung haben die Politik als Geldgeber, um überprüfen zu können, ob das System ihre Erwartungen erfüllt; die Universitäten und Fachhochschulen insoweit, als dieser Nachweis Voraussetzung für weitere Finanzierungen ist; schließlich aber vorrangig die Bürger als potenzielle Konsumenten des tertiären Bildungsangebotes, deren Nachfragemacht nur auf eingeschränkte Wahlmöglichkeiten trifft. Die Bürger wollen daher *value for money* garantiert wissen, und das mit einer möglichst günstigen Kosten-Nutzen-Relation.

Im Bildungsbereich – wie in vielen anderen Bereichen auch – sind daher die staatlichen Ausgaben dem Urteil zumindest einer eingeschränkten Öffentlichkeit ausgesetzt, während Forschung und Technologieentwicklung weit gehend nur von ihresgleichen beurteilt werden. Die Öffentlichkeit potenzieller WählerInnen, die zugleich auch Rechtsunterworfenen sind, bewirkt daher eine stärker normative, gesetzliche Verankerung der Evaluierungsfunktion im Universitäts- und Fachhochschulsektor. Die Gewährleistung dieser gesetzlichen Anforderungen erfolgt bei Universitäten und Fachhochschulen durchaus unterschiedlich.

Das Universitätsgesetz überlässt die Durchführung der gesetzlich geforderten Evaluierungen den Universitäten selbst. Diese regeln die Evaluierungsangelegenheiten näher in ihrer jeweiligen Satzung, ohne dass dem in allen Fällen eigene Organisationseinheiten innerhalb der Universitätsverwaltung gegenüberstehen. Ein Beispiel für eine solche Einheit ist das Büro für Qualitätssicherung der Universität Wien. Obwohl gesetzlich nicht grundgelegt, besteht in Form der Österreichischen Qualitätssicherungsagentur (AQA – Austrian Agency for Quality Assurance) eine bundesweit tätige Einrichtung. Die AQA bietet den Universitäten zur Erfüllung ihres gesetzlichen Auftrages, aber auch den Fachhochschulen, die Entwicklung und Durchführung von Evaluierungsverfahren an. Die AQA ist als gemeinnütziger Verein gestaltet, dessen Mitglieder die Österreichische Rektorenkonferenz, die Fachhochschulkonferenz, die Österreichische Hochschülerschaft und das Bundesministerium für Wissenschaft und Forschung sind. Im Gegensatz zur Plattform Forschungs- und Technologieevaluierung (deren Mitglied die AQA ist) kann die AQA damit selbst Evaluierungen

⁹ Der Begriff „Evaluierung“ findet sich im Aufgabenkatalog des Rates nicht. Der Rat lässt auch Studien erstellen, die sich mit den Möglichkeiten von Monitoring und Evaluierung befassen (zuletzt etwa Rhomberg et al. 2006).

¹⁰ Das Faktum des gemeinsamen Geldgebers beschränkt den von diesem z. T. „inszenierten“ Wettbewerb der Universitäten und Fachhochschulen untereinander.

durchführen, weist aber durch die Integration von Evaluierungs-Auftraggebern und -Betroffenen konsensuale Elemente auf.

Einen wesentlich höheren Verbindlichkeitsgrad weist die Situation im Fachhochschulbereich auf, wo sogar ein behördliches Verfahren vorliegt, weil die Fachhochschul-Trägerorganisationen im Gegensatz zu den Universitäten in der Regel keine öffentlich-rechtlichen Körperschaften sind. Die entsprechende Behörde ist der Fachhochschulrat, der für die Akkreditierung und Evaluierung der Fachhochschulen und Fachhochschul-Studiengänge zuständig ist und dafür Richtlinien und Verordnungen beschließt.¹¹ Aufgrund dieser Merkmale ist die partnerschaftliche Komponente der Evaluierungsfunktion im Fachhochschulbereich schwächer ausgeprägt.

Insgesamt ist also die institutionelle Verankerung der Evaluierungsfunktion im Bereich der tertiären Bildung durchaus von der Rechtfertigungsnotwendigkeit gegenüber den Steuerzahlern geprägt, die hier im Gegensatz zu Forschung und Technologieentwicklung auch in größerer Zahl potenzielle „Kunden“ und somit Nutznießer von Evaluierungsergebnissen darstellen. Die Evaluierungsfunktion ist stärker formalisiert, die Konsensdimension erfährt dadurch Einschränkungen.

3.3 Regionalentwicklung

Die Maßnahmen im Bereich der Regionalförderung sind primär durch die politisch-gesellschaftlich erwünschte Um- bzw. Gleichverteilung bestimmter Ressourcen motiviert. Evaluierung hat in erster Linie die Funktion, für die Steuerzahler, die hier keine unmittelbare Nachfragemacht besitzen, eine angemessene Kosten-Nutzen-Relation sicherzustellen. Während für die unmittelbaren Mittelallokationsentscheidungen Evaluierung eine nur untergeordnete Rolle spielt, stehen die Überprüfung von Effizienz und Effektivität im Vordergrund.

Die hauptsächlich an Evaluierung interessierte Akteursebene befindet sich entsprechend der makroökonomischen Begründung der regionalpolitischen Maßnahmen im EU-Kontext auf einer eher abstrakten Ebene, das heißt es handelt sich um die verantwortlichen Stellen der EU-Kommission, der Bundesregierung und der Landesregierungen. Diese stehen vor der Notwendigkeit, die durch die Regionalförderungen hervorgerufenen Markteingriffe zu rechtfertigen ohne sich wie bei den beiden anderen Beispielen auf das Argument des Marktversagens verlassen zu können. Natürlich haben auch die unmittelbaren Nutznießer ein Interesse am Nachweis der Zweckmäßigkeit der Maßnahmen.

Das durch die EU fest gefügte Regelwerk für die Evaluierung gibt den beteiligten Akteuren aber kaum Möglichkeiten, durch ihre Interessenkonstellation die institutionelle Ausprägung der Evaluierungsfunktion zu beeinflussen. Diese Möglichkeiten beschränken sich auf die Art und Weise, wie die EU-Vorgaben umgesetzt werden. Hierbei spielt die föderalistische Komponente eine vergleichsweise bedeutende Rolle.

Die Evaluierungsfunktion der aus den EU-Strukturfonds finanzierten Regionalförderungen ist sowohl auf Bundes- wie auch auf Länderebene verankert, weil einerseits die

¹¹ Ein ähnliches Verfahren gilt für die Privatuniversitäten unter der Zuständigkeit des Akkreditierungsrates (Universitäts-Akkreditierungsgesetz 1999).

Landesbehörden mit der Programmverwaltung betraut sind, andererseits die Republik Österreich gegenüber der EU vom Bund vertreten wird. Darüber hinaus übernehmen einige Förderungseinrichtungen des Bundes Teile der technischen Abwicklung. Um die Abwicklung der Regionalförderungsprogramme zu koordinieren, bedienen sich Bund und Länder der Österreichischen Raumordnungskonferenz (ÖROK). Die ÖROK ist ein gemeinsam von Bund, Ländern und Gemeinden getragenes Organ unter dem Vorsitz des Bundeskanzlers. Zur Unterstützung der Evaluierung der Strukturfondsprogramme wurde 2002 bei der ÖROK die „Koordinations- und Arbeitsplattform Evaluierung“ (KAP-EVA) eingerichtet, die auch ihrerseits bereits evaluiert wurde (Strohmeier/Holzinger 2004).

Als Auftraggeber der jeweils konkreten Evaluierungen fungieren die Verwaltungsbehörden (die Bundesländer), sodass die KAP-EVA tatsächlich nur eine koordinierende Rolle wahrnimmt. Die Evaluierungsfunktion in der EU-finanzierten Regionalförderung ist einerseits stark formalisiert, andererseits in organisatorischer Hinsicht schwach ausgeprägt. Die Außenwirkung ist vergleichsweise schwach, indem auch hier Evaluierung eine weit gehend systeminhärente Funktion hat, deren spezifische formelle und informelle Regeln kaum darüber hinausreichende Anknüpfungspunkte haben.

4 Abschließende Beobachtungen

Auf gesamtstaatlicher Ebene ist die Evaluierung staatlicher Interventionen eine in nur wenigen generellen, materienübergreifenden Normen und Institutionen verankerte Funktion. Im Bereich der gesetzlichen Normen konzentriert sich diese institutionelle Verankerung von Evaluierung oder verwandten Verfahren auf Aspekte, die eine wirtschaftliche Auswirkung auf den Bund selbst zur Folge haben. Erst in jüngster Zeit treten auch die wirtschaftlichen Auswirkungen auf Dritte bei solchen Erwägungen in den Vordergrund, ohne aber dass sich daraus ein ganzheitlicher Ansatz entwickelt hat. Teilansätze finden sich vor allem im Haushaltsrecht und in der Gebarungskontrolle, aber auch der Aspekt der statistischen Grundlagen bietet eine institutionelle Basis für Evaluierung. Demgegenüber geben die zunehmenden Evaluierungsverpflichtungen (anstelle früher üblicher, einfacher Berichtspflichten) in vielen Materiengesetzen noch ein recht uneinheitliches Bild ab.

Diese bundesgesetzlichen Grundlagen richten mitunter auch Institutionen (im Sinn von Organisationseinheiten) ein, die mit einzelnen Aspekten von Evaluierung betraut sind. Das gilt in erster Linie für den „klassischen“ Prüfungs- und Erhebungsbereich, wie er allgemein etwa vom Rechnungshof und der Bundesstatistik, speziell zum Beispiel vom Umweltbundesamt vertreten wird. Die organisatorische Institutionalisierung von Evaluierung ist ansonsten aber eher schwach ausgeprägt: Die Bundesministerien verfügen allesamt nicht über dezidierte Evaluierungsabteilungen, trotz der den BundesministerInnen auferlegten Evaluierungsverpflichtungen. Evaluierungen werden daher in der Regel in Auftrag gegeben, wobei die entsprechenden Nachfrage- und Angebotsroutinen den institutionellen Rahmen abgeben, der im Wesentlichen durch die Umsetzung des europäischen Vergaberechts in Form des Bundesvergabegesetzes 2006 (BVerG) bestimmt ist.

Das weit gehende Fehlen eigener fachlicher Evaluierungskapazitäten in der Bundesverwaltung wird teilweise durch mehr oder weniger informelle Plattformen ausgeglichen,

die von Auftraggebern von Evaluierungen, den Evaluierten sowie Evaluierenden getragen werden (vgl. die oben genannten Beispiele). Gegenüber einer intramuralen Institutionalisierung von Evaluierung im Sinn von Organisationseinheiten bieten solche Foren den Vorteil, das Bewusstsein für und die Akzeptanz von Evaluierung zu erhöhen. Darüber hinaus wird der größte Nachteil des Beauftragungsprinzips – nämlich Defizite bei Kontinuität und Aufbau von institutionellem Wissen – damit in Teilbereichen wettgemacht. Mehr oder weniger stark formalisierte Interessenausgleichsmechanismen, die mehrere betroffene Seiten integrieren, können in Österreich auf eine starke Tradition zurückblicken, hauptsächlich in Form der – auch durch viele gesetzliche Mitwirkungsrechte verankerten – Sozialpartnerschaft. Dieser Ansatz ist aber eben auch in Sachgebieten, in denen die Sozialpartner nur sekundär betroffen sind, vorzufinden und verweist auf eine starke Klammer zwischen Staat und Zivilgesellschaft (siehe Rathkolb 2005 zu den Grundmustern der Zweiten Republik). Nicht zuletzt durch die Tendenz, die formellen Mitspracherechte der Sozialpartner einzuschränken (Tálos 2006; z. B. auch im Bereich der Forschungspolitik: Gottweis/Latzer 2006: 721), können alternative Formen des Interessenausgleichs an Bedeutung gewinnen.

Jenseits einzelner Politikfelder ist schließlich zu beobachten, dass ein wesentlicher Einfluss auf die institutionelle Verankerung von Evaluierung im weitesten Sinn von den EU-Institutionen ausgeht. Dem durch den Beitritt zur EU entstandenen Kompetenzzuwachs der Bundesverwaltung steht eine entsprechende Prägung (und auch Modernisierung) der österreichischen Bürokratie gegenüber, deren Handeln auch durch die Umsetzung europäischer Ziele mitbestimmt ist (Falkner 2006: 90). In diesem Kontext spielt die Analyse und vergleichende Bewertung der Mitgliedstaaten eine immer größere Rolle. Besonders die Umsetzung der Lissabon-Strategie für Wachstum und Beschäftigung enthält evaluierungsähnliche Elemente in einer breit angelegten Querschnittsdimension, wie sie bisher allenfalls im Rahmen der OECD üblich waren. Das von jedem Mitgliedstaat zu erstellende Nationale Reformprogramm, welches im Rahmen der jährlichen Frühjahrsgipfel-Prozedur Gegenstand eines Fortschrittsberichtes und von Empfehlungen der Kommission (aktuell KOM(2006) 816 endg.) ist, erhöht auch in Österreich den Druck, Evaluierung im Politikzyklus stärker zu verankern. Das Bewusstsein der Politik dafür dürfte jedenfalls stärker werden: Allein im Programm der seit Jänner 2007 amtierenden Bundesregierung kommt – quer über alle Bereiche – der Begriff „Evaluierung“ 38 mal vor (Regierungsprogramm für die XXIII. Gesetzgebungsperiode 2007), mehr als eine Verdoppelung seit 2003.

Die institutionelle Einbettung der Evaluationsfunktion in der Schweiz

1 Einleitung

Nach Artikel 170 der Bundesverfassung (BV) der Schweiz muss das Parlament dafür sorgen, dass die Massnahmen des Bundes auf ihre Wirksamkeit überprüft werden. Damit wird die Pflicht zur Evaluation der Gesetzgebung in weltweit ziemlich einzigartiger Weise auf Verfassungsstufe verankert.

Im folgenden Beitrag gehe ich auf die Entwicklung und den heutigen Stand der Gesetzesevaluation in der Schweiz ein (siehe auch Mader 2003; 2005; 2006; 2007). Dabei werde ich insbesondere die verschiedenen Formen der Institutionalisierung der Gesetzesevaluation¹ auf der Ebene des Bundes darstellen. Ich werde auch auf Aspekte hinweisen, welche die Situation in der Schweiz von derjenigen in andern Ländern unterscheiden. Zudem werde ich aufzuzeigen versuchen, welche Faktoren die Entwicklung der Gesetzesevaluation in der Schweiz besonders begünstigt haben. Auch werde ich einige Ausführungen zum Stand der Gesetzesevaluation auf der Ebene der Kantone machen. Zunächst scheint es mir jedoch unerlässlich, einige begriffliche Klärungen vorzunehmen, um sicherzustellen, dass ein gemeinsames Grundverständnis der Thematik besteht.

2 Begriffliche Klärungen

Gesetze werden erlassen, um bestimmte Wirkungen zu erzielen. Durch die Beeinflussung oder Steuerung des Verhaltens der Regelungsadressaten (Einzelpersonen, Unternehmen, Behörden) will der Gesetzgeber bestimmte Ziele erreichen beziehungsweise gesellschaftliche Probleme lösen. Die Gesetzgebung ist in diesem Sinne ein Instrument. Aber erfüllt sie in der Praxis ihre Funktion überhaupt? Werden die gesetzten Ziele erreicht? Welches sind die Wirkungen der Gesetzgebung?

Die Gesetzesevaluation befasst sich mit ebendiesen Fragen und stützt sich dabei auf die getroffenen Massnahmen des Staates. Bei diesem Ansatz handelt es sich um die methodi-

¹ Ich verwende hier den Begriff der „Gesetzesevaluation“. Namentlich mit Blick auf die Frage der institutionellen Einbettung der Evaluationsfunktion besteht meines Erachtens kein wesentlicher Unterschied zwischen Evaluation staatlicher Massnahmen und derjenigen öffentlicher Politiken (public policies). In einer Rechtsordnung, die dem Legalitätsprinzip verpflichtet ist, gibt es keine wichtigen staatlichen Massnahmen oder öffentliche Politiken, die nicht auf gesetzlichen Grundlagen beruhen. Auch in methodologischer Hinsicht bestehen kaum Unterschiede.

sche Ermittlung und Beurteilung der Wirkungen der Gesetzgebung. Methodischen Charakter haben Wirkungsermittlungen und -beurteilungen dann, wenn sie:

- sich auf nachvollziehbare Überlegungen und Aussagen stützen, deren Grundlagen bekannt, zugänglich und überprüfbar sind;
- auf einem systematischen, den Einbezug möglichst aller relevanten Wirkungen anstrebenden Vorgehen beruhen;
- möglichst objektive Wirkungsermittlungen und -beurteilungen anstreben, d.h. nicht einseitig durch spezifische, partikuläre Interessen geprägt sind (Eidgenössisches Justiz- und Polizeidepartement 1991: 13).

Solcherart definierte Evaluationen haben einerseits nicht nur impressionistischen, intuitiven Charakter, wie dies bei rein politischen Wirkungsbeurteilungen oft der Fall ist; und andererseits handelt es sich nicht zwingend um Analysen, die in allen Teilen strengen wissenschaftlichen Anforderungen genügen. Sie bewegen sich in der Praxis häufig zwischen diesen beiden Polen und versuchen, sowohl Alltagswissen und Verwaltungserfahrungen zu erfassen und systematisch auszuwerten als auch wissenschaftliche Erkenntnisse zu nutzen oder wissenschaftliche Vorgehensweisen zur Beschaffung von Informationen zu verwenden. Gesetzesevaluationen, die diesen methodischen Charakter aufweisen, finden in der Regel ihren Niederschlag in der Form von verwaltungsinternen oder veröffentlichten Berichten.²

Je nachdem in welcher Phase des Prozesses der Gesetzesvorbereitung und -anwendung sowie zu welchen Zwecken Evaluationen durchgeführt werden, unterscheidet man zwischen prospektiven, begleitenden und retrospektiven Gesetzesevaluationen:

- prospektive Evaluationen (Wirkungsprognosen) werden im Vorfeld gesetzgeberischer Entscheide durchgeführt; sie sind eine wesentliche Entscheidungshilfe für den Gesetzgeber;
- begleitende Evaluationen (Vollzugskontrollen; Monitoring) werden in der Umsetzungs- oder Vollzugsphase durchgeführt und dienen vor allem der Steuerung des Vollzugs;
- retrospektive Evaluationen (Wirkungsüberprüfungen; Wirkungskontrollen) werden einige Zeit nach Erlass der Gesetzgebung durchgeführt; sie dienen vor allem der Kontrolle und Aufsicht und sind eine Grundlage für die Beurteilung eines allfälligen Korrekturbedarfs der Gesetzgebung.

Diese unterschiedlichen Evaluationstypen haben komplementären Charakter. Sie müssen zwar nicht unbedingt kombiniert werden, aber es ist offensichtlich, dass eine vertiefte und differenzierte explizite Wirkungsprognose im Rahmen der Gesetzesvorbereitung ein späteres Monitoring oder eine spätere Wirkungskontrolle erheblich erleichtert, ja zum Teil vielleicht sogar überhaupt erst ermöglicht.

² Nach Artikel 8 Absatz 5 des Bundesgesetzes vom 17. Dezember 2004 über das Öffentlichkeitsprinzip der Verwaltung (Öffentlichkeitsgesetz; SR 152.3) besteht ein Recht auf Zugang zu Evaluationsberichten.

Gesetzesevaluationen interessieren sich für die Gesamtheit aller relevanten Wirkungsbezüge und Wirkungen. Im Vordergrund stehen dabei vor allem die folgenden drei Wirkungskriterien: Effektivität, Wirksamkeit und Effizienz.

- Das Kriterium der Effektivität betrifft die Frage, ob gesetzliche Bestimmungen überhaupt befolgt und angewendet werden. Dies erfordert einen Vergleich zwischen dem normativen Modell einerseits und dem Verhalten der Normadressaten bzw. der Vollzugsinstanzen andererseits, wobei nur dann von Wirkungen gesprochen werden kann, wenn zwischen den beiden Elementen (normatives Modell und Verhalten) eine kausale Beziehung besteht oder, anders gesagt, wenn das Verhalten auf das normative Modell zurückzuführen ist. Je nach Normtyp oder -inhalt geht es darum, den Befolgungsgrad einer Vorschrift (etwa eines Verbots oder eines Gebots), den Grad der Inanspruchnahme einer Bestimmung (etwa bei Finanzhilfebestimmungen oder bei dispositivem Recht) oder den Vollzug bzw. die Durchsetzung (etwa bei Sanktionsbestimmungen oder generell bei Normen, die der Umsetzung und Anwendung durch eine Behörde bedürfen) zu eruieren.
- Mit dem Kriterium der Wirksamkeit wird die Frage des Zielerreichungsgrades der Gesetzgebung angesprochen. Die Effektivität einer Norm ist nicht unbedingt eine Garantie für deren Wirksamkeit mit Blick auf die Realisierung der vom Gesetzgeber angestrebten Ziele; sie ist es nur dann, wenn die Wirkungshypothesen, die dem gesetzgeberischen Entscheid und dem normativen Modell zu Grunde liegen, zutreffend sind. Auch das Kriterium der Wirksamkeit setzt also das Bestehen eines Wirkungszusammenhangs zwischen dem Zielerreichungsgrad und den gesetzgeberischen Massnahmen voraus. Werden die Ziele erreicht, ohne dass dies auf die Gesetzgebung zurückzuführen ist, kann nicht von Wirksamkeit gesprochen werden. Es ist nahe liegend, dass dieses Kriterium aus politischer Sicht im Vordergrund steht. Deshalb wird es in Artikel 170 BV auch besonders genannt.
- Das Kriterium der Effizienz schliesslich beinhaltet einen Kosten-Nutzen-Vergleich; bei den Kosten werden dabei einerseits der verwaltungsinterne Aufwand sowie die verwaltungsexternen Belastungen und andererseits die negativen Nebenwirkungen, das heisst die nicht-zielkonformen Wirkungen ermittelt. Beim Nutzen geht es um die Ermittlung einerseits der zielkonformen Wirkungen (Zielerreichungsgrad) und andererseits der positiven Nebenwirkungen. Erfasst werden dabei nicht nur materielle, sondern auch immaterielle Kosten und Nutzen; soweit möglich, sollten Kosten und Nutzen quantifiziert, unter Umständen sogar monetarisiert werden.

3 Zum Stand der Institutionalisierung der Gesetzesevaluation auf Bundesebene

3.1 Auf Verfassungsstufe

Wie eingangs erwähnt, ist die Pflicht, für die Evaluation der Wirksamkeit der staatlichen Massnahmen zu sorgen, in Artikel 170 BV verankert. Anders als es der Wortlaut der Bundesverfassung vermuten lassen könnte, muss dabei das Parlament die Evaluationen nicht unbedingt selbst veranlassen oder durchführen. Es muss aber zumindest sicherstellen, dass Evaluationen durchgeführt werden (allenfalls auch durch die Regierung oder die Verwaltung bzw. durch externe Stellen im Auftrag des Parlaments, der Regierung oder der Verwaltung). Gegenstand der Evaluation sind generell alle Massnahmen des Bundes. Darin eingeschlossen ist jegliches Handeln, jegliche Tätigkeit von Bundesorganen oder von Organen oder Personen, welche Massnahmen des Bundes umsetzen oder vollziehen, d.h. auch von Privaten, die mit der Erfüllung von Bundesaufgaben betraut sind, und von kantonalen Behörden, denen der Vollzug der Bundesgesetzgebung obliegt. Die verfassungsrechtliche Verpflichtung bezieht sich dabei auch nicht bloss auf das Kriterium der Wirksamkeit; gemeint ist eine umfassende Evaluation der Wirkungen. Und schliesslich bezieht sich Artikel 170 nicht nur auf die Wirkungskontrolle, sondern auch auf die Wirkungsprognose (Mader 2005; Morand 2001; Mastronardi 2002). Die Aufnahme dieser Bestimmung in die neue Bundesverfassung von 1999 war einerseits Kulminationspunkt einer längeren Entwicklung in der Praxis und in der Gesetzgebung, andererseits aber auch zusätzlicher Impuls für eine weitergehende Stärkung und Intensivierung der Evaluationstätigkeit auf der Ebene des Bundes.

3.2 Auf Gesetzesstufe

Die Pflicht, staatliche Massnahmen zu evaluieren, ist in der Schweiz keineswegs bloss in der Bundesverfassung festgeschrieben. Sie wird in vielfältiger Weise auf Gesetzes- und Verordnungsebene sowie in verwaltungsinternen Richtlinien umgesetzt und konkretisiert. Im Folgenden werde ich – ohne Anspruch auf Vollständigkeit – die in diesem Zusammenhang wichtigen Bestimmungen und Vorschriften namentlich auf Gesetzesstufe nennen.

Die wichtigsten Bestimmungen finden sich im Bundesgesetz über die Bundesversammlung (Parlamentsgesetz, SR 171.10). Nach Artikel 26 Absatz 3 dieses Gesetzes gehört die Wirksamkeit zu den zentralen Kriterien für die Ausübung der Oberaufsicht durch das Parlament. Nach Artikel 27 können die parlamentarischen Kommissionen verlangen, dass die Regierung Wirksamkeitsüberprüfungen durchführen lässt, sie können die im Auftrag der Regierung durchgeführten Wirksamkeitsüberprüfungen prüfen, und sie können selbst solche Überprüfungen in Auftrag geben. Die Legislativkommissionen müssen zudem nach Artikel 44 Absatz 1 die Resultate von Wirksamkeitsüberprüfungen berücksichtigen.

Nach Artikel 54 Absätze 3 und 4 des Parlamentsgesetzes muss die Konferenz der Präsidien der Aufsichtskommissionen und -delegationen für die Koordination und Kohärenz der Tätigkeiten im Bereich der Evaluation staatlicher Massnahmen besorgt sein. Sie hat

dabei mit dem Bundesrat zusammenzuarbeiten. Gemäss Artikel 54 Absatz 4 entscheidet die Konferenz über die Anträge der Kommissionen, die Wirksamkeit von Erlassen durch die mit dieser Aufgabe betraute Dienststelle der Parlamentsdienste, die Eidgenössische Finanzkontrolle oder Dritte überprüfen zu lassen.

Das Parlamentsgesetz erwähnt die Parlamentarische Verwaltungskontrolle (PVK). Nach Artikel 10 der Parlamentsverwaltungsverordnung (SR 171.115) führt die PVK Evaluationen im Rahmen der parlamentarischen Oberaufsicht durch und überprüft die von der Bundesverwaltung durchgeführten Evaluationen sowie deren Verwendung in Entscheidungsprozessen (Janett 2003; 2004). Sie überprüft auf Antrag von parlamentarischen Kommissionen die Wirksamkeit von Massnahmen des Bundes. Die personellen Ressourcen der PVK sind ziemlich beschränkt (etwa fünf Personen). Daneben verfügt sie aber selbständig über einen Expertencredit, der ihr den Beizug externer Experten ermöglicht, und sie ist bei der Bearbeitung ihrer Aufträge fachlich unabhängig.

Wichtiger noch ist die Eidgenössische Finanzkontrolle, welche die Bundesversammlung und die Regierung bei der Ausübung ihrer Finanz- und Aufsichtskompetenzen unterstützt (Grüter 1999; Hangartner 2006)³. Ihre zentrale Aufgabe ist es, staatliche Massnahmen nicht nur auf ihre Ordnungsmässigkeit und Rechtmässigkeit, sondern auch auf ihre Wirtschaftlichkeit zu überprüfen. Im Rahmen der Wirtschaftlichkeitsprüfungen klärt sie nach Artikel 5 Absatz 2 des Finanzkontrollgesetzes (SR 614.0), ob die finanziellen Aufwendungen, die mit staatlichen Massnahmen verbunden sind, die erwartete Wirkung haben. Ähnlich wie die Rechnungshöfe in andern Ländern hat die Eidgenössische Finanzkontrolle in den letzten zwei Jahrzehnten ihre Kontrolltätigkeit ausgeweitet und schenkt Fragen der Wirksamkeit und der Wirtschaftlichkeit mehr Beachtung.

Für die nachträglichen Wirksamkeitsüberprüfungen sind im Weiteren auch das Subventionsgesetz und die Evaluationsklauseln in zahlreichen neueren Gesetzen zu erwähnen. Nach Artikel 5 des Subventionsgesetzes (SR 616.1) muss die Regierung periodisch prüfen, ob die Bestimmungen über Finanzhilfen und Abgeltungen den Grundsätzen des Gesetzes entsprechen. Dazu gehört nach Artikel 6 dieses Gesetzes auch die Prüfung, ob die mit Finanzhilfen des Bundes unterstützten Aufgaben nicht einfacher, wirksamer und rationeller erfüllt werden könnten. In Ausführung dieses Gesetzes, das somit eine instrumentenspezifische Evaluationspflicht statuiert, sind bereits mehrere Subventionsberichte erstellt worden.⁴

Relativ zahlreich sind die bereichs- oder erlassspezifischen Evaluationsklauseln, welche die Regierung oder bestimmte Verwaltungsstellen verpflichten, den Vollzug und die Wirksamkeit der Massnahmen sowie deren Kosten zu überprüfen.⁵ Ein Beispiel für eine solche Klausel findet sich etwa in Artikel 19 des Öffentlichkeitsgesetzes (SR 152.3). Ein wei-

³ Siehe den Überblick über die Stellung und Tätigkeit der Eidgenössischen Finanzkontrolle (EFK), unter www.efk.admin.ch/pdf/efk_stellung_d.pdf [27.3.2008].

⁴ Siehe dazu die im Bundesblatt publizierten Berichte des Bundesrates: BBl 1998, S. 2051-2334, und BBl 1999, S. 8013-8369. Inzwischen gibt es fünf Controlling-Berichte des Eidgenössischen Finanzdepartements; vgl. Webseite der Eidgenössischen Finanzverwaltung unter <http://www.efv.admin.ch/d/themen/bundesfinanzen/subventionen/index.php>. [Stand 17.3.2008].

⁵ Siehe die Liste auf der Webseite des Bundesamtes für Justiz unter http://www.bj.admin.ch/bj/de/home/themen/staat_und_buerger/evaluation/materialien/_typen_und_terminologie.html. [Stand 9.4.2008].

teres Beispiel aus jüngster Zeit ist Artikel 20 des Bundesgesetzes gegen die Schwarzarbeit (AS 2007 359; SR 822.41), das am 1. Januar 2008 in Kraft getreten ist.

Schliesslich ist hervorzuheben, dass sich die Institutionalisierung der Evaluation nicht auf die Wirkungskontrolle beschränkt. Sie erfasst auch prospektive Evaluationen im Vorfeld gesetzgeberischer Entscheide. Artikel 141 Absatz 2 des Parlamentsgesetzes sieht vor, dass die Regierung sich in ihren Botschaften zu Erlassentwürfen auch zur Vollzugstauglichkeit und zu den voraussichtlichen Auswirkungen äussern muss. Damit sind keineswegs nur die personellen und finanziellen Auswirkungen sowie die Auswirkungen auf die Kantone und Gemeinden gemeint, sondern auch die Auswirkungen auf die Wirtschaft, die Gesellschaft, die Umwelt et cetera.

3.3 Verwaltungsinterne Vorschriften und Vorkehrungen

Zahlreiche Evaluationspflichten werden auf Verordnungsebene noch konkretisiert. Als Beispiel sei hier etwa Artikel 21 der Öffentlichkeitsverordnung (SR 152.31) erwähnt, der namentlich präzisiert, welche Informationen die zuständigen Behörden als Grundlagen für die Evaluation des Gesetzes jährlich erheben und der für die Evaluation verantwortlichen Amtsstelle, das heisst dem Eidgenössischen Datenschutz- und Öffentlichkeitsbeauftragten mitteilen müssen. Zum Teil erfolgt diese Konkretisierung auch in der Form verwaltungsinterner Vorschriften. Dies gilt insbesondere für die Pflicht zur prospektiven Wirkungsbeurteilung (Artikel 141 Absatz 2 des Parlamentsgesetzes): Der so genannte Botschaftsleitfaden der Bundeskanzlei verdeutlicht, welche Auswirkungen darzustellen sind und präzisiert, dass die Grundlagen angegeben werden müssen, die für die prospektive Beurteilung der Auswirkungen herangezogen werden. Damit sollen diese Beurteilungen überprüfbar und nachvollziehbar werden.

Erwähnt seien in diesem Zusammenhang auch die Richtlinien der Regierung und das Handbuch des Eidgenössischen Volkswirtschaftsdepartements zur so genannten Regulierungsfolgenabschätzung, die sich vor allem mit der prospektiven Beurteilung der Auswirkungen auf die Wirtschaft, insbesondere auch auf die kleineren und mittleren Unternehmen (KMU), befassen.

Um die Umsetzung der in Artikel 170 der Bundesverfassung enthaltenen Pflicht sicherzustellen, hat der Bundesrat in einem Beschluss vom 3. November 2004 verschiedene durch die Departemente und Ämter zu ergreifende Massnahmen beschlossen. Diese Massnahmen beziehen sich in erster Linie auf die retrospektive Evaluation. Der Bericht macht insbesondere die verwaltungsinternen Verantwortlichkeiten im Bereich der Gesetzesevaluation deutlich:

- hauptverantwortlich sind die Bundesämter, die aufgefordert werden, eine Amtsstrategie für die Durchführung von Wirksamkeitsüberprüfungen in ihrem Zuständigkeitsbereich zu entwickeln und für die notwendigen Ressourcen besorgt zu sein;
- die Departemente sollen sicherstellen, dass die Ämter ihren Verpflichtungen nachkommen; zudem sollen sie dem Bundesrat die für seine Entscheide relevanten Evaluationsergebnisse unterbreiten;

- der Bundesrat soll im Rahmen der Jahres- und der Legislaturplanung vermehrt Evaluationsschwerpunkte setzen und über Evaluationen mit interdepartementaler Tragweite beschliessen;
- verschiedene Ämter mit Querschnittsaufgaben sollen für die andern Ämter im Hinblick auf die Verstärkung der Evaluationsaufgabe Dienstleistungen erbringen.

Eine dieser Querschnittsaufgaben wird vom Bundesamt für Justiz erfüllt, welches ein verwaltungsinternes Netzwerk betreibt, das einen regelmässigen Erfahrungsaustausch zwischen den verschiedenen an Evaluationen interessierten Amtsstellen des Bundes sicherstellt und gewisse Fortbildungsaktivitäten organisiert. Zudem berät das Bundesamt für Justiz andere Ämter bei der Vorbereitung von Evaluationsvorhaben.

Erwähnt sei schliesslich, dass die Bundeskanzlei seit 2004 jährlich eine Erhebung zu den politisch wichtigen Wirksamkeitsüberprüfungen durchführt und dass diese durch die projektverantwortlichen Stellen in die Datenbank ARAMIS aufgenommen werden. Gestützt auf diese Erhebung informiert der Bundesrat das Parlament seit 2006 in einem Anhang zum jährlichen Geschäftsbericht über den Stand der Umsetzung des Beschlusses vom 3. November 2004 und über die durchgeführten Evaluationen. Bislang gibt es hierzu keine systematische Übersicht (Widmer 2007: 82–83; Bussmann 2007: 15–16). Wie im Geschäftsbericht 2007 festgehalten, führten von 31 aktiven Bundesämtern rund zwanzig Wirksamkeitsüberprüfungen durch. Bei 17 Bundesämtern bestanden Amtsstrategien oder zumindest Teilstrategien für Wirksamkeitsüberprüfungen (Geschäftsbericht des Bundesrates 2008: Anhang 4). Für die politisch weniger bedeutsamen Projektevaluationen (z. B. in der Entwicklungshilfe, in der Energieforschung und im Gesundheitsbereich) und Controllingaufgaben ist der Umfang nur sehr schwierig zu beziffern; Werner Bussmann schätzt, dass die Zahl der Projekte bei über 100 liegen könnte (Bussmann 2007: 15). Die Departemente sorgen mehrheitlich im Rahmen bestehender Instrumente (Planung, Controlling, Reporting) für die Umsetzung von Wirksamkeitsüberprüfungen durch ihre Ämter (Geschäftsbericht des Bundesrates 2008: Anhang 4). Noch liegt uns aber keine publizierte Departementsstrategie zur Wirksamkeitsüberprüfung vor.

3.4 Würdigung der verschiedenen Institutionalisierungsmassnahmen

Es zeigt sich, dass auf Bundesebene vor allem in den letzten zwei Jahrzehnten viele unterschiedliche Institutionalisierungsmassnahmen getroffen worden sind: prozedurale Sicherungen (besonders Evaluationspflichten im Vorfeld gesetzgeberischer Entscheide und im Nachgang dazu sowie Überprüfung der Beachtung dieser Pflichten im verwaltungsinternen Verfahren der Entscheidvorbereitung), organisatorische Massnahmen (Schaffung neuer Organe wie der PVK; Erweiterung der Aufgaben bestehender Organe wie der Eidgenössischen Finanzkontrolle) sowie finanzielle Vorkehrungen. Insgesamt ist damit in der Schweiz auf Bundesebene ein vergleichsweise sehr hoher Stand der Institutionalisierung erreicht worden.⁶

⁶ Gemäss einer vergleichenden Studie von Steve Jacob und Frédéric Varone aus dem Jahre 2003 belegt die Schweiz diesbezüglich zusammen mit Australien sogar den Spitzenplatz (Jacob/Varone 2003).

Dieser Umstand scheint mir in verschiedener Hinsicht vorteilhaft: Die Institutionalisierung trägt dazu bei, dass dem Anliegen der Evaluation in den staatlichen Entscheidungsverfahren mehr Beachtung geschenkt wird; rein politische Opportunitätsüberlegungen werden beim Entscheid darüber, ob eine Evaluation durchgeführt werden soll, in den Hintergrund gedrängt; die Abstimmung der zeitlichen Bedürfnisse für die Durchführung von Evaluationen auf die zeitlichen Rahmenbedingungen und Zwänge politischer Entscheidungsprozesse wird erleichtert; die für die Durchführung von Evaluationen unabdingbare Kooperation der betroffenen Verwaltungsstellen wird gefördert; Gleiches gilt für die Valorisierung der Evaluationsergebnisse im Gesetzgebungsprozess. Der Aufbau und die Aufrechterhaltung der notwendigen Evaluationskapazität und -kompetenz innerhalb und ausserhalb der Verwaltung werden unterstützt. Andererseits könnte ein hoher Institutionalisierungsgrad mit der Gefahr verbunden sein, dass die Durchführung von Evaluationen zur blossen Verwaltungsroutine, ja zu Verwaltungsleerlauf verkommt. Diese Gefahr sollte nicht überschätzt werden. Nicht zuletzt dürften finanzielle Restriktionen auch bei einem hohen Institutionalisierungsgrad gewährleisten, dass Evaluationen mit der gebotenen Vernunft und selektiv durchgeführt werden.

4 Zum Stand der Institutionalisierung der Gesetzesevaluation auf Kantonsebene

Ansätze für die Institutionalisierung der Evaluation staatlicher Massnahmen bestehen nicht nur auf Bundesebene, sondern auch in zahlreichen Kantonen und in grösseren Gemeinden oder Städten. Diese Ansätze sind vor allem in Zusammenhang mit Bestrebungen für die Einführung der wirkungsorientierten Verwaltung beziehungsweise des so genannten New Public Management entwickelt worden. Als Beispiele seien hier etwa die Kantone Bern, Luzern, Solothurn und Aargau genannt. Die Institutionalisierungsansätze weisen erhebliche Unterschiede auf, und auch der Entwicklungsstand auf kantonaler Ebene ist unterschiedlich.

Der Kanton Bern etwa verfügt seit 1992 über ausgedehnte Erfahrungen mit Evaluationen, wobei diese im Wesentlichen auf die Erfolgskontrolle von finanziellen Leistungen des Kantons (Subventionen) ausgerichtet sind. Sie dienen vor allem dazu abzuklären, ob und wieweit die kantonalen Subventionen wirksam sind und wirtschaftlich eingesetzt werden (Nuspliger 2005; Kettiger 2005). Diese Ausrichtung hat wohl auch damit zu tun, dass nach Artikel 101 Absatz 4 der Verfassung des Kantons Bern vom 6. Juni 1993 alle Aufgaben des Kantons „periodisch auf ihre Notwendigkeit und Zweckmässigkeit sowie auf ihre finanziellen Auswirkungen und deren Tragbarkeit zu überprüfen“ sind.

Ein Sonderfall, der hier besondere Erwähnung verdient, ist der Kanton Genf (Delley 2001; 2007: 113). Dort besteht seit 1995 eine durch ein kantonales Gesetz geschaffene, aus 16 verwaltungsexternen Fachleuten zusammengesetzte unabhängige Evaluationskommission, die im Auftrag der Regierung oder der Finanzkommission und der Geschäftsprüfungskommission des Parlaments sowie aus eigener Initiative Gesetzesevaluationen durchführt. Die Mitglieder der Kommission werden jeweils für acht Jahre gewählt und dürfen nicht der

kantonalen Verwaltung oder den politischen Behörden des Kantons angehören. Sie werden durch ein permanentes Sekretariat unterstützt, das über eigene Evaluationskapazitäten verfügt und zusätzlich ad hoc Evaluationsaufträge an Experten erteilen kann. Seit ihrer Schaffung hat die Evaluationskommission rund zwanzig Evaluationen zu verschiedenen kantonalen Gesetzgebungen durchgeführt (z. B. Wohnbauförderung, Berufsbildung, Steuern, Arbeitslosigkeit, Förderung der Musikerziehung, Stipendien an Studierende, öffentlicher Verkehr, Jugendschutz, Weiterbildung, internationale Adoptionen etc.).

5 Wie sieht die Praxis aus?

Die Institutionalisierung der Evaluation hat zumindest auf Bundesebene im Verlauf der letzten zwei Jahrzehnte beachtliche Fortschritte gemacht. Sie hat heute einen insgesamt sehr hohen Grad erreicht. Im internationalen Vergleich nimmt die Schweiz einen Spitzenplatz ein. Es wäre jedoch verfehlt, daraus abzuleiten, dass in der Schweiz praktisch alle staatlichen Massnahmen flächendeckend und regelmässig oder gar dauerhaft evaluiert würden. Die Praxis ist unbestreitbar deutlich bescheidener.

Zum praktischen Entwicklungsstand liefern insbesondere die Untersuchungen von Thomas Widmer, Erwin Rüegg und Peter Neuenschwander (2001) sowie von Andreas Balthasar (2007) wertvolle Angaben. Andreas Balthasar analysiert gestützt auf eine Schätzung der Ausgaben der verschiedenen Departemente und Ämter der Bundesverwaltung die Bedeutung der Gesetzes- oder der Politikevaluation in den verschiedenen Aufgabenbereichen des Bundes. Nach ihm wurden in den Jahren 1999 bis 2002 in der schweizerischen Bundesverwaltung etwas mehr als 300 Evaluationen durchgeführt.

Der Befund von Andreas Balthasar, dass (noch) nicht von einer flächendeckenden gefestigten Evaluationskultur in der Bundesverwaltung gesprochen werden kann, ist nicht zu bestreiten (vgl. auch den Beitrag von Balthasar in diesem Band). Die Evaluationspraxis ist aber meines Erachtens zu recht selektiv und konzentriert sich auf jene Fälle, wo Evaluationen mit vertretbarem, vernünftigem Aufwand nützliche, ja zum Teil unabdingbare Zusatzgrundlagen für gesetzgeberische Entscheide oder für die Ausübung von Aufsichtsfunktionen liefern können.

Drei aktuelle Beispiele aus dem Jahr 2007 mögen dies illustrieren:

- Am 8. März hat der Nationalrat über die Gleichstellung von Frau und Mann debattiert. Grundlage für die Debatte war – neben einer Vielzahl parlamentarischer Vorstösse – auch ein Bericht der Regierung über die Evaluation der Wirksamkeit des Gleichstellungsgesetzes. Der Bericht beruht auf einer umfassenden Evaluation des Gesetzes, die in Zusammenarbeit mit verwaltungsexternen Expertinnen und Experten durchgeführt worden ist (Bericht des Bundesrates 2006).
- Am 14. März hat der Ständerat sich mit einer Totalrevision des Opferhilfegesetzes befasst, die nicht zuletzt durch eine Evaluation des Vollzugs und der Wirksamkeit dieses Gesetzes ausgelöst worden ist und die den Ergebnissen dieser Evaluation weitgehend Rechnung trägt (Hilfe an Opfer von Straftaten 2000).

- Und schliesslich hat sich das Parlament im Verlauf der Frühlingssession mit einer Vorlage zur Unternehmenssteuerreform befasst, für die umfangreiche prospektive Studien nicht nur zu den finanziellen Konsequenzen, sondern vor allem auch zu den volkswirtschaftlichen Auswirkungen erstellt worden waren.

Die drei – abgesehen vom zeitlichen Kriterium (Frühjahrssession 2007 des Parlaments) – ziemlich zufällig ausgewählte Beispiele zeigen stellvertretend, dass (prospektiven und retrospektiven) Evaluationen auf Bundesebene in der Praxis durchaus eine gewisse Bedeutung zukommt.

6 Schweizerische Besonderheiten

Wenn wir die Institutionalisierungsansätze und die Praxis im Bereich der Evaluation auf Bundesebene betrachten, fallen ein paar Besonderheiten auf:

Ein erstes Merkmal ist der Pluralismus in organisatorischer Hinsicht. Es gibt kein zentrales Evaluationsorgan, sondern eine Vielzahl von Institutionen, die sich aus unterschiedlicher Perspektive mit der Problematik der Wirkungen staatlichen Handelns auseinandersetzen, Evaluationen durchführen oder durchführen lassen oder zumindest deren Ergebnisse nutzen. Dieser dezentrale Ansatz fördert insbesondere eine sehr breite Sensibilisierung für Evaluationsanliegen und sichert eine bessere Verankerung: Erfolg und Misserfolg bzw. die Akzeptanz der Evaluation hängt nicht von der Tätigkeit einer einzigen Fachstelle oder eines einzigen Organs ab. Andererseits erfordert dies einen erhöhten Koordinationsaufwand, dem bislang kaum entsprochen wird.

Zweitens fällt auf, dass sehr wesentliche Impulse vom Parlament kommen. Das Parlament ist häufig der Auftraggeber, sei es bei der Durchführung von Evaluationen in Zusammenhang mit der Ausübung von Kontrollfunktionen, sei es durch die Einfügung von Evaluationsklauseln in Gesetze. Die Durchführung von Evaluationen obliegt dann aber in vielen Fällen der Regierung und der Verwaltung, meist unter Beizug externen Sachverständs. Regierung und Verwaltung sind somit keineswegs passiv in diesem Bereich.

Der Umstand, dass die Verantwortung für die praktische Durchführung von Evaluationen in den meisten Fällen den zuständigen Departementen und Ämtern der Verwaltung obliegt, ist – drittens – ein Element, das den Einbezug der Evaluationsergebnisse in die Vorbereitung politischer Entscheide und bei der Ausübung von Aufsichtsfunktionen begünstigt.

Viertens scheint mir ganz entscheidend, dass die Nachteile, die mit einem dezentralen Institutionalisierungsansatz verbunden sein könnten, etwa die Verzettlung der Evaluationsaktivitäten und das Fehlen einer kritischen Masse, durch die gute Vernetzung innerhalb der Bundesverwaltung (unter Einbezug der Parlamentsdienste und der Finanzkontrolle) kompensiert werden. Das vom Bundesamt für Justiz organisierte „Netzwerk Evaluation“ der Bundesverwaltung hat diesbezüglich eine wichtige Aufgabe. Es hat zwar keine eigentliche Koordinationsfunktion, fördert aber die gegenseitige Information aller interessierten Amtstellen des Bundes.

Und schliesslich scheint mir auch die Tradition der so genannten „Milizverwaltung“ (d. h. eine Verwaltung, welche sich stark auf externen Sachverstand abstützt) eine gewisse Rolle zu spielen. Sie hat wohl dazu beigetragen, dass auf den Aufbau besonderer verwaltungsinterner Kapazitäten weitgehend verzichtet und dem Beizug externer Evaluatorinnen und Evaluatoren der Vorzug gegeben worden ist. Die verwaltungsintern vorhandenen Evaluationskapazitäten werden vor allem für die Vergabe von Evaluationsaufträgen an externe Experten sowie für die Begleitung der Evaluationsvorhaben und für die Valorisierung ihrer Ergebnisse eingesetzt.

7 Begünstigende Entwicklungsfaktoren

Worauf ist zurückzuführen, dass die Gesetzesevaluation sich in den letzten zwanzig Jahren in der Schweiz vergleichsweise gut entwickelt hat? Treibende Kräfte in dieser Entwicklung waren nach meiner Einschätzung nicht zuletzt die Bemühungen um eine qualitativ gute Gesetzgebung (Stichwort Gesetzgebungsmethodik) sowie um eine wirkungsorientierte Verwaltungsführung und -kontrolle (Stichwort New Public Management). Qualitativ gute Gesetzgebung ist notwendigerweise evidenzgestützt und kommt nicht aus ohne Gesetzesevaluation als ein wichtiger Teilaspekt methodischer Ansätze für die Gesetzesvorbereitung und -anwendung. Und der Übergang von der input- zur outputorientierten Verwaltung ruft zwingend nach einer zuverlässigeren Ermittlung und Beurteilung der Wirkungen.

Dazu kommt, dass das letzte Jahrzehnt vielleicht ein relativ günstiger Zeitpunkt, ein „window of opportunity“, für – wenn auch noch so bescheidene – institutionelle Reformen war. Die Verfassungsreform sowie die Reform des Finanzausgleichs und der Aufgabenteilung zwischen Bund und Kantonen (Projekt NFA) machen dies deutlich. Nicht zuletzt sei hier daran erinnert, dass die Aufnahme des neuen Artikels 170 BV auf Vorschläge zurückzuführen ist, die ihm Rahmen von Vorarbeiten für eine Parlamentsreform geleistet worden sind (Bericht von den Staatspolitischen Kommissionen 1996).

Sodann denke ich, dass das schweizerische Konkordanzsystem insofern ein begünstigender Faktor ist, als es einem gewissen Mass an Sachlichkeit in der politischen Auseinandersetzung und Kontrolle förderlich sein dürfte.

Auch die Langsamkeit des politischen Prozesses in der Schweiz scheint mir ein wichtiges Element zu sein. Sie schafft Möglichkeiten für die zeitliche Abstimmung der Entscheidungsvorbereitung mit Evaluationstätigkeiten, die bei rascher verlaufenden politischen Prozessen nicht im gleichen Ausmass bestehen.

Wahrscheinlich spielt auch eine Rolle, dass wir uns im politischen System der Schweiz sehr stark entwickelte Mitwirkungs- und Rückkoppelungsmechanismen innerhalb und ausserhalb der Verwaltung gewohnt sind. Evaluationen können diese Mechanismen zusätzlich verstärken, sie sind aber diesbezüglich nicht ein grundlegend neues oder anderes Element, was wohl die Offenheit für Evaluationen fördert.

Und schliesslich sei auch der sich aus der Milizverwaltung ergebende rege Austausch und Dialog zwischen Praktikern und Wissenschaftern erwähnt, der ein gutes Umfeld für Evaluationen und für eine über die Verwaltung hinausreichende Vernetzung schafft. Diese

findet positiven Niederschlag in breit gefächerten Ausbildungsaktivitäten, in der Entwicklung von „good practices“ oder Standards und in Fachpublikationen.

8 Schlussbemerkungen

Evaluationen sind mit Aufwand verbunden. Sie kosten Geld, unabhängig davon, ob sie verwaltungsintern oder -extern durchgeführt werden. Und sie beanspruchen manchmal Zeit, die im politischen Entscheidungsprozess oftmals fehlt.

Fehlentscheide des Gesetzgebers können jedoch auch Geld kosten; dies vor allem, wenn sie nicht als solche erkannt werden – und wenn sie erkannt werden, erfordern sie oftmals langwierige Korrekturen.

Staatliche Behörden, die ihre Verantwortung für die Folgen ihres Entscheidens und Handelns wahrnehmen wollen, sind in vielen Fällen auf Evaluationen als Grundlage für die Entscheidvorbereitung, für die Steuerung des Vollzugs sowie für die Ausübung von Aufsichts- und Kontrollfunktionen angewiesen. Sie dürfen dabei aber drei Dinge nicht aus den Augen verlieren:

- Evaluationen sind keine Allheilmittel; sie sind gezielt nur dann einzusetzen, wenn die realistische Aussicht besteht, dass sie einen zusätzlichen und erheblichen Erkenntnisgewinn verschaffen.
- Übertriebene Erwartungen sind verfehlt; Evaluationen liefern kaum eigentliche Beweise für Wirkungszusammenhänge und Wirkungen; die Wirkungszusammenhänge zwischen gesetzgeberischen Massnahmen und gesellschaftlichen Realitäten sind ausserordentlich komplex; es geht deshalb mehr um einigermassen gesicherte Plausibilitäten als um Gewissheiten.
- Evaluationen sind kein Ersatz für politische Entscheide; sie können aber dazu beitragen, dass gesetzgeberische Entscheide besser begründet und zielsicherer getroffen werden, und leisten damit einen wesentlichen Beitrag zur Qualität der Gesetzgebung.

Damit die Gesetzesevaluation die in sie gesetzten Erwartungen erfüllen kann, müssen die Evaluatorinnen und Evaluatoren praktische Relevanz anstreben und dürfen sich nicht damit begnügen, für die Schublade zu produzieren. Wenn Gesetzesevaluationen durch ihre fachliche Qualität zu überzeugen vermögen, können die in der Schweiz bestehenden Institutionalisierungsansätze sich als tragfähig für eine sinnvolle Praxis und für eine evidenzgestützte Gesetzgebung erweisen. Dies gilt, sofern es der Verwaltung und den politischen Behörden gelingt, das Minimum an Kapazitäten und Fachkompetenz zu bewahren, das notwendig ist, um sinnvolle Evaluationen durchzuführen oder zu veranlassen und um deren Ergebnisse in den legislatorischen Entscheidungsprozess einzubringen. Nicht zuletzt setzt dies schliesslich auch voraus, dass die Vergabe von Aufträgen an verwaltungsexterne Evaluatorinnen und Evaluatoren nach sachlichen Kriterien und transparent erfolgt. Andernfalls ist die Gefahr gross, dass den Evaluationsergebnissen die Glaubwürdigkeit fehlt, ohne die sie im politischen Entscheidungsprozess kaum berücksichtigt werden.

Die institutionelle Einbettung der Evaluationsfunktion: Deutschland, Österreich und die Schweiz im Vergleich

1 Einleitung

Der vorliegende Beitrag vergleicht die institutionelle Verankerung der Evaluationsfunktion in den drei Ländern Deutschland, Österreich und der Schweiz. Dabei sind die in jedem politischen System individuell ausgestalteten Rahmenbedingungen mit zu berücksichtigen. Die Gegenüberstellung profitiert aber vom Umstand, dass die drei Systeme einige gemeinsame Merkmale aufweisen, die den komparativen Zugang erleichtern. In erster Linie ist dabei auf den föderalistischen Staatsaufbau zu verweisen, der zwar je unterschiedlich ausgestaltet ist, sich aber doch grundsätzlich von jenem eines unitaristischen Systems unterscheidet.

Die drei politischen Systeme weisen auch grundlegende Unterschiede auf, die im vorliegenden Zusammenhang von zentraler Bedeutung sind. An erster Stelle ist dabei auf die Europäische Union zu verweisen, deren Einfluss bei den Mitgliedstaaten Deutschland und Österreich erheblich stärker zum Tragen kommt als beim Nicht-Mitgliedsland Schweiz, das sich in erster Linie mit den Aussenwirkungen des europäischen Integrationsprozesses konfrontiert sieht. Das sogenannte zweite Gesicht der Europäisierung (Mair 2004: 341) betrifft in eingeschränkter Weise auch die Schweiz, wie etwa am Beispiel der Bologna-Reformen im Hochschulsystem zu erkennen. Weiter ist die Schweiz als mehrsprachiges Land eher den Einflüssen von Entwicklungen im romanischen Sprachraum ausgesetzt als die Vergleichsländer. Von besonderer Bedeutung sind hier die Differenzen in den Regierungssystemen; namentlich unterscheidet sich die Schweiz von den Vergleichsländern durch die breit etablierten direkt-demokratischen Instrumente, durch das System der politischen Konkordanz und durch die stärker dezentralisierte Struktur, sowohl in vertikaler wie horizontaler Perspektive (Widmer 2007: 70-74).

Die drei diesem Artikel zugrundeliegenden Beiträge (Konzendorf zu Deutschland, Pichler zu Österreich und Mader zur Schweiz, alle in diesem Band) setzen ihre je eigenen Schwerpunkte in der Darstellung der Entwicklung der institutionellen Einbettung. Dies mag einerseits mit dem übergreifenden Charakter der zu behandelnden Thematik und mit der jeweiligen Perspektive der Autoren zusammenhängen, andererseits kommen darin auch materielle Differenzen zum Ausdruck, die beim vorliegenden Vergleich von vorrangigem Interesse sind.

Der Begriff der Institution respektive der institutionellen Einbettung oder Verankerung und der Institutionalisierung hat viele Facetten, weshalb an dieser Stelle eine Eingrenzung vorgenommen werden soll. Mit dem Begriff der institutionellen Einbettung ist hier gemeint,

wie die Beziehungen zwischen den Evaluationsaktivitäten und den staatlichen Instanzen strukturiert sind (vgl. dazu auch Davies et al. 2006). Grundsätzlich werden in diesem Beitrag Institutionen formeller wie informeller Art berücksichtigt, wobei Evaluationsaktivitäten mit Bezug zu staatlichen Instanzen und Aktivitäten im Zentrum stehen. Es erfolgt eine Beschränkung auf ausgewählte Aspekte, erstens auf die Rechtsgrundlagen, zweitens auf organisationale Vorkehrungen und drittens auf informelle Charakteristiken, die hier als Evaluationskultur bezeichnet werden können. Diesen drei Dimensionen folgt auch die Darlegung im nachfolgenden Abschnitt. Der Beitrag wird abgeschlossen mit weiterführenden Überlegungen zur Bedeutung von institutionellen Arrangements im Kontext der Evaluation.

2 Die institutionelle Einbettung im Ländervergleich

2.1 Rechtsgrundlagen

Auf der Ebene des Grundgesetzes resp. der Verfassung stellt die Schweiz einen Sonderfall dar, da die Schweizerische Bundesverfassung mit Artikel 170 einen verfassungsrechtlichen Evaluationsauftrag kennt. Im Gegensatz dazu finden sich in den entsprechenden Rechtssätzen in Deutschland und Österreich lediglich Normen, welche den Staat dazu anhalten, neben der Recht- und Ordnungsmässigkeit sowie der Wirtschaftlichkeit auch die Wirksamkeit staatlichen Handelns zu berücksichtigen. In der Schweiz besteht diese Verpflichtung auch, ergänzend wird das Parlament im genannten Verfassungsartikel aufgerufen, die Überprüfung der Wirksamkeit der Massnahmen des Bundes sicherzustellen. Damit besteht hier eine weitgefasste Verfassungsgrundlage für Evaluationen in allen Bereichen und auf allen Ebenen bundesstaatlichen Handelns. Auffällig ist länderübergreifend, dass Informationen zu den Rechtsgrundlagen auf subnationaler Ebene (Länder respektive Kantone sowie Kommunen) nur punktuell zur Verfügung stehen.

Die Ebene des Grundgesetzes respektive der Verfassung ist sicherlich nicht die einzige und bei weitem nicht die wichtigste Rechtsgrundlage für Evaluationsaktivitäten in den drei Ländern. In Deutschland und Österreich von weitreichender Bedeutung sind die entsprechenden normativen Grundlagen der Europäischen Union, die in den entsprechenden Politikfeldern (wie Struktur- und Regionalpolitik, Agrarpolitik etc.) massgeblich die Evaluationsaktivitäten auf nationaler und regionaler Ebene anleiten. Dieser Impuls besteht im Fall der Schweiz in erheblich kleinerem Ausmass, etwa bei Programmen der Europäischen Kommission mit grenzüberschreitendem Charakter, an denen Schweizer Akteure teilweise mitwirken.

Neben diesen Grundlagen wohl der relevanteste Bereich von Evaluationsverpflichtungen in rechtlicher Form sind speziell für die Schweiz (Bussmann 2007; Widmer 2007), aber auch in den beiden anderen Ländern, die spezialgesetzlichen Regelungen, in denen sektorbezogenen Vorgaben hinsichtlich der Durchführung von Evaluationen festgelegt sind. Dabei sind unterschiedliche Formulierungen anzutreffen, die verschiedene Bewertungskriterien (wie Wirksamkeit, Wirkung, Zielerreichung, Erfolg etc.) auf hohem Abstraktionsniveau nennen und zumeist auch eine Kompetenzzuweisung zur Evaluationsdurchführung formulieren. In der Regel enthalten diese Rechtsgrundlagen auf Ebene Gesetz oder Verordnung

keine Angaben dazu, zu welchem Zweck Evaluationen durchgeführt werden sollten oder wie und durch wen die Evaluationen zu nutzen wären. Den zuständigen Stellen wird somit ein erheblicher Ermessensspielraum bei der Erfüllung der rechtlichen Evaluationsverpflichtung überlassen. Es ist zu vermuten, dass die rechtsetzenden Instanzen zwar den Bedarf nach einer Überprüfung etwa der Wirksamkeit der entsprechenden Rechtsgrundlagen erkannt haben, eine funktionale Sichtweise dabei aber weitgehend fehlt. Auf diesem Hintergrund überrascht es kaum, dass die für die Realisierung der entsprechenden Evaluationen zuständigen Stellen je individuelle Interpretationen der Evaluationsklauseln vornehmen, die stark von Eigeninteressen geprägt sind – und vielfach aus einer Gesamtsicht betrachtet als dysfunktional bezeichnet werden müssen (Braun 2008; Frey 2008; Widmer 2008a). Die spezialgesetzlichen Regelungen führen damit zu einer disparaten, bereichsspezifischen institutionellen Einbettung der Evaluation.

2.2 Organisationen

In allen drei Ländern wurden bislang keine zentralen Evaluationsagenturen geschaffen, die sektor- und gewaltenübergreifend über eine allgemeine Evaluationskompetenz verfügen. In dieser Hinsicht ähneln die drei Vergleichsländer anderen Staaten (vgl. Leeuw 2006a), wie etwa den Niederlanden. Dort bestehen verschiedene Einrichtungen, die Evaluationsaufgaben auf Ebene der Zentralregierung wahrnehmen. Zu nennen wäre etwa das Büro für soziale und kulturelle Planung (SCP), das in Bereichen wie soziale Wohlfahrt und Gesundheit tätig ist, das Nationale Büro für Wirtschaftsforschung (CPB) sowie das Zentrum für wissenschaftliche Forschung und Dokumentation (WODC) beim Justizministerium, das sich mit Fragen der Kriminalität und Justiz befasst (Leeuw 2006b; Leeuw et al. 2007).¹ Die Evaluationsaktivitäten in den drei Vergleichsländern verteilen sich auf verschiedene Evaluationsinstanzen, die in der Regel sehr nahe bei den für den Evaluationsgegenstand zuständigen Stellen angesiedelt oder teilweise auch mit diesen identisch sind. Es besteht zwar eine Reihe von institutionellen Einheiten, die relativ ressortfern sind; der Grossteil der Evaluationsaktivitäten findet aber bei den entsprechenden Ressorts statt. In allen drei Ländern bilden die Rechnungshöfe (für Deutschland und Österreich) respektive die Eidgenössische Finanzkontrolle (für die Schweiz) eine Ausnahme. In allen drei Ländern wurden – einem internationalen Trend folgend – die herkömmlichen Prüfkriterien der Finanzaufsicht um Aspekte der Wirtschaftlichkeit und Effizienz ergänzt (Summa 2002: 16-17). In allen drei Ländern sind damit die Finanzaufsichtsorgane zumindest potentiell zu Evaluationsagenten geworden, wobei im Falle der Schweiz diese Rolle in Relation zu den klassischen Funktionen der Finanzaufsicht stärker entwickelt ist als in den beiden Vergleichsländern (vgl. dazu Bundesrechnungshof 2005).

Jenseits der Gesetzgebung und Finanzaufsicht scheint auf Seiten der Legislative das Interesse an Evaluation in Deutschland und Österreich bisher wenig zum Tragen gekommen

¹ Neben den genannten Einrichtungen bestehen in den Niederlanden zahlreiche weitere Stellen, die im Rahmen ihrer Inspektions- und Aufsichtsfunktionen Evaluationen durchführen. Aufgrund des erheblichen Ressourcenverbrauchs aller dieser Einrichtungen stehen sie unter zunehmendem Spardruck, der in den nächsten Jahren zu einer Reduktion der Evaluationsaktivitäten führen könnte.

zu sein. Auch wenn in seinem Umfang beschränkt, hat das schweizerische Bundesparlament mit der Parlamentarischen Verwaltungskontrolle (PVK) eine Evaluationseinheit zur Verfügung, die im Rahmen beschränkter Mittel im Auftrag des Parlaments im Rahmen der parlamentarischen Oberaufsicht Evaluationsfunktionen wahrnimmt. In den Vergleichsländern fehlt eine solche Organisation auf Seiten der Legislative. Dies mag damit in Zusammenhang stehen, dass entsprechende Aufsichtsfunktionen in Konkurrenzsystemen durch die Opposition wahrgenommen werden, wobei es zahlreiche Beispiele gibt, in denen entsprechende, mit Evaluation befasste parlamentarische Einrichtungen auch in Konkurrenzsystemen bestehen (vgl. Mayne et al. 1999); das prominenteste ist wohl das frühere United States Government Accounting Office des amerikanischen Kongresses (heute: Government Accountability Office; vgl. dazu Chelimsky 1990; Rist 1992; Mathiasen 1993; Rist/Paliokas 2002; Chelimsky 2006). Das Büro für Technikfolgenabschätzung beim Deutschen Bundestag ist zwar auch der Legislative zugeordnet, hat aber mit seinem Fokus auf den Folgen technologischer Innovationen eine deutlich andere Ausrichtung. Ansonsten wird aus Deutschland und Österreich über keine entsprechende Einrichtungen berichtet. An dieser Stelle ist aber auch zu bemerken, dass das schweizerische Bundesparlament hinsichtlich seiner allgemeinen Ressourcen deutlich geringer ausgestattet ist, als jene in Deutschland und Österreich. Neben der schweizerischen Eigenheit, dass Parlamentsmandate im Nebenamt wahrgenommen werden (sog. Milizparlament), sind auch die Dienste zur Unterstützung des Parlaments erheblich bescheidener ausgestattet. So müssen etwa für Deutschland auch der ausgebaut wissenschaftliche Apparat des Bundestags sowie die wissenschaftlichen Assistenzen zur Unterstützung der Bundestagsabgeordneten berücksichtigt werden, die partiell auch als funktionales Äquivalent zu einem parlamentarischen Evaluationsdienst gesehen werden können.

In der Schweiz wie auch in Deutschland und Österreich ist dessen ungeachtet der Grossteil der Evaluationsaktivitäten dezentral bei den entsprechenden Ressorts zu beobachten. Hierbei fällt auf, dass sich die Vorgehensweisen beim evaluationsspezifischen Kompetenzaufbau zwischen den drei Ländern wesentlich unterscheiden. In Deutschland sind traditionell bestehende oder speziell geschaffene, sektorspezifische Forschungseinrichtungen wie in den Niederlanden (siehe oben sowie Leeuw 2006b; Leeuw et al. 2007) mit der Durchführung von Evaluationen betraut worden. Erst in jüngerer Zeit werden häufiger auch Dritte, namentlich private Forschungseinrichtungen und Hochschulen mit der Durchführung von Evaluationen beauftragt. Während vermittelnd tätige Einrichtungen in Deutschland auch eine Rolle spielen, sind diese in Österreich in der für dieses Land typischen sozialpartnerschaftlichen Form von zentraler Bedeutung. Auch hier sind die entsprechenden Einrichtungen aber sektorspezifisch verfasst. Im Gegensatz dazu hat sich die (vergleichsweise sehr kleine) schweizerische Bundesverwaltung von Beginn an auf die Auftragsvergabe an externe Stellen (vornehmlich private Forschungsinstitute und universitäre Hochschulen, neuerdings vermehrt auch Fachhochschulen) konzentriert. Dabei haben verschiedene Stellen der schweizerischen Bundesverwaltung, namentlich solche, die häufiger Evaluationen durchführen lassen, amtsinterne Evaluationseinheiten geschaffen, die mittlerweile über eine hohe Kompetenz in der Vergabe von Evaluationsprojekten verfügen. Entsprechende Stellen sind in Deutschland und Österreich hingegen nur sehr punktuell vorhanden. Abweichend von diesem Muster bestehen auch in der Schweiz in Einzelfällen der Bundesverwaltung zuge-

ordnete Forschungseinrichtungen (etwa im Agrarbereich); diese stellen aber Ausnahmen dar. Nicht zu unrecht wird im entsprechenden Länderbeitrag auf die Tradition der „Milizverwaltung“ (Germann 1981) verwiesen, die in der Regel auf externe Expertise zurückgreift und nur bei Vorliegen zwingender Gründe verwaltungsinterne Kapazitäten schafft. Dies steht in deutlichem Gegensatz zur deutschen und österreichischen Tradition, in welcher der interne Aufbau von Kapazitäten im Vordergrund steht und mit dafür verantwortlich zeichnet, dass die deutschen und österreichischen Bürokratien (auch in Relation zur Grösse des Landes) deutlich umfangreicher sind.

Die drei Länderbeiträge fokussieren ihre Darstellungen auf die Bundesebene. So erscheinen die subnationalen Ebenen in allen drei Beiträgen nur sehr punktuell. Im Falle der Schweiz (siehe Beitrag von Mader in diesem Band) wird auf Aktivitäten in wenigen ausgewählten Kantonen verwiesen. In besonderer Weise fällt dabei der Hinweis auf den Kanton Genf auf, der mit der „Commission externe d'évaluation des politiques publiques“ (CEPP; vgl. auch Delley 2001; 2007; Horber-Papazian/Bützer 2008) eine unabhängige Evaluationsagentur geschaffen hat. Im deutschen Beitrag wird die in einigen Bundesländern bestehende Regelungen einer ‚sunset legislation‘ genannt, wobei hier keine mit dem Kanton Genf vergleichbare organisatorische Entwicklung berichtet wird (siehe Beitrag von Konzendorf in diesem Band). Für Österreich wird auf den interessanten Aspekt des höheren Rechtfertigungsbedarfs der Ländergesetzgebung verwiesen (vgl. Beitrag von Pichler in diesem Band). Insgesamt betrachtet bleibt das Gewicht der subnationalen Ebene in den Darstellungen gering. Dies mag zum einen mit dem gewählten Fokus zu tun haben, es scheint aber auch in Zusammenhang damit zu stehen, dass die Informationsdichte auf der Ebene der Bundesländer respektive der Kantone und besonders der Kommunen deutlich tiefer ist.

2.3 Evaluationskultur

Bei der Darstellung der institutionellen Einbettung der Evaluationsfunktion in den drei Ländern fällt auf, dass sich die Praktiken in verschiedenen Aspekten unterscheiden, die sich nur beschränkt auf die formalen Institutionen zurückführen lassen. Besonders auffällig scheint dies bei den bestehenden Grundlagen zur Evaluationstätigkeit in den drei Ländern zu sein. Während für die Schweiz eine recht grosse Anzahl von Beiträgen und Studien zur Evaluationstätigkeit und zu institutionellen Fragen vorliegen (siehe exemplarisch Widmer/Neuenschwander 2004; Balthasar 2007; Bussmann 2007; Widmer 2007), ist dies bei Deutschland und Österreich nicht der Fall. Diese Grundlage führt dazu, dass in der Schweiz deutlich mehr an Kenntnissen zu Evaluationsaktivitäten und deren Einbindung bestehen als in Deutschland und Österreich. Dies erstaunt besonders im Vergleich zu Deutschland, weil aufgrund der Grösse erwartet werden könnte, dass der Forschungsstand auf Bundesebene besser ist. Offenbar scheint in der Schweiz ein reflexiver Umgang mit der Materie Evaluation und deren Verankerung stärker etabliert als in den Vergleichsländern. Besonders gilt dies für Betrachtungen, die einen themenfeldübergreifenden Charakter haben, wo die bestehenden Grundlagen in der Schweiz weit umfangreicher sind als in Deutschland und Österreich. Besonders in Österreich fällt auf, dass die Grundlagen stark sektoral ausgerichtet sind (etwa Forschung und Technologie, Nachhaltigkeit etc.), während in Deutschland eher

ein Fokus auf bestimmte Instrumente und Verfahren (zum Beispiel Gesetzesfolgenabschätzung, Standard-Kosten-Modell) zu dominieren scheint.

Aufgrund der zahlreichen für die Schweiz vorliegenden Studien, welche sich in reflexiver und analytischer Form mit Evaluationsaktivitäten und deren institutioneller Verankerung auseinandersetzen (siehe etwa Horber-Papazian 1990; Bussmann 1995; 1996; 2007; Fornerod 2001; Mader 2003; Widmer/Neuenschwander 2004; Kettiger 2005; Widmer 2005; Good 2006; Balthasar 2007; Widmer 2007; Horber-Papazian/Bützer 2008; Spinatsch 2008), ist der Informationsstand – zumindest was die Bundesebene betrifft – deutlich besser als in Deutschland und in Österreich. Dies mag auch mit der Grösse der Verwaltungsapparate in Zusammenhang stehen. Darüber hinaus scheint das Interesse der öffentlichen Stellen wie auch der Forschung an diesen übergreifenden Fragen in Österreich und Deutschland verhältnismässig gering zu sein.

In allen drei Ländern besteht eine Herausforderung für die Bundesebene darin, die Evaluationsaktivitäten zu koordinieren. Während in Deutschland erste Bestrebungen festzustellen sind, über eine Internet-Plattform zumindest den Informationsaustausch zu verbessern (vgl. Beitrag von Konzendorf in diesem Band), bestehen in Österreich auf einzelne Themenfelder beschränkte Plattformen bereits (vgl. Beitrag von Pichler in diesem Band). In der Schweiz wird der Informationsaustausch durch die Vernetzung der im Evaluationsbereich tätigen Verwaltungsstellen unterstützt. Es gibt zudem in der Schweiz auch Verpflichtungen zur Koordination der Evaluationsaktivitäten, wobei diese bisher nur teilweise umgesetzt sind (vgl. Beitrag von Mader in diesem Band). Es scheint generell in allen drei Ländern nicht ganz einfach, die Evaluationsaktivitäten verschiedener Stellen ressortübergreifend zu koordinieren. Der Informationsaustausch, der für eine angemessene Koordination eine notwendige Voraussetzung darstellt, ist im Falle der Schweiz ansatzweise, in Österreich für ausgewählte Themenfelder und in Deutschland noch kaum realisiert.

3 Schlussbemerkungen

Trotz der grundsätzlichen Ähnlichkeit der drei Länder macht ein Vergleich der Institutionalisierung der Evaluationsfunktionen deutlich, dass sich erhebliche Differenzen zwischen den drei Ländern zeigen. In vielerlei Hinsicht scheint die Institutionalisierung der Evaluationsfunktion in der Schweiz weiter vorangeschritten als in Deutschland und Österreich. Dieser Befund steht im Einklang mit anderen Analysen des Stands der Institutionalisierung (Jacob/Varone 2003: 122-125), weicht jedoch ab von früheren Beobachtungen (Furubo/Sandahl 2002²). Aber auch in Deutschland und Österreich bestehen Tendenzen, Institutionen der Evaluation zu etablieren. Dies ist bei weitem kein spezifisches Phänomen in den drei hier behandelten Ländern, sondern steht in Zusammenhang mit allgemeinen Entwicklungen in demokratischen Staaten westlichen Zuschnitts. Durch Phänomene der zunehmenden Autonomisierung staatlicher oder staatlich delegierter Tätigkeit, durch den Bedeutungszu-

² Dieser Beitrag befasst sich mit Evaluationspraktiken und nicht spezifisch mit Fragen der Institutionalisierung und schliesst zwar Deutschland und die Schweiz, nicht aber Österreich mit ein. Vgl. dazu auch Derlien 1990b und Leeuw 2006a.

wachs von Mehrebenensystemen und durch vermehrte staatliche Steuerung in Netzwerkstrukturen erhöht sich der Bedarf an entsprechenden Wirkungsinformationen (siehe Widmer 2002; Widmer/Frey 2006; Verhoest et al. 2007; für den Hochschulbereich Lincoln 1990; Braun 1999). Die Expansion von Evaluationserfordernissen ist demnach nicht auf die Evaluation selbst zurückzuführen, sondern auf externe Entwicklungen. Dies soll jedoch nicht darüber hinwegtäuschen, dass Evaluationssysteme auch eine Eigendynamik entwickeln können (Leeuw/Furubo 2008: 164-165).

Zum Abschluss dieses Beitrags stellt sich die Frage, welche Konsequenzen eine verstärkte Institutionalisierung mit sich bringt. In dieser Hinsicht sind formelle und informelle Institutionen erheblich anders zu beurteilen. Zum einen ist zu betonen, dass Institutionen, seien diese nun formeller oder informeller Art, wesentliche Vorteile mit sich bringen. Gewisse Institutionen sind gar notwendige Voraussetzungen für eine systematische Evaluationsstätigkeit und für die Etablierung von Evaluationssystemen (Leeuw/Furubo 2008). Andererseits sind besonders formelle Institutionen auch dem Risiko ausgesetzt, negative Entwicklungen auszulösen. Ein Problem stellt etwa der Umstand dar, dass Evaluationsklauseln zwar Evaluationen auslösen. Dies bietet jedoch keinerlei Gewähr, dass die ausgelösten Evaluationen auch sinnvoll eingesetzt und genutzt werden. Einerseits können Evaluationsklauseln dazu führen, dass Evaluationen alleine zum Zweck durchgeführt werden, die rechtlichen Anforderungen zu erfüllen und somit „Phänomene der Routinisierung (Bürokratisierung) und Ritualisierung (Evaluation als symbolisches Ritual)“ (Widmer 2008a: 277) fördern. In derartigen Situationen werden Evaluationen kaum einen Nutzen stiften, sondern vornehmlich Kosten verursachen, was verschiedene Autoren dazu motivierte, den Begriff der ‚Evalutis‘ zu prägen (siehe etwa Simon 2000; Umbach 2002; Frey 2008). Ebenso können Evaluationsklauseln dazu führen, dass Evaluationen zu häufig durchgeführt werden oder zu wenig fokussiert erfolgen (Döring 2005; Widmer 2008a). Da die Qualitätsanforderungen, die an Evaluationen gestellt werden müssen, hoch sind, sind der jeweiligen Aufgabe entsprechende Mittel aufzuwenden (Widmer et al. 1996; Widmer 2005). Sind die Mittel insgesamt beschränkt – was wohl die Regel darstellt – werden bei einer Steigerung der Zahl an durchzuführenden Evaluationen die Mittel für die einzelne Evaluationen geringer ausfallen (vgl. Alkin/Solmon 1983). Dies kann zu wesentlichen Abstrichen bei der Evaluationsqualität führen, falls nicht spezifische Voraussetzungen erfüllt sind (siehe Widmer et al. 1996). Aber auch auf Seiten der von Evaluationsaktivitäten betroffenen Kreise wird sich aufgrund einer hohen Evaluationsdichte eine Evaluationsmüdigkeit (‚evaluation fatigue‘; Dahler-Larsen 2006: 146) bemerkbar machen, welche der Evaluationsqualität ebenfalls abträglich ist. Zu guter Letzt ist auch an die Adressaten von Evaluationen zu denken, die beschränkte Kapazitäten der Informationsverarbeitung aufweisen. Ein Überangebot an entsprechenden Evaluationeninformationen ist deren Nutzung kaum förderlich. Kommt hinzu, dass bei widersprüchlichen oder ambivalenten Resultaten der Einfluss empirischer Befunde erheblich abnimmt (Esterling 2004; Gormley 2007: 310-311). Evaluationstätigkeiten verfügen deshalb – ceteris paribus – über einen sinkenden Grenznutzen, der im Extremfall auch im negativen Bereich liegen kann. Damit soll jedoch nicht die These vertreten werden, dass sich die Evaluationsqualität alleine aufgrund der Zahl an Evaluationen und den dafür einsetzbaren Mitteln bestimmen würde; beides sind aber Rahmenbedingungen, welche die Evaluationsqualität beeinträchtigen können.

Diese Fehlentwicklungen sind in erster Linie bei einer hohen Dichte an verbindlichen Evaluationsklauseln zu erwarten, die keine funktionale Ausrichtung aufweisen, eine regelmässig wiederkehrende, flächendeckende Evaluationsdurchführung einfordern und Stellen mit der Evaluationsdurchführung betrauen, die über keine oder zu geringe Evaluationskompetenzen verfügen. In diesem Sinne kann durchaus von der Janusköpfigkeit der formellen Institutionalisierung der Evaluationsfunktion gesprochen werden. Gerade die Form der institutionellen Einbettung der Evaluationsfunktion spielt besonders für die Nutzung eine zentrale Rolle (Widmer/Neuenschwander 2004), wie dies für die Forschung schon früh festgestellt wurde (Merton 1936). Die Institutionalisierung sollte damit mit Bedacht in einer Weise erfolgen, die eine gezielte sowie selektive (und damit sinnvolle) Evaluationsaktivität fördert und Fehlentwicklungen vermeidet.

Agrarpolitik

Evaluation in der Agrarpolitik in Deutschland

1 Einleitung

Agrarpolitik in Deutschland umfasst eine Vielzahl von Handlungsfeldern und Kompetenzebenen, in denen Evaluationsaktivitäten unterschiedlich verankert sind.

Finanziell bedeutendste Bereiche der Agrarpolitik in Deutschland sind die Agrarhandelspolitik und die Markt- und Einkommenspolitik (zusammengefasst in der 1. Säule der Gemeinsamen Agrarpolitik [GAP] der Europäischen Union [EU]). Diese wird derzeit stark reformiert, zum Beispiel in Form der Entkopplung der Direktzahlungen und der Liberalisierung der Gemeinsamen Marktorganisationen (European Commission 2004). Im Zuge dieser Reformen ist die Bedeutung der Strukturpolitik für die Landwirtschaft und den ländlichen Raum sowie der Agrarumweltpolitik gestiegen. Diese ursprünglich separierten Handlungsfelder wurden im Jahr 2000 in der 2. Säule der GAP zusammengefasst und finanziell aufgestockt. Überwiegend national gestaltet ist die Agrarsozialpolitik und die sektorbezogene Steuerpolitik¹.

Schwerpunktmäßig wird in diesem Beitrag die Evaluation im Handlungsfeld der 2. Säule der GAP betrachtet. Dieses Handlungsfeld ist von einer Mehrebenenverflechtung (EU, Bund, Land) geprägt. Die EU setzt den Rahmen für die Formulierung, Implementierung und Finanzierung der Politik sowie deren Evaluation. Es werden im Folgenden die Impulse, welche die EU für Evaluation in Deutschland im Vergleich zu den im nationalen Kontext existierenden Vorgaben und deren Umsetzung gibt, dargestellt. Abschließend wird der Frage der Nutzung von Evaluationen auf den verschiedenen Ebenen nachgegangen.

2 Agrarpolitik in Deutschland: Handlungsfelder und Kompetenzverteilung

2.1 1. Säule der Gemeinsamen Agrarpolitik

Die 1. Säule der GAP wird von internationalen Vereinbarungen im Rahmen der Welthandelsorganisation und den Vorgaben der EU bestimmt. Wesentliche Instrumente in der EU sind Exportsubventionen und Marktunterstützungen sowie Direktzahlungen, deren Finan-

¹ Nicht behandelt werden in diesem Beitrag Politikfelder außerhalb des Landwirtschaftsressorts, auch wenn diese einen maßgeblichen Einfluss auf die Entwicklung des Agrarsektors haben. Dazu zählt beispielsweise das Erneuerbare-Energien-Gesetz (EEG) in der Zuständigkeit des Umweltministeriums, das den Anbau und die Verwendung von nachwachsenden Rohstoffen deutlich stimuliert hat.

zierung zu hundert Prozent aus dem EU-Haushalt erfolgt. Der Bund setzt, mit begrenzten Mitwirkungsmöglichkeiten der Bundesländer, die EU-Verordnungen der 1. Säule in nationales Recht um und wählt die möglichen Ausgestaltungsalternativen aus, während die Bundesländer für den Vollzug zuständig sind (Mehl/Plankl 2002: 3).

2.2 2. Säule der Gemeinsamen Agrarpolitik

Die 2. Säule der GAP umfasst die Förderung sektoraler Maßnahmen, von Ausgleichs- und Agrarumweltmaßnahmen und ländlichen Entwicklungsmaßnahmen. Im Gegensatz zur 1. Säule beteiligt sich die EU an der Finanzierung der 2. Säule nur anteilig. Sie setzt in den Verordnungen den groben Rahmen und überlässt den Mitgliedstaaten die konkrete Ausgestaltung der Programme.

Die Umsetzung liegt im föderalen System Deutschlands in der Zuständigkeit der Bundesländer. Aufgrund der Länderzuständigkeiten und der unterschiedlichen Programmierungs- und Finanzierungsvorgaben zwischen 2000 und 2006 gab es in Deutschland eine Vielzahl von Programmen².

Inhaltlich wird die Politik für ländliche Räume im Wesentlichen von der Bund-Länder-Gemeinschaftsaufgabe zur Verbesserung der Agrarstruktur und Küstenschutz (GAK) bestimmt. Die GAK beinhaltet agrarstrukturelle und landeskulturelle Maßnahmen sowie seit Mitte der 1990er Jahre auch Agrarumweltmaßnahmen. Daneben verfügen die Bundesländer über eigene Landesmaßnahmen im Agrarstruktur- und Umweltbereich.

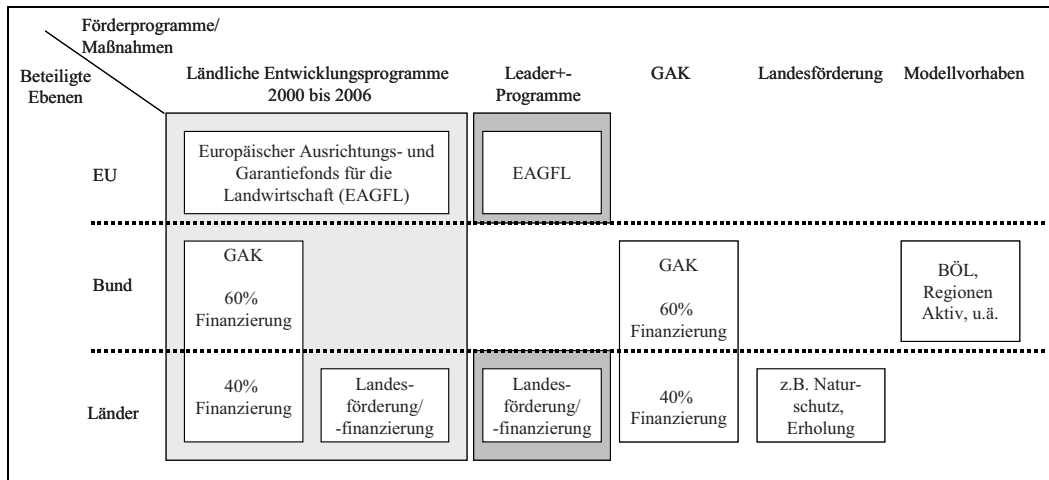
Sowohl die GAK als auch die Landesmaßnahmen sind überwiegend in die EU-Programme für den ländlichen Raum integriert, so dass Festlegung der Politikinhalte, Durchführung und Finanzierung von einer Mehrebenenverflechtung gekennzeichnet sind, an denen EU, Bund, Länder und zumindest finanzierungsseitig auch die Kommunen beteiligt sind (Mehl/Plankl 2002).

GAK- und Landesmaßnahmen werden in geringem Umfang auch außerhalb der EU-Programme angeboten. Darüber hinaus gibt es auf der Ebene des Bundes noch Modell- und Demonstrationsvorhaben zu ausgewählten Themen³ (siehe Abbildung 1).

² 16 Entwicklungspläne für den ländlichen Raum, sechs Operationelle Programme mit Maßnahmen des Europäischen Ausrichtungs- und Garantiefonds der Landwirtschaft (EAGFL), Abteilung Ausrichtung sowie ein länderübergreifendes Gemeinschaftliches Förderkonzept (GFK) für die neuen Bundesländer, 13 LEADER+-Programme und ein nationales Programm für die Vernetzungsstelle LEADER. Bei LEADER handelt es sich um eine EU-Initiative, die seit 1991 modellhaft innovativ Aktionen im ländlichen Raum fördert.

³ Zu nennen sind hier beispielsweise das Bundesprogramm Ökologischer Landbau (<http://www.bundesprogramm-oekolandbau.de/>) oder Regionen Aktiv (<http://www.nova-institut.de/modellregionen/>).

Abbildung 1: Struktur der ländlichen Entwicklungsprogramme und -maßnahmen in Deutschland 2000 bis 2006



2.3 Agrarsozial- und sektorbezogene Steuerpolitik

Die Sozialpolitik für landwirtschaftliche Unternehmer und deren Familien sowie die landwirtschaftliche Steuerpolitik ist in Deutschland auf Bundesebene angesiedelt, wobei allerdings die EU-Wettbewerbsregeln zu beachten sind (Mehl/Plankl 2002).

3 Evaluationen im Themenfeld Agrarpolitik aufgrund von nationalen Vorgaben

Der Stellenwert von Evaluation in Deutschland hat im letzten Jahrzehnt insgesamt zugenommen und erfasst, in unterschiedlichem Ausmaß, inzwischen alle Politikbereiche. Auch im Bereich Agrarpolitik ist ein Zuwachs an Evaluationsaktivitäten zu beobachten; allerdings fast ausschließlich durch EU-Vorgaben veranlasst. Es gibt nur wenige Ansätze zu einer eigenständigen nationalen Evaluation, und auch die möglichen Spielräume zur Verknüpfung eigener Fragestellungen mit den EU-Evaluationsvorgaben werden kaum genutzt (siehe Kapitel 4).

Auf nationaler Ebene ist Evaluation partiell in der Gesetzesfolgenabschätzung und der gemäß Bundeshaushaltsordnung (BHO) vorgesehenen Erfolgskontrolle verankert. Im Bereich der 1. Säule der GAP gibt es anlassbezogen seitens des Bundes und auch der Länder in Auftrag gegebene Studien im Sinne von Politikfolgenabschätzungen als Grundlage für den politischen Entscheidungsprozess und zur Festlegung der Ausgestaltungsvarianten der rahmensetzenden EU-Verordnungen (siehe beispielsweise Bertelsmeier et al. 2004; Isermeyer et al. 2005; Kleinhanß et al. 2003). In Deutschland kommt den ökonomischen Instituten der Ressortforschung des Bundesministeriums für Ernährung, Landwirtschaft und Verbraucherschutz (BMELV) eine wichtige Rolle zu. Dies gilt auch für das Handlungsfeld Agrarsozialpo-

litik. Am Institut für Ländliche Räume der Bundesforschungsanstalt für Landwirtschaft (FAL) wurden in jüngster Zeit einige Studien zur Reform des agrarsozialen Sicherungssystems angefertigt, welche die verschiedenen Gesetzesvorschläge in ihren Folgen beurteilten (Mehl 2003; 2006). Bei diesen Studien handelt es sich nicht explizit um Evaluationen. Sie erfüllen aber wesentliche Kriterien von Evaluationen, indem sie Ausgestaltungsvarianten mit ihren Vor- und Nachteilen bewerten und steuerungsrelevantes Wissen generieren.

Die Evaluation nationaler Politiken im Sinne einer Erfolgskontrolle ist von der BHO vorgegeben (zu anderen Evaluationsansätzen auf Bundesebene siehe Beitrag von Konzen-dorf in diesem Band). § 7 der BHO verlangt vor der Durchführung einer finanzwirksamen Maßnahme im Rahmen der Planung eine Wirtschaftlichkeitsuntersuchung und nach der (Teil-)Realisierung dieser Maßnahme eine Erfolgskontrolle (Ergebnisprüfung).

Trotz klarer gesetzlicher Vorgaben ist allerdings ein deutliches Umsetzungsdefizit festzustellen, die Erfolgskontrolle systematisch in der Bundesverwaltung zu verankern und die ggf. gewonnenen Ergebnisse in den politischen Entscheidungsprozess rückzukoppeln (BWV 1998). Im Bereich der GAK wird das Defizit in der Anwendung der Vorgaben der BHO v. a. auf mehrdeutige Förderziele und zersplitterte Zuständigkeiten zurückgeführt, die dazu führen, dass Zielerreichungs- und Wirkungskontrollen sich nahezu ausschließlich auf die Erstellung und Fortschreibung von Statistiken beschränken (BWV 2002: 123). Zudem ist in § 9 des Gesetzes über die GAK lediglich festgelegt: „Die Durchführung des Rahmenplanes ist Sache der Länder. Die Landesregierungen unterrichten die Bundesregierung und den Bundesrat auf Verlangen über die Durchführung des Rahmenplanes und der Gemeinschaftsaufgabe“. Es wird somit nicht vorgegeben, dass über den Vollzug hinausgehend auch Wirkungsanalysen durchzuführen sind. Darüber hinaus erfüllen einzelne Komponenten der GAK den Charakter von Dauersubventionen (z. B. die Ausgleichszulage für die Förderung landwirtschaftlicher Betriebe in benachteiligten Gebieten sowie die Förderung von Leistungsprüfungen in der tierischen Erzeugung), sodass kein eindeutiges und zeitlich wie inhaltlich erreichbares nationales Ziel vorhanden ist (BWV 2002: 124). Es gibt demnach keine systematische Erfolgskontrolle für die GAK, abgesehen von einzelnen anlassbezogenen Studien der 1990er Jahre (siehe Forstner 2000; Striewe et al. 1996).

Neue, eher formative Ansätze in der Evaluation werden im Rahmen der Modellvorhaben des Bundes verfolgt. Hier steht die Prozessorientierung im Vordergrund, das heißt die Ergebnisse werden unmittelbar in den laufenden Prozess rückgekoppelt. Die Gesamtevaluation des Bundesprogramms Ökologischer Landbau (BÖL) hatte den Auftrag, die Umsetzung des Programms kritisch zu begleiten und möglichst schnell Ergebnisse zur Ziel(gruppen)erreichung, Wirksamkeit und Effizienz der Einzelmaßnahmen zu liefern. Ziel war es, noch im laufenden Programm und für Planungsarbeiten im Falle seiner Verlängerung Empfehlungen zur Optimierung einzelner Bereiche bzw. zur Korrektur von Fehlentwicklungen auszusprechen (GIB 2003; GIB 2004). Die Begleitforschung des Modellvorhabens „Regionen Aktiv“ zielt auf Wissenstransfer und Initiierung von Lernprozessen ab. Das heißt, die Einbettung des Modellvorhabens in einen theoretisch-wissenschaftlichen Rahmen und die empirische Überprüfung sind wesentliche Bausteine der Untersuchung (Elbe 2006)⁴.

⁴ Die Berichte, die im Rahmen der Begleitforschung zu Regionen aktiv erstellt wurden, sind im Internet unter <http://www.nova-institut.de> verfügbar.

4 Evaluation des Themenfeldes Agrarpolitik aufgrund von EU-Vorgaben

Schon in den 1990er Jahren bestand eine Verpflichtung der Mitgliedstaaten zur Evaluation EU-kofinanzierter Programme und Maßnahmen. Die inhaltlichen Vorgaben und die Verknüpfung mit dem Politikzyklus waren allerdings wenig formalisiert, sodass die Evaluationsberichte sowohl innerhalb Deutschlands als auch EU-weit kaum vergleichbar waren. Seit 1998 gibt es innerhalb der Generaldirektion für Landwirtschaft und ländliche Entwicklung (DG Agri) eine Evaluierungseinheit, die sowohl für die Vergabe von eigenen Evaluationen (v. a. im Bereich der Gemeinsamen Marktorganisation) als auch für die Entwicklung von Leitfäden für die dezentrale Durchführung von Evaluationen vor allem ländlicher Entwicklungsprogramme in den Mitgliedstaaten zuständig ist. Bei dezentralen Evaluationen besteht ein wesentliches Interesse der EU-Kommission (EU-KOM) darin, eine EU-weite Vergleichbarkeit der Evaluierungsergebnisse sicherzustellen (Williams et al. 2002). Dieses Interesse findet seinen Ausdruck in strikten zeitlichen Vorgaben und inhaltlich-methodischen Leitlinien zur Evaluation durch die EU-KOM für den Programmplanungszeitraum 2000 bis 2006 (EU-KOM 1999a; EU-KOM 1999b; EU-KOM 1999c; EU-KOM 2000a; EU-KOM 2000b).

Im Wesentlichen handelt es sich bei dem Evaluationsansatz für die Förderperiode 2000 bis 2006 um ein indikatorgestütztes System, mit dem bottom-up Ergebnisse und Wirkungen über alle angebotenen EU-kofinanzierten Maßnahmen erfasst werden sollen. Dabei werden die Output- und teilweise auch Ergebnisindikatoren im Rahmen des Begleitsystems erfasst, während die Abschätzung der Wirkungsindikatoren Aufgabe der Evaluation ist (EU-KOM 1999c). Die Quantifizierung von Indikatoren soll die Basis für Kosten-Wirksamkeitsanalysen bilden („value for money“) (Eser 2001: 328).

Im Rahmen einer Evaluation sind drei Phasen⁵ vorgesehen, die eng an den Politikzyklus gekoppelt sein sollen. Eng mit der Programmierung verknüpft ist die Ex-ante-Bewertung, deren Ergebnisse in den Programmplanungsprozess zurückfließen und dort direkt qualitätsverbessernd wirken sollen. Die Halbzeitbewertung soll erste Ergebnisse darstellen und zu diesem frühen Zeitpunkt noch die Möglichkeit eröffnen, durch Programmänderungen unerwünschten Entwicklungen gegenzusteuern. Die Ex-post-Bewertung, die zwei Jahre nach Abschluss des Programms erstellt werden soll, hat vorrangig eine summative Funktion.

Die Veröffentlichung der Evaluationsberichte (entweder in der vollständigen Fassung oder in einer Zusammenfassung) ist von der DG Agri vorgegeben. Sie erfolgt, sobald die Einrichtung, die den Auftrag zur Bewertung erteilt hatte, den Bewertungsbericht angenommen hat (EU-KOM 1999a: 27).

4.1 Aktivitäten im Bereich der Evaluation EU-kofinanzierter Programme

Aufgrund der zersplitterten Zuständigkeiten für die Umsetzung der Maßnahmen der 2. Säule in Deutschland gibt es eine Vielzahl von Evaluationsaktivitäten. Tabelle 1 gibt einen

⁵ Optional vorgesehen war eine Aktualisierung der Halbzeitbewertung (Up-date) im Jahr 2005, die wesentliche Empfehlungen für die Programmierung der folgenden Förderperiode liefern sollte.

Überblick über abgeschlossene und laufende Evaluationsvorhaben im Zeitraum 2000 bis 2006 anhand der Zahl der jeweils erstellten Abschlussberichte⁶.

Tabelle 1: Überblick über abgeschlossene und laufende Evaluationsvorhaben der Programme zur ländlichen Entwicklung 2000 bis 2006 in Deutschland¹⁾

| | Entwicklungsprogramme (EPLR) 2000 bis 2006 | | | EPLR 2007 bis 2013 |
|---|--|------------------------|---|------------------------|
| | Ex-ante-Bewertung 1999/2000 | Halbzeitbewertung 2003 | Aktualisierung der Halbzeitbewertung 2005 | Ex-ante-Bewertung 2006 |
| Programmbewertung der Entwicklungsprogramme nach VO (EG) Nr. 1257/1999 bzw. 1698/2005 | 16 | 16 | 15 | 14 |
| Zentrale Evaluation von GAK-Maßnahmen | | 52 (4 GAK) | 40 (3 GAK) | - |
| OP-Ziel 1 | 6 | 6 | 6 | - |
| GFK-Ziel 1 | 1 | 1 | 1 | - |
| LEADER+ | 13 | 13 | 13 | - |
| Vernetzungsstelle LEADER | 1 | 1 | 1 | - |

¹⁾ In Klammern steht die Zahl der vergleichenden länderübergreifenden GAK-Berichte für Deutschland für das Agrarinvestitionsförderungsprogramm, die Ausgleichszulage für benachteiligte Gebiete, die Verarbeitung und Vermarktung landwirtschaftlicher Erzeugnisse und die Erstaufforstung (nicht 2005 und 2008). OP=Operationelles Programm, GFK=Gemeinschaftliches Förderkonzept.

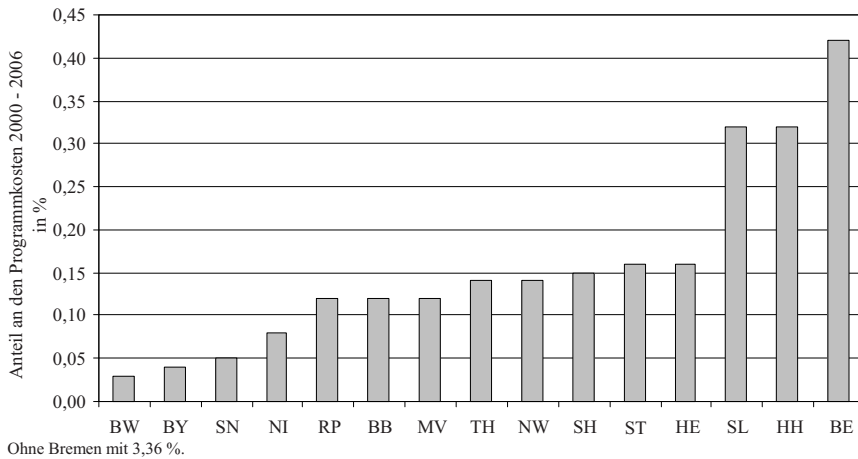
Im Vordergrund steht die Evaluation der Programme der Länder. Ausnahmen bilden die Evaluation der LEADER-Vernetzungsstelle sowie die zentrale, bundesweite Evaluation von GAK-Instrumenten (Agrarinvestitionsförderungsprogramm, Ausgleichszulage für benachteiligte Gebiete, Verbesserung der Verarbeitung und Vermarktung landwirtschaftlicher Produkte und Erstaufforstung), bei der neben Länderberichten auch zusammenfassende länderübergreifende Berichte erstellt wurden (Dirksmeyer et al. 2006; Plankl et al. 2006; Wendt et al. 2006). Dies stellt seitens des Bundes den Versuch dar, die EU-Evaluation zur gleichzeitigen Evaluation der eigenen Förderinstrumente zu nutzen (Forstner/Koch 2005), wobei allerdings über den Pflichtenkatalog der EU hinaus kaum eigene Fragestellungen entwickelt werden. Dieser gemeinsame Ansatz wird für die Förderperiode 2007 bis 2013 nicht weiter fortgeführt, obgleich die länderübergreifenden GAK-Berichte die einzigen Dokumente darstellen, die für Kernbereiche der GAK einen deutschlandweiten Überblick geben.

Insgesamt wurden für die Evaluation der Programme 2000 bis 2006 rund 15 Mio. Euro (geschätzt) an öffentlichen Mitteln eingesetzt; das sind im nationalen Durchschnitt rund 0,1 % an den öffentlichen Programmmitteln. Damit bleiben die Bundesländer deutlich unter dem von der EU-KOM vorgesehenen Maximalanteil von 1 % der Programmkosten. Die

⁶ Die Evaluationen für die Förderperiode 2000 bis 2006 haben sich überschneiden mit den Ex-post-Evaluationen, die den Förderzeitraum bis 1999 betrafen. Letztere sind in Tabelle 1 nicht berücksichtigt.

Anteile variieren deutlich zwischen den Bundesländern (siehe Abbildung 2). Der Anteil der Evaluationskosten an den Programmkosten ist tendenziell umso höher, je kleiner das Programmvolumen ist, da bei jeder Evaluation unabhängig vom Programmvolumen Fixkosten anfallen.

Abbildung 2: Anteil der Evaluierungsausgaben (geschätzt) in den deutschen Bundesländern¹⁾ an den öffentlichen Programmkosten 2000 bis 2006



¹⁾ BW-Baden-Württemberg, BY-Bayern, SN-Sachsen, NI-Niedersachsen, RP-Rheinland-Pfalz, BB-Brandenburg, MV-Mecklenburg-Vorpommern, TH-Thüringen, NW-Nordrhein-Westfalen, SH-Schleswig-Holstein, ST-Sachsen-Anhalt, HE-Hessen, SL-Saarland, HH-Hamburg, BE-Berlin

Quelle: Eigene Berechnung auf Grundlage der Finanztabellen der Bundesländerprogramme 2000-2006.

Die Auftragnehmerseite wird zahlenmäßig von privaten Anbietern und wissenschaftlichen Einrichtungen ohne öffentliche Grundfinanzierung dominiert (siehe Tabelle 2). Diese sind zumeist mit der Evaluation von Länderprogrammen beschäftigt.

Tabelle 2: Anzahl und Art der beteiligten Institutionen

| | EU-Programme 2000 bis 2006 | | EU-Programme 2007 bis 2013 |
|---|----------------------------|---|-----------------------------|
| | Halbzeitbewertung 2003 | Aktualisierung der Halbzeitbewertung 2005 | Ex-ante-Bewertung 2006/2007 |
| Universitäten | 3 | 1 | 2 |
| Wissenschaftliche Einrichtungen <u>mit</u> öffentlicher Grundfinanzierung | 3 | 2 | 2 |
| Wissenschaftliche Einrichtungen <u>ohne</u> öffentliche Grundfinanzierung | 8 | 9 | 5 |
| Private Beratungs-/Ingenieurbüros | 14 | 11 | 5 |

Quelle: Eigene Befragung

Die zentral über alle Länder durchgeführten Evaluationen von GAK-Maßnahmen sowie eine gemeinsame Evaluation von sechs Länderprogrammen wurden von den agrarökonomischen Instituten der FAL durchgeführt, denen damit in der Halbzeitbewertung und deren Aktualisierung eine wichtige Rolle zukam. Die Vor- und Nachteile dieser beiden spezifischen organisatorischen Ansätze beschreiben Forstner und Koch (2005). Die Vorteile liegen vor allem in Vergleichsmöglichkeiten, die Nachteile in einem erhöhten Abstimmungsaufwand und der Gefahr, Spezifika in den Bundesländern nicht ausreichend zu berücksichtigen.

Die Komplexität und der Umfang der Programme bestimmt die Art der evaluierenden Institutionen. Bei der Evaluation von Länderprogrammen gibt es noch eine vielfältige Anbieterstruktur; bei umfangreicheren Aufträgen beteiligen sich als Anbieter nur noch größere Forschungseinrichtungen und große, zumeist auch international operierende Beratungsunternehmen.

Die Veröffentlichung der Berichte erfolgte, teilweise mit erheblicher Verzögerung, auf den Internetseiten der federführenden Ministerien oder in Absprache mit den Auftraggebern auf den Internetseiten der Auftragnehmer. Die Struktur der Evaluationsberichte leidet unter den vielfältigen Ansprüchen der verschiedenen Adressaten, welche die Berichte insgesamt sehr umfangreich und zumindest für die breite Öffentlichkeit schwer lesbar machen. Besonders der Anspruch, wissenschaftlichen Standards in der Dokumentation der einzelnen Phasen des Untersuchungsdesigns genügen zu wollen, macht Berichte sehr umfangreich. Darüber hinaus wurde teilweise von der vorgegebenen Gliederung der DG Agri abgewichen und in Ergänzung zur Betrachtung ganzer Förderkapitel (z. B. Agrarumweltmaßnahmen) eine intensive Betrachtung einzelner Maßnahmen bzw. Richtlinien (z. B. Ökolandbau, Grünlandextensivierung, Feuchtgrünlandschutz, Magerrasen) durchgeführt. In einigen Bundesländern wurden zusätzlich die umfangreichen Evaluationsberichte für die breite Öffentlichkeit zusammengefasst und als Broschüren veröffentlicht.

4.2 Qualität der durchgeführten Evaluationen

Seitens der Auftraggeber (v. a. der Bundesländer) gibt es keine systematische Qualitätskontrolle der Evaluationsberichte. Die Abnahme der Berichte erfolgt durch die zuständigen Fachreferate und als Gesamtbericht durch die Programmkoordination beziehungsweise Verwaltungsbehörde. Schlussfolgerungen und Empfehlungen sind zumeist Gegenstand von Workshops und werden intensiv mit den Beteiligten (v. a. aus den Verwaltungen) diskutiert. Ob ein Bericht abgenommen wird, hängt angesichts des Fehlens von klaren Kriterien von der subjektiven Einschätzung der Verwaltung ab. Die DeGEval-Standards finden bislang kaum Beachtung bei Fragen der Qualitätssicherung. Dies ist sicher auch darauf zurückzuführen, dass der Umgang mit Evaluation insgesamt auf Auftraggeberseite wenig professionell ist und ein deutlicher Kapazitätsaufbau erforderlich wäre. Auf der EU-Ebene gibt es in den Leitlinien festgelegte Kriterien zur Beurteilung der Qualität der dezentral erstellten Bewertungsberichte (EU-KOM 1999a: 28). Zuständig für die Beurteilung ist der Programmverantwortliche, der einen Beurteilungsbogen ausfüllt und eine Gesamtnote vergibt. Von dieser Beurteilung macht die EU-KOM ihre Mitfinanzierung an der Evaluation

abhängig⁷. Das Verfahren ist wenig transparent, weil keine Begründung für die Bewertung in den einzelnen Kriterien gegeben wird. Auch fehlen Hinweise für den Evaluator, wo Verbesserungsbedarf gesehen wird.

Trotz klarer Vorgaben in den diversen Leitfäden unterscheiden sich die Evaluationsberichte sehr stark vor allem hinsichtlich des Kriteriums ‚umfassende Analyse‘ (EU-KOM 1999a: 29): „Wurden die quantitativen und qualitativen Informationen angemessen und systematisch nach neusten Methoden analysiert, sodass die Bewertungsfragen stichhaltig beantwortet werden?“

Unterschiede zwischen den Evaluationsberichten bestehen vor allem hinsichtlich der eingesetzten Methoden und der verwendeten Daten. Wissenschaftliche Einrichtungen setzen einen vielfältigen Methodenmix ein (siehe die Evaluationsberichte von FAL et al. 2005; Matzdorf et al. 2005; Schramek et al. 2005), der sowohl umfangreiche statistische Analysen (ökonometrische Analysen, räumliche Analysen mit GIS, Vorher-Nachher-Vergleiche, Mit-Ohne-Vergleiche), Primärdatenerhebungen und Fallstudien beinhaltet. Aufbereitet und ausgewertet werden große Datenmengen. Nicht vorrangig im Bereich der Forschung arbeitende Einrichtungen, wie zum Beispiel Beratungsunternehmen, legen einen größeren Schwerpunkt auf den Evaluations- und Beratungsprozess, setzen weniger zeit- und personalaufwendige Analysemethoden ein und verarbeiten auch nicht selbst die in den Ländern vorhandenen, oft nur mit großem Aufwand erschließbaren Sekundärdaten. Diese Unterschiede konnten auch in einem internationalen Workshop zur Evaluation der Agrarinvestitionsförderung herausgearbeitet werden (Bergschmidt et al. 2006: 244).

Da das Kriterium ‚umfassende Analyse‘ aber nur ein Kriterium unter vielen zur Beurteilung von Evaluationen ist, wäre es sinnvoll, die Vielfalt von durchgeführten Evaluationen im Bereich der 2. Säule einer Meta-Evaluation zu unterziehen, um die Stärken und Schwächen der verschiedenen Herangehensweisen an Evaluation systematischer herauszuarbeiten.

4.3 Bewertung der Evaluationsvorgaben der DG Agri

Das Evaluationsraster der DG Agri stellt gegenüber den vorherigen Leitfäden eine deutliche Weiterentwicklung dar und hat in Deutschland die Qualität der Evaluationsberichte – mit einer allerdings erheblichen Variabilität – deutlich verbessert. In der praktischen Anwendung sind allerdings vielfältige Probleme aufgetreten (Forstner et al. 2002; 2004), von denen hier einige dargestellt werden sollen.

- Evaluation ist auf geeignete Daten angewiesen. Die vorhandenen Begleitsysteme der Förderperiode 2000 bis 2006 konnten die Daten nur unzureichend bereitstellen. Entweder wurden keine für eine Evaluation erforderlichen Daten erfasst oder die Daten lagen in Formaten vor, die nur mit einem hohen Aufwand verarbeitbar waren. Schwierig zu klären sind auch datenschutzrechtliche und datensicherheitstechnische Aspekte (Reiter/Roggendorf 2007).

⁷ Soweit bekannt gab es bislang nur einen Bericht, bei dem die EU-KOM eine Mitfinanzierung abgelehnt hat. Diese Ablehnung führte zu einer Neuvergabe der Evaluation an einen anderen Auftragnehmer.

- Zu evaluieren waren 2000 bis 2006 ausschließlich die mit EU-Mitteln kofinanzierten Maßnahmen. Parallel erfolgte in Deutschland auch noch eine rein national finanzierte Förderung mit oft gleichen Inhalten (siehe Abbildung 1). Den Beitrag einzelner Maßnahmen und Akteure in einer Evaluation zu isolieren wird dadurch schwierig, wenn nicht ganz unmöglich.
- Erfasst werden sollen die Nettowirkungen der Förderung. Die Ermittlung von Nettowirkungen ist methodisch sehr anspruchsvoll und erfordert beispielsweise im Umweltbereich die Anwendung aufwendiger Dekompositionsanalysen, um die verschiedenen Einflussfaktoren zu separieren. Größtenteils sind noch methodische Ansätze zur Erfassung von Nettowirkungen zu entwickeln. Eine intensive Zusammenarbeit mit Forschungseinrichtungen ist erforderlich, da zum Beispiel die Entwicklung von geeigneten Modellen zur Wirkungsabschätzung langfristig anzulegen ist. Im Strukturfondsbereich sind die methodischen Probleme zur Erfassung von Nettowirkungen ähnlich gelagert (Schwab/Toepel 2000; Toepel 2000). Hier kommen teilweise makroökonomische Modelle zum Einsatz, die für das breite Anwendungsgebiet der ländlichen Entwicklung bislang fehlen.

Auf die vielfältigen kritischen Anmerkungen zu den Evaluationsanforderungen 2000 bis 2006 hat die Evaluationsabteilung der DG Agri in der Form reagiert, dass sie die Verbindlichkeit des Rasters abgeschwächt hat. „Das bestehende Bewertungsraster mit den darin enthaltenen Kernfragen, Kriterien und Indikatoren werde seitens der EU-Kommission als flexibel interpretiert und die Evaluierungskonzeption werde in einem prozessualen Kontext gesehen“ (Forstner et al. 2004: 5). Aufgrund der vielfältigen Diskussionen wurde das Bewertungsraster für die neue Förderperiode 2007 bis 2013 überarbeitet, wobei im Grundsatz der methodische und indikatorgestützte Ansatz beibehalten wurde. Im Ansatz neu ist

- eine stärkere Verknüpfung der Begleitungs- und Bewertungssysteme,
- die Vorgabe, die gesamte nationale Förderung unabhängig von der EU-Kofinanzierung zu evaluieren,
- die Festschreibung einer kontinuierlichen Bewertung und
- die Einrichtung eines Europäischen Evaluierungsnetzwerkes als „Help desk“ und Austauschplattform.

Hinzu kommt eine rechtliche Verpflichtung zur Lieferung von quantifizierten Indikatoren (auf Output-, Ergebnis- und Wirkungsebene), die in der Durchführungsverordnung festgelegt wurden.

5 Von der Nutzung der Evaluation ländlicher Entwicklungspolitik

Die Frage der Nutzung der Ergebnisse der Evaluation ländlicher Entwicklungspolitik in Deutschland kann nicht unabhängig von ihrer institutionellen Verankerung im Mehrebenensystem diskutiert werden.

Ländliche Entwicklungspolitik in Deutschland ist das einzige Handlungsfeld der Agrarpolitik, in dem es eine systematische, an den Politikzyklus gekoppelte Evaluation gibt. Der Anstoß zur Evaluation geht in erster Linie von der EU-KOM aus. National veranlasst gibt es – von wenigen Ausnahmen abgesehen – bis heute keine systematische Evaluation. Hier beschränkt sich die seitens der Haushaltsordnung vorgegebene Erfolgskontrolle zumeist auf die bloße Erstellung von Vollzugsstatistiken. Tiefergehende Analysen werden höchstens ad hoc durchgeführt. Neuere Impulse kommen lediglich aus den Evaluationsaktivitäten der Modellvorhaben des Bundes, die stark formativ ausgerichtet sind. Im Gegensatz zur mangelnden Umsetzung der eigentlich auch aufgrund von nationalen Rechtsvorschriften durchzuführenden Evaluationen setzt die EU klare und verbindliche Anforderungen an die Evaluation der von ihr mitfinanzierten Programme und setzt auch ihren Vollzug durch Androhung finanzieller Sanktionen durch.

Evaluation hat somit aufgrund der EU-Vorgaben sowohl quantitativ als auch qualitativ eine deutliche Aufwertung erfahren. In welcher Form Ergebnisse von Evaluationen genutzt werden, ist je nach beteiligter Ebene unterschiedlich. Wesentliche Adressaten der Evaluation sind die EU-KOM, die Bundesländer und der Bund.

Die EU-KOM als Veranlasserin von Evaluation nutzt weniger die Ergebnisse der einzelnen Programmevaluationen, sondern gibt eigene Aufträge zur Synthese aller Evaluationen in Auftrag, die in die generelle Politikgestaltung einfließen. Die Nutzung der Evaluationsergebnisse erfolgt mit erheblichem Zeitverzug. Der summative Zweck der Evaluation steht im Vordergrund; formative Aspekte spielen keine Rolle.

Die Bundesländer sind aufgrund von rechtlichen Verpflichtungen Auftraggeber der Evaluation im Bereich der ländlichen Entwicklung. Diese externe Verpflichtung wird verwaltungsseitig unterschiedlich wahrgenommen. Zunächst handelt es sich pragmatisch betrachtet um ein notwendiges „Übel“ zur Einwerbung von Kofinanzierungsmitteln der EU. „Evaluation ist eine externe Verpflichtung (Teil der Spielregel, um Förderquellen zu erschließen) mit geringer innerer Emotionalität. [...] Es besteht bei allen Beteiligten ein relativ geringes Interesse sowohl am Prozess der Evaluation als auch am Ergebnis. Wichtig allein ist, sie regelkonform durchzuführen.“ (Heintel 2004). Dies ist insbesondere die Sichtweise der Programmkoordinatoren beziehungsweise Verwaltungsbehörden. Bei Fachreferaten stößt Evaluation auf unterschiedliche Resonanz, da sich bei ihnen fachliche und Budgetsicherungsinteressen mischen. Die Offenheit gegenüber Evaluation nimmt in dem Maße ab, wie nicht nur Empfehlungen zur Umsteuerung, sondern zur finanziellen Gewichtung von Maßnahmen innerhalb der Programme gegeben werden. Allerdings sind solche Aussagen zur Programmstruktur aufgrund fehlender Programmziele zumeist nur schwierig im Rahmen der Evaluation zu treffen, sodass sich die meisten Empfehlungen eher auf die Feinsteuerung innerhalb der Maßnahmen beziehen. Auf der Ebene der einzelnen Maßnahmen werden diesbezügliche Evaluationsempfehlungen, so beispielsweise in den Länderprogrammen 2007 bis 2013, aufgegriffen. Die Empfehlungen zur Gesamtstrategie und zur Gewichtung der einzelnen Programmschwerpunkte der Bewertungsberichte des Jahres 2005 hingegen sind in den neuen Programmen kaum berücksichtigt. Gründe dafür sind, dass die Struktur eines Programms zumeist im Rahmen der Haushaltsverhandlungen bestimmt wird, vom Ausgleich der Interessen der wichtigen Stakeholder (Bauernverbände, Umweltverbände, Städte- und Gemeindetag) determiniert ist und damit einer anderen Logik als der von Eva-

lationsergebnissen folgt. Die Rückkopplung der Evaluationsergebnisse, bezogen auf grundsätzliche Fragen der Ausrichtung eines Programms, in politische Entscheidungsprozesse ist somit schwierig. Zu diesem Ergebnis kommt auch Zucker (2006) in ihrem Vergleich der Evaluationsansätze von Deutschland und Großbritannien. Die Nutzung von Evaluationsergebnissen bleibt demnach auf „kosmetische“ Anpassungen beschränkt; die grundsätzliche Ausgestaltung des Politikbereichs bleibt davon unberührt. Allerdings sind die Gestaltungsmöglichkeiten der Bundesländer in Deutschland durch den von der EU gesetzten Rechtsrahmen auch sehr begrenzt.

Der Bund profitiert kaum von den bisherigen Evaluierungsaktivitäten, da die Berichte viel zu heterogen sind, als dass diese eine Gesamtschau der ländlichen Entwicklungspolitik in Deutschland erlauben. Die bisher realisierten bundesweiten Evaluierungsansätze waren auf einzelne Maßnahmen beschränkt. Für die Weiterentwicklung der Politik für ländliche Räume auf Bundesebene sind die bisherigen Evaluationen kaum einsetzbar, und werden daher auf Bundesebene von den Fachreferaten auch kaum wahrgenommen. Hierfür wären zukünftig geeignete institutionelle Rahmenbedingungen zu schaffen, die zumindest durch einen intensiven Austausch zwischen Evaluatoren ein Mindestmaß an Vereinheitlichung sicherstellen.

Die bisherigen Erfahrungen in der Durchführung von Evaluationen der ländlichen Entwicklung in Deutschland zeigen, dass die Ergebnisse von Evaluationen umso stärker akzeptiert werden, je enger Verwaltung und Programmverantwortliche in den Prozess eingebunden sind, verbunden mit entsprechenden Rückkopplungsschleifen von der Planung bis zur Ergebnisdarstellung. Soll die Evaluation von EU-kofinanzierten Politikbereichen zukünftig stärker in administratives Handeln von Bund und Ländern eingebaut werden, so ist diesen formativen Aspekten ein größerer Stellenwert einzuräumen. Die Anregung von Eser (2001: 328), „ob mit Blick auf die qualitativen Aspekte der Politik und die Berücksichtigung des Lernens von Organisationen eventuell auf Kosten der Legitimationsfunktion und der statistischen Vergleichbarkeit langfristig eine Neuausrichtung der Evaluation avisiert werden sollte“, ist aber kritisch zu werten. Eine stärkere Prozessorientierung sollte die summativen Erfordernisse einer Evaluation nicht ersetzen. Es geht vor allem darum, sowohl den Prozessnutzen wie auch den Ergebnissenutzen einer Evaluation durch ein Aufgreifen von relevanten Fragestellungen des Bundes und der Länder zu erhöhen, die Ergebnisse zielorientierter aufzubereiten und zu operationalisieren. Eine nutzerorientierte Ausgestaltung von Evaluationen ist eine Grundvoraussetzung, um die Verwendung von Ergebnissen zu verbessern.

Evaluation in der österreichischen Agrarpolitik

1 Kontext der Evaluationen

1.1 Probleme der österreichischen Land- und Forstwirtschaft

Von den 3,26 Mio. ha landwirtschaftlich und 3,20 Mio. ha forstwirtschaftlich genutzten Flächen Österreichs liegen 5,87 Mio. ha in benachteiligten Gebieten, von denen neunzig Prozent zum Berggebiet gehören. Diese Topographie und Bodennutzung führt dazu, dass im internationalen Vergleich ziemlich kleine Betriebe vorherrschen, von denen viele auch Forstwirtschaft betreiben. Sie bedingt auch einen relativ hohen Arbeitsbesatz der Betriebe infolge eines Mangels an alternativen Beschäftigungsmöglichkeiten in der zumutbaren Umgebung.

Daher liegt das Einkommen der meisten Erwerbstätigen in der Landwirtschaft weit unter jenem, das in anderen Sektoren erzielt wird. Für die land- und forstwirtschaftlichen Betriebe bedeutet das, dass sie mehrheitlich im Nebenerwerb geführt oder aufgegeben werden müssen. Die Zahl der Haupterwerbsbetriebe hat sich in den letzten 25 Jahren fast halbiert, aber auch die Zahl der Nebenerwerbsbetriebe ist um mehr als ein Drittel gesunken (Tabelle 1).

Tabelle 1: Entwicklung der land- und forstwirtschaftlichen Betriebe und des Arbeitseinsatzes in der Land- und Forstwirtschaft (LFW)

| | | 1980 | 1990 | 1999 ¹ | 2005 ¹ |
|--------------------------------------|----------------------|-------|-------|-------------------|-------------------|
| Betriebe (in 1000) | gesamt | 318,1 | 281,9 | 217,5 | 189,6 |
| | Haupterwerb | 133,8 | 106,5 | 80,2 | 74,5 |
| | Nebenerwerb | 173,9 | 166,2 | 129,5 | 106,8 |
| | juristische Personen | 10,4 | 9,2 | 7,8 | 8,3 |
| Betriebsgröße in ha | Haupterwerb | 27,0 | 30,6 | 36,5 | 40,3 |
| | Nebenerwerb | 8,7 | 10,2 | 13,6 | 15,8 |
| | juristische Personen | 255,2 | 289,9 | 363,3 | 350,6 |
| Arbeitskräfte (in JAE ²) | in 1000 | 342,1 | 262,1 | 198,4 | 150,9 |

¹ Höhere Erfassungsgrenze: 1 ha Landw. bzw. 3 ha Forstw. ² Jahresarbeitseinheiten

Quelle: http://www.statistik.at/fachbereich_landwirtschaft/tab5.shtml [11.6.2007]

Die Bewirtschaftung, Pflege und Besiedlung eines großen Teils der österreichischen Kulturlandschaft kann nur mit staatlicher Unterstützung erhalten bleiben. Andererseits trägt dieser Lebensraum zur Verbesserung der Qualität von Oberflächen- und Grundwasser, zum

Schutz vor Naturgefahren und zur Erhaltung der Biodiversität bei. Die dadurch gewonnene Lebensqualität wird von Einheimischen und Touristen gleichermaßen geschätzt.

Wie groß die Abhängigkeit der österreichischen Landwirtschaft von Subventionen ist, zeigt sich daran, dass das Faktoreinkommen (d. i. die Summe von Arbeitslohn, Bodenrente und Kapitalverzinsung) von 2,35 Mrd. € im Jahr 2005 mit Hilfe von 1,44 Mrd. € „sonstigen (nicht güterbezogenen) Subventionen“ zustande kommt. Weitere 0,20 Mrd. € stammen aus Gütersubventionen, die in den Produktionswerten der Produkte zu Herstellungspreisen enthalten sind. Somit beruht die Entlohnung der in der österreichischen Landwirtschaft eingesetzten quasi-fixen Produktionsfaktoren zu etwa siebzig Prozent auf staatlichen Förderungen. Dennoch schreitet der Abbau von Arbeitskräften voran, denn die Arbeitsproduktivität liegt noch unter dem Durchschnitt der EU-25.

1.2 Schwerpunkte der österreichischen Agrarpolitik

Die Grundlagen der österreichischen Agrarpolitik stehen im österreichischen Landwirtschaftsgesetz 1992 (BMLFUW 2006: 312-315), im Marktordnungsgesetz 1985 (Bundesgesetzblatt Nr. 210/1985, 108/2001, 18/2006) und im AMA-Gesetz 1992 (BGBl. Nr. 376/1992, 108/2001). Die wesentlichen für die Land- und Forstwirtschaft relevanten Gesetze und Verordnungen sind in BMLFUW (2006: 300-311) aufgelistet und im Rechtsinformationssystem des Bundeskanzleramtes abrufbar (<http://www.ris.bka.gv.at/>).

In Österreich hat der Bereich Ländliche Entwicklung innerhalb der Agrarförderung eine wesentlich höhere Bedeutung als im Durchschnitt der EU. Das beruht einerseits auf einem relativ großzügigen Agrarumweltprogramm, an dem in Österreich viele Landwirte und Landwirtinnen freiwillig teilnehmen und sich verpflichten, gegen Entgelt entsprechende umweltschonende und -verbessernde Leistungen zu erbringen. Andererseits befindet sich in Österreich ein großer Teil der land- und forstwirtschaftlichen Betriebe in benachteiligten Gebieten, wo Bewirtschaftungserschwernisse teilweise ausgeglichen werden. Infolgedessen hat Österreich das fünftgrößte ländliche Entwicklungsprogramm der 15 alten EU-Mitgliedstaaten.

Im Jahr 2005 wurden in Österreich circa 2,4 Mrd. € an Agrarförderungen ausgezahlt. Fast 2,1 Mrd. € (85 %) davon wurden unter Beteiligung der EU ausgegeben, die 1,4 Mrd. € bereitstellte. Gliedert man die Agrarförderungen nach den Säulen der Gemeinsamen Agrarpolitik (GAP), dann zeigt sich für Österreich eine ungewöhnlich große Bedeutung des Ländlichen Entwicklungsprogramms (LEP), in dessen Rahmen 45 Prozent der Agrarförderungen abgewickelt werden. Das LEP stellt die so genannte zweite Säule der GAP dar und wird normalerweise durch EU, Bund und Länder im Verhältnis 50:30:20 finanziert. Weitere 35 Prozent der Agrarförderungen werden zu hundert Prozent aus dem EU-Budget bestritten; dazu gehören Betriebsprämien, Tierprämien, andere Produktprämien, Beihilfen für Verarbeitungsbetriebe, Ausfuhrerstattungen und Lagerhaltungskosten (Tabelle 2).

Tabelle 2: Ausgaben im Bereich der Agrarpolitik in Österreich nach Maßnahmen und Finanzierung im Jahr 2005 (in Mio. €)

| | Gesamt | EU | Bund | Land | Anteile in % |
|---------------------------------|--------------|--------------|------------|------------|--------------|
| Ländliches Entwicklungsprogramm | 1.071 | 481 | 351 | 239 | 45 |
| ÖPUL* | 654 | 323 | 198 | 133 | 27 |
| Ausgleichszulage | 275 | 86 | 112 | 77 | 12 |
| Übriges LEP | 142 | 72 | 41 | 29 | 6 |
| Betriebsprämie | 492 | 492 | - | - | 20 |
| Marktordnungsausgaben | 368 | 368 | - | - | 15 |
| Andere kofinanzierte Maßnahmen | 115 | 87 | 16 | 12 | 5 |
| nationale Maßnahmen | 357 | - | 107 | 250 | 15 |
| Gesamt | 2.403 | 1.428 | 474 | 501 | 100 |

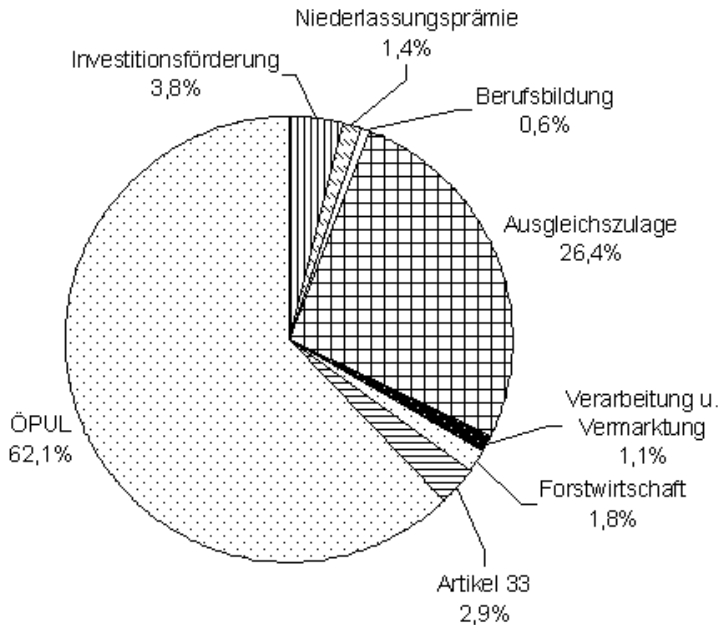
* Österreichisches Programm für eine umweltgerechte und extensive Landwirtschaft

Quelle: Bundesministerium für Land- und Forstwirtschaft, Umwelt und Wasserwirtschaft BMLFUW (2006); eigene Berechnungen.

Das LEP 2000-2006 wurde in ganz Österreich mit Ausnahme des Burgenlandes¹ durchgeführt. Sein Umfang hängt auch von den Teilnahmebedingungen ab, die Österreich und die Europäische Kommission gemeinsam festgelegt haben. In den Jahren 2000 bis 2004 entfielen ca. zwei Drittel der Ausgaben des LEP auf das Österreichische Programm für eine umweltgerechte und extensive Landwirtschaft (ÖPUL), an dem 2004 78,3 % der Betriebe mit landwirtschaftlichen Flächen und 88,8 % der landwirtschaftlichen Fläche teilnahmen. Ein weiteres Viertel des LEP entfiel auf die Ausgleichszulage in benachteiligten Gebieten und Berggebieten, wo sich 71,9 % der land- und forstwirtschaftlichen Betriebe und 79,1 % der Gesamtfläche Österreichs befinden. Weitere beachtliche Mittel fließen in die Investitionsförderung und in die verschiedenen Artikel 33 Maßnahmen der VO (EG) Nr. 1257/1999 (Abbildung 1).

¹ Burgenland war in dieser Periode Ziel-1-Gebiet mit eigenem Förderprogramm.

Abbildung 1: Agrarförderungen im Rahmen des Ländlichen Entwicklungsprogramms in Österreich, 2000-2004



Quelle: BMLFUW 2005.

2 Wer macht und wer evaluiert die Agrarpolitik in Österreich?

Die wichtigsten Akteure der österreichischen Agrarpolitik sind die politisch gewählten Vertreter der Landwirtschaftskammern, aus deren Reihen oder aus einer Landesregierung üblicherweise die Agrarminister hervorgehen. In allen bisherigen Koalitionsregierungen mit Beteiligung der Österreichischen Volkspartei (ÖVP) seit 1945 wurde der Agrarminister von der ÖVP gestellt, die bei den Kammerwahlen von etwa achtzig Prozent der abgegebenen Stimmen (ca. 370.000) gewählt wurde.

Dafür, dass auch andere Meinungen als die der Interessenvertretung der Bauern gehört und teilweise auch berücksichtigt werden, sorgt die so genannte Sozialpartnerschaft. Sie besteht darin, dass die Landwirtschaftskammer, Arbeiterkammer, Wirtschaftskammer und der Gewerkschaftsbund in bestimmten Angelegenheiten ein gesetzlich verankertes Mitspracherecht haben. Zum Beispiel legt das Landwirtschaftsgesetz fest, dass eine Kommission bestehend aus je einem Vertreter der Parlamentsparteien und der Sozialpartner Empfehlungen zur Agrarpolitik abgibt und an der Erarbeitung des jährlichen Grünen Berichts mitwirkt. Die Beratungen darüber erfolgen unter Teilnahme von Fachexperten, die der Landwirtschaftsminister einlädt. Da Empfehlungen nur mit einer Zweidrittelmehrheit zustande kommen, ergibt sich manchmal die Notwendigkeit, zur Klärung einer Streitfrage eine Studie in Auftrag zu geben.

Da die wirklich einschneidenden Entscheidungen bezüglich der Agrarpolitik in Brüssel getroffen werden und die Initiativen dazu von der Europäischen Kommission (und dem Rat) ausgehen, geschieht die Meinungsbildung dazu kurzfristig im Kreis der fachlich zuständigen Mitarbeiter des Lebensministeriums und der Landwirtschaftskammer Österreich sowie des Ministers und seiner politischen Vertrauenspersonen. Zu längerfristig absehbaren politischen Reformvorhaben werden von den fachlich zuständigen Mitarbeitern des Lebensministeriums und den von diesem (überwiegend) finanzierten Forschungsinstituten Forschungsprojekte und wissenschaftliche Arbeiten vorgeschlagen und, soweit möglich, auch durchgeführt. Externe Forschungsinstitute können weitere Forschungsprojekte einreichen; das dafür zur Verfügung stehende Budget ist bescheiden.

Ein starker Anreiz zur Durchführung agrarpolitisch relevanter Forschungsprojekte und Evaluationen geht von der EU aus, die nicht nur eine umfangreiche und detaillierte Evaluierung der ländlichen Entwicklungsprogramme und ihrer Maßnahmen vorschreibt, sondern seit 2007 auch die Möglichkeit schafft, sie aus diesem Programm zu finanzieren. Das erscheint durchaus notwendig, denn die Bereitschaft, politische Maßnahmen, die von außen (mit-)finanziert werden, im Land selbst in Frage zu stellen, ist gering. Denn auf nationaler Ebene befasst sich die Evaluation mit für die AgrarpolitikerInnen potenziell unangenehmen Fragen: mit der Ausgestaltung des Maßnahmenpakets, die sie zu verantworten haben, mit der Schätzung seiner Wirkungen, die zweifelhaft sind, und mit dem Hintergedanken, Einsparungsmöglichkeiten zu finden. Das Interesse an der Evaluation liegt daher vor allem auf Seiten des Finanzministers und der von ihm vertretenen Steuerzahler sowie der Europäischen Kommission.

Bisherige Aufträge für Evaluationen der Agrarpolitik auf subnationaler Ebene bestätigen den Trend zur Suche nach Einsparungsmöglichkeiten. Unter diesem Aspekt wurde zum Beispiel anlässlich der Bildung einer Koalitionsregierung für das Bundesland Salzburg im Jahr 2004 ein Auftrag zur Evaluation der in diesem Bundesland durchgeführten agrarpolitischen Maßnahmen erteilt (Salzburger Landesregierung 2004). Ein ähnlicher Auftrag erging vom Landesrechnungshof des Bundeslandes Oberösterreich anlässlich der Prüfung der Agrarabteilung der Landesregierung (Oberösterreichischer Landesrechnungshof 2006). Beiden Fällen gemeinsam war, dass sie in kürzester Zeit unter Beachtung politischer Sensitivitäten durchgeführt werden mussten, wobei auf vorhandenes Fachwissen zurückgegriffen werden sollte. Dieses lag aus Evaluationen des LEP in den damit befassten Instituten vor und konnte unter der Voraussetzung, dass die betreffenden EvaluatorenInnen vor politischer Einflussnahme geschützt sind, abgerufen werden. Sachkompetenz für wirtschaftliche Evaluationen der Agrarpolitik können in Österreich primär folgende Institute in die Waagschale werfen:

- Bundesanstalt für Agrarwirtschaft (AWI)
- Bundesanstalt für Bergbauernfragen (BABF)
- Institut für Nachhaltige Wirtschaftsentwicklung der Universität für Bodenkultur (NWE)
- Österreichisches Institut für Wirtschaftsforschung (WIFO)

Nachdem ein großer Teil der Förderungen für die österreichische Land- und Forstwirtschaft auf das Umweltprogramm entfällt, dessen Wirkungen primär ökologischer Natur sind, kommen als EvaluatorenInnen in diesem Bereich vor allem produktions- und umwelttechnische Institute zum Einsatz. Bei der Ex-ante Evaluierung des LEP 2007-2013 waren dies:²

- Umweltbundesamt (UBA)
- Österreichische Agentur für Gesundheit und Ernährungssicherheit (AGES)
- Bundesforschungs- und Ausbildungszentrum für Wald, Naturgefahren und Landschaft (BFW)
- HBLFA Raumberg-Gumpenstein (RG)

3 Was und wie wird evaluiert?

Eine Unterscheidung zwischen Forschungsprojekten und Evaluationen fällt mitunter schwer. Der Begriff Evaluierung wurde in die österreichische Agrarpolitik durch EU-Verordnungen eingeführt und hat sich daher erst ab dem Beitritt (1995) allmählich verbreitet. Studien, die bis dahin zur Agrarpolitik durchgeführt wurden, werden meistens als Forschungsberichte bezeichnet, auch wenn sie eine Evaluation zum Inhalt haben. Im Folgenden beziehe ich auch jüngere Forschungsprojekte, die sich mit Wirkungen der Agrarpolitik befassen, in die Darstellung ein, ohne Anspruch auf Vollständigkeit zu erheben. Außerdem sind folgende Studien zu erwähnen, die sich nicht in den folgenden Rahmen fügen:

- Biosprit in Österreich. Eine volkswirtschaftliche Analyse (Hofreither et al. 1987)
- Evaluierung der Politik zum Schutz vor Naturgefahren: Grundlagen der Evaluation und Forschungsansatz (Weiß 1997).
- Evaluierung der ökonomischen Effekte von Nationalparks (Kletzan 1999)
- Erneuerbare Energieträger in Österreichs Wirtschaft: Volkswirtschaftliche Evaluierung am Beispiel der Biomasse (Pichl et al. 1999)

3.1 Marktordnungsmaßnahmen

Der Beitritt Österreichs zur EU im Jahr 1995 führte zu einer drastischen Änderung der Kompetenzen für Marktordnungen, für die bis dahin Österreich allein zuständig war. Analysen bestehender und möglicher Änderungen von Marktordnungsregelungen waren vor dem Beitritt an der Tagesordnung. Sie beherrschten die agrarpolitische Diskussion besonders zu Beginn der 1990er Jahre, als die Verhandlungen zur Uruguay-Runde des GATT und Bei-

² Weitere Studien im Zusammenhang mit der Evaluation des Ländlichen Entwicklungsprogramms 2000-2006 wurden u. a. von folgenden Organisationen (mit) durchgeführt: Bundesamt für Wasserwirtschaft (BAW), Institut für Kulturtechnik und Bodenwasserhaushalt (IKT), Birdlife Österreich, Institut für Ökologie und Umweltplanung, Institut für Ökologie und Naturschutz der Universität Wien, Institut für Alpine Naturgefahren und Forstliches Ingenieurwesen der Universität für Bodenkultur, Österreichisches Kuratorium für Landtechnik und Landentwicklung.

trittsverhandlungen zur EU gleichzeitig geführt wurden. Beides begründete einen hohen Änderungsdruck in die gleiche Richtung, weil das agrarpolitische System Österreichs wesentlich protektionistischer war als jenes der EU und Überschüsse bei Getreide, Milch und Rindfleisch produzierte, die nur mit hohen Exportstützungen absetzbar waren (Hofreither 2002; Österreichischer Agrarverlag 1997; Ortner 2004).

Die OECD nimmt seit 1986 (jetzt nur mehr in zweijährigen Abständen) eine Beurteilung der agrarpolitischen Maßnahmen ihrer Mitgliedstaaten vor (OECD 2006). Sie schätzt unter anderem die Wirkungen der Marktordnungen der wichtigsten Agrarproduktmärkte auf die Einnahmen der Produzenten und die Ausgaben der Konsumenten und Steuerzahler mit Hilfe der so genannten Produzenten- und Konsumentenstützungsäquivalente. Diese enthalten die so genannte Marktpreisstützung (Preisabstand zum Weltmarktpreis) und die Produktprämien, die zur Verzerrung der Weltmärkte und des Wettbewerbs beitragen, aber auch alle anderen Subventionen und Maßnahmen, für die das zum Teil in wesentlich geringerem Maße oder überhaupt nicht zutrifft. Daher steht die Frage der Klassifizierung und Einordnung der Subventionen regelmäßig zur Diskussion, unter anderem auch bei der Doha-Runde der WTO, deren Fortschritt auch von einer Abgrenzung der handels- und weniger handelsverzerrenden Maßnahmen abhängt. Die OECD führt zahlreiche Studien zu Fragen der Agrarpolitik im Allgemeinen und über einzelne Mitgliedstaaten durch; sie werden aber teilweise nicht veröffentlicht und sind in diesem Fall nur den zuständigen Regierungsstellen zugänglich.

3.2 Methoden und Modelle

Die Zuständigkeit für die Gemeinsamen Marktordnungen der EU liegt in erster Linie bei der Europäischen Kommission und dem Rat. Sobald die Kommission einen Reformbedarf erkennt, startet sie einen Diskussionsprozess und gibt allenfalls Studien in Auftrag, um die Auswirkungen von Reformvorschlägen abzuschätzen. Das geschieht gewöhnlich mit Hilfe von Sektormodellen, zu deren Entwicklung und Weiterentwicklung Österreich derzeit, soviel ich weiß, nur am Rande beiträgt. Eine Ausnahme ist vielleicht das AGMEMOD, ein dynamisches Modell der Märkte für Agrarprodukte, differenziert nach Mitgliedstaaten (Teagasc 2006).

Auf nationaler Ebene werden Modelle für Agrarbetriebe, Buchführungsbetriebe und Agrarproduktmärkte an der Universität für Bodenkultur und der Bundesanstalt für Agrarwirtschaft entwickelt. Sie dienen in erster Linie zu Forschungszwecken und werden im Zusammenhang mit Forschungsprojekten des Lebensministeriums oder anderer Auftraggeber (z.B. dem Rechnungshof Oberösterreichs) eingesetzt. Zwei aktuelle Projekte betreffen die Auswirkungen der GAP-Reform (Tribl 2007) und Optionen der Milchquotenregelung ab 2014 (Pfungstner 2007b).

Die meisten auf europäischer Ebene zur Diskussion stehenden Agrarmarktordnungsmaßnahmen werden in Österreich mit Hilfe partieller Marktmodelle analysiert, die entsprechend dem Anlassfall entwickelt oder adaptiert, aktualisiert und verwendet werden (z.B. Ortner 1998). Parallel dazu werden häufig betriebswirtschaftliche Untersuchungen ange stellt, um den betroffenen Bauern mögliche Konsequenzen vor Augen zu führen und An-

passungsstrategien anzubieten (z.B. Kirner 2004). Beispiele dafür wären die Reform der gemeinsamen Marktordnung für Bananen (Hambrusch/Ortner 2005) und die Reform des Zuckermarktes (Kirner 2006). Auftraggeber solcher Studien sind meistens die betreffenden Abteilungen des Lebensministeriums. Ähnlich verhält es sich mit einer Begleitstudie zur „Exportoffensive“ des Lebensministeriums mit der Wirtschaftskammer Österreichs (WKÖ) und der Agrarmarkt Austria (Hambrusch 2006, Pfingstner 2007a).

3.3 Evaluationen im Rahmen des LEP 2000-2006

Umfang und Wirkungen der Maßnahmen des Ländlichen Entwicklungsprogramms (LEP) 2000-2006 müssen auf Basis der VO (EG) Nr. 1257/1999 und den Leitlinien der Europäischen Kommission evaluiert werden (Europäische Kommission 2007). Externe Evaluationen, die für Vorgängerprogramme in Auftrag gegeben worden waren, lieferten innerhalb der vorgegebenen Frist wertvolle Hinweise, hatten aber den Nachteil, dass es nach Ablauf des Vertrages oft noch einen dringenden Informationsbedarf gab, der nur mit Mühe befriedigt werden konnte. Daher und aus Kostengründen zieht es das Lebensministerium vor, die Evaluierung des LEP weitgehend an die eigenen Forschungsanstalten zu vergeben. Diese sind mit den Problemen und Maßnahmen vertraut und in der Lage, zusätzlichen Informationsbedarf in kurzer Frist, unbürokratisch und zufrieden stellend zu erfüllen. Soweit sie bestimmte Aufgaben der Evaluation nicht selbst erledigen können, unterstützen sie das Lebensministerium bei der Aufgabenstellung, Auswahl und der Koordination von damit im Zusammenhang stehenden nach außen zu vergebenden Forschungsprojekten.

Der beim Lebensministerium eingerichtete ÖPUL-Beirat, in dem Programmverantwortliche, Umweltorganisationen, EvaluatorenInnen und ExpertInnen vertreten sind, sorgt dafür, dass das Agrarumweltprogramm von verschiedenen Seiten beleuchtet wird. Der Beirat stärkt das Interesse an den Zielen des ÖPUL, seinen Wirkungen und der diesbezüglichen Forschung. Externe Studien zur Evaluierung des LEP wurden überwiegend zu ÖPUL-Maßnahmen in Auftrag gegeben (Tabelle 3). Dabei ging es vorwiegend darum, festzustellen, wie die Wirkungen bestimmter (Teil-) Maßnahmen optimiert werden können.

Tabelle 3: Beauftragte Studien zur Erstellung der Halbzeit- und Update-Evaluierungsberichte des Ländlichen Entwicklungsprogramms (2003/2005)

| Thema | Zahl der Studien |
|----------------------------|------------------|
| ÖPUL | |
| Kosten und Akzeptanz | 5 |
| Boden, Wasser, Erosion | 9 |
| Biodiversität | 13 |
| Landnutzung und Landschaft | 6 |
| Forst | 3 |
| Anderes | 3 |

Quelle: BMLFUW.

Die Berichte des BMLFUW zur Halbzeit- und Update-Evaluierung wurden veröffentlicht (BMLFUW 2005); darüber hinaus veröffentlichen die damit befassten Bundesanstalten teilweise ausführlichere Berichte über die Evaluation einzelner Maßnahmen (Bogner 2003; Hovorka 2004; 2005; Janetschek 2001; Oedl-Wieser 2004; Ortner 2005, Pfusterschmid 2004; Sinabell 2006; Sinabell et al. 2006; Tamme 2004a; 2004b). 2007/08 wird eine Ex-Post-Evaluation für die gesamte Programmperiode stattfinden; da sie naturgemäß keine Bedeutung für die aktuelle Programmperiode mehr hat, wird sie nach dem Vorbild der Update-Evaluierung durchgeführt werden.

3.4 Maßnahmen und Evaluationen des LEP 2007-2013

Das ländliche Entwicklungsprogramm für die aktuelle Periode sieht Maßnahmen in vier Schwerpunkten vor. Auf den ersten Schwerpunkt, die Verbesserung der Wettbewerbsfähigkeit, werden in Österreich circa 14 Prozent der Förderungen entfallen. Darin enthalten sind in erster Linie die Modernisierung landwirtschaftlicher Betriebe und die Erhöhung der Wertschöpfung landwirtschaftlicher Erzeugnisse (Tabelle 4).

Für den Schwerpunkt 2, die Verbesserung der Umwelt und der Landschaft, sind circa 72 Prozent der Mittel vorgesehen. Damit verbunden ist eine deutliche finanzielle Einschränkung des ÖPUL, während die Ausgleichszulage für benachteiligte Gebiete im bisherigen Umfang weitergeführt werden soll. Neu sind Ausgleichszahlungen für Maßnahmen in Natura 2000-Gebieten, die sich auf 1,2 Mio. ha ausdehnen.

Im Schwerpunkt 3 werden Maßnahmen angeboten, die in größerem Umfang als bisher durch Lokale Aktionsgruppen im Schwerpunkt 4 (LEADER) umgesetzt werden sollen. Dazu gehören Maßnahmen zur Diversifizierung der ländlichen Wirtschaft, für Dorf- und regionale Entwicklungsprojekte, zur Erhaltung des ländlichen Erbes, zum Schutz vor Naturgefahren sowie zur Verkehrserschließung und Investitionen in erneuerbare Energiequellen.

Tabelle 4: Maßnahmen, Förderung und Evaluatoren im LEP 2007-2013

| Nr. | Maßnahme | Förderung in Mio. € | Evaluatoren |
|----------|--|---------------------|-------------------------------------|
| 111, 331 | Berufsbildung und Information | 85 | AWI |
| 112 | Niederlassung von Junglandwirten | 637 | |
| 121 | Modernisierung landw. Betriebe | | |
| 123 | Erhöhung der Wertschöpfung lfw. Erzeugnisse | 134 | |
| 124 | Neue Produkte und Verfahren | 155 | |
| 132 | Lebensmittelqualitätsregelungen | | |
| 133 | Information und Absatzförderung | | |
| 211 | Ausgleichszahlungen in Berggebieten | 1.932 | BABF |
| 212 | Ausgleichszahlungen in anderen benacht. Gebieten | 3.610 | UBA+, AGES, BMLFUW+, AWI+, RG |
| 213 | Natura 2000 und Richtlinie 2000/60/EG | | |
| 214 | Agrarumweltmaßnahmen | | |
| 215 | Tierschutzmaßnahmen | | |
| 311 | Diversifizierung | 43 | AWI |
| 312 | Kleinstunternehmen | | |
| 313 | Fremdenverkehr | | |
| 321 | Verkehrerschließung | 167 | BABF |
| | Erneuerbare Energiequellen | 333 | AWI+ |
| 322 | Dorferneuerung, Agenda 21, kommunale Dienstl. | | |
| 323 | Ländliches Erbe, Umwelt | | |
| 341 | Kompetenzentwicklung | 417 | BABF+ |
| 41 | LEADER | | |
| | Forstwirtschaftliche Maßnahmen | 71* | BFW |

+ plus Forschungsaufträge

* Verbesserung des wirtschaftlichen Wertes der Wälder, Erstaufforstung landwirtschaftlicher Flächen, Zahlungen für Waldumweltmaßnahmen; mit den die Forstwirtschaft betreffenden Teile der anderen Maßnahmen: 175 Mio. €.

Die Ex-Ante-Evaluierung des LEP 2007-2013 war in dessen Programmierung eingebunden. Die Ergebnisse dieser Evaluation müssen im LEP enthalten und ihre Empfehlungen im LEP berücksichtigt sein. Die dadurch bedingten Abgrenzungsprobleme zwischen Programmstellung und Evaluierung wurden durch eine entsprechende zeitliche Abfolge gelöst (Tabelle 5).

Tabelle 5: Entwicklung des Ländlichen Entwicklungsprogramms 2007-2013

| | |
|---|----------------|
| Programm – Vorbereitung – Arbeitsgruppenveranstaltungen | ab Herbst 2004 |
| Strategie und Programm – Entwurf | Februar 2006 |
| Öffentliche Diskussion, Internetplattform | Februar 2006 |
| Ex-ante Evaluierung, Strategische Umweltprüfung | 1. Hälfte 2006 |
| Programm – Überarbeitung | 2. Hälfte 2006 |
| Programm – Einreichung bei EK | Februar 2007 |
| Laufende Evaluierung – Beginn | Februar 2007 |
| Programm – Genehmigung durch Europäische Kommission | 2. Hälfte 2007 |

Die Ex-Ante-Evaluierung des LEP, dessen erster Entwurf vom Februar 2006 eine detaillierte Maßnahmenbeschreibung mit den jeweiligen Zielen, Förderungsgegenständen, Förderungswerbern und Förderungsbedingungen enthielt, wurde von jenen EvaluatorInnen durchgeführt, welche die jeweiligen oder ähnliche Maßnahmen des Vorgängerprogramms evaluiert hatten. Dabei mussten die Vorgaben der Europäischen Kommission – Indikatoren und Wirkungsketten zwischen diesen für die jeweiligen Maßnahmen – eingehalten werden.

Die Evaluation wird vom Referat „Evaluierung“ des Lebensministeriums organisatorisch und dem Autor dieser Studie (AWI) fachlich koordiniert. An der Erstellung der Strategischen Umweltprüfung (SUP), die als Teil der Ex-Ante-Evaluierung durchgeführt wurde, waren auch andere Abteilungen des BMLFUW sowie Joanneum Research, UBA und BFW beteiligt (BMLFUW 2007).

Im Februar 2007 wurde die laufende Evaluierung in Angriff genommen. Im Zusammenhang mit dem neuen ÖPUL wurden einige Forschungsprojekte zu Fragen der Biodiversität, Natura 2000, Grundwasser, Oberflächenwasser, Bodenschutz, Erosion und CO₂ konkretisiert. Für die anderen Maßnahmen wurden Evaluierungskonzepte entwickelt und zur Diskussion gestellt. Auf dieser Grundlage werden ein Arbeitsplan und ein Bericht erstellt, der das Monitoring- und Evaluierungssystem darlegt und 2008 an die Europäische Kommission zu liefern ist.

4 Zusammenfassung

Die Beurteilung der österreichischen Agrarpolitik im Sinn der Suche nach besseren Wirkungen oder ähnlich großen Wirkungen zu geringeren Kosten hat eine Tradition, deren Beginn man vielleicht mit der Gründung der Bundesanstalt für Agrarwirtschaft im Jahr 1960 gleichsetzen kann. Die diesbezüglich durchgeführten Studien sind allerdings nicht unter der Bezeichnung Evaluation bekannt. Dieser Begriff tauchte erstmals durch die jährlichen „Monitoring and Evaluation“ Reports der OECD auf und wurde erst durch die Strukturprogramme der EU auch in der Agrarpolitik bekannt. Evaluierungstätigkeiten in der Agrarpolitik, die auch so genannt und der Öffentlichkeit bekannt werden, erstreckten sich seit dem EU-Beitritt Österreichs fast ausschließlich auf die Ländlichen Entwicklungsprogramme und die damit geförderten Maßnahmen. Zwei Ausnahmen von dieser Regel sind hervorzuheben:

Die Evaluierung der Agrarförderungen in Salzburg (2004) und die Initiativprüfung der Agrarabteilung der Landesregierung durch den Landesrechnungshof Oberösterreichs (2007).

Es scheint, dass die Grundsätze der Sparsamkeit, Wirtschaftlichkeit und Zweckmäßigkeit bei der Verwendung öffentlicher Mittel, auf die viele Bundesgesetze hinweisen, zunehmend so verstanden werden, dass ihre Einhaltung angestrebt und überprüft werden sollte. Die Bereitschaft, diese Überprüfung mit großem ideellem und finanziellem Engagement auszustatten und die dafür benötigten Schätzungen und Wertungen wissenschaftlich ausreichend abzusichern, ist aber gering. Während Evaluationen von Maßnahmen im Bereich der Landwirtschaft auf der Ebene der Verwaltung durchaus als nützlich anerkannt werden und kritische Hinweise gewöhnlich zu Korrekturen einer Maßnahme führen, werden derartige Hinweise in veröffentlichten Berichten vor allem dann nicht gern gesehen, wenn sie politische Auseinandersetzungen nach sich ziehen könnten. Kontroverse Probleme können aber in wissenschaftlichen Fachpublikationen behandelt werden.

Schweizer Agrarpolitik: Stand der Evaluation

1 Zuständigkeiten, Entscheidungsfindung und Akteure in der Schweizer Agrarpolitik

Die Agrarpolitik ist in der Schweiz eine Bundesaufgabe. Die Eidgenossenschaft hat eine umfassende Rechtsetzungskompetenz und nimmt diese auch wahr. Die Kantone sind grundsätzlich für den Vollzug zuständig und können ergänzende Massnahmen treffen. Dasselbe gilt für die Gemeinden. Der agrarpolitische Rahmen bildet die Bundesverfassung (BV, vgl. Abschnitt 2.2 zum Landwirtschaftsartikel der BV). Das Volk kann mit dem Instrument der Volksinitiative direkt Verfassungsänderungen vorschlagen. Über alle Vorschläge für Verfassungsänderungen entscheiden in letzter Instanz Volk und Stände (Kantone).

Der Bundesrat konsultiert die betroffenen Akteure in der Phase der Erarbeitung von Vorschlägen für Gesetzesänderungen und Ausführungsbestimmungen im Rahmen des so genannten Vernehmlassungsverfahrens. Zu den wichtigsten Teilnehmern im Vernehmlassungsverfahren im Bereich Landwirtschaft gehören die Kantone, die politischen Parteien und die Verbände (bäuerliche Organisationen, Produktionsmittelbranche, Wirtschaftsverbände und Nahrungsmittelindustrie, Konsumenten-, Umwelt-, Natur- und Tierschutzorganisationen).

Die agrarpolitischen Entscheide des Bundes werden vom Bundesamt für Landwirtschaft (BLW) vorbereitet und in Zusammenarbeit mit den Kantonen vollzogen. Das BLW ist ein Amt des Eidgenössischen Volkswirtschaftsdepartements. Ihm unterstellt sind die drei eidgenössischen landwirtschaftlichen Forschungsanstalten.

2 Die Schweizer Agrarpolitik im Umbruch

2.1 Vier Reformetappen seit Anfang der neunziger Jahre

Die Schweizer Agrarpolitik wird seit fünfzehn Jahren konsequent umgebaut. Eine Reihe von Herausforderungen interner und externer Natur machten Anfang der 1990er Jahre eine Anpassung der damaligen staatlichen Rahmenbedingungen unumgänglich. Zu den wichtigsten gehörten drohende Marktanteilsverluste der Schweizer Landwirtschaft infolge hoher Preise, steigende durch den Staat getragene Verwertungskosten, eine hohe staatliche Reglungsdichte, ökologische Defizite sowie eine zu erwartende Lockerung des Grenzschutzes infolge sich abzeichnender GATT-Verpflichtungen (Bundesrat 1996: 14).

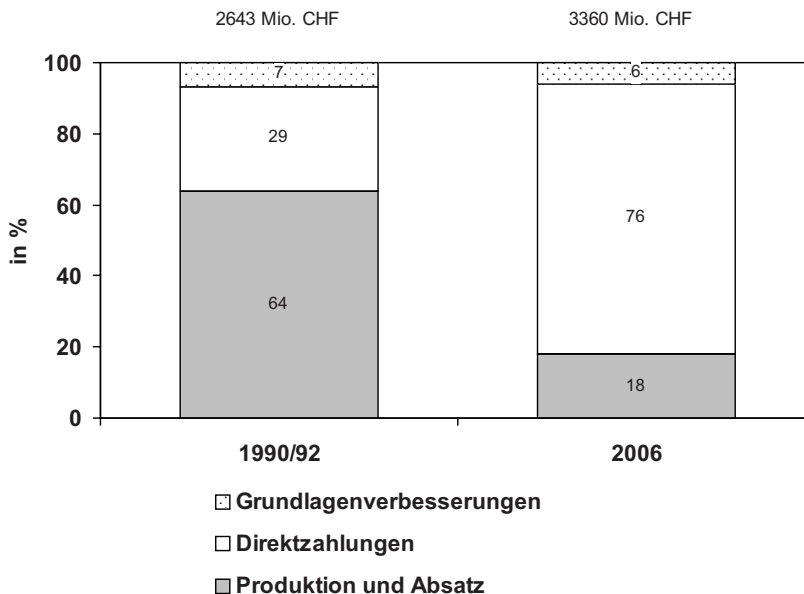
Als Antwort auf die Herausforderungen wurden in einer ersten Reformetappe 1993 die staatlich festgesetzten Produzentenpreise gesenkt und produktunabhängige Direktzahlungen verbunden mit ökologischen Anreizen eingeführt. Zudem wurde der Grenzschutz WTO-konform umgebaut (Überführung sämtlicher Grenzschutzmassnahmen in Zölle) (Bundesrat 1996: 22-23).

In einer zweiten Etappe (Agrarpolitik 2002) wurde der Umbau weitergeführt, indem sämtliche Preis- und Abnahmegarantien abgeschafft und die Marktstützung um einen Drittel reduziert wurde. Gleichzeitig wurde die Ausrichtung der Direktzahlungen an einen ökologischen Leistungsnachweis (ÖLN) geknüpft (konzeptionell mit der cross compliance der Gemeinsamen Agrarpolitik der EU (Nitsch/Osterburg 2005) vergleichbar). Die Umsetzung erfolgte ab 1999 (Bundesrat 1996).

Kernpunkt der dritten Etappe (Agrarpolitik 2007) war die gesetzliche Verankerung der schrittweisen Abschaffung der Milchkontingentierung (Milchquoten) bis 2009. Die Agrarpolitik 2007 trat 2005 in Kraft (Bundesrat 2002).

Mit der vierten Etappe (Agrarpolitik 2011) wurde die Umlagerung von der Marktstützung (Produktion und Absatz) zu den Direktzahlungen weiter vorangetrieben (Bundesrat 2006). Die Umlagerung der Agrarstützung seit 1990/1992 ist in Abbildung 1 dargestellt (BLW, diverse Jahrgänge).

Abbildung 1: Stützungsumlagerung



Das zentrale Ziel der bisherigen Reformetappen war die Verbesserung der Wettbewerbsfähigkeit der Schweizer Landwirtschaft. Die erste und zweite Etappe enthielten zudem wichtige ökologische Reformelemente. Das schrittweise Vorgehen bezweckt die wirtschaftliche und soziale Verträglichkeit der Reform. Die finanziellen Mittel für die Landwirtschaft werden vom Parlament seit dem Jahr 2000 jeweils in der Form von vierjährigen Zahlungsrah-

men beschlossen. Diese stellen eine Ausgabenobergrenze dar und schaffen für die Landwirtinnen und Landwirte eine gewisse Planungssicherheit.

2.2 Neue Verfassungsgrundlage

Besonders die erste Hälfte der 1990er Jahre war geprägt durch eine intensive agrarpolitische Diskussion. Verschiedene Volksinitiativen für eine Änderung der Verfassungsbestimmung zur Landwirtschaft zeigten das breite Meinungsspektrum in der Bevölkerung auf. In einer ersten Abstimmung lehnten Volk und Stände (Kantone) 1995 den vom Parlament ausgearbeiteten Gegenvorschlag zu einer Initiative des Schweizerischen Bauernverbandes (Volksinitiative „für eine umweltgerechte, leistungsfähige, bäuerliche Landwirtschaft“, SBV, zurückgezogen 1994) knapp ab. Als Gegenvorschlag zu einer zweiten Volksinitiative (Volksinitiative „Bauern und Konsumenten – für eine naturnahe Landwirtschaft“, BUK, zurückgezogen 1996) ergänzte das Parlament daraufhin seinen ursprünglichen Vorschlag mit dem ÖLN als Voraussetzung für die Entrichtung von Direktzahlungen. Diese Vorlage wurde am 9. Juni 1996 vom Schweizer Volk mit 78 Prozent der Stimmenden und von allen Ständen angenommen. In einer dritten Volksabstimmung 1998 lehnten schliesslich Volk und Stände eine weitere Initiative für eine Verfassungsänderung deutlich ab (Volksinitiative „für preisgünstige Nahrungsmittel und ökologische Bauernhöfe“). In diesem direktdemokratischen Prozess wurde ein neuer gemeinsamer Nenner für die Schweizer Agrarpolitik gefunden und damit eine tragfähige Grundlage für die Agrarpolitik des Bundes geschaffen.

Absatz 1 von Art. 104 BV umschreibt die grundsätzliche Zielsetzung der Agrarpolitik und nennt die Aufgaben der Landwirtschaft (vgl. Kasten und Bundesversammlung 1999). Die Absätze 2-4 enthalten Vorgaben für die Ausgestaltung der Massnahmen und die Finanzierung der Agrarpolitik. Grundlage für die Direktzahlungen und den ÖLN ist Abs. 3. Neben Art. 104 sind weitere Artikel der BV von agrarpolitischer Relevanz, zu nennen sind besonders Art. 73 (Nachhaltigkeit), 74 (Umweltschutz), 75 (Raumplanung), 76 (Wasser), 78 (Natur- und Heimatschutz), 80 (Tierschutz), 103 (Strukturpolitik), 41 (Sozialziele), 126 (Haushaltführung). Für die Evaluation von Bedeutung ist Art. 170 (Wirksamkeitsüberprüfung). Die Agrarpolitik hat vor allem Schnittstellen zu folgenden anderen Politiken: Wirtschaft (inkl. Regionalpolitik), Finanzen, Raumordnung, Umwelt, Tiergesundheit, Tierschutz, Lebensmittelsicherheit und Forschung.

Art. 104 Bundesverfassung

¹Der Bund sorgt dafür, dass die Landwirtschaft durch eine nachhaltige und auf den Markt ausgerichtete Produktion einen wesentlichen Beitrag leistet zur:

- a. sicheren Versorgung der Bevölkerung;
- b. Erhaltung der natürlichen Lebensgrundlagen und Pflege der Kulturlandschaft;
- c. dezentralen Besiedlung des Landes.

² (...)

³ (...)

- a. Er ergänzt das bäuerliche Einkommen durch Direktzahlungen zur Erzielung eines angemessenen Entgelts für die erbrachten Leistungen, unter der Voraussetzung eines ökologischen Leistungsnachweises.

(...)

⁴ (...)

3 Evaluationstätigkeit des Bundes: Drei Evaluationswellen seit 1994

Die meisten Evaluationsprojekte werden vom Bund in Auftrag gegeben. Dies ist darauf zurückzuführen, dass die Agrarpolitik eine Bundesaufgabe ist und der Bund von allen staatlichen Ebenen weitaus am meisten Geld für die Förderung der Landwirtschaft einsetzt.

Seitdem sich das BLW mit der Evaluation agrarpolitischer Massnahmen befasst, lag das Schwergewicht der Tätigkeit nacheinander bei den drei folgenden grossen Politikfeldern: ökologische Direktzahlungen, Marktmassnahmen und allgemeine Direktzahlungen. Daneben wurden selbstverständlich auch Massnahmen in anderen Bereichen evaluiert. Auf diese wird hier aber nicht näher eingegangen (v.a. Massnahmen im Bereich der Grundlagenverbesserungen).

Die Evaluationen des Bundes haben vor allem summativen Charakter. Aufgrund der Breite der Fragestellungen sind die verwendeten methodischen Ansätze sehr vielfältig, sie können aber in der Regel den quantitativen (empirisch-analytischen) Verfahren zugeordnet werden. Die zugrundeliegenden wissenschaftlichen Disziplinen sind die (Agrar-)ökonomie, Agronomie sowie im Bereich der ökologischen Fragestellungen zusätzlich die Umweltwissenschaften in einem weiten Sinn.

3.1 Ökologische Direktzahlungen

Die ersten systematischen Evaluationsarbeiten gab das BLW basierend auf der Evaluationsklausel in der Oeko-Beitragsverordnung (Bundesrat 1993) und nach umfangreichen Vorarbeiten im Bereich der neuen ökologischen und ethologischen Massnahmen 1997 in Auftrag (ÖLN, Ökobeiträge: Beiträge für ökologische Ausgleichsflächen, Extensioanbau, biologischer Landbau, kontrollierte Freilandhaltung von Nutztieren, besonders tierfreundliche Stallhaltungssysteme). Es ist kein Zufall, dass die ersten Evaluationsarbeiten in diesem Bereich vom Bundesrat in Auftrag gegeben wurden: Die Ökologie war in der agrarpolitischen Debatte

der 1990er Jahre ein zentrales Thema. Entsprechend bestand ein grosses Bedürfnis, die Entwicklung des Umweltzustandes systematischer zu dokumentieren und die Wirkung der neu eingeführten Massnahmen zu überprüfen.

Für sechs Bereiche (Biodiversität, Stickstoff, Phosphor, Pflanzenbehandlungsmittel, Tierhaltung und das Querschnittsthema Wirtschaftlichkeit) wurden Ziele definiert, ein Monitoring zur Programmbeteiligung aufgebaut und wissenschaftliche Arbeiten gestartet. Diese umfassten einerseits regionale Fallstudien, andererseits wurden Projekte lanciert, welche Aussagen erlauben sollten zur Entwicklung auf nationaler Ebene. So wurden beispielsweise für Stickstoff und Phosphor Bilanzrechnungen entwickelt.

Die durch das BLW definierten Arbeitsziele basierten auf wissenschaftlichen Arbeiten und Expertenberichten. In allen Begleitgruppen waren neben dem BLW die Forschung und das damalige Bundesamt für Umwelt, Wald und Landschaft (BUWAL, heute Bundesamt für Umwelt, BAFU) respektive im Bereich der Tierhaltung das Bundesamt für Veterinärwesen (BVET) beteiligt. Die Forschungsprojekte wurden vor allem durch die landwirtschaftlichen Forschungsanstalten des BLW durchgeführt und waren für einen Zeitraum von zehn Jahren angelegt (BLW 1999).

Im Jahr 2005 wurde Bilanz gezogen. Die Resultate zur Entwicklung in allen sechs Bereichen seit 1994 sind in einem Bericht zusammengefasst (Flury 2005). Ein Leitfaden, der nach der Methodik der Politikevaluation entwickelt wurde, gab ein inhaltliches Gerüst vor für die Beurteilung der einzelnen Massnahmen nach den Kriterien der Relevanz, Effektivität und Effizienz (Peter/Hefti 2004).

Die Evaluation kam zum Schluss, dass mit der Agrarreform seit 1994 die negativen Auswirkungen der landwirtschaftlichen Produktion auf die Umwelt stark reduziert und die positiven Leistungen bezüglich Erhaltung der Lebensgrundlagen verbessert werden konnten. Gleichzeitig wurde festgehalten, dass weiterhin Handlungsbedarf bestehe, speziell in den Bereichen Stickstoff und Biodiversität. Zudem gebe es noch ein beträchtliches Potenzial zur Steigerung der Effizienz des Ressourceneinsatzes. Entsprechend schlug der Bundesrat dem Parlament in seiner Botschaft zur Agrarpolitik 2011 vor, ein Programm „nachhaltige Ressourcennutzung“ einzuführen (Bundesrat 2006). Dieses hat zum Ziel, dass die von der landwirtschaftlichen Produktion benötigten Ressourcen effizienter genutzt, der Pflanzenschutz optimiert und der Boden sowie die biologische Vielfalt nachhaltiger genutzt respektive geschützt werden. Das Konzept basiert auf einem bottom-up Ansatz und erlaubt dem Bund, entsprechende regionale und branchenspezifische Projekte zu fördern.

Rückblickend betrachtet war das 1997 gestartete Evaluationsprogramm sehr ambitiös. Die Auswertung der Ergebnisse aus der Optik der Politikevaluation zeigt, dass auf der Impactebene (Verhaltensänderungen bei den Politikadressaten) relativ klare Aussagen gemacht werden können (Anteil ökologische Ausgleichsflächen an der LN, eingesetzte Wirkstoffmengen Pflanzenbehandlungsmittel etc.), während die Beurteilung der Wirkung von Massnahmen (Outcomeebene) aus folgenden Gründen eine grosse Herausforderung darstellt:

- Schwierigkeit, Wirkungen auf einzelne Massnahmen zurückzuführen, weil jeweils verschiedene Massnahmen auf einen Umweltzustand einwirken und dieser zudem durch eine Reihe weiterer Faktoren beeinflusst wird. So ist zum Beispiel der Eintrag

von Pflanzenbehandlungsmitteln (PBM) in Oberflächengewässern beeinflusst durch die agrarökologischen Massnahmen Extensio- und Bioanbau, ökologische Ausgleichsflächen und Auflagen im ÖLN wie Fruchtfolge, Bodenschutz und Anwendungsbeschränkungen für PBM, aber auch durch andere agrarpolitische Rahmenbedingungen (die z.B. einen Einfluss haben auf die Art und Intensität der Flächennutzung), das Verhalten der Landwirtinnen und Landwirte beim PBM-Einsatz (Mengen, Zahl der Applikationen, Verhalten bei Entsorgung, Reinigung von Spritzgeräten etc.) sowie die Niederschlagsmengen, welche die oberflächliche Abschwemmung beeinflussen (Flury 2005: 105-118).

- Schwierigkeit, nach zehn Jahren gültige Aussagen zu machen: Veränderungen zeigen sich oft erst mittel- bis langfristig (Bsp. Biodiversität).

Die Auswertung ergab zudem, dass Fragen nach der Effizienz von Massnahmen weniger gut beantwortet werden können als solche nach der Effektivität. Im Nachhinein wurde auch klar, dass man der Vollzugsevaluation zu wenig Aufmerksamkeit geschenkt hatte.

Das Evaluationsprojekt hatte schon während seiner Laufzeit politische Konsequenzen: So entschloss sich der Bundesrat, in seiner Botschaft zur Agrarpolitik (AP) 2007 an das Parlament 2002 erstmals offiziell agrarökologische Ziele zu formulieren (Bundesrat 2002: 4771). Diese basierten auf den zuvor vom BLW erarbeiteten Arbeitszielen, gesetzlichen Vorgaben und internationalen Verpflichtungen. Im Vergleich zu den BLW-Arbeitszielen waren die Bundesratsziele aber als Paket breiter abgestützt, da eine begleitende Expertengruppe bestehend aus Vertretern anderer Ämter und Interessengruppen an der Erarbeitung der Ziele beteiligt war.

Während die Evaluationsarbeiten im Gang waren, wurde das agrarökologische Instrumentarium bereits weiterentwickelt. So wurden 1998 im Bereich des Gewässerschutzes in Problemgebieten (Art. 62a Gewässerschutzgesetz) (Bundesversammlung 1991) und 2001 im Bereich der Biodiversität (Qualität der ökologischen Ausgleichsflächen) neue Instrumente eingeführt (Bundesrat 2001).

Die neue Massnahme im Gewässerschutz wurde als Novum bereits in der Pilotphase begleitend evaluiert. Der Grund dafür war, dass das federführende BLW noch keine Erfahrungen hatte mit regionalen Ökomassnahmen. Die Evaluation trug wesentlich zu einer praxisnahen Gestaltung des Vollzugs bei (Peter et al. 2001).

3.2 Marktmassnahmen

Eine intensive Evaluationstätigkeit wurde durch die AP 2002 (Bundesrat 1996) ausgelöst. Im Rahmen dieses Reformschritts wurde beschlossen, die Bundesmittel für die Marktstützung innerhalb von fünf Jahren um einen Drittel abzubauen. Dieser Abbau wurde im neuen Landwirtschaftsgesetz von 1998 festgeschrieben (Art. 187, Abs. 12) (Bundesversammlung 1998), zusammen mit dem Auftrag, die Instrumente der Marktstützung einer Evaluation zu unterziehen (Art. 187, Abs. 13). Die Evaluationen zu den Marktstützungen in den Bereichen Milch, Fleisch, Eier und Getreide wurden von der ETH Zürich nach einheitlichem Konzept durchgeführt und beinhalten zum Teil auch eine ex ante Komponente (Koch/Rieder 2002).

Zusätzlich wurde eine ex ante Evaluation in Auftrag gegeben zu den Auswirkungen einer Aufhebung der Milchkontingentierung und der Umlagerung der Milchpreisstützung in Direktzahlungen (Auswirkungen auf die Milch- und Fleischproduktion, alternative Betriebszweige, Strukturen, Einkommen und Bundesausgaben). Diese wurde von der ETH Zürich zusammen mit der damaligen Forschungsanstalt für Agrarwirtschaft und Landtechnik Tänikon (FAT, heute Agroscope Reckenholz-Tänikon ART) durchgeführt (Flury et al. 2003). Für die Arbeit wurde eine Begleitgruppe eingesetzt, in der neben der Wissenschaft das BLW und die betroffene Milch- und Viehwirtschaft vertreten waren. Der Bericht enthält separate Stellungnahmen dieser vier Akteurgruppen zu den Resultaten. 2004 wurde begonnen, auch die übrigen Marktmassnahmen (Pflanzenbau) zu evaluieren. Hauptauftragnehmer war hier die ART.

Die Arbeiten äusserten sich vorwiegend kritisch zu den Instrumenten der Marktstützung. Sie stiessen auf grosse Resonanz: Im Rahmen der AP 2007 wurde entschieden, die Milchkontingentierung bis 2009 abzuschaffen. Im Rahmen der AP 2011 hat der Bundesrat dem Parlament vorgeschlagen, alle verbleibenden Marktstützungsmassnahmen mit Ausnahme der Zulage für verkäste Milch (Milchbeihilfen, Siloverbotszulage, Verarbeitungsbeiträge für Zuckerrüben, Ölsaaten und Kartoffeln, Exportsubventionen) in produktunabhängige Direktzahlungen umzuwandeln. Das Parlament hat diesen Vorschlägen im Rahmen seiner Beratungen 2006/07 grundsätzlich zugestimmt. Einzig die Siloverbotszulage soll wie die Verkäsungszulage unbefristet weitergeführt werden.

3.3 Allgemeine Direktzahlungen

Die Evaluation der allgemeinen Direktzahlungen (Flächenbeiträge, Beiträge für Raufutter verzehrende Nutztiere, Beiträge für die Tierhaltung unter erschwerenden Produktionsbedingungen und Hangbeiträge) wurde im Jahr 2003 gestartet und an der ART Reckenholz-Tänikon durchgeführt (Mann/Mack 2004). Auf der Basis von ex post und ex ante Ansätzen wurde festgestellt, dass die allgemeinen Direktzahlungen effektiv sind, d.h. die angestrebten Ziele erreicht werden, dass andererseits einzelne Beiträge zuwenig genau den einzelnen Zielen zugeordnet werden können. Die Evaluation der allgemeinen Direktzahlungen wird im Rahmen der Beantwortung der Motion „Weiterentwicklung des Direktzahlungssystems“ weitergeführt (vgl. unten, Abschnitt 5).

3.4 Monitoring als Grundlage für die Evaluation

Das periodische Erheben, Aufbereiten und Darstellen von sektor- und massnahmenbezogenen Daten (Monitoring) ist eine wichtige Grundlage für die Evaluation. Mit der Zunahme der Bedeutung der Evaluation im Verlauf der 1990er Jahre wurde auch das Monitoring immer wichtiger. Die wichtigsten Ergebnisse des Monitorings werden jährlich im Agrarbericht publiziert (BLW, diverse Jahrgänge).

Im wirtschaftlichen Bereich hat das Monitoring bereits eine lange Tradition. Hervorzuheben sind die zentrale Auswertung von Buchhaltungsdaten der landwirtschaftlichen Be-

triebe durch die ART (diverse Jahrgänge), die in veränderter Form seit mehr als 100 Jahren durchgeführt wird (Hofer 1985: 48), und die sektorale landwirtschaftliche Gesamtrechnung, die vom Bundesamt für Statistik (BFS) in Zusammenarbeit mit dem Schweizerischen Bauernverband (Schweizerischer Bauernverband diverse Jahrgänge) erstellt wird. Zusammen erlauben die beiden Instrumente detaillierte Rückschlüsse auf die wirtschaftliche Lage der Schweizer Landwirtschaft.

Das Monitoringsystem im Bereich der Ökologie basiert auf einem Set von Agrarumweltindikatoren. Es befindet sich zur Zeit noch im Aufbau. Heute werden zum Teil noch provisorische Indikatoren verwendet. Folgende sechs Themenbereiche werden abgedeckt: Stickstoff, Phosphor, Energie/Klima, Wasser, Boden, Biodiversität/Landschaft. Der Agrarbericht des BLW widmet sich jährlich jeweils einem oder zwei dieser Themen (vgl. Konzept im Agrarbericht 2002). Zudem werden im Bericht jährlich Indikatoren zu Bodennutzung und Produktionsmitteleinsatz sowie zur Beteiligung an den Tierhaltungsprogrammen RAUS und BTS publiziert.

Die Berichterstattung im Bereich Soziales beruht auf einem Konzept, das 2001 entwickelt wurde (Agrarbericht 2001). Auf dieser Basis werden im Agrarbericht des BLW jährlich folgende Ergebnisse publiziert: Erstens wird die Entwicklung bei Einkommen und Verbrauch der landwirtschaftlichen Haushalte anhand zentraler Indikatoren dargestellt. Zweitens wird eine Bestandesaufnahme gemacht zu einem von fünf zentralen Themenbereichen (Lebensqualität, Ausbildung und Arbeit, Gesundheit, Einnahmen und Konsum, Inanspruchnahme sozialer Leistungen). Hier wird vor allem auf bestehende, institutionalisierte Befragungen abgestützt. Drittens werden Studien vorgestellt, die sich vertieft mit einem bestimmten Thema bzw. Projekt befassen (z.B. zur „Rolle der Frauen in der Landwirtschaft“ oder zur Thematik „Bäuerinnen und Bauern im Pensionsalter“). Neben dem Erfassen von harten Fakten hat sich das Durchführen solcher Studien bewährt. Sie verfolgen in der Regel einen soziologischen Ansatz und erlauben qualitative Aussagen.

Basierend auf dem Monitoring in den Bereichen Ökonomie, Ökologie und Soziales wird alle vier Jahre eine Nachhaltigkeitsbeurteilung vorgenommen. Erstmals wurden entsprechende Resultate im Jahr 2005 publiziert (vgl. Agrarbericht 2005).

3.5 Institutionalisierung

Im Laufe der neunziger Jahre hat der Bund die Anstrengungen zur systematischen Evaluation laufend verstärkt. Sie gipfelten in der neuen Bundesverfassung, die am 1. Januar 2000 in Kraft trat und einen Artikel 170 zur Wirksamkeitsüberprüfung enthält (Bundesversammlung 1999). Danach hat die Bundesversammlung dafür zu sorgen, dass die Massnahmen des Bundes auf ihre Wirksamkeit überprüft werden. Der Bundesrat gilt dabei als indirekter Adressat dieser Bestimmung. Auf Grund seiner Rolle als „oberste leitende und vollziehende Behörde des Bundes“ (Art. 174 BV) gehört die Wirksamkeitsüberprüfung zu seinen wichtigsten Führungsaufgaben (BJ 2004: 2). 2004 hat er beschlossen, dass bei der Umsetzung von Artikel 170 in der Bundesverwaltung die Ämter die Hauptverantwortung für die Wirksamkeitsüberprüfung tragen (Bundesrat 2004).

Die Evaluation war im BLW zwischen 1996 und 2000 in der früheren Hauptabteilung Direktzahlungen und Strukturen angesiedelt (heute: Direktionsbereich Direktzahlungen und Strukturen). Dies ist darauf zurückzuführen, dass sich die Evaluationstätigkeit zu Beginn fast ausschliesslich auf Massnahmen bezog, die in der Verantwortung dieser Hauptabteilung lagen (ökologische Direktzahlungen). Im Jahr 2000 wurde die Evaluation neu organisiert. Dies vor dem Hintergrund der neu geschaffenen verfassungsmässigen und gesetzlichen Rahmenbedingungen.

Es wurde entschieden, dass im BLW die Direktionsbereiche für die Evaluation der Massnahmen in ihrem Zuständigkeitsbereich verantwortlich sind und die Evaluationstätigkeit des Amtes durch eine Steuerungsgruppe koordiniert wird, in der alle Direktionsbereiche vertreten sind. Die Gruppe wird vom Direktionsbereich Strategie und Evaluation geleitet. In der Steuerungsgruppe wird alle vier Jahre eine Evaluationsstrategie erarbeitet. Sie legt die Schwerpunkte der Evaluationstätigkeit des Amtes für jeweils vier Jahre fest und wird von der Geschäftsleitung verabschiedet. Der Planungshorizont ist auf die Finanzplanung des Parlaments abgestimmt (vierjährige Zahlungsrahmen). Damit wird sichergestellt, dass die gewonnenen Erkenntnisse der Evaluationstätigkeit jeweils in die Diskussion um die Festlegung der nächsten Zahlungsrahmen einfließen. Die aktuelle Evaluationsstrategie bezieht sich auf die Jahre 2008/11. Der Direktionsbereich Strategie und Evaluation stellt den Evaluationsverantwortlichen in den Linien einen Evaluationsleitfaden als Arbeitsinstrument zur Verfügung (BLW 2003) und unterstützt diese bei Bedarf in methodischer Hinsicht.

4 Evaluationstätigkeit anderer Akteure

Kantone und Gemeinden geben selten Evaluationen in Auftrag. Dies ergibt sich aus der Tatsache, dass Agrarpolitik grundsätzlich eine Bundesaufgabe ist und die Kantone und Gemeinden relativ wenig eigene Fördermittel für die Landwirtschaft einsetzen. Selbstverständlich sind Kantone und Gemeinden aber dort an Evaluationen des Bundes beteiligt, wo der Vollzug untersucht wird oder es sich um bottom-up Massnahmen handelt. Beispiele dazu sind Evaluationen in den Bereichen ökologische Direktzahlungen und Strukturverbesserungen.

Das Sekretariat des Schweizerischen Bauernverbandes ist ein bedeutender Statistikproduzent und spielt damit auch für das Monitoring eine wichtige Rolle. Seine Erhebungen und Schätzungen dienen unter anderem der Berechnung der Landwirtschaftlichen Gesamtrechnung durch das Bundesamt für Statistik. In seinem jährlich erscheinenden Situationsbericht werden zu ausgewählten Themen wissenschaftliche Arbeiten publiziert, allerdings können diese nicht als Evaluationen im engeren Sinn bezeichnet werden. Die Universitäten und speziell die ETH Zürich bearbeiten agrarpolitische Fragestellungen, die Evaluationselemente enthalten, sie werden allerdings in der Regel nicht unter dem Stichwort „Evaluation“ publiziert. Eine Ausnahme bilden die Arbeiten der ETH im Rahmen des NFP 27 „Wirksamkeit staatlicher Massnahmen“ (Baur et al. 1994; Felder/Rieder 1994; Felder et al. 1994; Cavegn et al. 1994). Private Akteure treten selten als Auftraggeber von Evaluationen auf. Eine Ausnahme ist die Avenir Suisse, ein Think-tank der Schweizer Wirtschaft, die 2006 mit ihrer Publikation „Der befreite Bauer“ den Versuch unternommen hat, eine unabhängige

Globalbeurteilung der Wirkung der Schweizer Agrarpolitik vorzunehmen (Avenir Suisse 2006). Die OECD beurteilt in einem periodisch erscheinenden Bericht die Agrarpolitik ihrer Mitgliedsländer (OECD diverse Jahrgänge). Der Schwerpunkt liegt hier aber eher beim Monitoring der Höhe der finanziellen Stützung als bei der Evaluation im klassischen Sinn.

5 Stand der Evaluation beim Bund und Ausblick

Seit etwa zehn Jahren werden im BLW agrarpolitische Massnahmen evaluiert. Die Evaluationsstätigkeit kann heute als etabliert bezeichnet werden. Der Erlass von Verfassungs-, Gesetzes- und Verordnungsbestimmungen war dafür eine notwendige, jedoch nicht eine hinreichende Bedingung. Entscheidend war, dass sich bei den für die Massnahmen Verantwortlichen mit der Zeit die Überzeugung durchsetzte, dass eine periodische kritische Überprüfung der Instrumente zu einer besseren Politik führt. Wichtig war auch, dass im Rahmen der Steuerungsgruppe ein ständiger Dialog zum Thema Evaluation aufgenommen wurde.

Nach zehn Jahren Evaluationsstätigkeit sind die wichtigsten Instrumente der Schweizer Agrarpolitik evaluiert. Die Evaluation hat unbestritten zur Weiterentwicklung des Instrumentariums beigetragen. Es zeichnet sich ab, dass die Direktzahlungen in den nächsten Jahren noch stärker im Zentrum der Aufmerksamkeit stehen werden. Dies ist darauf zurückzuführen, dass sich die Direktzahlungen zur finanziell weitaus wichtigsten Massnahmenkategorie entwickelt haben und das Parlament den Bundesrat im Frühjahr 2007 mit einer Motion (06.3635) beauftragt hat, ihm bis 2009 einen Bericht über die Weiterentwicklung des Direktzahlungssystems vorzulegen. Die entsprechenden Arbeiten wurden im BLW in Zusammenarbeit mit anderen Bundesstellen aufgenommen. Zuerst soll das agrarpolitische Zielsystem, wie es im Landwirtschaftsartikel der Bundesverfassung festgehalten ist, erneut von Grund auf analysiert werden. Dies zeigt, dass nach vier Reformetappen und einem ersten zehnjährigen Evaluationszyklus wieder ganz grundsätzliche Fragen zur Relevanz des Instrumentariums im Vordergrund stehen.

Auch im Bereich der Ökologie ist der erste aufwändige Evaluationszyklus nach zehn Jahren Arbeit abgeschlossen. Anstatt sofort neue Evaluationsprojekte zu starten, werden die Ressourcen zuerst in den Ausbau des Monitorings investiert (System von Agrarumweltindikatoren). Aufbauend auf einem konzeptionell wohldurchdachten und umfassenden Monitoringsystem kann effizienter evaluiert werden als wenn für einzelne Evaluationsprojekte bei der Datenbeschaffung jedes Mal von Neuem begonnen werden muss.

Die Agrarpolitik unterliegt nach wie vor einer grossen Dynamik (weitere Liberalisierungsrunden, Energieproblematik, Klimawandel etc.). Dies und die Knappheit der zur Verfügung stehenden Bundesmittel machen eine seriöse Evaluationsstätigkeit auch in Zukunft unabdingbar.

Evaluation von Agrarpolitik im Ländervergleich

1 Agrarpolitik

Ehe man einen Überblick über die Evaluation von Agrarpolitik in der Schweiz, Österreich und Deutschland geben kann, sei Agrarpolitik näher definiert. Sie kann allgemein und umfassend als die Gesamtheit aller Bestrebungen und Maßnahmen bezeichnet werden, die darauf abzielen, die ordnungspolitischen Rahmenbedingungen für den Agrarsektor zu gestalten und den Ablauf der ökonomischen Prozesse im Agrarsektor zu beeinflussen (Henrichsmeyer/Witzke 1991). Die Begründungen für eine spezielle Agrarpolitik im Rahmen der allgemeinen Wirtschaftspolitik leiten sich in allen Ländern aus den besonderen Rahmenbedingungen der Landwirtschaft ab. Dies sind zum Beispiel Einfluss des Wetters, strukturelle Nachteile und Bodengebundenheit der Produktion. Auch die Bedeutung der Landwirtschaft für die Nahrungssicherung mag die Sonderstellung des Agrarsektors in den Ländern erklären.

Die grundsätzlichen Ziele der Agrarpolitik sind in allen drei Ländern in Gesetzen niedergelegt. In Deutschland und Österreich in einem Landwirtschaftsgesetz, in der Schweiz sogar in der Bundesverfassung. Daneben wirken viele andere politische Ziele in die Agrarpolitik hinein. Zu nennen sind hier besonders umweltpolitische Ziele. In Österreich und Deutschland wirken auch die in den Verträgen der Europäischen Union niedergelegten Ziele der gemeinsamen Agrarpolitik auf den Agrarsektor. Die gemeinsame Agrarpolitik hat in den beiden Ländern deshalb einen besonders großen Einfluss auf die Ausgestaltung agrarpolitischer Maßnahmen, weil die Agrarpolitik in der Europäischen Union der Politikbereich ist, der am weitesten gemeinschaftlich geregelt wird. Abzulesen ist dies auch am zwar sinkenden aber nach wie vor hohen Anteil der Agrarpolitik am Haushalt der Europäischen Union.

Abgeleitet aus den grundsätzlichen Besonderheiten des Agrarsektors und den Zielsetzungen der Agrarpolitik in den drei Ländern, hat sich ein weites Feld unterschiedlichster agrarpolitischer Maßnahmen entwickelt. Diese weisen in vielen Fällen Parallelen, Berührungspunkte und Überschneidungen zu anderen Politikfeldern wie zum Beispiel der allgemeinen Wirtschafts-, Umwelt- und Sozialpolitik auf. Im Wesentlichen lassen sich folgende Teilbereiche unterscheiden:

- Marktregulierung (zum Beispiel Preisstützung durch staatliche Aufkäufe von Agrarerzeugnissen, Zahlung von Erstattungen für den Export landwirtschaftlicher Güter, Festlegung von Produktionsquoten),
- direkte Einkommenssicherung (zum Beispiel Betriebsprämien und Flächenprämien an Landwirte),

- Agrarumweltpolitik (zum Beispiel Prämien für die Bewirtschaftung gemäß Kriterien des Ökolandbaus oder den Verzicht auf bestimmte Dünge- und Pflanzenschutzmittel),
- Agrarstrukturpolitik (zum Beispiel Investitionsförderung für Neubauten von Ställen, aber auch Verbesserung von Infrastrukturen im ländlichen Raum),
- Agrarsozialpolitik (zum Beispiel staatliche Zuschüsse für die landwirtschaftliche Alterssicherung).

In den drei betrachteten Ländern kommen grundsätzlich Instrumente aus allen diesen Teilbereichen zum Einsatz. Auch die Entwicklungslinien der Agrarpolitik in den letzten Jahrzehnten verlaufen ähnlich. So verlagerte sich das Gewicht von der Marktregulierung – nicht zuletzt aufgrund von unerwünschten Nebeneffekten (Überschüsse, „Preisdumping“) und steigender internationaler Kritik – tendenziell zu den anderen Teilbereichen. Nur die Agrarsozialpolitik bildet hier eine Ausnahme.

In der Schweiz ist der Bund im Wesentlichen für alle Teilbereiche der Agrarpolitik zuständig. In Deutschland und Österreich trägt die Europäische Union die Marktregulierung und die direkte Einkommenssicherung. Agrarstrukturpolitik und Agrarumweltpolitik werden in Österreich und insbesondere in Deutschland in einem stark verflochtenen System zwischen Europäischer Union, Bund und Ländern betrieben. Nur die Agrarsozialpolitik ist in allen drei Ländern eine ausschließliche Aufgabe der obersten nationalen Politikebene.

2 Grundlagen der Evaluation

Evaluation, Erfolgskontrolle oder vergleichbare Aufgaben finden sich in der Regel bei den für den jeweiligen Teilbereich der Agrarpolitik zuständigen Ebenen. Den höchsten Stellenwert dürfte die Evaluation in der Schweiz innehaben. Schon vor der Verankerung der entsprechenden Regelung in der Bundesverfassung zeigen sich hier bereits deutliche Bemühungen zur vorausschauend geplanten Evaluation der gesamten Agrarpolitik. Diese hatten zur Folge, dass in den letzten zehn Jahren alle wichtigen agrarpolitischen Instrumente der schweizerischen Agrarpolitik evaluiert wurden.

In Deutschland findet man die Grundlagen für Evaluationen in den jeweiligen Haushaltsordnungen der Länder und des Bundes. Ergänzt werden diese Regeln durch allgemeine Verwaltungsvorschriften, die auch Hinweise für die Gestaltung von Erfolgskontrolle, Wirtschaftlichkeitsanalysen und Soll-Ist-Vergleiche geben. Erfolgskontrollen sind demnach für jede finanzwirksame Maßnahme vorgeschrieben. Auch in Österreich gibt es vergleichbare Regelungen in verschiedenen Gesetzen. Die Häufigkeit solcher Bewertungen ist in den Regelungen nicht im Detail vorgeschrieben. Soweit also ein Instrument oder Gesetz in diesen Ländern eindeutig einer staatlichen Ebene zuzuschreiben ist, dürfte eine Bewertung der entsprechenden Maßnahme eindeutig geregelt sein.

Gleiches gilt auch für die Maßnahmen der Marktregulierung und direkten Einkommenssicherung, die in den Zuständigkeitsbereich der Europäischen Union fallen. Auch in den zugrunde liegenden Verordnungen gibt es Vorschriften, die eine Evaluation der Maßnahmen vorsehen. Diese erfolgen dann meist im Zusammenhag mit Gesetzesfolgenabschät-

zungen oder der regelmäßigen Berichterstattung über die ergriffenen Maßnahmen an das Parlament.

In den Teilbereichen der Agrarstruktur- und Agrarumweltpolitik, die von einer mehr oder minder großen Verflechtung verschiedener politischer Ebenen gekennzeichnet sind, stellt sich die Situation komplizierter dar. Verschiedene Vorschriften der Europäischen Union, des Bundes und der Länder überlagern sich. Auch die Interessen am Nachweis von Erfolg und Misserfolg sowie die Ansprüche an die Evaluationen und ihre vorrangigen Zwecksetzungen divergieren hier.

Während in der Schweiz offensichtlich der Bund der Hauptmotor für Evaluationen in allen Teilbereichen der Agrarpolitik ist, befördert in Deutschland und Österreich besonders die Europäische Union die Durchführung von Evaluationen im Teilbereich der Agrarstruktur- und Agrarumweltpolitik. Verantwortlich für die Durchführung der Evaluationen sind hier die jeweils verwaltenden Ebenen, das heißt die Länder oder der Bund.

Allen Ländern gemeinsam ist, dass die allgemeinen Regelungen zur Evaluation und Erfolgskontrolle durch ein regelmäßiges System der Berichterstattung über den Agrarsektor ergänzt werden. In Deutschland ist dies vor allem der agrarpolitische Bericht der Bundesregierung, in Österreich der „Grüne Bericht“ (<http://www.gruenerbericht.at>) und in der Schweiz der Agrarbericht des schweizerischen Bundesamtes für Landwirtschaft.

3 Evaluationsakteure

Hauptauftraggeber für Evaluationen in der Schweizer Agrarpolitik ist das Bundesamt für Landwirtschaft, das dem Eidgenössischen Volkswirtschaftsdepartment zugeordnet ist. Dem entspricht in Deutschland und Österreich etwa das Wirtschaftsministerium.

Dem schweizerischen Bundesamt für Landwirtschaft sind drei Forschungseinrichtungen unterstellt. Die bisher in Auftrag gegebenen Evaluationen wurden in der Regel zunächst thematisch scharf abgegrenzt und dann an öffentliche Forschungseinrichtungen im Bereich der Agrar- und Umweltwissenschaften vergeben, darunter auch die Ressortforschungseinrichtungen des Bundesamtes für Landwirtschaft. Den Aufträgen lag in der Regel ein mehrjähriges Arbeitsprogramm zugrunde, das eine tief reichende Befassung mit der agrarpolitischen Maßnahme ermöglichte.

In Deutschland werden die meisten Evaluationsaufträge aufgrund von Vorschriften der Europäischen Union durch die Länder alleine bzw. gemeinsam von Bund und Ländern vergeben. Inhaltlich handelt es sich dabei im Wesentlichen um die Programme für die Agrarumwelt- und Agrarstrukturpolitik. Die zeitlichen Vorgaben orientieren sich am Zeitraster der europäischen Vorschriften. Die Evaluationsstudien sollen dabei die gesamten Programme der Europäischen Union mit einer Vielzahl von einzelnen Maßnahmen abdecken. Als Auftragnehmer spielen öffentliche Forschungseinrichtungen, darunter auch Ressortforschungseinrichtungen des Bundesministeriums für Ernährung, Landwirtschaft und Verbraucherschutz eine wesentliche Rolle. Mittlerweile haben aber auch private Beratungsunternehmen und Ingenieurbüros den Zugang zur Evaluation von Maßnahmen zur ländlichen Entwicklung gefunden. In anderen Teilbereichen der Agrarpolitik greift das Bundesministerium für Ernährung, Landwirtschaft und Verbraucherschutz auf seine Ressortforschungs-

einrichtungen zurück. Die Wirkungsanalyse politischer Maßnahmen und die Politikfolgenabschätzung gehören zum regelmäßigen Aufgabengebiet dieser Einrichtungen.

In Österreich werden die meisten Evaluationsaufträge für die Agrarstruktur- und Agrarumweltpolitik durch das Bundesministerium für Land- und Forstwirtschaft, Umwelt und Wasserwirtschaft vergeben. Die Koordination obliegt der Bundesanstalt für Agrarwirtschaft. Der Evaluationsbereich ist sehr weit gefasst. Als Zeitraster dienen wie in Deutschland auch hier die Vorgaben der Europäischen Union. Auftragnehmer sind – auch für Teilbereiche einer Evaluation – in der Regel öffentliche Forschungseinrichtungen, darunter ebenfalls Ressortforschungseinrichtungen aus dem Geschäftsbereich des Ministeriums.

4 Evaluationen

In der Schweiz ist der Begriff Evaluation durch die Verankerung in der Verfassung weit verbreitet und es existiert zumindest in der verantwortlichen agrarpolitischen Verwaltung ein gemeinsames Verständnis davon. In Österreich und Deutschland wird der Evaluationsbegriff besonders durch die Vorgaben der Europäischen Union geprägt. Im Titel der entsprechenden Abschlussberichte muss dann selbstverständlich der Terminus Evaluation erscheinen. Außerhalb der Politik der Europäischen Union beobachtet man hier, dass anscheinend zahlreiche als Gutachten oder Forschungsarbeiten bezeichnete Evaluationen existieren, die den Begriff Evaluation im Titel verdienen würden, aber nicht so benannt werden. Letztlich sollte dies als vorübergehend angesehen werden, da sich dieser Begriff für systematisch-empirische Untersuchungen auch im Bereich der Agrarpolitik schließlich durchsetzen dürfte.

Die Beiträge aus den drei Ländern konzentrieren sich alle auf Evaluationen der Agrarumwelt- und Agrarstrukturpolitik. In der Europäischen Union fallen diese Maßnahmen unter den Begriff der Ländlichen Entwicklung bzw. der zweiten Säule der Agrarpolitik. Als gemeinsame Schlussfolgerung aus den drei Beiträgen lässt sich festhalten, dass für die Evaluationen ein geeignetes Monitoring als essentiell angesehen wird. Bestimmte Elemente des Monitoring und Ergebnisse von Evaluationen finden sich in den regelmäßigen Berichten der Regierungen zum Agrarsektor.

5 Fazit

Im Ländervergleich zwischen Österreich, Deutschland und der Schweiz lassen sich im Ergebnis einige Besonderheiten und Gemeinsamkeiten bei der Evaluation von Agrarpolitik identifizieren:

- In allen drei Ländern spielen Ressortforschungseinrichtungen eine wesentliche Rolle bei der Durchführung von Evaluationen.
- In Deutschland sind mehr private Unternehmen als Evaluationsanbieter auf dem Markt aktiv als in den anderen beiden Ländern.

- In der Schweiz wurde eher ein fachlich konzentrierter und längerfristiger Weg zur Evaluation der gesamten Agrarpolitik gewählt.
- In Deutschland und Österreich erübrigen sich hingegen angesichts der unterschiedlichen Zuständigkeiten und der teils starken Politikverflechtung Überlegungen zur Evaluation der gesamten Agrarpolitik aus einem Guss.
- In Österreich und in Deutschland scheinen die Regelungen der Europäischen Union auch zu einem Zeitraster zu führen, das nicht immer den Raum und die Zeit lässt für tiefer gehende Analysen.
- In allen drei Ländern ist eine internationale Vernetzung bei der Erstellung der Evaluationsstudien kaum sichtbar.

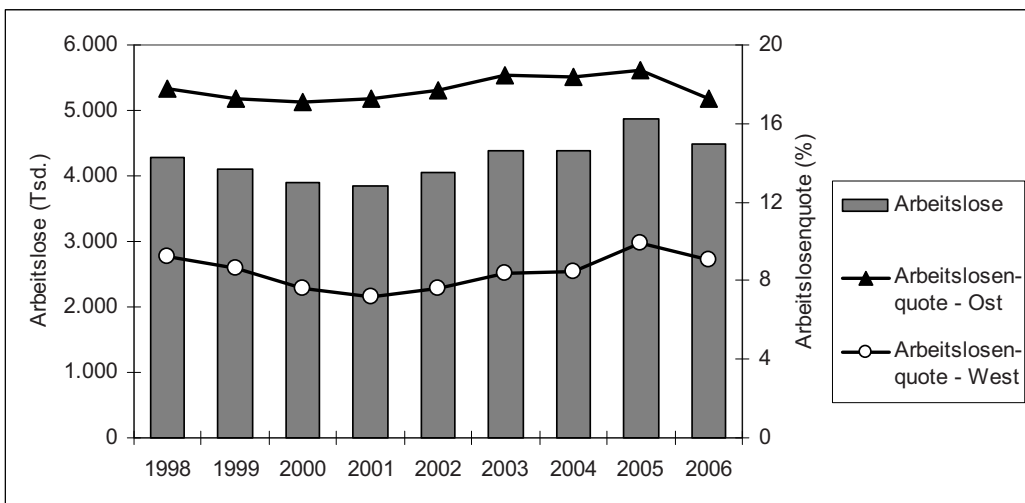
Arbeitsmarktpolitik

Evaluation in der Arbeitsmarktpolitik in Deutschland

1 Kontext der Evaluationstätigkeit

Deutschland wies in den zurückliegenden Jahren, bedingt u.a. durch den Transformationsprozess der neuen Bundesländer, einen anhaltend hohen Stand der Arbeitslosigkeit auf (Abbildung 1). Die Alimentierung dieser hohen Arbeitslosigkeit und umfangreiche Aktivitäten der aktiven Arbeitsförderung haben zu starken Belastungen der öffentlichen Haushalte geführt.

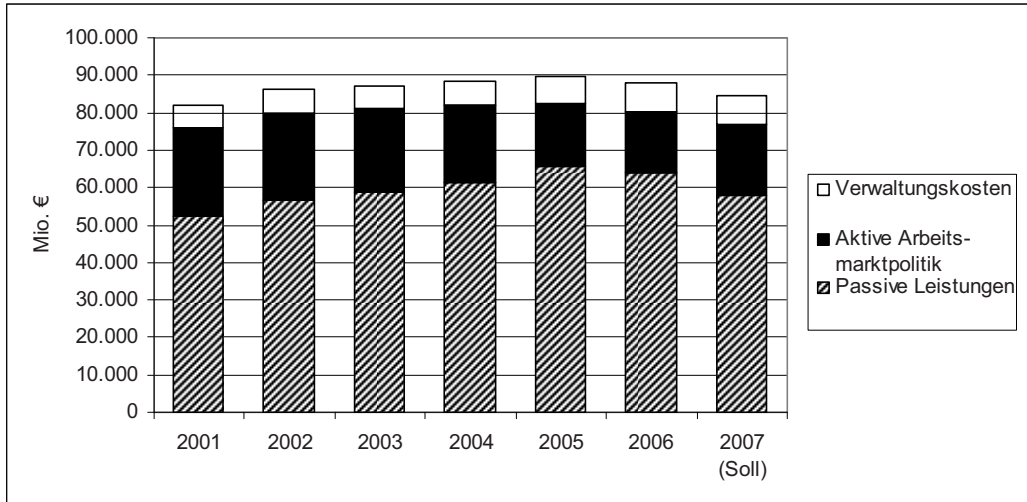
Abbildung 1: Entwicklung der Arbeitslosigkeit in Deutschland, 1998-2006



Quelle: Bundesministerium für Arbeit und Soziales

Die Ausgaben nicht nur für Lohnersatzleistungen, sondern auch für aktive Arbeitsförderung sind im Zuge der deutschen Wiedervereinigung stark gestiegen und haben sich seit Anfang des letzten Jahrzehnts auf einem sehr hohen Niveau eingependelt (Abbildung 2).

Abbildung 2: Kosten der Arbeitsmarktpolitik/ Förderung von Erwerbsfähigen 2001-2007



Quelle: Bundesministerium für Arbeit und Soziales

Vor diesem Hintergrund wurden in den zurückliegenden Jahren zwei prioritäre politische Ziele verfolgt: Die Bekämpfung von Arbeitslosigkeit und die Erhöhung der Kosteneffizienz der Arbeitsförderung. Hierzu wurden umfassende Reformen des arbeitsmarktpolitischen Anreiz- und Förderinstrumentariums auf den Weg gebracht. Seit Ende der 1990er Jahre befindet sich die Arbeitsmarktpolitik praktisch in einem anhaltenden Veränderungsprozess. Dessen vorläufiger Höhepunkt sind die ab dem Jahr 2003 schrittweise in Kraft getretenen so genannten „Hartz-Gesetze“. Die wichtigsten Meilensteine auf diesem Weg werden nachfolgend kurz skizziert.

(1) Ende der 1990er Jahre brachte die Bundesanstalt für Arbeit (heute: Bundesagentur für Arbeit – BA) das *Reformkonzept „Arbeitsamt 2000“* auf den Weg. Dieses Konzept zum Umbau der Arbeitsverwaltung folgte der Einsicht, dass nicht nur der Umfang des Förderbudgets und der „Instrumentenkasten“ zentrale Erfolgsfaktoren der Arbeitsförderung sind, sondern vor allem auch die Art und Weise, wie Förderpolitik unter den konkreten Arbeitsmarktbedingungen „vor Ort“ praktisch umgesetzt wird. Intention des Konzepts „Arbeitsamt 2000“ war es, die Handlungsspielräume der Arbeitsämter auf lokaler Ebene zu erweitern und die Organisation der örtlichen Arbeitsämter von einer funktionalen Gliederung zu kundenorientierten Arbeiterteams umzubauen.

(2) Mit der *Einführung des Sozialgesetzbuches (SGB) III im Jahr 1998* wurde eine stärkere Akzentverlagerung von passiver zu aktiver Arbeitsmarktpolitik angestrebt. Für die Evaluationsforschung in Deutschland war die Einführung des SGB III insofern von Bedeutung, als hier erstmals das Gebot zur Evaluierung der aktiven arbeitsmarktpolitischen Maßnahmen gesetzlich fixiert wurde.

(3) Auf die Feststellung einer wachsenden „Sockelarbeitslosigkeit“ in Deutschland reagierte die Politik im Jahr 1999 mit dem 2. SGB III – Änderungsgesetz. Es beinhaltete eine stärkere Ausrichtung der Arbeitsförderung auf benachteiligte Zielgruppen und auf die Bekämpfung der Langzeitarbeitslosigkeit.

Als ein wichtiger Lösungsansatz für dieses Problem gilt die stärkere Öffnung des Arbeitsmarktes für gering Qualifizierte beziehungsweise die Erweiterung des Angebots „einfacher“ Arbeitsplätze. In diesem Kontext wurde die Erprobung zweier unterschiedlicher Kombilohn-Modelle beschlossen, mit denen zusätzliche Beschäftigungspotenziale im Niedriglohnsektor erschlossen werden sollten: die Saar-Gemeinschaftsinitiative (SGI) und das so genannte „Mainzer Modell“.

(4) Im Jahr 2000 folgte mit dem Gesetz zur Verbesserung der Zusammenarbeit von Arbeitsämtern und Trägern der Sozialhilfe eine weitere wichtige Initiative. Die Bundesregierung förderte regionale Modellprojekte zur Verbesserung der Zusammenarbeit von Arbeitsämtern und Trägern der Sozialhilfe (Jobcenter, Kommunale Arbeitsvermittlungsagenturen). In diesem Kontext wurde die Möglichkeit eröffnet, die Zielgruppe „arbeitslose Sozialhilfeempfänger“ durch „Einstiegsgehlter“ und „Kombilöhne“ zu fördern.

(5) Weitere wichtige Veränderungen brachte im Jahr 2001 das so genannte Job-AQTIV-Gesetz, u.a.

- eine stärkere Betonung präventiver und vermittlungsorientierter Arbeitsmarktpolitik (Profiling, Eingliederungsvereinbarung, Möglichkeit zur Einschaltung Dritter bei der Vermittlung, Job-Rotation);
- die jährliche Festlegung von geschäftspolitischen Schwerpunkten (Bundeszielen) der Bundesanstalt für Arbeit;
- Einführung eines kontinuierlichen Controllings einschließlich der Eingliederungsbilanz, Verstärkung der Wirkungsforschung im Bereich der aktiven Arbeitsmarktpolitik sowie die Einrichtung eines Forschungsdatenzentrums zur Förderung des Wettbewerbs in der Arbeitsmarktforschung.

Bis zum 5. Meilenstein kann man alles in allem von einer *evolutionären Fortentwicklung der Arbeitsmarktpolitik* sprechen. In die so angelegten Veränderungsschritte fiel Anfang 2002 der sogenannte „Vermittlungsskandal“ der BA: Der Bundesrechnungshof hatte öffentlich gemacht, was Insider schon immer wussten – dass es unterschiedliche Möglichkeiten gibt, die durch die BA veröffentlichten Zahlen zur Arbeitsvermittlung zu interpretieren, und dass bei enger Auslegung des Kriteriums „Vermittlung“ diese Zahl deutlich unter der veröffentlichten lag. In der danach einsetzenden öffentlichen Diskussion wurde die Leistungsfähigkeit des deutschen Systems von Arbeitsvermittlung und Arbeitsförderung grundsätzlich in Frage gestellt.

Für die Bundesregierung, die mit dem Regierungswechsel im Jahr 1998 den Abbau der Arbeitslosigkeit zum zentralen Erfolgsmaßstab erklärt hatte und im Herbst 2002 mit einer bescheidenen Erfolgsbilanz zur nächsten Wahl antreten musste, eröffnete der Vermittlungs-

skandal ein „*window of opportunity*“. Die Bundesregierung nutzte dieses Fenster für eine umfassende Initiative für Strukturreformen in der Arbeitsmarktpolitik.

Es folge der „*Zweistufenplan*“ der Bundesregierung zur Reform der Arbeitsvermittlung: Als „Sofortmaßnahme“ wurde das Gesetz zur Reform der Arbeitsverwaltung und der Arbeitsvermittlung verabschiedet, verbunden mit einer „Vermittlungsoffensive“ der BA:

- Personalumschichtungen zur Verstärkung der Vermittlung;
- verstärkte Zusammenarbeit mit Dritten in der Arbeitsvermittlung;
- weitere Lockerung des Vermittlungsmonopols der BA;
- Einführung eines Rechtsanspruchs auf Vermittlungsgutscheine.

Zum anderen wurde die Kommission „Moderne Dienstleistungen am Arbeitsmarkt“ (Hartz-Kommission) mit dem Auftrag eingesetzt, bis zum Herbst 2002 Vorschläge für grundlegende Veränderungen in der Arbeitsmarktpolitik auszuarbeiten.

Nach einem Kabinettsbeschluss zur „1:1-Umsetzung“ der Kommissionsvorschläge traten bereits im Jahr 2003 die beiden ersten so genannten Hartz-Gesetze in Kraft (organisatorischer Umbau der BA, Neuordnung der Förderinstrumente). Hinzuweisen ist beispielsweise auf den Start der Personal-Service-Agenturen (PSA), des Förderinstruments „Ich-AG“ sowie der neu konzipierten Beschäftigungsverhältnisse Mini- und Midi-Jobs, auf den ersten Blick mit teilweise beachtlichen Anfangserfolgen: In den ersten hundert Tagen nach Inkrafttreten wurden fast 1 Mio. zusätzliche Minijobs angemeldet.

Mit dem IV. Hartz-Gesetz folgte im Jahr 2005 die Zusammenführung von Arbeitslosenhilfe und Sozialhilfe zur einheitlichen Grundsicherung für Arbeitslose. Die Betreuung Arbeitsloser ohne Versicherungsanspruch sowie aller Personen, die länger als ein Jahr arbeitslos sind, erfolgt seither nicht mehr durch die Arbeitsagenturen, sondern durch ARGEN – also Gemeinschaftseinrichtungen von Arbeitsagentur und Kommune, beziehungsweise optional und als „großer Feldversuch“ angelegt in alleiniger kommunaler Verantwortung.

Weitere arbeitsmarktpolitische Initiativen des Bundes starteten zuletzt im Kontext der Heraufsetzung des Renteneintrittsalters.

Schließlich sind mit Blick auf die neuen ESF-Programme für den Förderzeitraum 2007-13, die im Jahre 2006 erarbeitet wurden, einige bemerkenswerte Entwicklungen hervorzuheben. Zum einen zeichnet sich eine weitgehende *Abkoppelung* der Förderaktivitäten der Länder von der nationalen Arbeitsförderung ab. Zum anderen ist ein *Prioritätenwechsel* hin zu neuen Herausforderungen beziehungsweise Problemwahrnehmungen wie den Themen demografischer Wandel, Fachkräftesicherung, Verbesserung der Beschäftigungspotenziale Älterer etc. erkennbar.

Zusammenfassend lässt sich festhalten: Die Arbeitsmarktpolitik hat in den letzten Jahren einen Prozess permanenter Veränderungen und zuletzt gravierender struktureller Neuausrichtungen durchlaufen. Sie ist in ihren Umsetzungsstrukturen heute deutlich heterogener als Ende der 1990er Jahre. Sie ist, wie verschiedene Modellprojekte und Initiativen belegen, experimenteller geworden. Und schließlich stehen – u.a. im Kontext der Diskussion zu Kombi- und Mindestlöhnen – auch für die nähere Zukunft weitere grundlegende Neuorientierungen an.

2 Überblick über Evaluationsaktivitäten

2.1 Aktivitäten der Bundesregierung

Begleitend zu den eingangs skizzierten Reformbemühungen kam es zu einer deutlichen Intensivierung der Evaluierungsaktivitäten. Das Spektrum der Aktivitäten ist so breit, dass die hier vorgestellte Auswahl dieser Vielfalt nicht annähernd gerecht wird, sondern nur einen knappen Überblick geben kann.

Zu den wichtigen arbeitsmarktpolitischen Initiativen und Modellprojekten des Bundes wurden wissenschaftliche Begleitforschungen beziehungsweise Evaluierungen initiiert, zum Beispiel zu

- den Modellprojekten zur Verbesserung der Zusammenarbeit von Arbeitsämtern und Trägern der Sozialhilfe (MoZArT);
- den im Rahmen des Haushaltstitels „Förderung von Maßnahmen zur Erprobung zusätzlicher Wege in der Arbeitsmarktpolitik“ geförderten Arbeitsvermittlungsagenturen auf kommunaler Ebene;
- den Kombilohn-Modellen SGI und „Mainzer Modell“;
- der Initiative „Perspektive 50 plus“ (Förderung der längeren Beschäftigungsfähigkeit Älterer).

Für eine nähere Betrachtung soll allerdings ein anderes Projekt herangezogen werden: die umfassende Evaluierung der unter dem Titel „Hartz I“ bis Hartz III“ umgesetzten Arbeitsmarktreformen (vgl. Abschnitt 2.4).

Schließlich wird auch das vierte und letzte Hartz-Reformpaket, die Zusammenlegung von Arbeitslosen- und Sozialhilfe zur Grundsicherung für Arbeitslose, im Zeitraum 2006-08 evaluiert.

2.2 Institutionen auf Bundesebene

Nicht nur der Bund als Initiator und Finanzier von richtungsweisenden Modellprojekten und Strukturreformen hat Evaluierungsaktivitäten forciert. Gleiches gilt für die BA und ihr Institut für Arbeitsmarkt- und Berufsforschung (IAB) sowie das Bundesinstitut für Berufsbildung (BIBB). Das Themenspektrum ist sehr breit, es umfasst einzelne Förderinstrumente, Eingriffe in beschäftigungspolitische Rahmenbedingungen und auch die Simulation unterschiedlicher Politikoptionen in Sinne einer Ex-ante-Evaluierung.

Die Untersuchungsergebnisse des IAB finden in der Öffentlichkeit traditionell starke Resonanz. Weniger öffentlichkeitswirksam, aber mit Blick auf das Thema Evaluierung ebenfalls sehr wichtig sind die Anstrengungen zur Verbesserung der internen Erfolgskontrolle im Aktivitätsbereich der BA wie

- der Ausbau interner Controllinginstrumente und -ressourcen;
- der Ausbau regionaler Forschungsstützpunkte;

- die Entwicklung datenseitiger und methodischer Grundlagen für eine Verstetigung der internen Erfolgsbewertung. Hierzu zählt u.a. ein Projekt, das Analysen der Netto-Wirkungen für einzelne Instrumente der Arbeitsförderung auf kleinräumiger Ebene ermöglichen soll (Stephan et al. 2006).

Schließlich ist darauf hinzuweisen, dass IAB und BIBB nicht nur selbst Evaluierungsprojekte bearbeiten, sondern dabei fallweise auch mit Dritten – privaten Instituten oder Hochschulen – kooperieren und das BIBB auch stärker als Auftraggeber auftritt.

2.3 Weitere Institutionen

Ganz überwiegend werden Evaluierungsaufträge nach öffentlicher Ausschreibung an *private Anbieter beziehungsweise Konsortien* vergeben. *Hochschulen* sind als Auftragnehmer auf Bundesebene kaum aktiv. Häufiger sind sie beteiligt an der Evaluierung von Modellprojekten auf Landes- oder kommunaler Ebene. Dies könnte daran liegen, dass für derartige Projekte meist längere Bearbeitungszeiten und weniger stringente Terminsetzungen gelten. Zum Teil sind solche Evaluationen auch mit einem sehr hohen empirischen Aufwand verbunden; hier hat das Hochschulsystem durch den kostengünstigen Einsatz von Studenten und Doktoranden „Wettbewerbsvorteile“ gegenüber privaten Anbietern.

Die „großen“ *staatlich finanzierten Wirtschaftsforschungsinstitute* haben sich in der Vergangenheit nur sehr selektiv um Evaluationsaufträge beworben. Das liegt wohl vor allem daran, dass diese oft auf praktische Politikberatung gerichteten Projekte nur wenig mit den Zielvorstellungen des Wissenschaftsrates zusammenpassen, der die Institute regelmäßig evaluiert und über ihre weitere Förderwürdigkeit befindet.

Schließlich ist auf einige *private Initiativen* hinzuweisen, die unabhängig von öffentlichen Aufträgen versuchen, über Evaluierung Einfluss auf Arbeitsmarktpolitik zu nehmen: Zum einen ist es in Deutschland die Bertelsmann-Stiftung, die in den zurückliegenden Jahren zum Beispiel Projekte initiiert hat, die auf Dezentralisierung und Kommunalisierung der Arbeitsmarktpolitik abheben (Hackenberg 2003). Zum anderen hat sich das von privaten Unternehmen finanzierte „Forschungsinstitut zur Zukunft der Arbeit“ (IZA) als ein Think Tank der Arbeitsmarktforschung und Evaluation in Deutschland profiliert.

2.4 Exemplarisches Beispiel: Hartz-Evaluierung 2004 bis 2006

Die Vorschläge der Kommission „Moderne Dienstleistungen am Arbeitsmarkt“ (Bundesministerium für Arbeit und Sozialordnung 2002) waren Ausgangspunkt für einen tief greifenden Reformprozess der nationalen Arbeitsmarktpolitik in Deutschland. Kurz darauf wurde die Bundesregierung vom Deutschen Bundestag mit der Untersuchung von Umsetzung und Wirkungen der Strukturreformen am Arbeitsmarkt beauftragt (Deutscher Bundestag 2002).

Die so genannte „Hartz-Evaluation“ (Deutscher Bundestag 2006d) ist das bisher umfassendste Projekt der Arbeitsmarktevaluierung in Deutschland. Der Auftrag des Bundesta-

ges enthielt eine Reihe präziser Vorgaben für die Evaluation: Namentlich sollten untersucht werden:

- a. die *Neuorganisation* der ehemaligen Bundesanstalt zur nunmehrigen Bundesagentur für Arbeit;
- b. die Einführung neuer und die Weiterentwicklung bestehender arbeitsmarktpolitischer *Instrumente*;
- c. die Änderung *rechtlicher* beschäftigungspolitischer *Rahmenbedingungen*;
- d. die *Akzeptanz* des Dienstleisters BA bei den Kunden.

Schließlich hatte der Bundestag einen mit Blick auf die Dimension des Reformvorhabens ehrgeizigen Zeitplan vorgegeben: Innerhalb von drei Jahren sollten erste belastbare Ergebnisse vorliegen.

Die praktische Umsetzung des Untersuchungsauftrags erwies sich – wie das Reformkonzept selbst – als ein sehr komplexes Vorhaben. Das Projekt wurde zunächst grob in fünf Arbeitspakete strukturiert. Gegenstände dieser Arbeitspakete waren

1. die Wirkung der Förderinstrumente
2. der organisatorische Umbau der BA
3. die Akzeptanz der BA bei ihren Kunden (Arbeitgebern und Arbeitsuchenden)

sowie zwei „Dienstleistungspakete“:

1. die externe administrative und organisatorische Unterstützung des Evaluationsprozesses durch einen privaten Auftragnehmer und schließlich
2. die Bereitstellung und Verarbeitung von prozessgenerierten Daten der BA für die Evaluierung.

Begrenzung in den Erkenntniszielen

Die Analyse der Wirkung der Instrumente (Arbeitspaket 1) stellte – im Verständnis der meisten Beteiligten und auch des Auftraggebers – den „Kernbereich“ der Hartz-Evaluation dar. Zur Vorbereitung der in diesem Arbeitspaket vorgesehenen Untersuchungen wurden zunächst – ungewöhnlich genug – zwei konkurrierende Vorstudien durchgeführt (Fertig et al. 2004; Hagen/Spermann 2004). Ziel war es, Grundlagen für die Konzeption des Evaluationsdesigns und damit auch für die Leistungsprofile der nachfolgenden Ausschreibungen zu schaffen. Eine zentrale Schlussfolgerung aus den Vorstudien war schließlich, die Eingliederung der jeweils geförderten Personen in Erwerbsarbeit als maßgeblichen Erfolgsindikator für alle zu untersuchenden Instrumente festzulegen. Andere, aus Sicht der Arbeitsmarktforschung und der auf europäischer Ebene formulierten Politikziele ebenfalls relevante Kriterien – wie die Verbesserung von Beschäftigungsfähigkeit oder die Verhinderung sozialer Ausgrenzung – galten als wenig operationalisierbar und wurden daher nur punktuell in die Bewertung einbezogen.

Nach Auswertung der Vorstudien wurde das Arbeitspaket 1 schließlich in sechs Untersuchungsmodule aufgeteilt. Fünf der sechs Untersuchungsmodule repräsentieren jeweils einzelne Förderinstrumente oder Bündel von Instrumenten mit ähnlicher Ausrichtung:

1. Vermittlung, Einschaltung Dritter in Vermittlungsaktivitäten, Personal-Service-Agenturen (PSA);
2. Förderung der beruflichen Weiterbildung (FbW), Transferleistungen;
3. Arbeitsbeschaffungsmaßnahmen (ABM);
4. Eingliederungszuschüsse (EGZ), Beitragsbonus für Arbeitgeber, Entgeltsicherung;
5. Existenzgründungszuschuss („Ich-AG“), Überbrückungsgeld für Existenzgründer.

Im sechsten Modul finden sich unterschiedliche Analysedimensionen beziehungsweise methodische Ansätze:

- a. Wirkung der Reform beschäftigungspolitischer Rahmenbedingungen (Neuregelung der Arbeitnehmerüberlassung, Mini-/ Midi-Jobs, erleichterte Befristung für Ältere);
- b. Benchmarking erfolgreicher Arbeitsagenturen;
- c. makroökonomische Analysen.

Anknüpfend an diese Vorstrukturierung wurden im Jahr 2004 im Ergebnis eines Ausschreibungsverfahrens Untersuchungsaufträge für einzelne Arbeitspakete und Module vergeben. Die Kosten dafür beliefen sich auf gut zehn Millionen Euro. Erklärter Grundsatz des Bundesministeriums für Wirtschaft und Arbeit (BMWA) als Auftraggeber war es, „allen leistungsfähigen Forscherinnen und Forschern einen Zugang zu den Forschungsarbeiten zu ermöglichen, um europaweit das vorhandene Potenzial an Auftragnehmerinnen und -nehmern zu sichten und die Erfolg versprechendsten (...) Angebote auswählen zu können“ (Heyer 2006: 474).

Schließlich wurden mit den Untersuchungen insgesamt knapp zwanzig wissenschaftliche Einrichtungen und Institute mit rund hundert Forscherinnen und Forschern beauftragt – ausnahmslos Institute aus Deutschland. Erkennbar ist eine Tendenz zur Leistungserbringung durch *Forschungskonsortien* – insbesondere dann, wenn es sich um umfangreichere und komplexere Evaluierungsprojekte handelt. Dies liegt zum einen in der Logik der Sache, weil sich so unterschiedliche Kompetenzen spezialisierter Anbieter kombinieren und bündeln lassen. Andererseits bleibt durch die Einbindung nahezu aller führenden Institute wenig Raum für grundsätzliche Kritik am Evaluierungsprojekt.

Methoden und Begrenzungen

Die Hartz-Evaluierung repräsentiert einen breiten Kanon von methodischen Herangehensweisen an den komplexen Untersuchungsgegenstand Arbeitsmarktpolitik. Mikroökonomische Analysen bilden das übergreifende Kernelement aller Untersuchungen zu den Wirkungen der unterschiedlichen Förderinstrumente. Sie werden jeweils ergänzt durch eine

deskriptive Analyse der Förderaktivitäten und durch Untersuchungen zur Implementation und Steuerung der Umsetzung der jeweiligen Instrumente.

Darüber hinaus wurden mit dem Benchmarking-Ansatz (Analyse der Erfolgsfaktoren überdurchschnittlich erfolgreicher Arbeitsagenturen) sowie den Bemühungen um die Schätzung makroökonomischer Effekte weitere methodische Zugänge in das Evaluierungsdesign aufgenommen. Allerdings war der Benchmarking-Ansatz so angelegt, dass hier wiederum vor allem unterschiedliche Kombinationen des Förderinstrumenten-Einsatzes auf regionaler Ebene als zentraler Erklärungsfaktor des Erfolgs angesehen wurden.

Die Methodenvielfalt führte im Ergebnis bei einer Reihe von Fällen zu unterschiedlichen Einschätzungen hinsichtlich der Wirksamkeit der untersuchten Instrumente – je nachdem, welcher Ansatz der Bewertung zugrunde gelegt wurde. Dieser Mangel an Eindeutigkeit mag als unangenehm empfunden werden. Er ist jedoch insofern „nützlich“, weil damit ein deutlicher Hinweis auf die begrenzte Reichweite und Verallgemeinerungsfähigkeit der Evaluierungsergebnisse verbunden ist. Insgesamt zeigte sich, dass vor allem die Messung der Effizienz und die Bewertung der makroökonomischen Wirkungen methodisch noch am Anfang stehen.

Zusammenfassend lässt sich einschätzen, dass das Evaluierungsdesign einerseits methodisch breit angelegt ist. Von den Erkenntniszielen her war es allerdings eher begrenzt: Ein komplexes Zielsystem wurde weitgehend reduziert auf den einen Erfolgsindikator „Eingliederungswahrscheinlichkeit“. Bemerkenswert ist in diesem Zusammenhang die grundsätzliche Einschätzung der Evaluation, wonach die strikte Umsetzung einer auf betriebswirtschaftliche Kostenreduzierung sowie auf das Mengenziel „Integrationen“ gerichteten Strategie dazu führt, dass andere zentrale Ziele des SGB III (frühzeitige Intervention zur Vermeidung von Langzeitarbeitslosigkeit, Minimierung volkswirtschaftlicher Kosten der Arbeitslosigkeit) „zu weit aus dem Blickfeld der Bundesagentur für Arbeit geraten“ (Deutscher Bundestag 2006d: 14). Als künftige Herausforderung gilt darüber hinaus eine Individualisierung der Arbeitsförderung im Hinblick auf die Beurteilung der Beschäftigungsfähigkeit sowie zielgenaue und vor allem präventive Fördermaßnahmen.

Neben der recht engen Untersuchungsperspektive ist auf *zwei wichtige begrenzende* Faktoren des Evaluierungsprojekts hinzuweisen: Zum einen beschränkt sich die Untersuchung auf die Arbeitsförderung im *Rechtskreis des SGB III*. Sie zieht also eine eher willkürliche Trennlinie zum steuerfinanzierten Bereich der Arbeitsmarktpolitik für Arbeitslose mit Anspruch auf „Grundsicherung“ nach SGB II. Letzterem wird mittlerweile aber die Mehrzahl der Arbeitslosen in Deutschland zugeordnet.

Zweitens standen für die Analyse der Wirksamkeit der Instrumente *Prozessdaten* aus der Arbeitsförderung zur Verfügung, die in der Regel *bis Mitte 2005* reichten. Die Umsteuerung (Neuausrichtung bzw. Neueinführung der zu untersuchenden Instrumente) begann jedoch erst im Jahr 2003. Insofern war eine belastbare Abschätzung nachhaltiger, längerfristiger Integrationseffekte „datentechnisch“ kaum möglich. Der 2003 neu eingeführte Förderansatz „Ich-AG“ sah beispielsweise eine bis zu 3-jährige, individuelle Förderdauer vor.

2.5 Fazit

Der Überblick über wichtige Aktivitäten der zurückliegenden Jahre zeigt, dass Evaluierung im Handlungsfeld Arbeitsmarktpolitik fest etabliert ist. Die „Evaluierungsgemeinde“ ist – angesichts der vielfältigen Aktivitäten nicht überraschend – groß und diversifiziert. Das IAB nimmt im Hinblick auf Datenbereitstellung, umfangreiche eigene Forschungsaktivitäten, Ergebnistransfer und Erfahrungsaustausch eine Anker- und Leitfunktion wahr.

3 Verbindungen zu anderen Evaluationsfeldern und internationale Entwicklungen

Die deutsche Arbeitsmarktforschung und -evaluation hat von Aktivitäten in anderen Ländern erhebliche Impulse erhalten. Vor allem weil in diesen Ländern die Verfügbarkeit von relevanten Individualdaten für die Forschungsgemeinde früher gegeben war, konnten dort mikroökonometrische Evaluierungsdesigns entwickelt und umgesetzt werden, für die in Deutschland lange die datenseitigen Voraussetzungen fehlten.

Nachdem die Datengrundlagen auch in Deutschland in den letzten Jahren deutlich verbessert werden konnten, hat hier in der Evaluierungspraxis so etwas wie eine nachholende Entwicklung stattgefunden, die im Übrigen oft auch zu ähnlichen Erkenntnissen wie in anderen Ländern führte.

Beim Vergleich solcher Ergebnisse zeigte sich jedoch auch: International vergleichende Untersuchungen eröffnen interessante Perspektiven, können Denkanstöße geben, aber kaum mehr. Unterschiedliche gesellschaftliche Kontextbedingungen werden immer wieder zu unterschiedlichen Lösungsmodellen führen und bilden jeweils einen spezifischen Wirkungs- und Bewertungshintergrund (Konle-Seidl 2005: 8-9).

4 Stand der Evaluation und Bilanzierung

Evaluation von Arbeitsmarktpolitik hat inzwischen in Deutschland verbindliche rechtliche Grundlagen. Für die nationale Arbeitsmarktpolitik sind diese im SGB II und SGB III verankert, für die Evaluierung der ESF-kofinanzierten Arbeitsmarktpolitik gelten einschlägige Vorgaben der Europäischen Kommission.

Mit dem Umfang der Aktivitäten sind auch die „Potenziale“ der Evaluierung im Handlungsfeld gewachsen: Die Gemeinde der Evaluatoren, die sich des Themas annimmt, ist größer geworden, die datenseitigen Voraussetzungen haben sich deutlich verbessert – damit verbunden hat sich auch das Methodenspektrum verbreitert.

Blickt man auf die *Evaluatoren-gemeinde*, so bleibt jedoch festzuhalten, dass sie in der DeGEval bislang kein etabliertes Forum gefunden hat. Wesentliche Diskussionsprozesse vollziehen sich auf anderen Plattformen, beispielsweise im Arbeitskreis Sozialwissenschaftliche Arbeitsmarktforschung (SAMF) und bei regelmäßigen informellen Treffen von Evaluatoren, die im Bereich der ESF-Programme arbeiten.

Im Kontext der Bemühungen um die Verbesserung der *Datenlage* für Evaluierungen gab es in den letzten Jahren umfangreichen Anstrengungen zur Verknüpfung von unterschiedlichen prozessproduzierten Datensätzen der BA-Informationstechnologie. Im Ergebnis entstanden Datenbasen wie die „Maßnahme-Teilnehmer-Gesamtdatenbank“ (MTG), die „Integrierte Erwerbsbiografie“ (IEB) oder die „IAB-Beschäftigtenstichprobe“ (IABS).

Parallel dazu wurden Wege gefunden, die Datenbestände den interessierten Instituten und Wissenschaftlern in datenschutzrechtlich unbedenklicher Form zugänglich zu machen. In diesem Zusammenhang ist auf die Einrichtung des Forschungsdatenzentrums (FDZ) der Bundesagentur für Arbeit am IAB Ende 2003 zu verweisen. Damit wurde eine Infrastruktur geschaffen, die den Datenzugang für Forscher außerhalb des IAB wesentlich erleichtert.

Zur Verbesserung des Austauschs hat das IAB aus den bereits seit langem gepflegten Dokumentations-Datenbanken eine Internet-basierte Informationsplattform „Evaluation der Arbeitsmarktpolitik“ eingerichtet.¹

Mit Blick auf das *methodische* Instrumentarium ist einzuschätzen, dass ökonometrische Untersuchungsansätze sich zunehmend als Standard etabliert haben. Die Anfänge dieses Prozesses in den 1990er Jahren basierten noch auf sehr unzureichenden Datengrundlagen. Sie erlaubten zumeist nur recht undifferenzierte Aussagen in Bezug auf die durchschnittlichen (Eingliederungs-)Wirkungen der untersuchten Instrumente. Dabei blieben wichtige Erklärungsfaktoren für Erfolg oder Misserfolg der Maßnahmen wie die Heterogenität des Teilnehmerspektrums, Unterschiede in der konkreten Ausgestaltung der Fördermaßnahmen oder regional ungleiche Problemlagen am Arbeitsmarkt weitgehend unberücksichtigt. Demzufolge konnten die Analysen wenig Handlungswissen für einen wirksameren und kosteneffizienteren Einsatz der Instrumente produzieren. Erst mit der Verbreiterung der Datengrundlagen durch Erschließung der administrativen Daten für Evaluationszwecke waren hier die Voraussetzungen für mehr „Tiefenschärfe“ in der Bewertung gegeben.

Mit Blick auf Stand und aktuelle Trends der Evaluierung erscheint ein zentraler Erfolgsfaktor der Arbeitsmarktpolitik noch immer unterbelichtet: die *Leistungsfähigkeit der Institutionen*, die für die Umsetzung der Anreiz- und Förderinstrumente „vor Ort“ zuständig sind. Der Trend zur ökonometrisch gestützten Erfolgsbewertung für einzelne Instrumente hat solcherart statistisch schwer kontrollierbare Einflussfaktoren etwas in den Hintergrund gedrängt.

Drei der vier wichtigsten Einflussfaktoren werden beim gegenwärtigen Stand der Kunst der Erfolgsbewertung in die Modelle einbezogen:

- a. die Ausgestaltung der Maßnahmen (Art, Dauer, Kosten...);
- b. die Voraussetzungen der Teilnehmer (soziodemografische und erwerbsbiografische Merkmale);
- c. die unterschiedlichen regionalen Arbeitsmarktbedingungen (Angebot-Nachfrage-Relation, Dynamik).

Was fehlt, ist die systematische Bewertung der Leistungsfähigkeit derjenigen, die in der Praxis über Auswahl und Kombination der Faktoren (a) und (b) entscheiden. Diejenigen, die

¹ Siehe <http://iab.de/infoplattform>.

– um ein Bild aus einem anderen Lebensbereich zu benutzen – für den Patienten die richtige Diagnose stellen und die richtige Therapie auswählen sollen.

Dabei erscheint die Hypothese plausibel, dass Faktoren wie Fachkompetenz, Betreuungsintensität, Entscheidungsspielraum und Sozialkompetenz maßgeblich dafür verantwortlich sind, ob für das jeweilige Problem auch die wirksamste Mischung aus Fordern und Fördern gefunden wird.

Konsequenz dieser Argumentation für die Evaluierung ist es, die Bemühungen nicht zu sehr auf eine möglichst exakte Bewertung der „Leistungsfähigkeit der Förderinstrumente“ zu konzentrieren. In einem sich ständig wandelnden Umfeld hat die Suche nach verlässlichen, auch für die Zukunft gültigen Größen für Treatment-Wirkungs-Zusammenhänge wenig Chancen. Je nach Anwendungsfall und Dosierung kann ein Instrument stark oder schwach, positiv oder negativ wirken.

Mindestens ebenso wichtig ist die Leistungsfähigkeit derjenigen, die über den Instrumenteneinsatz entscheiden, also der Fallmanager, der beauftragten Dritten etc. Diesen Erfolgsfaktor in Evaluierungsdesigns einzubeziehen, vielleicht sogar in den Mittelpunkt der Aufmerksamkeit zu rücken, dürfte eine der zentralen Herausforderungen für zukünftige Evaluierungen im Bereich der Arbeitsmarktpolitik sein.

Evaluation von arbeitsmarktpolitischen Programmen und Maßnahmen in Österreich

1 Hintergrund

Die Wurzeln der Evaluation im arbeitsmarktpolitischen Bereich liegen in den 1980er Jahren, als sich das Ende der Vollbeschäftigungsära herauskristallisierte und sich strukturelle Arbeitslosigkeit verfestigte. Unter Sozialminister Alfred Dallinger wurde 1983 durch eine Novelle des Arbeitsmarktförderungsgesetzes die „Experimentelle Arbeitsmarktpolitik“ begründet. Ziel war es, neue und innovative Instrumente im Kampf gegen die Arbeitslosigkeit zu entwickeln, für den die traditionellen Instrumente der aktiven Arbeitsmarktpolitik nicht ausreichend erschienen. Dallinger definierte die Aufgabe der aktiven Arbeitsmarktpolitik wie folgt: „Vor dem Hintergrund, dass Wirtschaftswachstum nicht mehr unmittelbar Beschäftigung bewirken kann (...), heißt es für die aktive Arbeitsmarktpolitik alle Möglichkeiten initiativ zu nutzen, um mögliche Beschäftigung auszuloten und entsprechend realisieren zu helfen.“ (Lechner et al. 1988: 5) In dieser Phase stellte experimentelle Arbeitsmarktpolitik einen innovativ zu nutzenden Freiraum dar, innerhalb dessen Mittel bereit gestellt wurden, um Neues auszuprobieren. Einer der bekanntesten Ansätze in dieser Richtung war die „Aktion 8000“, deren Ziel es war, „neue, zusätzliche und kreative Arbeitsplätze im so genannten „non-profit-Bereich“ bei Ländern, Gemeinden, gemeinnützigen Einrichtungen und eigenständigen Trägern für am Arbeitsmarkt benachteiligte und langzeitarbeitslose Personen zu schaffen“ (Lechner et al. 1988: 6).

Neben der Aktion 8000 entwickelte sich in den 1980er Jahren im Rahmen der aktiven Arbeitsmarktpolitik der Ansatz einer zielgruppenspezifischen Maßnahmenkonzeption und Förderung. So wurden beispielsweise Berufsorientierungsmaßnahmen speziell für Jugendliche, Frauen nach längeren Berufsunterbrechungen und Langzeitarbeitslose entwickelt (z. B. Pinter 1988). Darüber hinaus wurden Sozialökonomische Betriebe und Gemeinnützige Beschäftigungsprojekte ins Leben gerufen, um besondere Problemgruppen (v.a. langzeitarbeitslose Personen mit psychosozialen Problemen) via eines „dritten Arbeitsmarktes“ bei der Reintegration in den Arbeitsmarkt zu unterstützen. Bei diesem Maßnahmentyp wird die Förderung von Beschäftigung mit psychosozialer Betreuung gekoppelt. Zielsetzung ist, die Geförderten wieder an den „Arbeitsalltag“ zu gewöhnen und für eine Beschäftigungsaufnahme im regulären Arbeitsmarkt vorzubereiten. Es handelt sich dabei zumeist um kleine Projekte, in denen pro Jahr eine überschaubare Zahl von Personen betreut wird.

Diese in den 1980er Jahren geschaffenen Maßnahmentypen wurden evaluiert und in den meisten Fällen während der 1990er Jahre – in adaptierter Form – fortgeführt. Dabei wurde Evaluation primär in Form von Forschungsprojekten an SozialwissenschaftlerInnen

vergeben. Evaluation hatte die Aufgabe, Informationen bereitzustellen, die zur Weiterentwicklung der Maßnahme beitragen und eine Einschätzung der Wirkungsweise erlauben. Die Evaluationen dieser Maßnahmen verfolgen zumeist einen „ganzheitlichen“ Ansatz, d.h. berücksichtigen die Maßnahme in ihrem Kontext, die Programmtheorie sowie die den Wirkungen zugrunde liegenden Wirkungsmechanismen. Diese Evaluationen kennzeichnen sich durch einen Mix an qualitativen und quantitativen Methoden, die sich gegenseitig ergänzen (z. B. BMSV 1982; 1984; Aichholzer et al. 1984; Köppl 1984; Lechner et al. 1988; Pinter 1988).

2 Veränderungen in den 1990er Jahren

In den 1990er Jahren war aufgrund der deutlich steigenden quantitativen Betroffenheit von Arbeitslosigkeit weniger Raum für derartige Innovationen in der Arbeitsmarktpolitik, auch wenn das arbeitsmarktpolitische Instrument der „Stiftungen“ einen deutlichen Boom erfuhr.¹ Insgesamt verlor die aktive Arbeitsmarktpolitik in den 1990er Jahren das nahe Verhältnis zur Sozialpolitik; diese beiden Bereiche wurden stärker voneinander getrennt.

Zentraler Akteur im Bereich der Arbeitsmarktpolitik des Bundes ist das Arbeitsmarktservice (AMS), ein Dienstleistungsunternehmen öffentlichen Rechts. Andere Akteure/-innen, wie zum Beispiel Länder oder Gemeinden sind zum Teil auch im Bereich der Arbeitsmarktpolitik aktiv, zumeist aber in Kooperation mit dem AMS (z. B. der Wiener ArbeitnehmerInnen Förderungsfonds). Das AMS wurde 1994 aus der hoheitlichen Verwaltung ausgegliedert. Trotzdem wird die Ausrichtung der Arbeitsmarktpolitik vom/von der zuständigen Bundesminister/in vorgegeben, der/die für einen mittelfristigen Zeitraum (rund fünf Jahre) Zielvorgaben für das AMS formuliert, wobei diese innerhalb des von der EU vorgegebenen Rahmens gesetzt werden (d.h. unter Berücksichtigung der Leitlinien für Beschäftigung). Diese Ziele werden vom AMS konkretisiert und jährlich an die aktuellen Arbeitsmarktprognosen angepasst.

Innerhalb des AMS wurde seit der Dezentralisierung (ab 1995) ein österreichweites Zielsystem etabliert, das durch quantitative Vorgaben die Umsetzung der arbeitsmarktpolitischen Ziele steuert (BMWA 2005; 2006). Dieses „management by objectives“ sichert innerhalb des AMS, dass die Zielvorgaben des Bundesministers/der Bundesministerin von den neun Landesorganisationen und 96 regionalen Geschäftsstellen verfolgt werden. Im Zuge der Etablierung dieses Zielsystems wurde mit dem Aufbau eines einheitlichen Monitoringsystems, dem Datawarehouse des AMS (DWH), begonnen (Mahringer 2004). Das Data Warehouse (DWH) im AMS liefert Managementinformationen und stellt die Grundlage für das arbeitsmarktpolitische Controlling dar. Seit dem Jahr 1998 wurden schrittweise alle zentralen Aufgabenbereiche des AMS in das DWH integriert, zum Beispiel die Arbeitslosenstatistik, die Förderinformationen und die Projektdatenbank. In den Jahren 2000 bis 2004 wurde in einem gemeinsamen Projekt von Bundesministerium für Wirtschaft und Arbeit (BMWA)

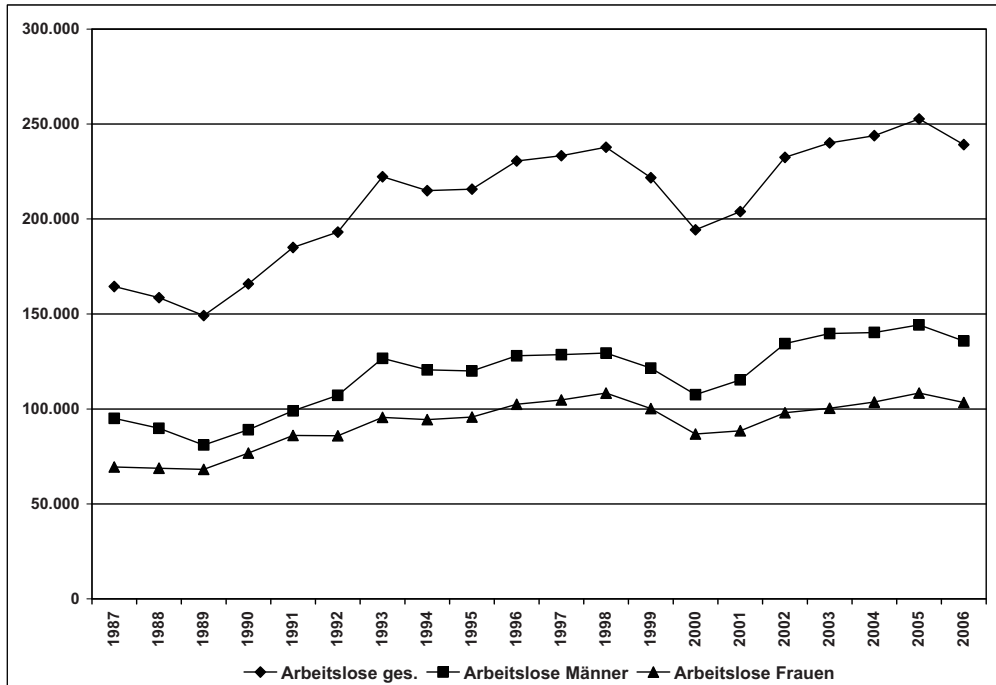
¹ Stiftungen sind komplexe Maßnahmen, die langzeitarbeitslosen Personen oder Personen mit besonderer Arbeitsmarktpolitik beim Wiedereinstieg unterstützen sollen. Stiftungen unterstützen arbeitslose Personen durch eine Kombination von Orientierung, Qualifizierung (zumeist auch inklusive Praktikum) sowie einer Unterstützung bei der Arbeitsplatzsuche. Die Dauer der Maßnahme kann je nach konkretem Einzelprojekt bis zu drei Jahren betragen.

und AMS das Monitoring der Arbeitsmarktförderung mit den Informationen der allgemeinen Arbeitsmarktbeobachtung zusammengebracht und so die Voraussetzungen für eine Beobachtung des gesamten österreichischen Arbeitsmarktgeschehens geschaffen. Das DWH ist ein Dokumentations-, Controlling- und Planungswerkzeug, das sowohl strategische Planungsfunktionen als auch arbeitsmarktstatistische Funktionen unterstützt. Dabei werden die Informationen unterschiedlicher Datenquellen in einer Datenbank zusammengeführt und einer schnellen, standardisierten Auswertung zugänglich gemacht. Die unterschiedlichen Datenquellen, die im DWH zusammengeführt werden, sind: Personenstandsregister der Arbeitslosen, Arbeits- und Lehrstellensuchenden; Informationen über die Arbeitsmarktförderung (z. B. Art, Dauer, Kosten von Maßnahmen der aktiven Arbeitsmarktpolitik); Informationen über Betriebe, die Leistungen des AMS in Anspruch nehmen; Informationen zur Ausländerbeschäftigung und zur Leistungsvergabe durch das AMS (Bezug von Arbeitslosengeld, Notstandshilfe etc.). Diese Daten werden mit den Informationen des Hauptverbandes der Sozialversicherungsträger, in denen alle sozialversicherungsrelevanten Stati einer Person erfasst werden (z. B. Beschäftigung, Arbeitslosigkeit, Pension, Karenz, sonstige Sozialversicherungszeiten), verknüpft.²

Die zweite einschneidende Veränderung in der Arbeitsmarktpolitik, die in den 1990er Jahren stattfand, war der Beitritt Österreichs zur Europäischen Union und die Übernahme der beschäftigungspolitischen Zielsetzungen der EU. Durch die Strategie des Benchmarking wurde der Prozess der Quantifizierung der Zielsetzungen verstärkt. So erfolgte in der zweiten Hälfte der 1990er Jahre beispielsweise eine massive Ausweitung der aktiven Arbeitsmarktpolitik – sowohl gemessen an der Zahl der TeilnehmerInnen wie auch der finanziellen Mittel, unter anderem bedingt durch die Zielsetzung, zwanzig Prozent aller Arbeitslosen in Schulungen zu integrieren. Damit ging auch eine Veränderung des Maßnahmenpektrums einher: Die Ausweitung der Schulungstätigkeit erfolgte primär durch tendenziell eher kürzere Qualifizierungsmaßnahmen, die pro Teilnahme auch geringere Kosten verursachen, als die eingangs beschriebenen Maßnahmen der experimentellen Arbeitsmarktpolitik. Umfassende und intensive Maßnahmen werden nach wie vor durchgeführt, doch kommt ihnen quantitativ ein geringerer Stellenwert zu (Wroblewski 2004).

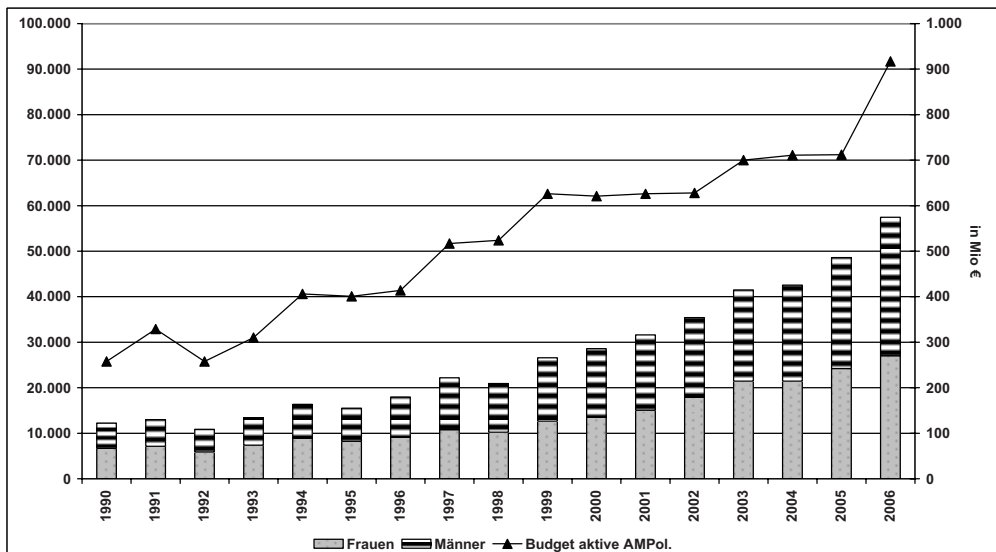
² Da hier Individualdaten verknüpft werden, ergab sich in diesem Zusammenhang auch eine intensive Diskussion von Datenschutzfragen und wurde der Zugang zum DWH unter Einbeziehung der Datenschutzkommission geregelt. Zugang zum DWH haben auf der einen Seite ausgewählte VertreterInnen von Ministerien oder öffentlichen Einrichtungen (wie z. B. BMWA oder AMS), die den gesetzlichen Vorschriften zum Datenschutz unterliegen. Zum anderen erhalten ForscherInnen den Zugang zu Individualdaten nur, wenn sie eine Vereinbarung zur Nutzung der Daten und zur Einhaltung des Datenschutzes unterzeichnen. Die leicht zugänglichen Auswertungen aus dem DWH (in Form von standardisierten Tabellenabfragen) enthalten keine Individualdaten, sondern Auswertungen auf aggregierter Ebene.

Abbildung 1: Entwicklung der Arbeitslosigkeit 1987-2006



Quelle: AMS

Abbildung 2: Entwicklung der Schulungstätigkeit des AMS 1990-2006



Quelle: AMS

3 Veränderter Stellenwert von Evaluation

Der Beitritt zur EU veränderte auch die Rolle und den Stellenwert von Evaluation im arbeitsmarktpolitischen Kontext. Evaluationsaktivitäten werden hier nunmehr zentral durch die EU-Vorgaben, konkret die verpflichtende Evaluierung aller mit EU-Geldern kofinanzierten Programme und Maßnahmen, sowie die jährliche Berichterstattung über die Fortschritte bei der Erreichung der gemeinsamen Zielsetzungen (z. B. Lissabon-Prozess), bestimmt. Die Evaluierungen von arbeitsmarktpolitischen Programmen (z. B. ESF Ziel 3, EQUAL) erfolgt üblicherweise durch Auftragsvergabe an ein Konsortium mehrerer (in- und ausländischer) Forschungseinrichtungen. Dabei dominieren größere Forschungseinrichtungen, die eine entsprechende Infrastruktur aufweisen. EvaluatorInnen werden – wie es ein hochrangiger Beamter und Auftraggeber von Evaluationen im Rahmen einer Diskussionsveranstaltung³ einmal ausdrückte – als Teil der Abteilung im Ministerium gesehen, das heißt mehr als weisungsgebundene ArbeitnehmerInnen (Beamte) denn als unabhängige ExpertInnen.

Evaluierungen von Einzelmaßnahmen werden in den letzten Jahren seltener in Form von Forschungsaufträgen durchgeführt. Hier wird vielmehr auf das neu entwickelte Monitoringsystem (DWH) zurückgegriffen, d.h. die Evaluation erfolgt nunmehr primär intern. Bei der Berichterstattung an die Europäische Kommission bezüglich der Fortschritte in Hinblick auf die Erreichung gemeinsamer Zielsetzungen (wie z. B. im Rahmen der nationalen Aktionspläne für Beschäftigung [1998–2004] beziehungsweise den Nationalen Reformprogrammen [seit 2005]) wird ebenfalls auf Monitoringinformationen (DWH) zurückgegriffen, allerdings erfolgt die Analyse durch extern beauftragte ArbeitsmarktexpertInnen.

Evaluationsaktivitäten im Zusammenhang mit EU-Programmen oder der Berichterstattung an die EU werden primär durch das zuständige Bundesministerium finanziert und beauftragt. Der Fokus der Evaluationsaktivitäten liegt auf der Wirkungsanalyse, d.h. es sollen nach einer bestimmten Zeitspanne der Umsetzung (Mid-term-Evaluation) die durch das Programm erzielten Wirkungen analysiert werden. Dabei wird das Konzept des Programms an sich nur in Ausnahmefällen hinterfragt und die Umsetzung selbst verbleibt ebenfalls in einer Blackbox. Gegenüber den eingangs beschriebenen Fragestellungen für die Evaluation von arbeitsmarktpolitischen Maßnahmen im Rahmen der experimentellen Arbeitsmarktpolitik handelt es sich nunmehr um ein deutlich eingeschränktes Spektrum an Fragestellungen.

Darüber hinaus werden auch Evaluationen von Einzelmaßnahmen oder Pilotprojekten durchgeführt, die mit EU-Geldern (ko-)finanziert wurden, beispielsweise im Rahmen von ESF-Ziel 3 oder EQUAL. In diesem Fall werden die Evaluationsaufträge von jenen Institutionen beauftragt, welche die Maßnahme durchführen. Die Kosten der Evaluation werden dabei zumeist bereits im Projekt selbst mitkalkuliert und die Fragestellungen, methodischen Vorgangsweisen sowie geforderten Qualitätskriterien jeweils im Einzelfall festgesetzt. Dadurch ergibt sich ein sehr breites Tätigkeitsfeld für Evaluation im arbeitsmarktpolitischen Kontext, das stark durch die EU-Vorgaben definiert ist. Außerhalb des EU-Kontexts werden nur in Einzelfällen Evaluationen von arbeitsmarktpolitischen Programmen durchgeführt.

³ ÖIR Plattform „6 Jahre Evaluation in Österreich“, 6.9.2001, Universität Wien.

Dabei handelt es sich zumeist um Maßnahmen, die bei ihrer Einführung politisch sehr umstritten waren und eine verpflichtende Evaluation als Teil des Kompromisses vereinbart wurde (wie z. B. bei der Einführung des Kinderbetreuungsgeldes, das sowohl als arbeitsmarkt- wie auch familienpolitische Maßnahme wahrgenommen wurde)⁴.

4 Institutionalisation/Professionalisation

Evaluation im arbeitsmarktpolitischen Bereich war zu Beginn des neuen Jahrtausends ein boomendes Forschungsgebiet, besonders bedingt durch die Evaluationsanforderung im Rahmen von EU-Programmen. Es entstand dadurch eine sehr heterogene „Evaluations-Szene“, die einige größere Forschungsinstitute sowie eine Vielzahl kleiner Unternehmen (bis hin zu Ein-Mann/Frau-Unternehmen) umfasst. Dabei bestehen große Unterschiede in Hinblick auf methodische Zugänge, Qualitätsstandards und disziplinären Hintergrund, was u.a. auch darauf zurückzuführen ist, dass es keine spezifische Ausbildung für EvaluatorInnen gibt. Qualifikationen werden in diversen Studiengängen erworben (z. B. Ökonomie, Soziologie, Psychologie, Erziehungswissenschaften) und weisen einen empirisch-methodischen Schwerpunkt auf. Die größeren Institutionen, die primär im Bereich der Evaluation von EU-Programmen tätig sind, zumeist dabei miteinander kooperieren, haben mittlerweile einen einheitlichen Standard für die Analyse von Wirkungen ausgebildet. So ist es beispielsweise mittlerweile Minimalanforderung bei Wirkungsanalysen mit quasi-experimentellen Evaluationsdesigns zu arbeiten. In diesem von ÖkonomInnen dominierten Bereich hat sich mittlerweile auch ein Set an Kernindikatoren herausgebildet, das für Evaluationen herangezogen wird (vgl. dazu z. B. Lutz et al. 2005a; 2005b). Dies ist u.a. darauf zurückzuführen, dass generell auf das DWH als Datengrundlage zurückgegriffen wird und auch AuftraggeberInnen (Bundesministerien bzw. AMS) ein Interesse an der Vergleichbarkeit von Evaluationsergebnissen haben (Mahringer 2004). Dieses Interesse an vergleichbaren Ergebnissen führt u.a. dazu, dass verstärkt Maßnahmenbündel oder umfassende Programme evaluiert werden, wie z. B. die im Rahmen des Territorialen Beschäftigungspaktes in Wien durchgeführten Maßnahmen (Leitner et al. 2003) oder die Evaluation der Gesamtheit aller Maßnahmen der aktiven Arbeitsmarktpolitik (Lechner et al. 2007).

Auf Ebene der Evaluation von Einzelmaßnahmen stellt sich die Situation anders dar. Hier wurde noch kein vergleichbarer Standard erzielt. So gibt es beispielsweise keine einheitliche Definition von „TeilnehmerInnen“ an einer Maßnahme, d.h. je nach verwendetem Konzept unterscheiden sich TeilnehmerInnenzahlen, Drop-Out-Raten u.ä. deutlich voneinander (Lassnigg et al. 2000). Problematisch ist dabei, daß diese Evaluationen nur eine eingeschränkte Aussagekraft aufweisen, da sie kaum mit anderen Studien vergleichbar sind. Es wurde bereits darauf hingewiesen, dass in den letzten Jahren weniger Evaluationen von Einzelmaßnahmen ausgeschrieben wurden, als dies noch Ende der 1990er Jahre und um den Jahrtausendwechsel der Fall war. In diesem Bereich, der über einige Jahre hinweg auch massiv boomte, war ein extrem starker Konkurrenzdruck und damit auch Preiskampf anzu-

⁴ Vgl. dazu u.a. Rille-Pfeiffer und Kapella (2007) für die familienpolitische und Riesenfelder et al. (2006) für die arbeitsmarktpolitische Sichtweise.

treffen. Eine Evaluation einer Einzelmaßnahme, möglichst umfassend, wurde in der Regel mit 30.000 Euro ausgeschrieben.

Es gibt in geringem Ausmaß eine Vernetzung zwischen den im Bereich der arbeitsmarktpolitischen Evaluation tätigen Institutionen beziehungsweise Personen, die aus Kooperationen im Rahmen der Evaluation von Programmen herrührt oder sich durch Vernetzungstreffen, die vom BMWA beziehungsweise AMS organisiert werden, ergibt. Diese Vernetzungsaktivitäten umfassen nur einen Teil der „Szene“. Es handelt sich dabei primär um ÖkonomInnen beziehungsweise VertreterInnen größerer Einrichtungen. Derzeit besteht keine Berufsvertretung für EvaluatorInnen in Österreich, es gibt auch keine Ansätze in diese Richtung. Eine Handvoll EvaluatorInnen respektive Institutionen sind im Rahmen der DGEval organisiert. Auch hier ist mein Eindruck, dass eher größere Institutionen vernetzt sind, da sich kleinere Forschungseinrichtungen derartige Aktivitäten kaum leisten können.

5 Bilanzierung

Zu den Stärken zählt in Österreich das gut ausgebaute und zu moderaten Preisen einer breiten Forschungsöffentlichkeit zugängliche Monitoringsystem (DWH). Damit wurde in den letzten Jahren eine deutliche Verbesserung der Datenqualität erreicht und die Grundlage für methodische Weiterentwicklung gelegt. Dies schlägt sich unter anderem in einer Diskussion beziehungsweise Gegenüberstellung der Ergebnisse unterschiedlicher Evaluationen zur Wirkungsweise arbeitsmarktpolitischer Maßnahmen nieder (z. B. Lechner et. al 2007; Lutz et al. 2005a; 2005b). Auch die mit dem DWH mögliche Verschränkung von Arbeitsmarktbeobachtung und Evaluation stellt ein großes Potenzial dar, da damit die Voraussetzung geschaffen wurde, Evaluationsergebnisse in bestehende Abläufe und Steuerungslogiken zu integrieren.

Als Schwäche sehe ich die starke Fragmentierung des Feldes, vor allem die Trennung in größere Einrichtungen respektive Konsortien, die mit der Evaluierung von Programmen beauftragt werden und in kleinere Institutionen, die Einzelmaßnahmen evaluieren. Problematisch dabei sind m.E. die unterschiedlichen Qualitätsstandards, die in beiden Bereichen vorherrschen, die sich auch in der Preisgestaltung niederschlagen. So basieren beispielsweise nicht alle Evaluationen von Einzelmaßnahmen auf einem Vergleichsgruppenansatz, selbst wenn Kosten-Nutzen-Analysen durchgeführt werden. Problematisch ist auch das Fehlen eines einheitlichen Sets von Indikatoren, wodurch die Vergleichbarkeit der Ergebnisse unterschiedlicher Evaluationsstudien erschwert wird. Bei größeren Institutionen, die zum einen regelmäßig Evaluationen durchführen und zum anderen häufig miteinander kooperieren, hat sich – auch aufgrund der zunehmenden Verfügbarkeit der im DWH verfügbaren Individualdaten – in den letzten Jahren ein einheitliches und über weite Strecken abgestimmtes Set an Indikatoren herausgebildet. Unterschiede zwischen den beiden angesprochenen Segmenten bestehen auch in der Nutzung und Mitgestaltung von nationalen und internationalen Vernetzungsmöglichkeiten. Die DeGEval ersetzt die notwendige Vernetzung nur zum Teil, da auch hier wiederum eher größere Einrichtungen vertreten sind.

Insgesamt werden die Möglichkeiten von Evaluation im arbeitsmarktpolitischen Kontext derzeit nicht ausgeschöpft. Durch die starke Konzentration auf die zu erfüllenden EU-

Vorgaben wird Evaluation primär als etwas von außen Vorgegebenes gesehen und in ihr vor allem eine Rechtfertigungsfunktion. Damit wird das Potenzial von Evaluation, zur Weiterentwicklung von Maßnahmen beizutragen, nicht genutzt beziehungsweise übersehen. Ähnliches gilt auch für die aktuellen Bestrebungen, von externen ExpertInnen durchgeführte Evaluation durch internes Controlling zu ersetzen, insbesondere im Bereich der Evaluation von Einzelmaßnahmen. Es sind derzeit keine Ansätze zu erkennen, diese beiden Bereiche, denen ja auch unterschiedliche Funktionen zukommen, zu verbinden. Damit spiegeln sich im Stellenwert, der Evaluation zugeschrieben wird, die Veränderungen im Bereich der aktiven Arbeitsmarktpolitik der letzten zehn Jahre.

Evaluation der Arbeitsmarktpolitik in der Schweiz

Wir leben in einer Arbeitsgesellschaft. Dies nicht nur weil wir beträchtliche Anteile der Lebenszeit mit Erwerbsarbeit verbringen und unsere Existenz über das generierte Einkommen sichern, sondern auch weil wir unsere Identität wesentlich an unsere Erwerbsbiographie knüpfen. Auf staatlicher Ebene zeigt sich diese gesellschaftliche Zentrierung um die Erwerbsarbeit darin, dass insbesondere die Arbeitsmarkt- und Sozialpolitik nicht voneinander trennbar sind und in vielfältiger Weise mit der Bildungs-, Gesundheits-, Finanz- und Steuerpolitik verwoben sind. Für Evaluationen in diesen Bereichen hat dies die Konsequenz, oftmals zu einer ganzheitlichen Sichtweise gezwungen zu sein.

Auf gesellschaftlicher Ebene ist die Gewährleistung der sozialen Sicherheit in einem weiten Sinn verstanden eine Grundvoraussetzung, dass die am Wettbewerbsprinzip ausgerichtete Arbeitswelt auf Dauer funktionieren kann. Umgekehrt ist die Ausübung einer Erwerbstätigkeit, mit der ein Erwerbseinkommen erzielt wird, der Grundstein für die Finanzierung der sozialen Sicherheit. Die möglichst produktive Erwerbstätigkeit breiter Teile der Bevölkerung ist daher im Interesse des Sozialstaates. Die ergänzende Arbeitsmarktpolitik versucht darum die Vermittlung zwischen Arbeitsangebot und -nachfrage zu verbessern. Dies kann über Vorschriften, finanzielle Anreize, Subventionen oder auch Informations- und Beratungstätigkeit und Unterstützung der Jobvermittlung geschehen. Die verantwortlichen Akteure benötigen daher Informationen zur Wirksamkeit und Effizienz der eingeleiteten Massnahmen. Die Evaluation der Arbeitsmarktpolitik soll diese Informationen bereitstellen. Sie dient einerseits der Legitimation des staatlichen Handelns und bildet andererseits die Grundlage für künftige Entscheide in diesem Politikbereich.

1 Kontexte und Inhalte der Evaluationen im Bereich der Arbeitsmarktpolitik

Evaluationen werden vorwiegend dann durchgeführt, wenn das staatliche Handeln durch neue gesetzliche Rahmenbedingungen verändert oder durch politische Akteure in Frage gestellt wird.

Wir zeigen im Folgenden auf, wie die schweizerische Arbeitsmarktpolitik nach Evaluationen verlangt und deren Ergebnisse in die Gestaltung der Arbeitsmarktpolitik einfließen. Dabei erheben wir nicht den Anspruch einer umfassenden Übersicht der Evaluationstätigkeit im Bereich der Arbeitsmarktpolitik, sondern versuchen vielmehr, einen Überblick zur thematischen und methodischen Ausrichtung sowie zu den relevanten Akteuren zu geben.

Die Kantone sind zwar über die Führung ihrer regionalen Arbeitsvermittlungszentren (RAV) zuständig für die Umsetzung der Arbeitsmarktpolitik, weil aber für die Finanzie-

rung, Steuerung und Wirkungskontrolle der Arbeitsmarktpolitik eine Bundeskompetenz besteht, spielen die gesetzliche Veränderungen im Arbeitslosenversicherungsgesetz sowie die konzeptionelle Ausrichtung der Bundespolitik eine zentrale Rolle. Aus diesem Grund gibt es im Bereich der Arbeitsmarktpolitik vergleichsweise wenige Evaluationen auf kantonalen und kommunaler Ebene. Dadurch unterscheidet sich die Arbeitsmarktpolitik von anderen Politikbereichen wie etwa der Bildungspolitik, wo die föderalen Strukturen und Hoheitsrechte nach Evaluationen auf Kantonsebene verlangen und Bundesaktivitäten bloss ergänzenden Charakter besitzen.

1.1 Die explorative Forschungsphase

Das am 1. Januar 1984 in Kraft getretene Bundesgesetz über die obligatorische Arbeitslosenversicherung und die Insolvenzenschädigung (AVIG) brachte unter anderem einen Leistungsausbau im Bereich der Präventivmassnahmen gegen die Arbeitslosigkeit. Dazu gehörten Massnahmen wie die Entrichtung von Kurstaggeldern, Beiträge an die Kosten im Zusammenhang mit der Mobilität der Arbeitskräfte (Pendlerkosten und Wochenaufenthaltsbeiträge), Einarbeitungszuschüsse sowie finanzielle Beiträge an Organisationen, die Umschulungs- und Weiterbildungskurse oder Beschäftigungsprogramme organisieren. Die Auswirkungen der Massnahmen wurden von Freiburghaus (1987) im Auftrag des Bundesamtes für Industrie, Gewerbe und Arbeit (BIGA) evaluiert. Dabei hat sich Freiburghaus (1987: 171-200) auf methodischer Ebene intensiv mit den Möglichkeiten und Grenzen der Wirkungsanalyse der eingeleiteten Präventivmassnahmen auseinandergesetzt und gibt einen – immer noch aktuellen – Überblick zu Problemen und Lösungsmöglichkeiten der Wirkungsanalyse von staatlichen Massnahmen zur Bekämpfung der Arbeitslosigkeit.

1.2 Reine Wirksamkeitsevaluationen

Der Anstieg der Arbeitslosigkeit zu Beginn der 1990er Jahre hat vom Bund eine Anpassung der bisherigen Arbeitsmarktpolitik erfordert. Das Arbeitslosenversicherungsgesetz (AVIG) wurde daher einer Revision unterzogen. Das neue AVIG setzte die Akzente noch stärker auf eine aktive Wiedereingliederungspolitik der Stellensuchenden. Das bisherige Instrument der Präventivmassnahmen ist durch die aktiven *arbeitsmarktlichen Massnahmen* (AMM) ersetzt worden. Der Einsatz der arbeitsmarktlichen Massnahmen soll die Vermittlungsfähigkeit der Stellensuchenden erhöhen und zu einer raschen und dauerhaften Wiedereingliederung beitragen (vgl. AVIG Artikel 59 Absatz 2). An diesen grundlegenden Zielen hat sich bis heute nichts verändert. Im Rahmen der erwähnten Gesetzesrevision sind die Kantone verpflichtet worden, regionale Arbeitsvermittlungszentren (RAV) zu schaffen. Damit wechselte die Verantwortung für den Vollzug der Betreuung von Stellensuchenden von den Gemeinden, welche bisher eigene Arbeitsämter geführt hatten, zu den Kantonen. Letztere sind verpflichtet worden, Beschäftigungsprogramme, Weiterbildungs- und Umschulungskurse anzubieten, um die Vermittelbarkeit der Arbeitslosen zu verbessern. Die Investitions- und Betriebskosten der RAV werden seither vollständig von der Arbeitslosenversicherung ge-

tragen. Die Regionalisierung der Arbeitsvermittlung wurde vom Bundesamte für Industrie, Gewerbe und Arbeit (BIGA) bereits zu Beginn der Achtzigerjahre konzeptuell geplant und wurde anschliessend evaluiert (vgl. BIGA: 1993).

Ausgehend von der neuen institutionellen Basis der regionalen Arbeitsvermittlungszentren (RAV) und der Akzentverschiebung hin zur aktiven Wiedereingliederungspolitik mit arbeitsmarktlichen Massnahmen (AMM) hat das Bundesamt für Wirtschaft und Arbeit (BWA, früher BIGA) eine *erste Welle* von wissenschaftlichen RAV- und AMM-Evaluationsstudien in Auftrag gegeben, welche die effektiv erzielten Wirkungseffekte (Impacts) auf die erwünschten Ziele valide identifizieren sollten. Hunold (1998) hat die Qualität der RAV und der arbeitsmarktlichen Massnahmen mittels einer telefonischen Befragung von Stellensuchenden und Firmen beurteilt. Ebenfalls über eine telefonische Befragung haben Martinovits-Wiesendanger und Ganzaroli (2000) die Wirkungen von EDV- und Gastronomie-Kurse sowie Beschäftigungsprogramme bei Teilnehmenden und einer Kontrollgruppe evaluiert. Bauer et al. (1999) haben die Auswirkungen des Zwischenverdienstes (einer Art der Lohnsubvention) anhand von Gesprächen mit Personalberatern der RAV und Stellensuchenden sowie der Verknüpfung von Prozessdaten der Arbeitslosenversicherung zu ermitteln versucht. Sheldon ermittelte in unregelmässigen Abständen die Effizienz der öffentlichen Arbeitsvermittlung (vgl. dazu Sheldon 2000; 2005) und kam zum Schluss, dass die Einführung regionaler Arbeitsvermittlungszentren (RAV) die Effizienz der Arbeitsmarktpolitik in der Schweiz erhöht hat. Ebenso in unregelmässigen Abständen werden weitere Aktivitäten, Massnahmen und Instrumente staatlichen Handelns überprüft. Während zum Beispiel „Zufriedenheits“-Evaluationen der RAV-Aktivitäten – erfragt bei Stellensuchenden, Arbeitgebern, anderen sozialstaatlichen Institutionen – vor allem von verwaltungsinternem Interesse sind, haben die übrigen wirksamkeitsorientierten Studien, wie etwa die wiederholten Studien der Konjunkturforschungsstelle der ETH Zürich (KOF) zur Kurzarbeitsentschädigung (vgl. Frick/Lampart 2005), eher eine legitimatorische Funktion gegenüber den politischen Instanzen, dann aber auch eine an Effizienzverbesserung orientierte pragmatische Funktion. In diesen Kontext gehört zum Beispiel auch die Studie von Gerfin und Lechner (1999). Die Autoren schätzten Effekte verschiedener arbeitsmarktlicher Massnahmen auf die individuellen Erwerbswahrscheinlichkeiten mit Hilfe der administrativen Daten der Arbeitslosenversicherung und der AHV. Da Teilnehmende an Massnahmen nicht repräsentativ sind für alle Stellensuchende (Selektionseffekt), können sie nicht direkt mit Nicht-Teilnehmenden an arbeitsmarktlichen Massnahmen verglichen werden, um den „separaten“ Wirkungseffekt der Teilnahme zu identifizieren. Weil in der Regel keine experimentell „sauberen“ Kontrollgruppen verfügbar sind (qua Randomisierung), werden ökonometrische Verfahren eingesetzt, um eine Randomisierung möglichst gut zu simulieren. Als innovativ kann der Ansatz von Falk et al. (2002) beurteilt werden. Der Effekt der Teilnahme an Computerkursen wurde zu messen versucht, in dem Lebensläufe von Kursteilnehmern einmal mit und einmal ohne Erwähnung der Kursteilnahme an verschiedene Arbeitgeber geschickt wurden.

1.3 Die Wirkung der Arbeitsmarktpolitik als Querschnittsthema

In diesem Abschnitt möchten wir kurz auf einige Arbeiten hinweisen, welche sich aus einem weiteren gesellschaftlichen Blickwinkel mit der Arbeitsmarktpolitik auseinandersetzen. In der Schweiz geschieht dies oftmals im Rahmen der Nationalen Forschungsprogramme (NFP 43, 45 und 51). So haben beispielsweise Lechner et al. (2004) im NFP 43 mit dem Schwerpunkt „Bildung und Beschäftigung“ mit einem quasi-experimentellen Ansatz abgeschätzt, wie sich die Beschäftigungschancen durch die Teilnahme an einer bestimmten arbeitsmarktlischen Massnahme verbessern oder verschlechtern. Auch im NFP 45 mit dem Schwerpunkt „Probleme des Sozialstaates“ sind unter dem Themenblock „Transformation des Arbeitsmarktes und Arbeitsmarktpartizipation“ einige Studien verfasst worden, die sich mit den Wirkungen des staatlichen Handelns im Bereich der Arbeitslosenversicherung befassen (vgl. für einen umfassenden Überblick Gärtner/Flückiger 2005).

Ferro Luzzi et al. (2004) untersuchen die verschiedenen Methoden zur Messung der Effizienz der RAV und kommen zum Ergebnis, dass die Effizienz durch besseres Management erhöht werden kann, aber zu grossen Teilen von externen Faktoren bestimmt wird. Aepli et al. (2004) untersuchen die Wirkungen von Beschäftigungsprogrammen bei ausgesteuerten Arbeitslosen in den Kantonen Basel-Stadt, Zürich und Genf. Die Analyse stützt sich auf quantitative und qualitative Methoden. Dabei wurde eine Personengruppe, welche an einem Beschäftigungsprogramm teilgenommen hat, mit einer Referenzgruppe verglichen, die kein entsprechendes Programm durchlaufen hat. Weiter ist die Arbeit von Magnin (2005) zu erwähnen, welche die Beratungspraxis verschiedener RAV mit Hilfe eines qualitativen Ansatzes untersucht hat. Dabei weist die Forscherin auf die Doppelfunktion der RAV hin, welche die Arbeitsbemühungen der Stellensuchenden einerseits *kontrollieren* müssen und andererseits bei der Stellensuche behilflich sein und *beraten* sollten.

Das Nationale Forschungsprogramm NFP 51 zum Thema „Integration und Ausschluss“ geht in seinem vierten Modul auf die Thematik der Erwerbstätigkeit und der Existenzsicherung ein. In diesem Kontext sind zwei Arbeiten zu erwähnen. Einerseits untersuchen Nadai und Maeder (2006) die Aktivierung und Integration von Arbeitslosen in einer Übungsfirma, einer Werkstatt für unqualifizierte Arbeitslose und einem Motivationskurs für erwerbslose Jugendliche. Dazu haben die Autoren an Personalsitzungen und Beratungsgesprächen teilgenommen, qualitative Interviews mit Teilnehmenden, dem Personal und Experten geführt. In ihrem kritischen Fazit stellen sie unter anderem fest, dass die Programme und Massnahmen nach der Logik der Disziplinierung der Teilnehmenden wirken.

Die Grenzen zwischen „Evaluationen“ und „angewandter Forschung“ in der Arbeitsmarkt- und Sozialpolitik werden hier zunehmend fließend. So sind beispielsweise die Untersuchungen zu den Ursachen und Konsequenzen der „prekären Arbeitsverhältnisse“ (Marti et al. 2003) und spezifisch der „Arbeit auf Abruf“ (Henneberger et al. 2004) implizit immer auch Analysen, ob und inwiefern sich staatliches Handeln im entsprechenden Bereich aufdrängt oder nicht. Dasselbe gilt auch für die Problematik der „Working Poor“ (vgl. Gerfin et al. 2002).

1.4 Evaluationstätigkeit auf kantonaler und kommunaler Ebene

Als Folge der stark angestiegenen Arbeitslosenzahlen sowie Veränderungen der gesetzlichen Rahmenbedingungen hat die Anzahl der „ausgesteuerten Arbeitslosen“ beziehungsweise der Stellensuchenden ohne Versicherungsanspruch in den 1990er Jahren zugenommen. Seit 1997 können ausgesteuerte Personen nicht mehr an arbeitsmarktlichen Massnahmen teilnehmen, und die vorübergehende Beschäftigung im Rahmen von arbeitsmarktlichen Massnahmen wird nicht mehr als Beitragszeit für den Anspruch auf Taggelder der Arbeitslosenversicherung angerechnet. In verschiedenen gesamtschweizerischen Situationsanalysen werden die Lage und die zahlenmässige Entwicklung von ausgesteuerten Personen beschrieben und analysiert (vgl. Aeppli et al. 1996; 1998; Aeppli 2000). Diese Analysen finden ihre Fortsetzung in kantonalen Vertiefungsstudien. So zum Beispiel durch Friboulet et al. (1997) im Kanton Freiburg oder Bachmann et al. (1998) im Kanton Luzern.

Ausgehend von den düsteren Prognosen zur Situation von ausgesteuerten Stellensuchenden und dem wachsenden finanziellen Druck auf die kommunalen Sozialämter, sind in den Kantonen Zürich, Bern, Solothurn und Genf aber auch in einigen Städten Programme initiiert worden, welche die Bezügerinnen und Bezüger von Sozialhilfe wieder in den Arbeitsmarkt integrieren sollen. Dabei wurde eine doppelte Strategie verfolgt: Sollte die Integration in den ersten Arbeitsmarkt nicht gelingen, wurde versucht, die Programmteilnehmenden über Lohnzahlungen mindestens ein Jahr zu beschäftigen, damit sie eine neue Rahmenfrist „erarbeiten“ können (das heisst ihren Anspruch auf Gelder der Arbeitslosenversicherung erneuern). Auf Grund der Kosten und des erwarteten Nutzens für die Gemeinden und die Kantone wurden diese Projekte meistens begleitend evaluiert. Dies geschah in einigen Fällen durch die verantwortlichen Vollzugsstellen selber. Oft wurde aber auch auf das im Bereich der sozialen Arbeit vorhandene Know-how der regionalen Fachhochschulen abgestützt.

So haben zum Beispiel Inderbitzin und Erni (2000) im Auftrag des kantonalen Sozialamtes Zug eine Erfolgskontrolle der kantonalen Soziallohn-Projekte durchgeführt. Zu erwähnen sind hier auch die Arbeiten der Commission externe d'évaluation des politiques publiques (CEPP) aus den Jahren 1998 und 2000, welche die kantonale Politik für ausgesteuerte Personen evaluiert hat. Merz und Walser (1996; 1997) haben im Auftrag des Kantons Zürich eine Wirkungsanalyse zum Beschäftigungsprogramm „Soliwork“ durchgeführt und Jans (1999) hat die Beschäftigungsmassnahmen für Sozialhilfeempfänger in der Gemeinde Baar im Kanton Zug evaluiert. Auch Arnold et al. (2003) evaluieren im Auftrag der Stadt Basel die Wirkungen eines Projektes zur Integration von Klientinnen und Klienten der Sozialhilfe in den regulären Arbeitsmarkt und stützen sich dabei auf Interviews mit den Programmleitenden sowie den Teilnehmenden und führen eine Kosten-Nutzen Analyse durch. Flückiger et al. (2002) haben Effekte identifiziert, welche die Effizienz im Sinne des Zweckartikels im Arbeitslosenversicherungsgesetz eindeutig verschlechtern (vgl. auch Flückiger/Morales 1994).

Im weiteren Zusammenhang mit den Evaluationsaktivitäten auf kantonaler und kommunaler Ebene sind die Evaluationen im Bereich der interinstitutionellen Zusammenarbeit (IIZ) und MAMAC (medizinisch-arbeitsmarktliches Assessment mit Case Management) zu erwähnen. Mit dem zunehmenden finanziellen Druck auf die verschiedenen Zweige der

sozialen Sicherung, dem Drehtüreffekt beziehungsweise der mangelnden Koordination unter den verschiedenen Sicherungssystemen, sind ab dem Jahr 2000 einige Projekte mit dem Ziel von umfassend koordinierten Massnahmen zur beruflichen und sozialen Integration entstanden, welche in der Folge evaluiert worden sind.

1.5 Evaluation der wirkungsorientierten Steuerung

Auf der Grundlage der Ergebnisse zu den Wirkungen der arbeitsmarktlichen Massnahmen, welche bis ins Jahr 2000 über ein vorgeschriebenes Mindestangebot gesteuert wurden (sogenannte Inputsteuerung), hat der Bund entschieden, das Mindestangebot für arbeitsmarktliche Massnahmen (AMM) zu streichen und in Anlehnung an das New Public Management zu einer wirkungsorientierten Steuerung des kantonalen Vollzugs zu wechseln. Auf Grund der Ergebnisse der erwähnten Evaluationsstudien zu den effektiv erzielten Wirkungen der RAV ist ein System von Wirkungsindikatoren aufgebaut worden, welches die Erreichung der Zielsetzungen der Arbeitslosenversicherung misst. Die Wirkungssteuerung wurde auf operativer Ebene zum ersten Mal im Jahr 2000 anfänglich mit dreijähriger (heute vierjähriger) Geltungsdauer in einer *Wirkungsvereinbarung* des Eidgenössischen Volkswirtschaftsdepartements mit den kantonalen Regierungen umgesetzt. Die Kantone sind bei der Art der Umsetzung des AVIG im Rahmen der jeweils geltenden Wirkungsvereinbarung völlig frei. Ihre Infrastruktur- und Fallkosten werden (bis zu einem meist nicht bindend werdenden oberen Plafond) durch die Arbeitslosenversicherung voll refinanziert. Die beiden Kernelemente der Wirkungsvereinbarung bilden:

1. Die *vier Wirkungsindikatoren* (vgl. nachfolgende Tabelle), auf deren Grundlage regelmässig eine RAV- und daraus eine Kantonsrangliste erstellt und veröffentlicht wird. Weil der so resultierende Reputationseffekt in den meisten Kantonen stark wirkt, konnte in der zweiten Wirkungsvereinbarung für die Jahre 2003 bis 2005 auf das bisherige finanzielle Bonus-Malus-Anreizsystem ganz verzichtet werden.
2. Ein einfaches *ökometrisches Regressionsmodell* versucht, exogene Störeinflüsse herauszufiltern, um die RAV auf faire Weise auch wirklich untereinander vergleichbar zu machen. Es berücksichtigt als exogene Variablen die regionale Zugangsrate zur Arbeitslosigkeit, die Saisonalität, die Grösse des Ortschaft in welcher sich ein RAV befindet und den regionalen Ausländeranteil.

Das Ergebnis der Wirkungsmessung ist ein gewichteter korrigierter Index, der auf den Wert 100 für das durchschnittlich wirksam funktionierende RAV geeicht ist. Als relativer Benchmark interpretiert, bedeutet der korrigierte Indexwert 100, dass der um die exogenen Einflüsse bereinigte Erwartungswert für ein RAV bei diesem Wert liegen sollte.

Tabelle 1: Wirkungsindikatoren mit Gewichtungsfaktoren

| Wirkung | Indikator | Gewichtung |
|---|--|------------|
| 1. Rasche Wiedereingliederung | Durchschnittliche Anzahl der Bezugstage der abgemeldeten Bezüger von Arbeitslosenentschädigungen in der laufenden Rahmenfrist, bzw. von Personen, die an das Ende ihrer Rahmenfrist gekommen sind. | 0.50 |
| 2. Langzeitarbeitslosigkeit vermeiden/senken | Zugänge zur Langzeitstellensuche dividiert durch die Anzahl Personen, die vor 13 Monaten eine neue Rahmenfrist mit Anspruchscode 1 (= anspruchsberechtigt) eröffnet haben. | 0.20 |
| 3. Aussteuerungen vermeiden/senken | Anzahl Aussteuerungen im Berichtsmonat dividiert durch die Anzahl Personen, die vor 2 Jahren eine neue Rahmenfrist mit Anspruchscode 1 (= anspruchsberechtigt) eröffnet haben. | 0.20 |
| 4. Wiederanmeldungen vermeiden/senken | Anzahl Wiederanmeldungen im Berichtsmonat dividiert durch die Anzahl Personen, welche in einem der vorangehenden vier Monaten abgemeldet worden sind. | 0.10 |

Die Ausgestaltung der Beziehung zwischen dem Staatssekretariat für Wirtschaft (Seco) und den Kantonen über Zielvereinbarungen findet eine Entsprechung auf der Ebene zwischen den Kantonen und den Anbietern von arbeitsmarktliche Massnahme (AMM). Gemäss der Verordnung zum AVIG müssen die Kantone mit den Anbietern von kollektiven arbeitsmarktlichen Massnahmen (AMM) Zielvereinbarungen treffen (vgl. AVIV Artikel 81d). Neben den Zielen und der Dauer der Massnahme müssen darin auch die Zielgruppen sowie Zielwerte und Indikatoren der Massnahmen definiert sein. In diesem Zusammenhang evaluieren die kantonalen Amtsstellen die arbeitsmarktlichen Massnahmen (AMM). Dies geschieht in der Regel durch die Verantwortlichen der Kantone und fliesst in die interne Berichterstattung ein. Es handelt sich daher nicht um umfassende Evaluationsberichte, jedoch werden hier deutlich Elemente der Wirkungsmessung angewandt.

Weil die Kantone im Rahmen der jeweils geltenden Wirkungsvereinbarung weitgehend frei in der Umsetzung der gesetzlichen Grundlagen sind, haben die Vollzugsverantwortlichen nicht primär Interesse daran zu erfahren wie gut oder schlecht sie in der Vergangenheit die gesetzlichen Vorgaben in die Organisationspraxis regionalen Arbeitsvermittlungszentren (RAV) umgesetzt haben, sondern sie wollen wissen, was sie in Zukunft ändern müssen, um dadurch möglichst viel Wirkungsverbesserung zu erreichen. Dies erfordert jedoch methodisch und empirisch anspruchsvolle *Kausalanalysen*, die überdies *zeitlos gültig beziehungsweise generalisierbar* sein sollten.

2 Verbindungen zu anderen Evaluationsfeldern und zum internationalen Kontext

Die schweizerische Arbeitsmarktpolitik und deren Evaluation sind eng mit der europäischen Arbeitsmarktpolitik verknüpft. Mitte der 1980er Jahre startete im Rahmen der OECD eine intensive Diskussion der Möglichkeiten zur Reformierung der – angeblich – „euroskeptisch“ verhärteten arbeitsmarktlichen Institutionen und der untauglich gewordenen „keynesianischen“ Beschäftigungspolitik. Zumindest im Bereich der *öffentlichen Arbeitsvermittlung* ist eine OECD-weite robuste „*unité de doctrine*“ – inklusive ihrer Evaluierung – entstanden (das heisst jenseits der *staatlichen* Institutionen, zum Beispiel in der Mindestlohn- und Kündigungsschutzpolitik, und jenseits auch der *sozialpartnerschaftlichen* Regulierungen). An sie knüpft konzeptionell auch die EU mit ihren Evaluationsaktivitäten der aktiven Arbeitsmarktpolitik an. An ihren Veranstaltungen sind regelmässig auch Nicht-EU-Länder wie Neuseeland, Australien, die USA und die Schweiz vertreten. Die Schweiz hat aus diesem internationalen Austauschprozess die wesentlichen Einsichten übernommen.

Bisher haben wir die konkrete Evaluations- und Forschungstätigkeit im engeren Sinn der Arbeitsmarktpolitik dargestellt. Natürlich kann die „Arbeitsmarkt- und Beschäftigungspolitik“ nicht nur in einem engen institutionalistischen Sinn der *Politics*, sondern auch in einem breiten funktionalistischen Sinn der umfassenden gesellschaftspolitischen *Policy* verstanden werden. In einem auf diese Art geöffneten Blickfeld erweitern sich, wie einleitend angesprochen, die Aktivitätsbereiche, welche Instrumente, Massnahmen und Programme zu evaluieren sein können, die mit dem Problemfeld der „(Erwerbs-)Arbeit“ direkt oder indirekt zusammenhängen. Dazu gehören die Schnittstellen zu den Bereichen Gesundheit, Migration, Bildung, Alter und Gleichstellung. Diese Problemfelder werden von verschiedenen Institutionen bearbeitet. Neben den Forschungsprojekten des „Nationalen Forschungsprogramms“ (NFP) spielen hier das Bundesamt für Sozialversicherungen (BSV), das Staatssekretariat für Wirtschaft (Seco) (vgl. dazu Elias/Ragni 2007) sowie das Bundesamt für Berufsbildung und Technologie (BBT) eine zentrale Rolle. Auf internationaler Ebene sind vor allem die von der OECD institutionalisierten Kooperationen zu nennen sowie die von der EU initiierten Aktivitäten zur „Beschäftigungspolitik“, in die auch die Schweiz einbezogen ist.

3 Stand der Evaluationen im Bereich der Arbeitsmarktpolitik

Wie wir gezeigt haben, wird die Evaluation der Arbeitsmarktpolitik durch den Bund beziehungsweise die zuständige Aufsichtsstelle massgebend geprägt. Ausgehend von einer explorativen Forschungsphase wurde im Zusammenhang mit der strategischen Neuausrichtung der regionalen Arbeitsvermittlungszentren (RAV) und den arbeitsmarktlichen Massnahmen (AMM) eine *erste Welle* mit RAV- und AMM-Evaluationsstudien durchgeführt. Dabei handelte es sich mehrheitlich um Arbeiten, welche die damals neuesten mikroökonomischen Evaluationsmethoden berücksichtigten, um methodische Probleme – wie zum Beispiel „Selektionseffekte“ – zu neutralisieren. Diese Arbeiten wurden vornehmlich durch

Ökonomen aus dem Umfeld privater Beratungsfirmen und Universitäten verfasst. Aus methodischer Sicht handelte es sich hierbei ausnahmslos um so genannte „Impact evaluations“, bei denen das effektiv erzielte Resultat unter Berücksichtigung der eingesetzten Ressourcen im Nachhinein auf die zielorientierte Wirksamkeit hin überprüft wurde. Das staatliche Handeln auf der Ebene des Vollzugs blieb bis dahin eine „black box“. Zudem war die Datenbasis für diese Studien auf der Zeitachse noch sehr schmal, sodass die Robustheit der empirischen Ergebnisse – insbesondere für grössere Zeiträume – fraglich war. Dies hatte zur Folge, dass unter den Praktikern Enttäuschung über die Resultate der Evaluationen dominierte. Lediglich bei der „unechten“ arbeitsmarktlichen Massnahme (AMM) des Zwischenverdienstes (Aufnahme einer vorübergehenden Tätigkeit) konnten wenn auch nur schwache, aber einigermaßen robuste Wirkungen der Arbeitsmarktpolitik nachgewiesen werden. Ironischerweise wurde in einer Studie der jüngsten Evaluationswelle ausgerechnet dieses Resultat in Frage gestellt: Lalive d'Épinay et al. (2006) demonstrierten anhand eines auf den schweizerischen Arbeitsmarkt kalibrierten Gleichgewichtsmodells, dass der Zwischenverdienst bedeutende negativ wirkende indirekte Substitutionseffekte zeitigt, welche die scheinbar deutlich positiven Effekte auf einzelwirtschaftlicher Ebene im gesamtwirtschaftlichen Kontext weit gehend neutralisieren. Die schlechten Ergebnisse bezüglich der mikroökonomischen Wirkungen wurden damit begründet, dass die erst 1995 eingeführte *völlig neuartige* Philosophie der „Aktivierung“ bis 1999 noch nicht ausgereift war und daher nicht auf wirksame Weise umgesetzt werden konnte.

Im Rahmen der Revision des Arbeitslosenversicherungsgesetzes in den Jahren 2001 und 2002 wollte man sich aus Sicht des Staatssekretariats für Wirtschaft (Seco) mit dem „Follow-up“ der Wirkungsevaluation der schweizerischen Arbeitsmarktpolitik nicht auf eine blossе Wiederholung der Übungsanlage der ersten Welle beschränken, sondern einerseits die Wirkungen über verfeinerte Forschungsdesigns (konkretere, problemorientierte Fragestellungen, erweiterte Datenbasis, aktualisierte Prüfmethode, gesamtwirtschaftliche Effekte) und andererseits durch den Einbezug der betrieblichen Abläufe der regionalen Arbeitsvermittlungszentren (RAV) erweitern.

Letzteres geschah vor allem deswegen, weil die Praxis „hilflos“ gegenüber Ergebnissen reiner „Impact evaluations“ ist und nicht nur wissen möchte, *wie wirksam* (im Sinne der Wirkungsvereinbarung) sie in der Vergangenheit agiert hat, sondern auch wissen möchte, *warum* sie mehr oder weniger wirksam agiert hat, beziehungsweise auf Grund welcher Massnahmen Veränderungen bei den Wirkungen erzielt werden können. Das sind statistisch und methodisch ausgesprochen anspruchsvolle Zielsetzungen, welche profunde Kenntnisse der mikroökonomischen *Kausalanalysen* voraussetzen und gleichermassen die innerbetrieblichen Prozesse der regionalen Arbeitsvermittlungszentren (RAV) und die Wirkungen bei den Stellensuchenden berücksichtigen.

Daher wurden im „Follow-up“ *kombinierte* Strategien und Prozesse und nicht bloss einzelne Massnahmen und Instrumente isoliert betrachtet. Die arbeitsmarktlichen Massnahmen (AMM) werden somit im Gegensatz zur der ersten Welle der Evaluationen nicht mehr als eigenständiger Beeinflussungsfaktor der Effizienzverbesserung gesehen, sondern wurden eingebettet in die Vollzugspraxis der kantonalen Arbeitsstellen analysiert. Dabei gilt zu berücksichtigen, dass in den letzten Jahren im Durchschnitt lediglich ein Drittel der Stellensuchenden an einer arbeitsmarktlichen Massnahme (AMM) teilgenommen haben. Gleichwohl

können die arbeitsmarktlichen Massnahmen (AMM) auf ihre Wirkungen hin „separat“ evaluiert werden, das heisst konditional auf die existierende Zuweisungspraxis in den Kantonen, den regionalen Arbeitsvermittlungszentren (RAV) und den Personalberatenden gemessen werden. Man muss sich dabei aber bewusst sein, dass auf diese Weise die Wirkungen der arbeitsmarktlichen Massnahmen (AMM) *nicht* auf eine kausal interpretierbare Weise identifiziert werden können. „Einzelne“ Massnahmen sind nie per se „wirksam“ oder „unwirksam“, selbst dann nicht, wenn sie auf bestimmte Gruppen der Stellensuchenden fokussiert werden. Nur in Kombination mit den übrigen RAV-Aktivitäten, den institutionellen Rahmenbedingungen des Bundes und des Kantons sowie den sozioökonomischen Umweltbedingungen kommen *generalisierbare, konsistente RAV-Strategien* zustande. In der ersten Welle der Evaluationen ist diese wichtige Einsicht zumindest in den Forschungsdesigns noch völlig ignoriert worden.

Im Gegensatz zur beschriebenen Evaluationsforschung im Auftrag des Staatssekretariats für Wirtschaft (Seco), sind im Rahmen der nationalen Forschungsprogramme aber auch im Auftrag von Kantonen und Gemeinden Evaluationen entstanden, welche sich weniger mit den quantitativen Wirkungen der initiierten Massnahmen im Sinne des Zweckartikels des AVIG sondern mit ihren qualitativen Einflüssen auf die betroffenen Zielgruppen (Stellensuchende oder Ausgesteuerte) befassen. Dementsprechend kommen bei diesen Evaluationen vermehrt qualitative Forschungsmethoden wie Interviews oder Gruppengespräche zur Anwendung. Neben Universitäten haben sich hier vor allem Fachhochschulen etabliert. Diese verfolgen ausgehend von ihrer an der Praxis der sozialen Arbeit orientierten Sichtweise meistens einen kundenbezogenen Zugang zu den Evaluationsthemen und geben wichtige Hinweise auf Auswirkungen der Konzeption und des Vollzugs von staatlichen Massnahmen auf die betroffenen Individuen beziehungsweise die Zielgruppe der Massnahmen.

4 Fazit

Der Fokus der Evaluation im Bereich der Arbeitsmarktpolitik hat sich von der Analyse einzelner Massnahmen hin zur Analyse der Vollzugstätigkeit und komplexen Wirkungszusammenhängen entwickelt. Die Evaluation von isolierten Instrumenten und Massnahmen hat damit ihre zentrale Stellung in den Evaluationsdesigns verloren.

Mit der Übernahme von Elementen des New Public Management (NPM) in den Vereinbarungen zwischen Bund, Kantonen und Anbietern von arbeitsmarktlichen Massnahmen hat die Überprüfbarkeit der Erreichung gesetzlicher Vorgaben und Ziele eine stark erhöhte Bedeutung erlangt. Entsprechend sind Evaluationen des zielgenauen, wirksamen und effizienten Mitteleinsatzes wichtig geworden.

Nach wie vor bleibt die Wirkungsmessung aber schwierig, weil diese durch eine Vielzahl von Faktoren wie die Organisation der regionalen Arbeitsvermittlungszentren (RAV), die unterschiedlichen Strategien Personalberatender, die Heterogenität der Stellensuchenden, konjunkturelle Schwankungen, regionale Differenzen, Unschärfen in den Datenbasen, Endogenitäts- und Selektionseffekten etc. beeinflusst werden. Die Identifikation von Wirkungen „isolierter“ und kombinierter Massnahmen und Strategien ist daher im Bereich der Arbeitsmarktpolitik oft sehr kompliziert und aufwändig. Aus diesem Grund hat sich das

Staatssekretariat für Wirtschaft (Seco) auf die Messung der möglichst *schnellen Wiedereingliederung* der Stellensuchenden in den Arbeitsmarkt konzentriert. Das bedeutet, dass die beiden anderen gesetzlichen Ziele, die dauerhafte Vermittlung sowie die Prävention der Arbeitslosigkeit, unterbelichtet bleiben. Es ist daher wenig überraschend, dass beinahe sämtliche Kursarten bei fast allen Personenkreisen keine Wirkungen oder sogar signifikant negative Wirkungen aufweisen, weil der sogenannte *Lock-in-Effekt* fast immer empirisch dominiert (das heisst die unwillkürliche Effort-Minderung bei der Stellensuche seitens Stellensuchender und bei der Stellenvermittlung seitens Personalberater nach Zuweisung in eine arbeitsmarktliche Massnahme). Wenn bei den Programmen zur vorübergehenden Beschäftigung auf dem zweiten Arbeitsmarkt eine per saldo sogar positive Wirkung festgestellt wird, dann nicht, weil die langfristig wirkenden Arbeitsmarktchancen in „an sich“ erwünschter Weise verbessert wurden (Humankapitaleffekt, soziale Integration und Tagesstrukturen intakt gehalten, etc.), sondern schlicht weil sie in der kurzen Frist einen empirisch überwiegenden *Droheffekt* auf die Stellensuchenden ausüben und so die Bereitschaft, einen relativ unattraktiven Job anzunehmen, vergrössern.

Die methodische Komplexität zeigt, dass es einerseits weiterhin Evaluationen braucht, welche sich auf die Wirkungen einzelner (neuer) Instrumente der Arbeitsmarktpolitik konzentrieren. Damit aber komplexere Wirkungszusammenhänge verstanden werden können, die dann auch in kausaler Weise interpretierbar sind, braucht es Evaluationsdesigns, die möglichst grosse Teile der gesamte Wirkungskette von der Konzeption über den Vollzug, den effizienten Einsatz von Ressourcen, die Wirkungen bei den Zielgruppen und die Auswirkungen auf die volkswirtschaftlichen Kosten berücksichtigen. Dazu sind methodisch konsistente Konzepte nötig, die eine kluge Kombination von quantitativen und qualitativen Forschungsinstrumenten erlauben.

Ländervergleich – Evaluation Arbeitsmarktpolitik

Sören Aabye Kierkegaard hat den Spruch geprägt: „Das Vergleichen ist das Ende des Glücks und der Anfang der Unzufriedenheit.“ Auch wenn dies nicht als Leitspruch für einen Ländervergleich taugt, mahnt er doch, zunächst die Voraussetzungen einer vergleichenden Analyse zu klären. Dazu gehört zum Ersten die Arbeitsmarktpolitik und ihr wirtschaftliches Umfeld, zum Zweiten die Bedeutung der Arbeitsmarktpolitik, schließlich die Normierung in den drei Ländern.

1 Rahmenbedingungen für die Evaluierung der Arbeitsmarktpolitiken

1.1 Wirtschaftliche Entwicklung und Situation auf dem Arbeitsmarkt

Die Arbeitsmarktpolitik und deren instrumentelle Umsetzung unterliegt in allen drei Ländern in den letzten zehn bis 15 Jahren einem kontinuierlichen Wandel. Die Arbeitsmarktreformen weisen Parallelen auf, wobei die wirtschaftliche Ausgangssituation und die Ausprägung und Struktur der Arbeitslosigkeit sehr unterschiedlich sind.

In der Schweiz ist die Arbeitslosigkeit nach wie vor die relativ niedrigste (OECD 2007). Die Arbeitslosenquote liegt selbst in konjunkturell schwachen Jahren (2005) mit 4,5 % deutlich niedriger als in Deutschland (9,4 %) oder Österreich (5,2 %). Aktuell führt die insgesamt gute konjunkturelle Lage in allen Ländern zu deutlichen Rückgängen der Arbeitslosigkeit. Deutschland liegt mit seiner Arbeitslosenquote über dem Durchschnitt in der Eurozone, Österreich liegt angesichts guter konjunktureller Verläufe deutlich darunter.

Die höhere Arbeitslosigkeit in Deutschland ist die Folge der Wiedervereinigung, denn in Ostdeutschland ist die Arbeitslosigkeit – auch nach 17 Jahren – doppelt so hoch wie in Westdeutschland. Außerdem sind in Deutschland ArbeitnehmerInnen mit geringen Qualifikationen und ältere Arbeitnehmende besonders betroffen von Langzeitarbeitslosigkeit. Das Risiko, arbeitslos zu werden und lange zu bleiben, hat sich in den vergangenen Jahren in allen OECD-Staaten erhöht. Allerdings sind die Arbeitsmarktchancen für wenig qualifizierte ArbeitnehmerInnen in Deutschland besonders schlecht.

Gemessen an der deutschen Situation stellt sich die österreichische Arbeitsmarktlage günstig dar. Auch die Langzeitarbeitslosigkeit ist im EU-Vergleich weiterhin relativ niedrig. Probleme bereitet hier die Beschäftigung älterer ArbeitnehmerInnen. Besondere Aufmerksamkeit erfährt außerdem die Jugendarbeitslosigkeit, auch wenn sie immer noch unter den Durchschnittswerten der EU-25 liegt.

Bemerkenswert an den guten Arbeitsmarktdaten der Schweiz ist, dass sie nicht mit den Erwartungen der Bevölkerung korrespondieren, die nach Umfragen in der Arbeitslosigkeit

nach wie vor ihre größte Sorge sehen. Verständlich wird dies, wenn man bedenkt, dass die Quote Anfang der 1990er Jahre noch bei einem Prozent lag. Auch die regionalen Unterschiede innerhalb der Schweiz – mit höheren Quoten im Tessin und Teilen der Westschweiz – dürften zur engagierten Diskussion über Funktion und Wirkung von Maßnahmen der aktiven Arbeitsmarktpolitik beitragen.

1.2 Ausgaben für die aktive Arbeitsmarktpolitik

Die unterschiedlichen Arbeitsmarktbedingungen führen beim Ausgabeverhalten für Maßnahmen der aktiven Arbeitsmarktpolitik zu Unterschieden. Deutschland steht nach vergleichenden Daten der OECD an der Spitze der drei Länder (OECD 2005): Hier werden für aktive Maßnahmen ein Prozent und mehr des BIP eingesetzt. Auch bei den passiven Leistungen zur Überbrückung von Arbeitslosenzeiten ist der deutsche Anteil höher als in den beiden anderen Ländern, wobei in der Schweiz erst seit etwa 1995/96 die Rolle und Funktion der Arbeitsmarktpolitik zur Bekämpfung der Arbeitslosigkeit betont werden. Seitdem hat es dort einen leichten Anstieg bei der Finanzierung der Arbeitsmarktpolitik gegeben. Der Anteil am BIP ist mit 0,74% im Vergleich zu Österreich ähnlich. In Österreich werden etwa 0,6% des BIP für aktivierende Maßnahmen investiert. Die hohen Ausgaben in Deutschland hängen insbesondere mit der politischen Entscheidung zu Beginn der 1990er Jahre zusammen, mit einem substanziellen Einsatz von Sozialversicherungsmitteln den Transformationsprozess zu flankieren. Das Ausgabenniveau Deutschlands wird nur von Ländern wie Dänemark, den Niederlanden oder Schweden noch übertroffen.

Die Verteilung der Mittel für Maßnahmen der aktiven Arbeitsmarktpolitik ist in allen Ländern grundsätzlich ähnlich. Die wichtigsten Instrumente sind hier die Wiedereingliederung von Arbeitslosen durch Lohnkostenzuschüsse, die Förderung der beruflichen Qualifizierung (Training), die Förderung von Existenzgründungen sowie spezielle Angebote für besonders benachteiligte Personengruppen (z. B. Behinderte).

In Österreich wird ein stärkeres Gewicht auf Trainingsprogramme gelegt. Der Anteil liegt deutlich über dem EU-Durchschnitt (Hofer/Weber 2006). In der Schweiz ist die Einführung des Zwischenverdienstes als eine Lohnsubvention in dieser Form einmalig und steht auch in anderen Ländern Pate, mit Hilfe von Kombilohnmodellen Anreize zur Arbeitsaufnahme zu setzen (Straubhaar/Werner 2003). Die Ausgaben für öffentlich finanzierte Arbeit sind ein Spezifikum Deutschlands und hier insbesondere in Ostdeutschland anzutreffen; nach den Arbeitsmarktreformen der vergangenen Jahre allerdings mit deutlich zurückgehender Bedeutung.

1.3 Rechtliche Rahmenbedingungen für die Evaluationstätigkeit

Eine deutliche Parallele weisen die Schweiz und Deutschland hinsichtlich der Normierung der Arbeitsmarktpolitik-Evaluation auf. In Deutschland ist die Evaluation integraler Bestandteil des Sozialgesetzbuchs III (SGB III – Arbeitsförderung). Mit der Integration der gesetzlichen Bestimmungen der Arbeitsförderung in die Sozialgesetzbücher (1998) ging

zugleich eine Stärkung und Präzisierung des gesetzlichen Auftrages der Bundesagentur für Arbeit einher. Gemäß § 280 ff. SGB III hat die Bundesagentur für Arbeit „die Lage und Entwicklung der Beschäftigung und des Arbeitsmarktes im Allgemeinen und nach Berufen, Wirtschaftszweigen und Regionen sowie die Wirkungen der aktiven Arbeitsförderung zu beobachten, zu untersuchen und auszuwerten.“ Das angeschlossene Institut für Arbeitsmarkt- und Berufsforschung (IAB) hat hier eine zentrale Stellung.

Die organisatorischen Voraussetzungen sind in Österreich ähnlich, denn die Arbeitsmarktbeobachtung obliegt dem Arbeitsmarktservice, der mit der Bundesagentur für Arbeit vergleichbar ist. Im österreichischen Bundesgesetz über den Arbeitsmarktservice (§ 4 AMSG) wird auch die Evaluation geregelt. Die Bundesorganisation soll Vorsorge für Arbeitsmarktbeobachtung und -statistik sowie für Grundlagen- und Entwicklungsarbeit und für die Forschung in den Bereichen Arbeitsmarkt, Beschäftigung und Berufswelt treffen. Der rechtliche Rahmen ist gegeben, aber anders als in der Schweiz beschränkt sich der Gesetzgeber in Österreich darauf, Grundzüge zu beschreiben.

Im Arbeitslosenversicherungsgesetz (AVIG) der Schweiz hingegen wurde vor etwa zehn Jahren eine umfassende Finanzierung eines Evaluationsprogramms beschlossen und in diesem Kontext auch die Normierung der Evaluation forciert. Demnach soll der Bedarf an arbeitsmarktlichen Maßnahmen systematisch und dabei auch in Bezug auf ihre geschlechtsspezifischen Auswirkungen analysiert werden, der Erfolg der geförderten Maßnahmen kontrolliert und bei der Vorbereitung und Durchführung weiterer Maßnahmen berücksichtigt werden. Auch die Erfahrungen im In- und Ausland sollen ausgewertet und den für die Durchführung zuständigen Amtsstellen entsprechende konkrete Maßnahmen empfohlen werden. Mit dem neuen Gesetz kann die für den Arbeitsmarkt zuständige, schweizerische Bundesbehörde selbst Forschungsaufträge erteilen, um die Maßnahmen der Versicherung auf ihre Wirksamkeit hin zu überprüfen.

2 Trends der Evaluation der Arbeitsmarktpolitiken

2.1 Evaluationen als Reaktion auf die Entwicklung am Arbeitsmarkt

Wie bereits oben erwähnt, reagierte die Schweiz auf steigende Arbeitslosenzahlen mit einem stärkeren finanziellen Engagement in der aktiven Arbeitsmarktpolitik. Der Umfang des Maßnahmeneinsatzes zwischen 1993 und 1997 wurde mehr als verzehnfacht. Zugleich vollzog die Schweiz mit der zweiten Revision des AVIG (1996) einen Paradigmenwechsel in der Arbeitsmarktpolitik, der mit einer Intensivierung der Evaluation einherging. Der Übergang von einer passiven zu einer aktiven, die Integrationschancen der Stellensuchenden verbessernden Arbeitsmarktpolitik, führte neben der Schaffung neuer Instrumente auch zur institutionellen Weiterentwicklung. Dazu zählen namentlich die Regionalen Arbeitsvermittlungszentren (RAV). Die Kantone sind seither verpflichtet, solche RAVs zu betreiben. Dieser Strategiewechsel hatte unmittelbare Auswirkungen auf das Evaluationsgeschehen. Wissenschaftler wurden erstmalig aus dem System der Arbeitslosenversicherung mit umfangreichen Datensätzen versorgt. Die Evaluation konzentrierte sich hierbei auf Matching-Ver-

fahren, um die Wirkung von Programmen auf die Integration in den Arbeitsmarkt zu messen (Lechner 2006).

Zudem wurde der Vollzug – unter Federführung des Seco – systematisch evaluiert. Dabei konnten verschiedene Verbesserungspotenziale lokalisiert werden. Ein zentraler Meilenstein dieses Prozesses waren dabei die Zielvereinbarungen für die regionalen Arbeitsvermittlungszentren, die Logistikstellen für arbeitsmarktliche Maßnahmen und die Arbeitslosenkassen.

Schon früher als in der Schweiz hat in Österreich der Prozess der Evaluierung eingesetzt. In Österreich stand zunächst die Evaluierung einer experimentell orientierten Arbeitsmarktpolitik im Vordergrund. Den Kern bildeten programmbegleitende Evaluationen, die eher geringe Fallzahlen aufwiesen und nur auf der Maßnahmenebene Informationen lieferten. Aber auch in Österreich deuteten sich aufgrund von steigenden Arbeitslosenzahlen rasch Reformen der Arbeitsmarktpolitik an. Die Folgen dieser Entwicklung waren auch für die Evaluation sofort spürbar. Die Dezentralisierung der Umsetzung der Arbeitsmarktpolitik führte zum Aufbau eines umfangreichen Datensystems (Data Warehouse), vergleichbar mit dem sozioökonomischen Panel in Deutschland und dem Aufbau zeitraumbezogener und verknüpfbarer Datenbasen beim IAB.

2.2 Evaluationen im Kontext der Politik

Jenseits der Lage auf dem Arbeitsmarkt bestimmen politische Entscheidungen unmittelbar den Trend des Evaluationsgeschehens. In Deutschland war ein politischer Skandal um Statistiken der Bundesanstalt für Arbeit Auslöser einer umfangreichen Arbeitsmarktreform, die in der Folgezeit zu einem Boom der Evaluation in Deutschland führte. Die von der Bundesregierung initiierte Kommission „Moderne Dienstleistungen am Arbeitsmarkt“ – zusammengesetzt aus Persönlichkeiten aus Politik, Wirtschaft, Gewerkschaften und Wissenschaft – hat innerhalb weniger Monate einen neuen Zuschnitt für die Organisation der Arbeitsmarktverwaltung vorgelegt (Schmid 2003). Mit den Ergebnissen der Kommission wurde zudem die Vorlage geliefert, das System der Grundsicherung für erwerbsfähige Hilfebedürftige in Deutschland neu zu regeln. Die politische Grundlage für den Evaluationsauftrag wurde direkt vom Deutschen Bundestag erteilt und anschließend in einem komplexen und aufwändigen Vorhaben in verschiedenen Arbeitspaketen ausgeschrieben. Der rechtliche Rahmen sowie die seit vielen Jahren geübte Evaluationspraxis haben erheblich dazu beigetragen, die Evaluation als Bestandteil des Reformprozesses mitzudenken.

Die Schwerpunkte der Arbeitsmarktevaluation wurden in Deutschland 2002 im Arbeitsförderungsgesetz (SGB III) genauer definiert. Danach soll u.a. untersucht werden, in welchem Ausmaß die Teilnahme an einer Maßnahme die Vermittlungsaussichten der Teilnehmer verbessert und ihre Beschäftigungsfähigkeit erhöht, wie die Kosten von Maßnahmen in Relation zu ihrem Nutzen stehen und worin die volkswirtschaftlichen Nettoeffekte beim Einsatz arbeitsmarktpolitischer Instrumente bestehen. Die intensive, monatelang andauernde politische Diskussion führte letztlich dazu, dass das Auftragsvolumen dieser Evaluation bisherige programmbegleitende Evaluationen der Vergangenheit in den Schatten stellte. Zugleich lässt sich in Deutschland besonders gut studieren, dass erhebliche Dis-

krepanzen zwischen den zeitlichen Vorgaben der Politik, den Wirkungszeiträumen von arbeitsmarktpolitischen Maßnahmen sowie den für Evaluationen erforderlichen Untersuchungszeiträumen bestehen.

2.3 Evaluationen und internationale Einflüsse

Neben nationalen Faktoren spielen auf dem Gebiet der Arbeitsmarktpolitik auch zunehmend internationale Einflüsse eine Rolle. Zunächst wird durch die beschäftigungspolitischen Leitlinien der EU, verstärkt durch den Lissabon-Prozess (EU 2000), ein Augenmerk auf quantifizierbare Entwicklungen gelegt, was auch für die nationalen Arbeitsmarktpolitiken und Evaluationsaufträge nicht ohne Folgen bleibt. Zudem sind Evaluationen meist integraler Bestandteil beim Einsatz von Mitteln des Europäischen Sozialfonds, sowohl auf Programm- als auch auf Maßnahmenebene. Insbesondere Österreich ist ein Beispiel dafür, wie der Beitritt zur EU (1995) sowohl das Ausmaß als auch die Richtung der Evaluierung arbeitsmarktpolitischer Programme erheblich beeinflusst hat. Auf der Programmebene überwiegen Studien, bei denen ökonomische Fragestellungen im Fokus stehen, auf der Ebene einzelner Maßnahmen sind es eher ad hoc Evaluierungen, die in der Regel von kleineren Anbietern durchgeführt werden.

Ein weiterer Punkt, der nicht zuletzt durch den internationalen Diskurs in die nationale Politik getragen wird, ist die Forderung nach einer Weiterentwicklung von der aktiven zur aktivierenden Arbeitsmarktpolitik. Die Diskussion war zunächst damit verbunden, passive Leistungen zugunsten von aktiven umzuwidmen. Damit sollten passive Lohnersatzleistungen reduziert und durch die Finanzierung von (aktiven) Maßnahmen ersetzt werden. In der Folgezeit wurde dieser Ansatz nochmals verändert und der Akzent auf die Aktivierung und die Verantwortung des Einzelnen verschoben. Mit der Aktivierung war neben dem „Fördern und Fordern“ insbesondere eine noch stärkere Orientierung auf den ersten Arbeitsmarkt verbunden.

Das Primat der Evaluation hat sich seitdem geändert. In Österreich wurde der experimentelle Charakter verdrängt und unmittelbare Wirkungen werden stärker betont. Auch in Deutschland wurde schon durch den internationalen Benchmark einer wissenschaftlichen Autorengruppe im Rahmen des „Bündnisses für Arbeit“ (1999) diesem Trend gefolgt, indem der raschen Eingliederung Arbeitsloser durch die Schaffung eines Niedriglohnsektors das Wort geredet wurde. Fragen nach der Beschäftigungsfähigkeit und der sozialen Stabilisierung von bestimmten Personengruppen werden bei der Evaluation der Hartz-Gesetze zwar angesprochen, treten aber in den vorliegenden Berichten in den Hintergrund (Deutscher Bundestag 2006d). In dieser Tradition steht auch die Schweiz, die wiederum das Wording der OECD übernommen hat und bei Evaluierungen insbesondere die (betriebswirtschaftliche) Optimierung und die Überprüfung von Wirkungsindikatoren betont.

3 Fazit

Die Funktion der Arbeitsmarktpolitik ist in allen Ländern strittig oder zumindest nicht geklärt. Zudem dürfte es – vielleicht mit Ausnahme des Bildungssektors – keinen Politikbereich geben, in dem in den vergangenen Jahren mehr reformiert und ausprobiert wurde. Je nach gesellschaftlichen Kräfteverhältnissen wird der Arbeitsmarktpolitik entweder die Verbesserung der Passung zwischen Angebot und Nachfrage auf dem Arbeitsmarkt zugewiesen, oder sie soll den sozialen beziehungsweise regionalen Ausgleich oder die Verbesserung der Beschäftigungsfähigkeit unterstützen. Für das Evaluationsgeschehen sind die Folgen dieses Spagats evident, die (Erfolgs-)Kriterien entsprechend unterschiedlich.

Der Trend ist in allen drei Ländern insofern ähnlich, als es eine deutliche Hinwendung zu einer „aktivierenden“ oder „investierenden“ Arbeitsmarktpolitik gibt. Diese Entwicklung folgt dem internationalen Mainstream (OECD, EU), der in den vergangenen sieben bis zehn Jahren durch Leitlinien und Parameter den Weg vorgezeichnet hat. Die Sichtung internationaler Evaluierungsergebnisse zeigt, dass es in der Regel um den Vergleich von direkten und indirekten Einflüssen von messbaren Effekten der Arbeitsmarktpolitik auf die individuelle Wiedereingliederung geht (Konle-Seidl 2005). Der Kontext öffentlich finanzierter Beschäftigung und die Funktion der sozialen Stabilisierung treten in den Hintergrund.

Mit den Evaluationen der Arbeitsmarktreformen in Deutschland wurden kurzfristige Erfolge (z. B. Trainingsmaßnahmen, Fort- und Weiterbildung) in den Mittelpunkt gestellt, wohingegen für langfristige Entwicklungen (politisch) kaum Zeit eingeräumt wurde. Österreich hat sich in den vergangenen Jahren darauf konzentriert, die Datenlage für prozessorientierte Erhebungen zu verbessern, um Erwerbsverläufe besser zu rekonstruieren. Durch die Orientierung der Evaluation an diesem System treten weiter gehende Fragen nach Zusammenhängen und Wirkungsketten in den Hintergrund. In der Schweiz ist die Tradition der arbeitsmarktpolitischen Evaluation noch relativ jung und noch prägender als in den anderen beiden Ländern auf die schnelle Vermittlung orientiert. Unbehagen über diese Entwicklung ist bei WissenschaftlerInnen in allen drei Ländern spürbar.

Die Schweiz hat inzwischen die Wirkungsanalyse der Arbeitsmarktpolitik auch auf institutionelle Verflechtungen erweitert, ähnlich wie in Deutschland, wo mit einer weiteren breit angelegten Untersuchung bei der Grundsicherung für Arbeitsuchende der Einfluss der unterschiedlichen Trägerschaft (Bund oder Kommune) untersucht wird. Ein solches Vorhaben, um kontextabhängige Zusammenhänge und Wirkungen zu untersuchen, steht für Österreich noch aus.

Das in vielen Fällen ausgeklügelte und plausible System von mikro- und makroökonomischen Evaluationsstudien, breit angelegten Untersuchungen über Prozess- und Implementationsanalysen kann nicht darüber hinwegtäuschen, dass ein öffentlicher Diskurs zu führen ist, welche Funktion die aktive Arbeitsmarktpolitik hat und welches zentrale, den Evaluationen zugrunde liegende Bewertungskriterien sind. Es wird in erster Linie Aufgabe der Politik sein, mit unterschiedlichen programmatischen Ansätzen darauf Einfluss zu nehmen. Auch die Protagonisten der Evaluation – wie die DeGEval und SEVAL – sollten ihre Konzepte und internationalen Erkenntnisse bündeln und in die Diskussion einbringen.

Bildung

Evaluation im Bildungsbereich Schule in Deutschland

1 Kontext der Evaluationstätigkeit

Das deutsche Bildungswesen ist föderal und entlang von verschiedenen Schulstufen und Schulformen gegliedert. Bei gewissen Überschneidungen bestehen in Deutschland länderspezifische Unterschiede in den Angebotsstrukturen, wobei die Vielfalt durch das Selbstverwaltungsrecht der Kommunen verstärkt wird (Konsortium Bildungsberichterstattung 2006).

Entscheide über die Gestaltung des Bildungswesens liegen weitestgehend in der Kompetenz der Bundesländer. Dem gegenüber liegen die Aufgaben des Bundesministeriums für Bildung und Forschung (BMBF) unter anderem in der Rechtssetzung für die außerschulische berufliche Bildung und Weiterbildung, in der Forschungsförderung und in der Gesetzgebung zur Ausbildungsförderung. Hinsichtlich der Evaluation ist die Ständige Konferenz der Kultusminister der Länder (KMK) ein bedeutsamerer Akteur. Sie ist ein Zusammenschluss der für Bildung und Erziehung, Hochschulen und Forschung sowie kulturelle Angelegenheiten zuständigen Minister/innen beziehungsweise Senator/innen der Länder und beruht auf einem Übereinkommen der Länder. Sie stimmt die Aufgaben in diesen Bereichen, die von länderübergreifender Bedeutung sind, konsensuell und kooperativ zwischen den Bundesländern ab. Dies betrifft insbesondere auch die Thematik der Evaluation. So hat die KMK im Nachgang zu den ersten PISA-Ergebnissen (Deutsches PISA-Konsortium 2001) als eines von sieben politischen Handlungsfeldern formuliert: „Maßnahmen zur konsequenten Weiterentwicklung und Sicherung der Qualität von Unterricht und Schule auf der Grundlage von verbindlichen Standards sowie eine ergebnisorientierte Evaluation“ (Kultusministerkonferenz Deutschland 2004).

Eine Untersuchung zu den bildungspolitischen Zielsetzungen und Maßnahmen im Jahre 2007 bestätigt, dass das Thema Evaluation – und damit zusammenhängend die Themen Qualitätssicherung, Rechenschaftslegung/Kontrolle sowie Erweiterung schulischer Eigenverantwortung – zuoberst auf der Agenda der Bundesländer steht (Emmerich/Maag Merki 2007). Die aktuellen Entwicklungen im Bereich Evaluation, wie sie nachfolgend vorgestellt werden, sind somit im Spannungsfeld zwischen länderspezifischen Zielsetzungen und Maßnahmenplanungen sowie länderübergreifenden Abstimmungen und Beschlüssen zu verorten.

2 Evaluationsaktivitäten

Evaluation ist in der Erziehungswissenschaft kein neuer Begriff, sondern kann bereits auf eine jahrzehntelange Geschichte zurückblicken, auch wenn das Thema nicht immer mit gleicher Intensität und Bedeutung sowie mit gleichen Inhalten diskutiert wurde. Erstmals wurde die Breite der (schul-)evaluationstheoretischen Literatur, insbesondere in den USA und der BRD, von Beywl (1988) aufbereitet. Es lassen sich zumindest drei verschiedene Phasen der Begriffsbestimmung unterscheiden: 1) Evaluation als Systemevaluation auf der Makroebene, 2) Evaluation im Kontext qualitätsorientierter Schulentwicklung auf der Mesoebene, 3) Evaluation als Instrument des Qualitätsmanagements auf der Meso- und Makroebene.

2.1 Evaluation als Systemevaluation auf der Makroebene

In einer ersten Phase (1970er und 1980er Jahre) wird Evaluation als Systemevaluation auf der Makroebene verstanden. Dabei ging es um Fragen zur Struktur und zum Aufbau des Schulsystems, wobei unter anderem Ziel war, die Effektivität von Ganztagschulen (Ipfling 1981) oder des dreigliederten Schulsystems im Vergleich zur integrierten Gesamtschule zu untersuchen (Fend 1982; Raschert 1974). Damit sind erstmals für die bundesdeutsche Erziehungswissenschaft komplexe quantitative Verfahren in großen Stichprobenstudien (ca. 11.000 Schüler/innen) eingesetzt worden. Die Evaluationsforschung wurde in dieser Zeit in einzelnen Bundesländern mit der Gründung verschiedener Arbeitsstellen institutionell verortet, beispielsweise durch den Aufbau des „Zentrums für empirische Bildungsforschung“ in Konstanz. Vorherrschendes theoretisches Paradigma der Steuerung im Bildungswesen war die so genannte Input-Steuerung. Ein gutes Bildungssystem wird in dieser Perspektive garantiert über die Qualität und Quantität von verschiedenen Inputfaktoren wie Ausbildung der Lehrpersonen, Lehrpläne, Schulstruktur et cetera. Evaluationsforschung fokussierte damit insbesondere die Qualität dieser Faktoren.

2.2 Evaluation im Kontext qualitätsorientierter Schulentwicklung auf der Mesoebene

In einer zweiten Phase (1990er Jahre) wurde Evaluation stärker im Kontext der qualitätsorientierten Schulentwicklung auf der Mesoebene (Einzelschule) verortet. Hintergrund waren empirische Ergebnisse, die aufgezeigt haben, dass sich Schulen stärker *innerhalb* eines spezifischen Bildungssystems (z. B. Dreigliedrigkeit) unterscheiden und die Unterschiede *zwischen* verschiedenen Bildungssystemen geringer sind (Fend 1986). Hohe Schulqualität wird demnach „vor Ort“, also in den einzelnen Schulen erzeugt und weniger über die Perfektionierung von Gesetzen und Verordnungen. Somit müssen Schulen gestärkt werden, ein für das eigene Umfeld und die eigenen Schüler/innen optimales Lernangebot aufzubauen, welches sich unterscheiden kann von anderen Schulen (Rolff 1993). In diesem Kontext ist Evaluation Instrument der Organisationsentwicklung von Einzelschulen beziehungsweise lernenden Schulen (Argyris/Schön 1999) und kann dem Entwicklungsparadigma zugeordnet

werden (Kromrey 2006b). Es wurden relevante Bezüge zu theoretischen Modellen der Organisations- und Institutionsforschung hergestellt und diese auf den schulischen Kontext transferiert (Rolff 1991). Hierzu ist eine Vielzahl an Literatur über Selbstevaluation oder interne Evaluation entstanden (z. B. Buhren et al. 1999; Dalin et al. 1996; Landesinstitut für Schule und Weiterbildung 1996; Rolff 1996). Die Modellversuche dieser Zeit wurden einerseits in einzelnen Bundesländern durchgeführt (z. B. Nordrhein-Westfalen), andererseits international verankert (z. B. das Projekt INIS der Bertelsmann Stiftung; siehe Czerwanski 2003).

2.3 Evaluation als Instrument des Qualitätsmanagements auf der Meso- und Makroebene

Die dritte Phase (ab 2000) wurde beeinflusst von den großen Leistungsvergleichsstudien PISA (Deutsches PISA-Konsortium 2001) und TIMSS (Baumert et al. 2000). Neben der stärkeren internationalen Anbindung der empirischen Forschung und der Berücksichtigung international vergleichender Analysen war für die Evaluation erkenntnisleitend, dass diese sich nicht auf die Evaluation der Einzelschule beschränken kann, sondern dieser Ansatz ergänzt werden muss durch weitere, extern gesteuerte Evaluationsformen (siehe unten). Evaluation wird damit in stärkerem Ausmaß im Kontext des Kontroll- und Legitimationsparadigmas verortet (Kromrey 2006b) und hat zum Ziel, im Sinne eines Instrumentes des Qualitätsmanagements Informationen zur Qualität des Bildungssystems zu gewinnen. Gleichzeitig werden die in den 1990er Jahren eingeleiteten Entwicklungen zu mehr Gestaltungsfreiheiten für die einzelnen Schulen in allen Bundesländern in stärkerem Maße systematisiert, gesetzlich verankert und Verfahren der Rechenschaftspflicht für die Schulen eingeführt. Innerschulische und externe Evaluationsformen werden so erstmals systematisch aufeinander bezogen, wobei in einzelnen Fällen die externen Verfahren gezielt auf den internen Verfahren (Selbstevaluation) aufbauen und Vorgaben hinsichtlich der Verwendung der Evaluationsinstrumente gemacht werden. Es wird damit ein integrales Konzept von Evaluation (Klieme 2005) implementiert, welches kompatibel ist mit internationalen Entwicklungen (Döbert 2003), sich aber an gewissen Stellen (insbesondere in der Frage der Konsequenzen für die Akteure) systematisch davon abgrenzt. Kein Bundesland in Deutschland hat sich diesen Entwicklungen entzogen, wenn auch im Detail zwischen den Bundesländern Unterschiede auszumachen sind (Bos et al. 2006; Van Ackeren/Bellenberg 2004). Folgende externen Evaluationsformen können in Deutschland identifiziert werden:

- a. *Internationale Leistungsvergleichsstudien:* Die Länder nehmen im Zusammenwirken mit dem Bund (BMBF) an den internationalen Leistungsvergleichsstudien wie PISA (Deutsches PISA-Konsortium 2001) oder TIMSS (Baumert et al. 2000) teil, um die Leistungsfähigkeit des deutschen Bildungswesens im internationalen Vergleich feststellen zu können (Kultusministerkonferenz Deutschland 2006).
- b. *Überprüfung von Bildungsstandards sowie Diagnose- und Vergleichsarbeiten:* Die Länder haben sich auf gemeinsame Bildungsstandards geeinigt, welche die Schüler/innen zu gewissen Zeitpunkten in ihrer Bildungsbiographie erreichen müssen (Kultusminister-

konferenz Deutschland 2006). Basis dazu war die über die Landesgrenze hinaus rezipierte Expertise von Klieme et al. (2003). Das Erreichen der Bildungsstandards wird auf nationaler Ebene auf der Grundlage repräsentativer Stichproben wie auch in den einzelnen Bundesländern mittels zentraler Vergleichs- und Diagnosearbeiten überprüft (Van Ackeren/Bellenberg 2004). Ab 2008 werden sich die Vergleichsarbeiten der Länder an den Bildungsstandards der KMK ausrichten (Kultusministerkonferenz Deutschland 2006). Eine institutionelle Verankerung dieses Evaluationsverfahrens findet sowohl im auf nationaler Ebene eigens für die wissenschaftlich fundierte Entwicklung der Testmaterialien gegründeten „Institut für Qualitätsentwicklung im Bildungswesen“ (IQB) wie auch in den teilweise neu gebildeten Referaten und Arbeitsstellen in den Ministerien der einzelnen Bundesländer statt.

- c. *Durchführung von zentralen Abschlussprüfungen:* Um Informationen über das Ausmaß der Zielerreichung bei den Schüler/innen im eigenen Bundesland zu erhalten, Abschlüsse vergleichbar über verschiedene Schulen zu halten sowie insgesamt bessere Leistungen bei den Schüler/innen zu erreichen, werden in praktisch allen Bundesländern zentrale Abschlussprüfungen auf unterschiedlichen Schulstufen (Sekundarstufe I und II) eingeführt (Van Ackeren/Bellenberg 2004), so beispielsweise zentrale Abiturprüfungen.
- d. *Durchführung von Schulinspektionen:* Während die Einführung und Überprüfung von Bildungsstandards sowie die Durchführung von zentralen Abschlussprüfungen die Überprüfung des Outputs sowie das Generieren von Informationen zur Systemqualität zum Ziel haben, geht es bei den Schulinspektionen in erster Linie um die Überprüfung der schulischen Prozesse, also des Handelns der Lehrer/innen und Schulleitungen in der Einzelschule. Bis auf das Saarland wurden in allen Bundesländern in den letzten Jahren dazu Verfahren entwickelt (Bos et al. 2006; 2007). Die Konzepte der Bundesländer sind teilweise wenig ausdifferenziert (z. B. hinsichtlich der Konsequenzen für die Schulen nach Ablauf der Schulinspektion) und variieren in Bezug auf die eingesetzten Evaluationsinstrumente oder der institutionellen Anbindung der Evaluationsteams beträchtlich. Neben der Einführung von Bildungsstandards ist die Einführung von Schulinspektionen durch den Aufbau neuer Agenturen (z. B. das Institut für Qualitätsentwicklung in Hessen) am stärksten institutionell verankert worden.
- e. *Bildungsberichterstattung:* Die Beurteilung des Bildungssystems erfolgt in Deutschland zudem mittels einer von Ländern und Bund getragenen gemeinsamen Bildungsberichterstattung auf der Grundlage gesicherter Daten (Konsortium Bildungsberichterstattung 2006; Kultusministerkonferenz Deutschland 2003). Dabei wird die Anschlussfähigkeit an internationale Berichtssysteme, beispielsweise der Organisation for Economic Cooperation and Development (OECD), angestrebt (Kultusministerkonferenz Deutschland 2006). In einzelnen Ländern (z. B. Nordrhein-Westfalen) werde zudem regionale Bildungsberichte erstellt, die spezifisch die Situation im eigenen Bundesland darlegen.

3 Aktueller Stand der Evaluation und Bilanzierung

In den letzten Jahren kann eine deutlich stärkere Fokussierung der Thematik und der Institutionalisierung auf der Ebene des Bildungssystems und der Einzelschule festgestellt werden. Zudem zeigt der wissenschaftliche Diskurs eine erhöhte Aufmerksamkeit gegenüber der Thematik, beispielsweise hinsichtlich der Diskussion über Standards für Evaluation (DeGEval 2002) und deren Anwendung in der Selbstevaluation (DeGEval 2004b) oder für Schulevaluationen (Abs/Klieme 2005; Abs et al. 2006). Dennoch kann teilweise eine (noch) eingeschränkte Anschlussfähigkeit der Thematik an die Schulpraxis und an die Praxis der Bildungsadministration beobachtet werden:

- a. Aufgrund der raschen und zum Teil gleichzeitigen Einführung der verschiedenen Evaluationsinstrumente sind die Schulen häufig überfordert, einerseits in der Erfüllung des Auftrags zur Selbstevaluation, andererseits in der Verarbeitung der Ergebnisse der verschiedenen externen und internen Evaluationen (Hosenfeld/Groß Ophoff 2007; Ditton/Arnoldt 2004).
- b. Die Bildungsadministrationen, beispielsweise die KMK oder die Ministerien in den Bundesländern, entwickelten innert kürzester Zeit ein Konzept für die Durchführung einzelner externer Evaluationsverfahren, wofür aus wissenschaftlicher Perspektive mehr Zeit notwendig gewesen wäre. Dies kann insbesondere in der Entwicklung und Überprüfung von Bildungsstandards beobachtet werden (Klieme et al. 2003).
- c. Der Aufbau von Evaluationsagenturen für die Durchführung von Schulinspektionen fiel mit der gleichzeitigen Entwicklung des Konzeptes sowie der Rekrutierung und Schulung der Evaluationsteams zusammen. Zudem musste der zentrale Aushandlungsprozess mit wichtigen anderen Akteursgruppen, insbesondere mit der Schulaufsicht, zeitgleich geführt werden. Dabei war aber nicht im notwendigen Umfang klar, wie mögliche Konsequenzen für Schulen, die gewisse Standards nicht erreichen, aussehen könnten oder welche Mittel die Schulaufsicht zur Verfügung hat, um die Schulen und Lehrpersonen zu möglichen Handlungsänderungen zu verpflichten – aufgrund des Beamtenstatus der Lehrpersonen eine nicht zu vernachlässigende Problematik.
- d. Von der Evaluation erhoffen sich verschiedene Akteure im Bildungsbereich eine deutliche Qualitätssteigerung und das Ende der Bildungsmisere, die im Anschluss an die Veröffentlichung der PISA-Ergebnisse die Diskussion in Deutschland geprägt hat. Hierzu ist zu fragen, ob die hohen Erwartungen erfüllt werden können, insbesondere auch in der Zeit, in der sich dies die Bildungspolitik erhofft. Die Erkenntnisse aus der Schuleffektivitäts- und Schulentwicklungsforschung lassen aber vermuten, dass hier einzig langfristige Aktivitäten zu den gewünschten Ergebnissen führen werden (Fullan/Watson 2000). Zudem ist es notwendig, dass die verschiedenen offenen Baustellen produktiv bearbeitet werden, so dass die Einzelschule in der Lage sein wird, über die Verarbeitung der Ergebnisse aus internen und externen Evaluationen eine für die Schüler/innen bessere Lernumgebung anzubieten, die diese dann in optimaler Weise nutzen können. Wichtig in diesem Zusammenhang sind die Entwicklungen an verschiedenen Hochschulen in Deutschland, die zunehmend Studiengänge anbieten, in denen Evaluation zentrales Thema ist. Zudem sind an verschiedenen Orten Bestrebungen im Gang,

welche die Ausbildung der Schulleitungen professionalisieren, sodass die Schulleitungen zukünftig in der Lage sein werden, die ihnen neu zugestandenen Kompetenzen, beispielsweise im Bereich der Personal- und Organisationsentwicklung, zu übernehmen. Dazu muss aber auch die Politik die notwendigen Rahmenbedingungen, insbesondere im Bereich der Verwaltungsvorschriften, entsprechend dem implementierten Qualitätsmanagementkonzept, verändern.

Evaluation an deutschen Hochschulen – Entwicklung, Stand und Perspektiven¹

1 Vorbemerkung

Das Hochschulsystem in Deutschland ist geprägt durch die föderale Struktur und die damit verbundene Hoheit der Länder in Fragen der Hochschulausbildung. Durchschnittlich studierten an den insgesamt 383 Hochschulen in den vergangenen Jahren circa zwei Millionen Studierende, von denen etwa siebzig Prozent an Universitäten und knapp dreißig Prozent an Fachhochschulen immatrikuliert waren.²

Evaluation ist seit der Novelle des Hochschulrahmengesetzes (HRG) im Jahre 1998 gesetzlich verankert, was auch seinen Niederschlag in der weit überwiegenden Mehrzahl der Hochschulgesetze auf Länderebene findet.

Durch die Akkreditierung von Bachelor- und Masterstudiengängen erhöht sich seit Ende der 1990er Jahre die rechtliche Verbindlichkeit von Qualitätssicherungsverfahren, wenngleich ausgehend von den entsprechenden länderübergreifenden Regelungen³ die konkrete Verankerung in Landesgesetzen unterschiedlich geregelt ist.

Evaluation an deutschen Hochschulen hat mit Blick auf die Bewertung von Forschungsleistungen eine bereits längere Tradition, wogegen sie im Bereich von Studium und Lehre sowie bezüglich der Evaluation von Fächern und Fachbereichen vergleichsweise jüngeren Datums ist.

Die zunächst mit erheblichen Ressentiments versehene Debatte um Fragen der Hochschulevaluation nahm in Deutschland ihren Anfang im Kontext der deutschen Wiedervereinigung. Evaluation war hierbei weitgehend mit der Frage befasst, ob und in welcher Weise (ostdeutsche) Hochschulen (bzw. Fächer, Lehrstühle) weitergeführt werden sollen oder inwieweit sich neue Bedarfe zeigen und Synergieeffekte zwischen benachbarten Hochschulen denkbar sind. Dies führte sowohl zu Neugründungen von Hochschulen als auch zu konfliktreichen Prozessen der Konzentration einzelner Fächer. Unter anderem diese Entwicklung führte dazu, dass Evaluation an deutschen Hochschulen eng mit dem Gedanken

¹ Die für den Bereich der Hochschulevaluation relevante Literatur auch nur ansatzweise hier darzustellen, würde den zur Verfügung stehenden Rahmen überschreiten. Verwiesen sei daher ausdrücklich auf die Literaturpools der HRk (ehemals Evonet) (http://www.hrk.de/de/projekte_und_initiativen/121_4074.php [8.7.2008]) sowie des Instituts für Hochschulforschung Wittenberg, HoF (<http://www.hof.uni-halle.de/biblio/> [8.7.2008]).

² Vgl. Statistisches Bundesamt unter: <http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Navigation/Statistiken/BildungForschungKultur/Hochschulen/Hochschulen.psml> [31.8.2007].

³ Insbesondere Gesetz zur Errichtung einer Stiftung zur Akkreditierung von Studiengängen in Deutschland und Länderübergreifende Strukturvorgaben für die Akkreditierung von Bachelor- und Masterstudiengängen (vgl. hierzu <http://www.akkreditierungsrat.de/index.php?id=20&L=0> [31.8.2007]).

der „Abwicklung“ assoziiert wurde, wobei die kritische Haltung gegenüber Evaluationen auch grundsätzlicher Natur war.

Dessen ungeachtet ist im Verlaufe der 1990er Jahre ein sukzessiver Ausbau von Hochschulevaluation vor allem im Bereich von Studium und Lehre zu beobachten (vgl. zusammenfassend u. a. Barz 1998; HRK 1998; 2007). Wenngleich Evaluation im Hochschulwesen bis zur Gegenwart immer wieder kritisch diskutiert wird – sei es aus methodischen, sei es aus Gründen der Effizienz –, so lässt sich doch festhalten, dass in Deutschland Evaluation, wenn auch je nach Bundesland und einzelnen Hochschulen in unterschiedlicher Intensität, inzwischen zu einem festen Bestandteil der Hochschulwirklichkeit geworden ist.

2 Evaluationsverständnis und -verfahren

Diese grundsätzliche Akzeptanz wird allerdings begleitet von einem zum Teil diffusen, zumindest aber heterogenen Begriffsverständnis. Das Spektrum dessen, was im Kontext von Evaluation thematisiert wird, reicht von internen und externen Evaluationen über Forschungs- und Lehrberichte, Lehrveranstaltungsbefragungen, Absolventenbefragungen, Rankings und Forschungsratings bis hin zu Akkreditierungen und einem umfassenden Qualitätsmanagementsystem.

Versucht man diese unterschiedlichen Zugangsweisen zu systematisieren, so bietet sich eine Unterscheidung nach Gegenständen, dem Bezug und den Funktionen der Evaluation an. Im Hinblick auf den Gegenstand ist zwischen den Bereichen Studium und Lehre sowie Forschung, bezüglich der Form zwischen Projekt-, Programm- und institutioneller Evaluation zu unterscheiden. Funktionen der Evaluation schließlich können nach solchen der Kontrolle, der Entwicklung, der Forschung sowie der Legitimation differenziert werden.⁴

Tabelle 1: Gegenstand, Bezug und Funktion von „Qualitätssicherungsverfahren“

| Verfahren | Vorwiegender Gegenstand | Bezug | Funktion |
|--|---------------------------------|-------------------------|------------------------|
| Institutionelle Evaluation/ Peer-Review | Studium und Lehre, Forschung | Institutionen/Programme | Entwicklung |
| Lehrveranstaltungsbefragung | Studium und Lehre | Programme | Entwicklung/Kontrolle |
| Berichte | Studium und Lehre, Forschung | Institutionen/Projekte | Kontrolle/Legitimation |
| Rankings/Ratings | Studium und Lehre, Forschung | Institutionen | Kontrolle/Legitimation |
| Akkreditierung | Studium und Lehre | Programme | Kontrolle/Entwicklung |
| Projektcontrolling | Forschung | Projekte | Kontrolle/Legitimation |

⁴ An dieser Stelle müssen mit Blick auf den zur Verfügung stehenden Raum diese kursorischen Beschreibungen genügen. Vgl. hierzu weiterführend Schmidt 2005a; 2005b; Kromrey 2006; Stockmann 2006a. Letzterer differenziert nach Erkenntnis-, Kontroll-, Dialog- und Legitimitätsfunktion.

In diesem Zusammenhang von Qualitätssicherungsverfahren zu sprechen, deutet an, dass einige der Verfahren keine Evaluationen im eigentlichen Sinne sind. Selbst Lehrveranstaltungsbefragungen und Hochschulrankings entsprechen nur in Teilen einem strengen Verständnis von Evaluation. Weitgehend unumstritten sind hingegen sogenannte interne und externe Evaluationen beziehungsweise peer-reviews, die – wenn auch in unterschiedlicher Ausgestaltung – standardmäßig an vielen Hochschulen durchgeführt werden. Dies entspricht auch dem Selbstverständnis der Evaluationsverbünde in Deutschland, die auf Verfahren der internen und externen Evaluation rekurrieren. Und auch die Akkreditierungsverfahren sind in Deutschland mit Blick auf die Verfahrensschritte eng an interne und externe Evaluationen angelehnt. Wenn auch die Verfahrensgestaltung auf den ersten Blick in allen Verbünden gleich erscheint, zeigen sich bei näherem Hinsehen Unterschiede. Diese lassen sich im Wesentlichen im Hinblick auf die Einbindung des Verbundes beziehungsweise der Agentur im Rahmen der internen Evaluation, die Form der externen Evaluation, die Beteiligung an der Implementierung und Begleitung von Zielvereinbarungen und Maßnahmen sowie den Gegenstand der Evaluation festmachen.

Tabelle 2: Schwerpunkte der Evaluationsverfahren in Agenturen und Verbünden

| Verbund/Agentur | Interne Evaluation | Externe Evaluation | Beteiligung Maßnahmen | Gegenstand |
|---|-------------------------------|---------------------------|--|-------------------------------|
| Zentrale Evaluations- und Akkreditierungsagentur (ZEVA) | Leitfaden Selbstreport | Externe Fachgutachter | In der Regel liegt die Umsetzung in Händen der Hochschulen | Studium und Lehre |
| Verbund Norddeutscher Universitäten | | | | |
| Evaluationsagentur Baden-Württemberg (evalag) | | | | |
| Hochschulevaluierungsverbund | Stärken- und Schwächenanalyse | | Moderation der Zielvereinbarung | Forschung, Studium und Lehre, |
| Evaluationsnetzwerk Wissenschaft (ENWISS) | Organisation der Vernetzung | Gegenseitige Begutachtung | Umsetzung in Hochschulen | Studium und Lehre |
| Verbund Technischer Universitäten | Leitfaden Selbstreport | | | Forschung, Studium und Lehre, |
| Wissenschaftliche Kommission Niedersachsen | | Externe Fachgutachter | Umsetzung durch Ministerium und Hochschule, beratende Unterstützung durch WK | Forschung |

3 Evaluationseinrichtungen

Dieser Überblick deutet bereits an, dass Hochschulevaluation mit Bezug auf die Tätigkeit von Verbünden innerhalb einzelner Bundesländer eine sehr unterschiedliche Verbreitung erfährt. Eine lange Tradition ist mit der Zentralen Evaluations- und Akkreditierungsagentur sowie der Wissenschaftlichen Kommission für Niedersachsen zu konstatieren. Auch Baden-Württemberg (evalag), Rheinland-Pfalz, das Saarland und Hessen (Hochschulevaluierungsverbund, ENWISS) sowie Hamburg, Bremen und Teile von Mecklenburg-Vorpommern sind im Bereich Evaluation breit aufgestellt,⁵ wohingegen die Mehrzahl der ostdeutschen Bundesländer sowie Bayern und die Universitäten in Nordrhein-Westfalen über keine ausgewiesenen Verbünde verfügen, wohl aber zum Teil über Evaluationseinrichtungen in einzelnen Hochschulen.⁶

Insgesamt ist nicht zuletzt unter der Maßgabe von Akkreditierungsauflagen zu beobachten, dass der Ausbau von qualitätssichernden Einrichtungen innerhalb von Hochschulen zunimmt, wenn auch vielfach noch unklar ist, in welcher Funktion diese Einrichtungen oder einzelne Akteure im Status von Qualitätssicherungsbeauftragten gesehen werden beziehungsweise mit welchen Befugnissen sie innerhalb der Hochschule ausgestattet sind. In diesem Zusammenhang wäre interessant zu wissen, wie hoch die Anzahl derjenigen an Hochschulen ist, die sich in einem engeren Sinne beruflich mit Fragen der Evaluation befassen. Augenfällig ist, dass allmählich die einzelnen Instrumente der Qualitätssicherung zusammengeführt werden sollen und zunehmend eine sinnvolle Koppelung von Akkreditierung, Evaluation und weiteren Steuerungsinstrumenten versucht wird. Die vom Projekt Q – einem durch die HRK initiierten und getragenen Projekt, das in erster Linie die Funktion hat, ein Forum des Austauschs und der Wissensakkumulation für die Evaluation in Hochschulen auszubilden – initiierte Veranstaltungsreihe „Von der Qualitätssicherung der Lehre zur Qualitätsentwicklung als Prinzip der Hochschulsteuerung“ deutet diese Entwicklung an (vgl. HRK 2006a). Dem entspricht, dass die in den vergangenen Jahren dominante Programmakkreditierung – so die neuesten Beschlüsse von KMK und Akkreditierungsrat – ergänzt wird durch Verfahren, die auf eine Integration unterschiedlicher Qualitätssicherungsinstrumente abstellen (vgl. HRK/ACQUIN 2007).⁷ Verbunden hiermit wird – so ist zu vermuten – eine neue Rolle von Akkreditierungsagenturen,⁸ Evaluationsverbünden und den jeweiligen hochschulinternen Einrichtungen sein, die sich nicht zuletzt an den im Rahmen des Bologna-Prozesses entwickelten Standards orientieren muss (HRK 2006b), um unter

⁵ Vgl. zur Arbeit der einzelnen Verbünde und Agenturen zusammenfassend das Portal der (http://www.hrk.de/de/projekte_und_initiativen/121_4078.php [8.7.2008]). Legt man die Evaluationsverbünde zugrunde, so ist davon auszugehen, dass ca. zwischen einem Drittel und der Hälfte der deutschen Hochschulen mehr oder weniger intensiv in die Arbeit von Evaluationsagenturen eingebunden sind.

⁶ So z. B. die Universitäten Bielefeld, Bonn und Duisburg-Essen. Der Evaluationsverbund der Universitäten Halle, Leipzig und Jena wurde inzwischen aufgelöst und durch die Evaluationstätigkeit an den jeweiligen Hochschulen selbst ersetzt.

⁷ Vgl. zum Projekt Systemakkreditierung an der Universität Mainz http://www.zq.uni-mainz.de/sys_akk [31.8.2007].

⁸ Zur Arbeit der Akkreditierungsagenturen und des Akkreditierungsrates vgl. u. a. <http://www.akkreditierungsrat.de> [31.8.2007].

anderem die zur Zeit virulente Frage der gegenseitigen Anerkennung von Studienleistungen zwischen einzelnen Hochschulen zu regeln.

Gleichzeitig aber ist nach wie vor eine nicht unerhebliche, auch institutionelle Differenzierung zwischen Evaluationen auf Ebene von Fächern und Fachbereichen mit Schwerpunkt auf dem Bereich der Lehre und in der Forschung zu beobachten. Zeichnet man eine Landkarte der Forschungsevaluation, so sind neben den bereits genannten Agenturen und Verbänden weitere institutionelle Akteure, wie der Wissenschaftsrat⁹ und die jeweiligen Evaluationseinrichtungen der Leibniz-Gemeinschaft, der Helmholtz-Gesellschaft, der Max-Planck-Gesellschaft, der Fraunhofer-Gesellschaft und einiger Bundesforschungseinrichtungen sowie das jüngst von der DFG gegründete Institut für Forschungsinformation und Qualitätssicherung (IFQ) zu nennen.¹⁰ Diese institutionelle Trennung zwischen Einrichtungen, die sich in erster Linie mit der Evaluation von Studium und Lehre befassen und solchen, die sich schwerpunktmäßig der Forschungsevaluation widmen, entspricht den zu weiten Teilen voneinander entkoppelten Diskursen in der Hochschulevaluation, was unter anderem auch in der Differenzierung der Arbeitskreise ‚Hochschule‘ sowie ‚Forschung und Technologietransfer‘ innerhalb der DeGEval zum Ausdruck kommt (vgl. den Beitrag von Stefan Kuhlmann in diesem Band).

4 Perspektiven der Hochschulevaluation in Deutschland

Die fehlende Verknüpfung von Forschungs- und Lehrevaluation sowie der Evaluationsverfahren an Hochschulen insgesamt lässt sich als ein Indiz für eine Evaluationstradition werten, der es bislang nicht gelungen ist, wesentliche Elemente und Verfahren der Evaluation, der Hochschul- und Organisationsentwicklung sowie der Hochschulsteuerung zusammenzuführen (vgl. Hennen/Häuser 2002). Diese fehlende Integration hat unterschiedliche Ursachen, auf die an dieser Stelle nicht im Einzelnen eingegangen werden kann. Kennzeichnend jedoch ist an vielen Orten eine strukturelle Schwäche von Evaluation, die an das mehr oder weniger starke Engagement der mit Evaluation befassten Akteure gebunden ist. Die in vielen Hochschulen nach wie vor fehlende institutionelle Verankerung von Evaluation führt mithin auch dazu, dass bereits vorhandene qualitätssichernde Potentiale nicht zusammengeführt werden. Der durch die anstehenden Akkreditierungen und Reakkreditierungen zunehmende Druck, interne Strukturen der Qualitätssicherung zu implementieren, könnte für einige Hochschulen ein weiterer Impuls sein, nochmals die Frage der jeweiligen Struktur im Bereich von Qualitätssicherung und Hochschulentwicklung aufzugreifen. Die Projekte zur Prozess- und Systemakkreditierung zumindest legen solche Perspektiven nahe. In diesem Zusammenhang wird auch zu diskutieren sein, wie Evaluation strukturell innerhalb von Hochschulen eingebunden ist und insbesondere, in welchem Maße den Evaluierenden

⁹ Wenngleich der Wissenschaftsrat auch institutionelle Evaluationen an privaten Hochschulen durchführt, die sich sowohl auf Forschung als auch auf Lehre beziehen.

¹⁰ Vgl. hierzu <http://www.forschungsinfo.de> [31.8.2007].

eine strukturelle Unabhängigkeit zugestanden wird, die Bewertungen auf unterschiedlichen Ebenen der Institution sowie der einzelnen Prozesse zulässt.¹¹

Ein in gleicher Weise grundsätzliches Problem lässt sich in der deutschen Hochschulevaluation an der bislang fehlenden Berücksichtigung sowie Verbindung unterschiedlicher Qualitätsebenen festmachen. Betrachtet man die typisierende Differenzierung nach Struktur-, Prozess- und Ergebnisqualität, so ist zu beobachten, dass die deutsche Hochschulevaluation – unabhängig von den eingesetzten Verfahren – im Bereich von Studium und Lehre wesentlich auf Fragen der Prozessqualität fokussiert. Zwar werden die Ergebnisse, wie beispielsweise Fachstudiendauer und Studienabbruchquoten, als Kontext in die Betrachtung einbezogen, doch bleibt die Analyse auf die Prozesse der Vermittlungsqualität, der Lehrorganisation und so weiter begrenzt. Insbesondere – und dies wird meines Erachtens die Hochschulevaluation in den kommenden Jahren nachhaltig beschäftigen – fehlt es an kausalen Erklärungen hinsichtlich des Zusammenhangs zwischen den erzielten Ergebnissen und den evaluierten Prozessen.

Tendenziell entgegengesetzt ist die Entwicklung in der Forschungsevaluation. Hier lässt sich – wenn auch mit vielfältigen Brüchen – die Ergebnisqualität über etablierte Indikatoren in Form von Publikationen, Promotionen und Drittmitteln darstellen, doch es fehlt ebenfalls an einer kausalen Verknüpfung zur Prozess- und Strukturanalyse, die beispielsweise Aussagen darüber zulassen, welche Prädiktoren relevant sind für die Qualität von Forschungsleistungen oder inwiefern beispielsweise Vernetzung von Forschung tatsächlich den avisierten Mehrwert mit sich bringt.

Eine weitere für die deutsche Hochschulevaluation in Zukunft zentrale Frage wird sein, ob und inwieweit es gelingt, einen Qualitätsbegriff einzuführen, der nicht mehr nur darauf rekurriert, dass eine verbindliche Definition von Qualität kaum herzustellen sei und entsprechend das den Evaluationen zugrunde liegende Qualitätsverständnis zum einen auf Plausibilitätsannahmen beruht und daraus folgernd zum anderen in seinen Grundgedanken nicht expliziert und transparent gemacht werden kann. Angesprochen ist hiermit das weitgehende Fehlen einer theorie- oder zumindest modellgestützten Evaluation (Schmidt 2005a). Zu beobachten ist in diesem Zusammenhang, dass im Bereich der Hochschulevaluation beispielsweise die Joint-Committee-Standards (Joint Committee 2006) oder Standards der DeGEval (DeGEval 2002) eine untergeordnete Rolle spielen. Die Standards and Guidelines for Quality Assurance in the European Higher Education Area wurden ohne direkten Bezug auf diese Veröffentlichungen entwickelt (European Association for Quality Assurance in Higher Education 2005).

¹¹ Die Erfahrungen des Modellversuchs Systemakkreditierung an der Universität Mainz zeigen auf, dass eine strukturelle Unabhängigkeit der Evaluationsfunktion von Administration sowie Hochschulleitung erreicht werden konnte, indem das Zentrum für Qualitätssicherung und -entwicklung (ZQ) als fachübergreifende wissenschaftliche Einrichtung errichtet worden ist.

5 Zusammenfassung

Fasst man den hier nur kursorisch dargestellten Stand der deutschen Hochschulevaluation zusammen, so lassen sich folgende Thesen festhalten:

1. Hochschulevaluation erfährt in Deutschland insgesamt eine breite Akzeptanz in dem Sinne, dass ihre Existenz nicht grundlegend in Frage gestellt wird.
2. Nach wie vor prägend für die Hochschulevaluation sind die jeweiligen Verbünde – eine Entwicklung, die durch die Einrichtung von Akkreditierungsagenturen nochmals verstärkt wurde. Die Entwicklung elaborierter hochschulinterner Qualitätssicherungssysteme ist mit wenigen Ausnahmen erst allmählich zu beobachten.
3. Die deutsche Hochschulevaluation zeichnet sich durch eine hohe Differenzierung der eingesetzten Verfahren sowie der evaluierten Gegenstände und beteiligten Akteure aus. Insgesamt ist ein anhaltendes Defizit hinsichtlich der Koppelung von Evaluationsverfahren, daraus abzuleitenden Maßnahmen und Hochschulentwicklung zu resümieren.
4. Hiermit einher geht eine weitgehend fehlende theoretische Fundierung von Evaluation an Hochschulen.
5. Die gegenwärtige Diskussion macht insofern zuversichtlich, als im Anschluss an den breiten Aufbruch in den vergangenen zwei Jahrzehnten verstärkt darauf abgestellt wird, die bislang fehlenden Verbindungen beispielsweise zwischen Evaluation, der Steuerung von Hochschulen sowie den analytischen Qualitätsebenen herzustellen. Diese Entwicklung – so bleibt zu hoffen – könnte auch zu einer besseren Verknüpfung von Akkreditierung und Evaluation sowie einzelner Evaluationsverfahren führen.

Österreich: Evaluation im Bereich der Schule

1 Kontext

Das österreichische Schulwesen ist wie das der anderen deutschsprachigen Länder in sich gegliedert. Im Bereich der Schule der 10-14-jährigen gibt es im Wesentlichen zwei Schulzweige, die Hauptschule und die „Allgemeinbildenden Höheren Schule (AHS)“. Letztere stellt das Äquivalent des deutschen Gymnasiums dar. Ein mittlerer Zweig wie die Realschule fehlt in der österreichischen Schule.¹

Das österreichische Schulwesen ist weitaus zentralistischer verwaltet als das deutsche oder das schweizerische. Grundsatzentscheidungen werden im Parlament in Wien getroffen und gelten für das ganze Land. Auf der anderen Seite zeigen sich auf der Schulebene Vielfalt und Beweglichkeit, weil seit 1993 viele der notwendigen Veränderungen unter dem Stichwort der Autonomisierung in die Verantwortung der Einzelschulen gestellt wurden (Pelinka 1996).

Das Spannungsfeld zwischen politischem Zentralismus und pädagogischer Autonomie der Schulen prägt auch den Bereich der Evaluation.

1.1 Evaluation

Im österreichischen Schulsystem dominiert bis heute reine Inputsteuerung durch gesetzliche Vorgaben, zentrale Lehrpläne, Ressourcen und eine einheitliche Lehrerbildung. Dieses Modell beruht auf den Annahmen,

- dass durch zentrale Regelwerke und Ressourcensteuerung die Schulwirklichkeit in relativ engen Grenzen standardisierbar ist und
- dass folglich auch die Ergebnisse des Bildungsprozesses direkt von den Steuerungsakten der Zentrale beeinflusst werden.

Diese Annahmen wurden bislang nicht systematisch überprüft: Bis vor Kurzem war vollständig unbekannt, ob und inwieweit die erwünschten Lernergebnisse tatsächlich erreicht werden, wie groß dabei die Unterschiede zwischen Schülern, Klassen, Schulen und Regionen sind, und wie das System auf der Ergebnisseite auf neue Gesetze, Veränderungen der Lehrpläne und der Ressourcen reagiert.

Evaluation kommt in der Schulgesetzgebung vor allem in zwei Zusammenhängen vor:

¹ Siehe <http://www.bmukk.gv.at/schulen/bw/uebersicht/Bildungswege4541.xml> [6.7.2007].

- zum einen als Aufgabenfeld der Schulaufsicht, besonders im Sinne der Rechtsaufsicht, aber auch der Prozesskontrolle über Unterricht und Schulentwicklung sowie als Metaevaluation der Aktivitäten der Selbstevaluation an den Schulen,²
- zum anderen im Zusammenhang mit Schulversuchen, die „zu betreuen, zu kontrollieren und auszuwerten“ sind. (§ 7 Abs. 6 Schulorganisationsgesetz)

Speziell im Zusammenhang mit der Auswertung von Schulversuchen gibt es eine ausgeprägte Tradition empirischer Evaluationsforschung, die in die 1970er Jahre zurück reicht.

2 Evaluationsaktivitäten

2.1 Historische Entwicklung

Der Beginn systematischer Evaluation im Schulbereich erfolgte zusammen mit den Ansätzen zu einer Systemreform in den 1970er Jahren. Ähnlich wie in Deutschland wurden im Bemühen um höhere Bildungsbeteiligung und verbesserte Chancengleichheit groß angelegte Schulversuche für eine „Gemeinsame Schule der 10-14jährigen“ (Gesamtschule) durchgeführt. Zur wissenschaftlichen Begleitung und Evaluation dieser Schulversuche wurde mit dem „Zentrum für Schulversuche und Schulentwicklung“ ein eigenes Institut gegründet.

Die Themen der Evaluation entsprachen weitgehend jenen der Großuntersuchungen, die zur selben Zeit in Deutschland von den Forschergruppen um Helmut Fend (1982) behandelt wurden. Es ging um Systemvergleiche in Bezug auf Schulleistungen, Durchlässigkeit und Mobilität, soziale Chancengleichheit sowie überfachliche Lernergebnisse.

In den 1970er und 1980er Jahren erschienen mehrere Bände mit Ergebnissen dieser Evaluationsstudien (Petri 1984a). Obwohl diese durchaus positive Effekte der integrativen Systemvariante nachwiesen, fanden sie kaum Berücksichtigung in den bildungspolitischen Entscheidungen. Die Kommunikation zwischen bildungspolitischen und wissenschaftlichen Instanzen war zu schwach ausgeprägt, Traditionen einer wissenschaftsgestützten Bildungspolitik fehlten weitgehend. Zudem fehlte der an dem regierungsnahen Institut angelagerten Evaluationsforschung die nachdrückliche Legitimation durch die Scientific Community. So mündeten die Gesamtschulversuche in eine „kleine Reform“ der Hauptschule (äußere Leistungsdifferenzierung), während der Bestand des Gymnasiums (AHS) unangetastet blieb.

Das „Zentrum für Schulversuche und Schulentwicklung“ blieb als Einrichtung zur Evaluation von Schulversuchen auf nationaler Ebene bestehen. Es profilierte sich mit zahlreichen Studien zur Evaluation von schulischen Neuerungen und blieb lange Zeit der einzige Ort, an dem zu schulischen Themen systematisch evaluiert wurde.

Wichtige Evaluationsuntersuchungen der Folgezeit betrafen unter anderem:

- die Neugestaltung der gymnasialen Oberstufe (Petri 1983; 1984b; 1985),

² Siehe Bundesministerium für Unterricht und Kunst (1999). Die hier beschriebene Aufgabe der Metaevaluation ist insofern paradox, als den Schulen selbst keine formellen Evaluationspflichten übertragen sind.

- die Schulversuche zum gemeinsamen Unterricht behinderter und nichtbehinderter Schüler (Gruber/Petri 1989; Specht 1993; 1997),
- die Effekte der Erweiterung von Autonomie und Eigenständigkeit der Schulen (Bachmann et al. 1996)

Bildungspolitisch am einflussreichsten waren die Studien zur Integration von behinderten Kindern. Sie trugen dazu bei, dass 1993 für die Grundschule und 1997 für die Sekundarstufe I der gemeinsame Unterricht vom Schulversuch in das Regelschulwesen übergeführt wurde.

2.2 Evaluation der Schulen

Nach dem Scheitern der Systemreform im Sinne der integrierten Gesamtschule veränderten sich Inhalt und Stellenwert der Evaluation.

1993 wurde die Erweiterung der Handlungs- und Entscheidungsbefugnisse der Einzelschulen („Schulautonomie“) im Nationalrat beschlossen. Dieser Beschluss folgte einem internationalen Trend, das Qualitätspotential standortbezogener Entwicklungsarbeit zu nutzen. Die „Innere Schulreform“ im Sinne der Gestaltung der „Schule als pädagogischer Handlungseinheit“ (Fend 1986) wurde zum bildungspolitischen Programm, „Schulentwicklung als Entwicklung von Einzelschulen“ (Rolff 1991) zum handlungsleitenden Paradigma.

In Österreich wurde diese Entwicklung durch zwei Gutachten von Peter Posch und Herbert Altrichter befördert (Posch/Altrichter 1993; 1997). Unter dem Titel „Möglichkeiten und Grenzen der Qualitätssicherung und Qualitätsentwicklung im Schulwesen“ plädierten die Autoren für ein System der Qualitätssicherung im Schulwesen, das primär auf schulinterne Formen des Qualitätsmanagements setzt. Dabei spielten die Konzepte des Schulprogramms (als schulinterne Entwicklungsplanung) und der Selbstevaluation in schulischen Handlungsfeldern eine zentrale Rolle.

Das so genannte „Qualitätsgutachten“ von Posch und Altrichter (1997) war insofern einflussreich für die Evaluationspraxis, als es klar Stellung bezog

- für interne gegenüber externen Formen der Evaluation der Schulen,
- für qualitative gegenüber quantitativen Methoden,
- für prozess- gegenüber ergebnisorientierten Analysekrterien.

Obwohl diese Empfehlungen nie rechtliche Verbindlichkeit erlangten, entfalteten sie an den Schulen doch beträchtliche Wirkungen. Vor allem im systeminternen Wettbewerb um Ansehen in der Bevölkerung und um Schüleranmeldungen versuchte eine wachsende Zahl an Schulen, unterschiedliche Formen der Selbstevaluation intern umzusetzen und nach außen zu dokumentieren.

Zur Abdeckung des Unterstützungs- und Beratungsbedarfs entstanden, teils unterstützt durch das Unterrichtsministerium, Qualifizierungsinitiativen für Schulentwicklung

und Evaluation,³ und eine Vielfalt an praxisorientierten Publikationen (Radnitzky/Schratz 1999; Schratz et al. 2000; Altrichter et al. 2004). Hingewiesen sei auf die bis heute auch international bekannte Website „Qualität in Schulen“ (siehe <http://www.qis.at>), die Schulen in Entwicklungsprozessen und Evaluationsmaßnahmen unterstützt.

2.3 Entwicklungstendenzen zu evidenzbasierter Steuerung in Bildungspolitik und Schulentwicklung

Gegen Ende der 1990er Jahre haben die internationalen Leistungsstudien (TIMSS, PISA) die Entwicklung neuer Sichtweisen auf Ziele und Funktionen bildungsbezogener Evaluation eingeleitet (Specht 2007: 206-210).

Zunehmend setzte sich die Erkenntnis durch, dass

- sich Evaluation im Schulwesen nicht auf die Ebene der Einzelschulen beschränken darf, sondern die unterschiedlichen Systemebenen adressieren muss;
- Prozessinformationen („Schulentwicklung“) keine hinlängliche Auskunft über die Qualität von Ergebnissen geben;
- die Evaluationsfunktion ein Maß an Professionalität erfordert, das an den Schulen meist nicht vorhanden ist;
- die Kernfrage der Qualität im Schulwesen jene nach den Lernergebnissen und den erworbenen Kompetenzen der Schüler ist.

Dieser Wandel wurde befördert durch eine internationale Tagung, die im Februar 1999 in der Hundertwasser-Therme im steirischen Blumau stattfand (Grogger/Specht 1999), sowie durch ein zweites, ebenfalls vom Bildungsministerium in Auftrag gegebenes Gutachten (Eder 2002), an dem mehrere empirisch arbeitende Bildungsforscher mitwirkten, die für ein Gesamtkonzept der Qualitätsentwicklung auch Systemdaten (Monitoring, Assessments) und die Berücksichtigung von Ergebniskriterien einforderten. Vor drei Jahren hat eine Reformkommission ihren Abschlussbericht vorgelegt, in dem Maßnahmen und Strategien der Evaluation auf allen Systemebenen eine wichtige Rolle spielen (Haider et al. 2005).

Wichtige Konsequenzen dieser Entwicklung sind bis heute:

1. Entwicklung und Überprüfung von Bildungsstandards:

Derzeit werden in Österreich Bildungsstandards für die Fächer Deutsch, Mathematik und Englisch und die Zeitpunkte der vierten und achten Jahrgangsstufe in groß angelegten Pilotphasen erprobt (Lucyshyn 2006). Vorgesehen ist, dass bis zum Ende des Jahrzehnts stichprobenbasierte Testungen der Schülerkompetenzen erfolgen, deren Ergebnisse sowohl den Steuerungsebenen, als auch den Schulen und Schulklassen rückgemeldet werden. Dies kann als Schritt in Richtung ergebnisorientierter Steuerung auf allen Ebenen angesehen werden.

³ Die wohl wichtigste Qualifikationsagentur ist der Verein EOS: <http://www.eos.at/index.html> [6.7.2007]. Siehe auch <http://www.pi-wien.at/schulentwicklungaps/seb.htm> [6.7.2007] für einen Überblick über Ziele und Selbstverständnis der Schulentwicklungsberatung in Österreich.

2. Nationale Bildungsberichterstattung:

Derzeit ist – teilweise den Vorbildern in Deutschland und der Schweiz folgend – ein nationaler Bildungsbericht in Erarbeitung, der Wissensgrundlagen für eine evidenzbasierte Bildungspolitik bereitstellen soll. Datengrundlage sind zum einen aufbereitete Ergebnisse der Bildungsstatistik und der vorliegenden internationalen Vergleichsstudien. In einem analytischen Teil werden zum anderen einzelne Themen- und Problemfelder forschungs- und evaluationsbasiert dargestellt und bildungspolitische Optionen erarbeitet. Es ist geplant, spezielle Programme fokussierter Evaluation zu einzelnen Schwerpunktthemen gezielt zu fördern.

3. Externe Evaluation von Schulen:

Aus der Erfahrung heraus, dass das schulische Lehrpersonal mit der Durchführung sachgerechter Evaluation der eigenen Praxis häufig überfordert ist, treten neben die bisher dominierenden Formen interner Selbstevaluation zunehmend Projekte und Verfahren professioneller externer Schulevaluation. Sie werden entweder von Schulen beziehungsweise Schulbehörden in Auftrag gegeben oder auf Initiative der forschenden Stellen selbst durchgeführt. Beispiele sind der „Modulansatz zur Selbstevaluation von Schulentwicklungsprojekten (MSS)“, der an der Universität Salzburg entwickelt wurde (Riffert/Paschon 2005), das Projekt EVIST am Pädagogischen Institut Oberösterreichs (Engleitner et al. 1998) oder das aktuelle Projekt „Schule BEWUSST“ in der Steiermark (Specht/Grabensberger 2007). Hemmend wirkt, dass den Schulen keine Finanzmittel für externe Evaluationsstudien zur Verfügung stehen.

3 Aktueller Stand und Bilanz

Evaluation im Bereich der Schule befindet sich in Österreich in einem deutlichen Wandel, namentlich unterhalb der formellen und gesetzlichen Ebene. Dieser lässt sich exemplarisch in vier Punkten beschreiben.

3.1 Institutionalisierung

Das gesteigerte Interesse an professionellen Formen der Evaluation auf allen Ebenen des Schulsystems führt derzeit zu neuen Institutionalisierungen. Das ehemalige „Zentrum für Schulentwicklung“ wird in ein neues Bundesinstitut übergeführt und aufgewertet. Es soll sich zum institutionellen Kern der Evaluation im Schulbereich entwickeln und dabei nicht nur selbst Aufträge durchführen, sondern Projekte an Universitäten, Pädagogischen Hochschulen und anderen Einrichtungen beauftragen und unterstützen. Ziel ist die Förderung einer Kultur evaluationsgestützter Entwicklung im gesamten Schulwesen.

3.2 Implementationsstudien

Die Aufwertung der Evaluationsfunktion dürfte auch damit zusammenhängen, dass sich ein neuer Typus von Evaluationsstudien jenseits klassischer Prozess- oder Produktevaluation entwickelt, der von Bildungspolitikern und Programmplanern als besonders wertvoll empfunden wird und am ehesten als Implementationsforschung umschrieben werden kann: In welcher Weise können zentral geplante Reformen umgesetzt werden, so dass erwünschte Effekte maximiert und unerwünschte Nebenwirkungen minimiert werden? Solche Evaluationsstudien sind im Zusammenhang mit der Implementation von Bildungsstandards (Freudenthaler/Specht 2006) und mit der Einführung eines neuen Lehrerdienstrechts (Seel et al. 2006) realisiert worden.

3.3 Überwindung alter Antagonismen

Im Zuge dieser Entwicklungen beginnen sich alte Antagonismen aufzulösen, welche die mit der Evaluationsfunktion verbundenen Wissenschaftstraditionen in Österreich über lange Zeit bestimmt (und ihre Wirksamkeit gelähmt) haben: Datengestützte Evaluation wurde im Spannungsfeld zu produktiver Entwicklungsorientierung gesehen, Prozessorientierung im Gegensatz zu Ergebnisorientierung, quantitativ-statistische Methoden als tendenziell unvereinbar mit qualitativ-ethnographischen Zugängen.⁴

Diese Positionen beginnen sich aneinander anzunähern. Insbesondere die Kontroverse über den Stellenwert diagnostischer Daten und ihre Entwicklung stimulierende Funktion beginnt sich aufzulösen. „Messen und Entwickeln“ wird nicht mehr als antagonistisches Verhältnis begriffen (Altrichter/Buhren 1997). Neuere Sichtweisen betonen eher Komplementarität diagnostischer Evaluationsdaten und zielgerichteter Entwicklungsprozesse.

3.4 Integrative Qualitätssicherung und Qualitätsentwicklung

Diese Rückkehr zu einem funktionalen Pragmatismus erleichtert die Realisierung von Konzepten der Qualitätssicherung und Qualitätsentwicklung, die in zweierlei Hinsicht als integrativ zu beschreiben sind. Sie verbinden Daten und Entwicklungsstrategien und fügen diese auf unterschiedlichen Handlungsebenen zusammen.

Ein solches Konzept ist derzeit im berufsbildenden Schulwesen unter dem Kürzel QIBB⁵ in einem fortgeschrittenen Stadium: Ausgehend von einer konzeptuellen Qualitätsmatrix werden auf allen Handlungsebenen (Unterricht, Schule, Region, System) Qualitätsdaten erhoben und rückgemeldet und dienen als diagnostische Ausgangspunkte für Initiativen der Weiterentwicklung. Gleichzeitig werden die Daten in aggregierter Form auf den übergeordneten Systemebenen als Steuerungsinformationen genutzt. Das Ganze ist einge-

⁴ Siehe dazu beispielhaft die Kontroverse zwischen Altrichter (1999a; 1999b) und Leutner (1999) in Thonhauser/Patry (1999).

⁵ Qualität im Berufsbildenden Schulwesen; siehe <http://www.qibb.at> [6.7.2007].

bettet in ein integriertes Berichtswesen, mit dem die Entwicklung für alle Handlungsebenen transparent gehalten wird. QIBB ist (noch) ein Modellprojekt, dem Vorbildfunktion für formelle Verfahrensregeln im gesamten Schulwesen zukommen könnte.

Evaluation im österreichischen Hochschulsystem

1 Das österreichische Hochschulsystem

1.1 Überblick

Bis Mitte der 1990er Jahre war das österreichische Hochschulsystem durch öffentliche Universitäten und Hochschulen geprägt. Die *öffentlichen Universitäten* erhielten mit dem Universitätsgesetz 2002 den Status vollrechtsfähiger Einrichtungen. Eine Erweiterung und Diversifizierung des österreichischen Hochschulsystems wurde durch die Schaffung von *Fachhochschulen* auf Grundlage des Fachhochschul-Studiengesetzes 1993 und von *Privatuniversitäten* auf der Grundlage des Universitäts-Akkreditierungsgesetzes 1999 erzielt. Mit der Donau-Universität Krems verfügt Österreich über eine staatliche *Universität für postgraduale Weiterbildung*, die in ihren Strukturen den öffentlichen Universitäten weitgehend entspricht. Zuletzt erhielten die österreichischen *Pädagogischen Akademien* den Hochschulstatus.¹

An den 22 öffentlichen Universitäten studieren rund 218.000 Personen (rund 80 % der Studierenden). An 18 Fachhochschulen beziehungsweise FH-Trägereinrichtungen werden fast 26.000 Studierende (rund 10 % der Studierenden) ausgebildet. Die Pädagogischen Hochschulen bilden etwa 13.600 Personen (rund 5 % aller Studierenden) aus. Derzeit bestehen zehn akkreditierte Privatuniversitäten, an welchen 2006 rund 3.600 Studierende inskribiert waren. Zusätzlich nehmen rund 10.000 HörerInnen universitäre Weiterbildungsangebote in Anspruch.

1.2 Stellenwert von Evaluierung und Qualitätssicherung

Die Evaluierung von Lehre und Forschung übernimmt seit Mitte der 1990er Jahre eine zunehmend wichtige Aufgabe für die Hochschulorganisation und Hochschulentwicklung. Evaluationsmaßnahmen waren anfangs eher durch spezifische Anlässe und einzelne Initiativen begründet als durch den Willen zur Schaffung eines systematischen Rahmens der Qualitätssicherung und Evaluierung.

Die Praxis der Qualitätssicherung und Evaluierung steht in einem engen Zusammenhang mit der Entwicklung der einzelnen Hochschulsektoren. So erforderte die Schaffung

¹ Vgl. die Bundesgesetze über die Organisation der Universitäten und ihre Studien (Universitätsgesetz 2002), BGBl. I 2002/120; über Fachhochschul-Studiengänge (Fachhochschul-Studiengesetz- FHStG), BGBl. I 1993/340; über die Akkreditierung von Bildungseinrichtungen als Privatuniversitäten (Universitäts-Akkreditierungsgesetz – UniAkkG), BGBl. I 1999/168; über die Universität für Weiterbildung Krems (DUK-Gesetz 2004), BGBl. I 2004/22; über die Organisation der Pädagogischen Hochschulen und ihre Studien (Hochschulgesetz 2005), BGBl. I 2006/30.

neuer Hochschulangebote eine externe Qualitätssicherung in Form der Akkreditierung. Die größere Autonomie der Hochschulen wurde andererseits mit dem Anspruch verbunden, dass die Hochschulen eigenständige Maßnahmen der Qualitätssicherung setzen.

Öffentliche Universitäten unterliegen keiner Verpflichtung zur Akkreditierung von Studienprogrammen oder zur institutionellen Akkreditierung. Universitäten sind ex lege eingerichtet und schließen seit dem Jahre 2006 Leistungsvereinbarungen mit dem Bund. In den Leistungsvereinbarungen verpflichten sich die Universitäten zu Zielen und Vorhaben der Qualitätssicherung und Evaluierung, deren Erreichung sie nach Ablauf der Vereinbarungsperiode von drei Jahren nachweisen. Dieser Nachweis kann durch die Zertifizierung des Qualitätsmanagementsystems der Universität erfolgen. Pädagogische Hochschulen sind ebenso wie öffentliche Universitäten per Gesetz anerkannt und unterliegen keiner nationalen Akkreditierungsverpflichtung.

Die Akkreditierungspflicht besteht ausschließlich für die noch jungen Hochschulektoren der Fachhochschulen und Privatuniversitäten. Hierfür wurden eigene Behörden eingerichtet, deren Kompetenzen sich ausschließlich auf diese beiden Sektoren beschränken (Fachhochschulrat, Akkreditierungsrat für Privatuniversitäten). Der Akkreditierungsentcheidung gehen externe Evaluierungsverfahren voraus, die durch international anerkannte Qualitätssicherungsagenturen durchgeführt werden.

Gemeinsam ist allen österreichischen Hochschulen die Verpflichtung zum Auf- und Ausbau hochschulinterner Qualitätsmanagementsysteme. Sie zieht sich als roter Faden durch die Gesetzesgrundlagen der öffentlichen und privaten Universitäten, der Fachhochschulen und Pädagogischen Hochschulen. Die einzelnen Hochschulgesetze treffen allerdings keine gestalterischen Vorgaben und überlassen somit den Hochschulen, welche Evaluierungsinstrumente eingesetzt werden und wie die interne Qualitätssicherung organisiert wird.

Gestärkt wird der Grundsatz der Eigenverantwortung der Hochschulen durch den Bologna-Prozess und die 2005 festgelegten Standards und Leitlinien der European Association for Quality Assurance in Higher Education (2005). Den Europäischen Standards folgend soll die externe Qualitätssicherung im Zuge von Zertifizierungs- und Akkreditierungsverfahren die Wirksamkeit der hochschulinternen Qualitätssicherung prüfen. Eine kontinuierliche Evaluation der Lehre und Forschung liegt in erster Linie in der Kompetenz der Hochschulen und gilt als Bestandteil eines funktionierenden internen Qualitätssicherungssystems.

2 Evaluationsaktivitäten

Die jüngeren Entwicklungen der Evaluation im österreichischen Hochschulsystem können durch die folgenden Besonderheiten charakterisiert werden:

- Evaluationsaktivitäten im österreichischen Hochschulsystem erfolgten in der Vergangenheit zumeist anlassbezogen.
- Evaluierungsverbünde bestanden ebenso wenig wie hochschulübergreifende Planungen zur kontinuierlichen Evaluation der Lehre und Forschung.

- Die Entwicklung einer hochschulinternen Feedback-Kultur wurde durch Studierendenvertretungen mitgetragen.

2.1 Hochschulübergreifende Evaluationen

Im Zuge der Profil- und Standortentwicklung der österreichischen Universitäten wurden *universitätsübergreifende Evaluierungen einzelner Fachbereiche* (z. B. Physik, Elektrotechnik, Architektur, Mathematik, Geowissenschaften) durchgeführt. Die in Form von Peer-Reviews durchgeführten Evaluationen wurden sowohl durch Fachgesellschaften als auch das zuständige Bundesministerium initiiert. Die hochschulübergreifende Bewertung von Fachbereichen sollte als Entscheidungsgrundlage für die Hochschulentwicklung an einzelnen Universitätsstandorten dienen.

Ein *fächerbezogener Vergleich* von Universitäten und Fachhochschulen wurde seit 2003 im Rahmen des CHE-Hochschulrankings durchgeführt. Das Verfahren sieht eine indikatorenbegleitende Ranggruppenzuordnung einzelner Fachbereiche vor. Grundlage für den Vergleich bilden vornehmlich Urteile von Studierenden. Zunehmend wurden auch Forschungsleistungen einzelner Fachbereiche in den Vergleich einbezogen. In Österreich wurde der Vergleich bis 2007 durch die Österreichische Qualitätssicherungsagentur koordiniert, die den Universitäten und Fachhochschulen über die veröffentlichten Ergebnisse hinausgehende Detailanalysen zur Verfügung stellte (Österreichische Qualitätssicherungsagentur 2007c).

Anlassbezogen wurden auch *Evaluationsverfahren zu Organisationsprozessen* der österreichischen Universitäten durchgeführt: Die Aufforderung des Europäischen Gerichtshofes zur Gleichbehandlung von Studienwerberinnen und -werbern war Ursache für die Einführung von Zugangsbeschränkungen in Studienfächern, für welche in Deutschland ein Numerus clausus besteht. Die entsprechende Novelle zum Universitätsgesetz 2002 sah die Verpflichtung zur hochschulübergreifenden Evaluierung der Auswirkungen von Auswahlverfahren vor. Die Evaluierung erfolgte anhand bislang vorliegender Daten und Fakten und wurde im Januar 2007 abgeschlossen (Bundesministerium für Wissenschaft und Forschung 2007).

Ein weiteres Beispiel für eine hochschulübergreifende thematische Evaluierung ist ein Verfahren zur Begutachtung der Frauenförderung und Gleichstellung an österreichischen Universitäten, in dem entsprechende Reformschritte des Universitätsgesetzes 2002 extern begutachtet wurden. Die Evaluierung hatte hochschulpolitische Handlungsempfehlungen sowie Empfehlungen für die einzelnen Universitäten zum Ergebnis (Österreichische Qualitätssicherungsagentur 2007a).

2.2 Evaluationen auf Ebene der Hochschulen und der Studienprogramme

Die österreichischen Universitäten und Fachhochschulen setzen im Rahmen ihrer internen Qualitätssicherungsprozesse verschiedene Evaluationsmaßnahmen.

Als bereits traditionelles Verfahren gilt die *Lehrveranstaltungsevaluierung* durch Studierende, die an allen Hochschulen durchgeführt wird. Unterschiede bestehen hinsichtlich der Regelmäßigkeit und des Umfangs sowie der Konsequenzen von Lehrveranstaltungsevaluie-

rungen. Einzelne Universitäten haben die Intensität der Verfahren verringert und begründen dies mit dem hohen Erhebungsaufwand. Studierende sehen die Auswirkungen von Lehrveranstaltungsevaluationen mit Skepsis.

Zunehmend beziehen die Hochschulen Alumni in die Bewertung ihres Ausbildungsangebots ein und führen Absolventenbefragungen durch.

Mehrere Universitäten haben ihre Studienrichtungen, Fachbereiche und Forschungsfelder *Peer Review-Analysen* unterzogen. Diese Begutachtungen werden entweder durch die Universitäten selbst organisiert oder durch externe Qualitätssicherungs- und Zertifizierungsagenturen durchgeführt. Zudem unterziehen sich österreichische Universitäten auch Evaluierungen durch internationale Fach-Communities (insb. Naturwissenschaften/Medizin, Wirtschaftswissenschaften, Ingenieurwissenschaften). Eine Verpflichtung zur Akkreditierung von Studienprogrammen besteht für öffentliche Universitäten nicht.

Den gesetzlich festgelegten Akkreditierungsverfahren im Sektor der Fachhochschulen und Privatuniversitäten gehen Peer-Review-Evaluierungen voraus, die in der Regel nicht durch die Akkreditierungsbehörde sondern durch nationale und internationale Evaluierungsagenturen durchgeführt werden.

Forschungsleistungen werden, abgesehen von Peer-Review-Verfahren, häufig anhand der *Berichts- und Informationssysteme* der Universitäten evaluiert. Die mit dem Universitätsgesetz 2002 eingeführte Wissensbilanz enthält verschiedene Indikatoren zur Darstellung der universitären Forschungsleistungen. Außerdem wurden im Zuge des bereits genannten Hochschulvergleichs für einzelne Fachbereiche (z. B. Geisteswissenschaften) erstmals bibliometrische Analysen durchgeführt.

Im Zusammenhang mit dem gesetzlichen Auftrag zur Einrichtung hochschulinterner Qualitätsmanagementsysteme gewinnen *Auditverfahren* an Bedeutung, in welchen die Hochschulen die Wirksamkeit ihrer internen Qualitätssicherung und Evaluierung nachweisen. Quality Audit-Verfahren werden bereits in anderen europäischen Ländern eingesetzt beziehungsweise entwickelt. In Österreich wird ein Audit-Verfahren seit 2007 durch die Österreichische Qualitätssicherungsagentur umgesetzt (Österreichische Qualitätssicherungsagentur 2007b).

Zur hochschulinternen Qualitätssicherung und Evaluierung haben die Universitäten und Fachhochschulen in den vergangenen Jahren Organisationseinheiten eingerichtet beziehungsweise weiter ausgebaut. Häufig sind diese als Stabstellen der Hochschulleitung zugeordnet und übernehmen koordinierende und konzeptive Aufgaben.

3 Verbindungen mit anderen Evaluationsfeldern und internationale Entwicklungen

Evaluation hat in den vergangenen Jahren in allen Bereichen des österreichischen Hochschulsystems an Bedeutung gewonnen. Anstoß war vor allem der Trend zur Profilbildung und Autonomie der Hochschulen. Diese erforderten einerseits Grundlagen für hochschul- und standortpolitische Entscheidungen. Andererseits sind die Hochschulen zu regelmäßi-

gen Nachweisen ihrer Leistungen in Lehre, Forschung und Organisation veranlasst – sei dies im Rahmen von Leistungsvereinbarungen oder Akkreditierungsverfahren.

Eine Zusammenarbeit von Hochschulen zu Fragen der Evaluation und Qualitätssicherung besteht nur eingeschränkt, meist bezogen auf Disziplinen. Nationale Evaluationsverbünde wurden bislang nicht geschaffen, sondern eher die Zusammenarbeit in internationalen Fachgemeinschaften gesucht.

Der Bologna-Prozess hat zur Aufwertung der Qualitätssicherung beigetragen, sei es durch die regelmäßigen Bestandsaufnahmen („stocktaking“) oder durch den Diskurs zu geeigneten Verfahren der externen Qualitätssicherung. In der nationalen Diskussion wird häufig eine gemeinsame Betrachtung von Lehre und Forschung beziehungsweise Erschließung der Künste eingefordert.

Mit der Österreichischen Qualitätssicherungsagentur wurde eine gemeinsame Einrichtung für alle Hochschulektoren geschaffen, die zu einem einheitlicheren Verständnis für Qualitätssicherung und Evaluierung beitragen soll (Hanft/Kohler 2007).

4 Stand der Evaluation

Die Initiative zur Evaluation ist in den vergangenen Jahren vom politischen System auf die Ebene der Hochschulen übergegangen. Sie steht gleichzeitig in engem Zusammenhang mit den Rahmenanforderungen der jeweiligen Hochschulektoren. Der vergleichsweise große Freiraum der öffentlichen Universitäten erklärt die sehr unterschiedlichen Fortschritte der Universitäten bei der Gestaltung von Evaluationsmaßnahmen. Während einzelne österreichische Universitäten in den vergangenen Jahren ein breites Spektrum an Evaluierungsinstrumenten entwickelt haben, stehen andere Universitäten noch eher am Anfang einer Kultur der kontinuierlichen Bewertung eigener Leistungen. Allen Universitäten ist jedoch ein deklarerter Bedarf nach externer Begleitung und Expertise zu Fragen der Qualitätssicherung und Evaluierung gemeinsam.

Evaluation erfolgt meist noch punktuell und anlassbezogen. Allerdings gibt der gesetzliche Anspruch zur Entwicklung hochschulinterner Qualitätsmanagementsysteme Anlass dazu, Evaluierungsmaßnahmen in einem System der Qualitätssteuerung und Qualitätsentwicklung zusammenzuführen. Dadurch können Stellenwert und Verbindlichkeit der Evaluation gestärkt werden.

5 Bilanz

Das Ausmaß der Evaluation in Lehre und Forschung scheint im österreichischen Hochschulsystem im geringeren Maße gegeben als in anderen (deutschsprachigen) Ländern. Damit kann zwar ein Nachholbedarf bei der Methodenentwicklung festgestellt werden, die Hochschulen zeigen zugleich auch Initiative zur selbstverantwortlichen Qualitätssicherung und Evaluierung. Evaluation wird als informatorische Grundlage für Steuerungsentscheidungen auf der politischen und hochschulinternen Ebene an Bedeutung gewinnen.

Evaluation im Bildungsbereich in der Schweiz

1 Einleitung

Die Steuerung des föderal aufgebauten Schweizer Bildungssystems ist geprägt durch das Subsidiaritätsprinzip: Jede föderale Ebene (Bund, Kantone, Sprachregionen, Gemeinden, Institutionen) verantwortet und regelt gewisse Bereiche und macht Vorgaben für die nächstfolgende Ebene, die wiederum über einen Handlungsspielraum verfügt und allenfalls Vorgaben für die nächste Ebene macht. Das Subsidiaritätsprinzip gilt auch bei der Qualitätssicherung und -entwicklung und strukturiert somit auch die Evaluationstätigkeiten im Bildungsbereich.

Qualitätssicherung und -entwicklung ist auf allen Bildungsstufen aktuell, wobei für die Vorschule und die obligatorische Schule grundsätzlich die Kantone und für alle anderen Bereiche Bund und Kantone gemeinsam verantwortlich sind. Evaluationen im Bildungsbereich sind in der Regel Teil von Qualitätsmodellen, die über die Einzelschule hinausreichen und lokalen, kantonalen und sprachregionalen Verhältnissen angepasst werden müssen.

Für die Qualitätsentwicklung auf allen Stufen werden spezifische Evaluationsinstrumente bereitgestellt. Weiterhin eine wichtige Rolle spielen traditionelle Instrumente wie Lehrpläne und Lehrmittel, einheitlich geregelte Bildungsabschlüsse oder Rahmenbedingungen im Bereich der personellen und finanziellen Ressourcen. Eine systematische Übersicht über die Evaluationen im Bildungsbereich liegt für die Schweiz nicht vor. Der Bildungsbericht Schweiz 2006 (vgl. SKBF 2007) kommt diesem Anliegen jedoch sehr nahe, weil das Wissen über das Bildungssystem aus Forschung (einschliesslich Evaluation), Statistik und Verwaltung unter dem Aspekt der Systemsteuerung erstmals synthetisiert worden ist. Ein Orientierungsrahmen für die Durchführung von Evaluationen regt auf interkantonaler Ebene an, die Qualität und Quantität der Evaluationen auf allen Ebenen zu verbessern und versucht über die Kantons- und Sprachgrenzen hinaus die Entwicklung einer „Evaluationskultur“ zu unterstützen (EDK 2008).

2 Evaluationsaktivitäten auf den Bildungsstufen

2.1 Obligatorische Schule (Primarstufe, Sekundarstufe I)

Gemäss Bundesverfassung (BV) sind die Kantone für die obligatorische Schule verantwortlich, die staatlicher Leitung und Aufsicht untersteht. Die Aufsicht wird in den kantonalen Schulgesetzen und Bestimmungen geregelt. Zahlreiche Kantone haben kantonale Konzepte für die Qualitätssicherung und -entwicklung an Schulen ausgearbeitet. Üblicherweise wird

die Qualitätssicherung und zum Teil die Qualitätsentwicklung in kantonalen Schulgesetzen, Bestimmungen und Lehrplänen geregelt.

Individuelle Ebene

Selbstbeurteilung

Es gehört zur professionellen Arbeit von Lehrpersonen, dass sie ihr eigenes Handeln laufend beobachten und entwickeln und dabei von unterschiedlichen Seiten Rückmeldungen einholen. Dazu gehören kollegiale Hospitationen, Individualfeedbacks, klassenbezogene Rückmeldungen und Leistungsstandsmessungen, anhand derer die Lehrpersonen den Leistungsstand ihrer Klassen im Vergleich mit demjenigen von Vergleichsstichproben einschätzen können. Als Instrumente stehen nebst anderem „Klassencockpit“ und „Stellwerk“ zur Verfügung (vgl. Landert 2002).

Schülerinnen und Schüler können ebenfalls anhand von klassenbezogenen Leistungsmessungen oder anhand von Vergleichsarbeiten ihre eigene Leistung beurteilen und Standortbestimmungen durchführen. Verbreitet wird bei der Leistungsbeurteilung und in Zeugnissen auch die Selbstbeurteilung von Schülerinnen und Schülern berücksichtigt.

Fremdbeurteilung

Hinweise für Qualitätsansprüche an Lehrpersonen und Unterricht sind in Stellenbeschrieben, Berufsleitbildern oder in Schul- und Personalgesetzen enthalten. Die Kantone haben unterschiedliche Konzepte entwickelt und praktizieren die Evaluation der Lehrpersonen verschieden. Einzelne Qualitätskonzepte, die sich auf Schulen beziehen, sind aus der Diskussion um taugliche Beurteilungsformen von Lehrpersonen entstanden.

Die Evaluation der Leistungen der Schülerinnen und Schüler gehört genuin zu einem wirkungsorientierten, didaktisch aufbereiteten Unterricht und gewinnt an Bedeutung, je individualisierter der Unterricht ausgerichtet wird.

Institutionen

Die vermehrte Eigenverantwortung der Einzelschule und die Einrichtung von Schulleitungen mit grösserem Entscheidungsspielraum (vgl. Rhyn et al. 2002) hat seit Mitte der 1990er Jahren zu entsprechenden Schulentwicklungs- und Qualitätsprojekten in der obligatorischen Schule und in Schulen der Sekundarstufe II geführt (vgl. Büeler 2006).

Verschiedene Kantone arbeiten auf interkantonaler Basis zusammen und führen Projekte zur Qualitätssicherung und -entwicklung durch. Je nach Kanton und Schulstufe ist die Durchführung einer internen Evaluation oder die Erstellung eines eigenen Qualitätskonzeptes für die eigene Schule gesetzlich vorgeschrieben.

Bei einer *internen Schulevaluation* steht die Entwicklungs- und Selbststeuerungsfähigkeit der einzelnen Schule im Zentrum; sie kann als Standortbestimmung einer nachfolgenden Schulentwicklung dienen. Zielsetzung, Gestaltung und Durchführung liegen in der Verantwortung der Schule respektive der Einrichtung. Je nach Kanton oder Stufe kann ein einrich-

tungs- beziehungsweise schulinternes Qualitätsmanagement obligatorisch sein und entsprechende standardisierte Instrumente können zur Verfügung gestellt werden.

- Das *Förderorientierte Qualitätsevaluations-System (FQS)* ist von der Pädagogischen Arbeitsstelle des Dachverbandes der Schweizer Lehrerinnen und Lehrer (LCH) entwickelt worden. FQS zielt auf eine breit angelegte Qualitätsentwicklung einer Schule durch die Selbstevaluation der Lehrpersonen. Elemente des FQS sind die Bildung von Qualitätsgruppen, die gemeinsame Vereinbarung von Qualitätsvorstellungen, das Führen einer gegenseitigen Feedback-Kultur oder Aktivitäten zur Untersuchung der Schulqualität. Prozesse, Erkenntnisse und Entwicklungen werden in Berichten dokumentiert.
- Das Modell *Qualität durch Evaluation und Entwicklung (Q2E)* ist ein ganzheitliches Orientierungsmodell für Schulen, die ein eigenes Qualitätskonzept aufbauen wollen. Es basiert auf sechs Komponenten, die den allgemeinen Rahmen eines schulinternen Qualitätsmanagements umreissen und jeweils schulspezifisch angepasst und ausgelegt werden müssen. Entwickelt wurde das Q2E-Modell im Rahmen des Projektes Qualitätsentwicklung und -sicherung auf der Sekundarstufe II der Nordwestschweizerischen Erziehungsdirektorenkonferenz.
- Das Modell *Qualität und Qualifizierung (2Q)* wurde ursprünglich für die Lehrpersonenqualifikation entwickelt. Die 2Q-Methode zielt auf die Entwicklung der Unterrichtsqualität durch Zielvereinbarungen im Mitarbeitergespräch: Es ist ein zielorientiertes Personalentwicklungs- und Führungsinstrument, bei dem Zielvereinbarungen und Beurteilungsgespräche mit den Lehrpersonen über die Erreichung der vereinbarten Ziele im Zentrum stehen. Elemente sind ein schuleigener Katalog von Entwicklungsoptionen und regelmässige Mitarbeitergespräche der Schulleitung. Die Lehrpersonen erarbeiten gemeinsam eine Liste mit didaktischen zum Teil team- und schulbezogenen Entwicklungsbedürfnissen und -möglichkeiten. Daraus trifft die Lehrperson eine Wahl und definiert Entwicklungsziele und Umsetzungsmöglichkeiten und bespricht diese im Mitarbeitergespräch mit der Schulleitung und setzt einen Zeitraum fest, nach welchem eine Überprüfung erfolgt.

Diese und andere gebräuchliche Instrumente (vgl. Gonon et al. 1998) können der Personalführung und der Schulevaluation dienen. Sie können teilweise auch flächendeckend eingesetzt werden und dienen dann der Systemevaluation auf kantonaler Ebene.

Systemebene

Viele Kantone beschäftigen sich mit der Entwicklung neuer Formen der Steuerung und Qualitätsentwicklung nicht nur der einzelnen Schulen, sondern auch des ganzen Schulsystems (Rhyn 2007). Dabei geht es um die Neudefinition der Kompetenzen und Handlungsspielräume der Schulen und Gemeinden, um die Reorganisation der Schulaufsicht und um eine Neubestimmung der Rollen der Inspektorate, der lokalen Schulbehörden und der geleiteten Schulen beziehungsweise der Schulleitungen. Der Entwicklungsstand zeigt sich je nach Kanton unterschiedlich (vgl. ARGEV 2005).

Systemevaluation ist in aller Regel externe Evaluation. Diese muss sich in Bezug zu anderen Funktionen und Aufgaben der kantonalen Systemsteuerung sowie des lokalen und schulischen Qualitätsmanagements stellen (vgl. Burkard 2005). Sie kann im Rahmen eines kantonalen Qualitätsmanagements als Instrument der Schulentwicklung oder als Instrument der Schulaufsicht gebraucht werden:

- Als *Instrument der Schulentwicklung* zielt sie auf die Unterstützung der einzelnen Schule. Durch eine externe Beobachtung erhält die Schule beispielsweise Rückmeldungen über Stärken und Schwächen oder über den Erfüllungsgrad von selber gesetzten oder vorgegebenen Qualitätsstandards. Im Zentrum steht ein beraterischer Ansatz, unabhängige Fachpersonen nehmen eine Bestandesaufnahme vor und schätzen den Entwicklungsbedarf ein.
- Als *Instrument der Schulaufsicht* legt die externe Schulevaluation Rechenschaft ab. Hauptadressaten sind die Behörden. Das Hauptziel liegt in der Kontrolle über die Qualität der Schule. Die übergeordneten Instanzen (lokale und kantonale Schulbehörden) erhalten Informationen, ob die vorgegebenen Qualitätsanforderungen erfüllt werden oder ob entsprechende Massnahmen getroffen werden müssen. In diesem Sinne dient die Rechenschaftslegung auch der Qualitätsverbesserung.

Die Nordwestschweizerische Erziehungsdirektorenkonferenz (NW EDK), die Erziehungsdirektorenkonferenz der Ostschweizer Kantone und des Fürstentums Liechtenstein (EDK-Ost) und die Bildungsdirektorenkonferenz Zentralschweiz (BKZ) führen die Fachkonferenz *Arbeitsgemeinschaft Externe Evaluation von Schulen (ARGEV)*, welche die sprachregionale Zusammenarbeit und Koordination für die obligatorische Schule sicherstellt. Die ARGEV verfolgt für die externe Schulevaluation unter anderem folgende Ziele:

- Unterstützung der Kantone bei der Wissensentwicklung;
- Professionalisierung und Qualitätssicherung;
- Zusammenarbeit und Koordination der deutschsprachigen Kantone;
- Unterstützung der Kantone bei Weiterentwicklungen im Bereich der obligatorischen Schule.

Auf regionaler Ebene werden breit angelegte Versuche mit der Basisstufe, das heisst die gemeinsame Unterrichtung der Vier- bis Achtjährigen evaluiert¹ und im Bereich des Fremdsprachenlernens in der Schule wird ein Instrument für die Evaluation von Fremdsprachenkompetenzen entwickelt (vgl. Studer et al. 2004).

Auf gesamtschweizerischer Ebene werden im Rahmen der interkantonalen Vereinbarung über die Harmonisierung der obligatorischen Schule (HarmoS-Konkordat) Bildungsstandards entwickelt, deren Überprüfung eine kantonal vergleichende Systemevaluation ermöglichen wird (vgl. EDK 2006). Die Bildungsstandards werden für die Fächer lokale Landessprache, erste Fremdsprache, Mathematik und Naturwissenschaften wissenschaftlich

¹ Vgl. http://www.edk-ost.sg.ch/home/projekte/grundstufe_basisstufe.html [27.11.2007].

entwickelt, indem Kompetenzmodelle und -niveaus sowie konkrete Aufgabenstellungen und Tests zu deren Überprüfung erarbeitet werden. (Maradan/Mangold 2005).

2.2 Sekundarstufe II

Auf der Sekundarstufe II werden für die allgemein bildenden Schulen (Gymnasien und Fachmittelschulen) die Aufsicht und gegebenenfalls Konzepte für die Qualitätssicherung und -entwicklung in den kantonalen Gesetzen und Bestimmungen geregelt. Auf nationaler Ebene haben Bund und Kantone durch die Verordnung des Bundesrates beziehungsweise das Reglement der EDK über die Anerkennung von gymnasialen Maturitätsausweisen (MAR) von 1995 die gymnasiale Maturität eidgenössisch geregelt – und somit einer gesamtschweizerischen Evaluation zugänglich gemacht (vgl. unten EVAMAR). Gemäss dieser Verwaltungsvereinbarung überprüft die Schweizerische Maturitätskommission (SMK) die Einhaltung der Anerkennungsbedingungen durch die anerkannten Maturitätsschulen.

Gemäss Bundesgesetz über die Berufsbildung (Berufsbildungsgesetz, BBG) und Verordnung der Berufsbildung (Berufsbildungsverordnung, BBV) müssen alle Anbieter von Berufsbildung (berufliche Grundbildung, höhere Berufsbildung, berufsorientierte Weiterbildung) die Qualitätsentwicklung sicherstellen. Der Bund stellt Qualitätsstandards auf; dabei wird das Bundesamt für Berufsbildung und Technologie (BBT) eine Liste von Methoden zur Qualitätsentwicklung erarbeiten (Übergangsfrist bis Ende 2008). Die Anbieter von Berufsbildung können unter den aufgeführten Methoden zur Qualitätsentwicklung frei wählen, wobei die Kantone für öffentlich-rechtliche Anbieter eine Methode vorschreiben können.

Evaluationen auf Institutionsebene

Schulen können ihr Qualitätsbestreben mittels *Zertifizierungsverfahren* von bestimmten Zertifizierungsstellen attestieren lassen. So steht die ISO-Norm 9001-2000 für die Zertifizierung von Qualitätsmanagementsystemen. Sie bestätigt, dass eine Institution ihr Qualitätsmanagement auf der Basis dieser Norm aufgebaut hat und laufend überprüft. Ursprünglich wurde die Qualitätsnorm ISO (International Organization for Standardization) für Produktionsbetriebe entwickelt, in der Folge auch für Dienstleistungsbetriebe erweitert und ist auch für Schulen verwendbar. Eine ISO-Zertifizierung verlangt die Dokumentation sämtlicher Organisationsstrukturen und Abläufe bis hin zu detaillierten Verfahrensanweisungen und Stellenbeschreibungen in einem Qualitätshandbuch. Die ISO-Norm untersucht Organisations- und Führungsaufgaben, macht jedoch keine inhaltlichen Aussagen zur Qualität einer Schule. Die ISO-Zertifizierung erfolgt durch die Schweizerische Vereinigung für Qualitäts- und Management-Systeme (SQS) und kommt vor allem auf der Sekundarstufe II, der Tertiärstufe und in der Weiterbildung zur Anwendung.

Seit Beginn der Qualitätsdebatten besitzt die interkantonale Zusammenarbeit für den Austausch und die Koordination der kantonalen Erfahrungen eine wichtige Bedeutung. Aus dem Projekt Qualität durch Evaluation und Entwicklung (Q2E) ist die *Interkantonale Fachstel-*

le für Externe Schulevaluation auf der Sekundarstufe II (IFES²) hervorgegangen. Diese führt externe Evaluationen für Berufsfachschulen und allgemein bildende Schulen der Sekundarstufe II der Nordwestschweizer Kantone mit Schwerpunkt beim Qualitätsmanagement der Schule durch.

Systemebene/Programme

Auf der Sekundarstufe II laufen auf Systemebene grosse Evaluationsprojekte. Zu nennen sind folgende:

Das Projekt Evaluation der Maturitätsreform (EVAMAR) evaluiert im Auftrag von Bund und Kantonen die 1995 erfolgte Reform der schweizerischen Maturitätsbildung. EVAMAR ist in zwei Phasen unterteilt:

EVAMAR I (2002-2004) untersuchte die Umsetzung der Reform in den Kantonen und Schulen sowie die Konsequenzen für den Erfolg beim Abschluss der Maturitätsausbildung (vgl. Ramseier et al. 2005). Im Wesentlichen wurden drei Themenbereiche bearbeitet:

1. die Passung von Wahlfachangebot und Interessen der Schülerinnen und Schüler sowie die Auswirkungen auf den Ausbildungserfolg, namentlich die subjektiv wahrgenommene Qualität der Vorbereitung auf ein Hochschulstudium;
2. die Umsetzung der fächerübergreifenden Ziele;
3. die Bewältigung der Reformen durch die Schulen.

EVAMAR II (2005-2008) legt das Schwergewicht der Evaluation auf den Ausbildungsstand der Schülerinnen und Schüler am Ende ihrer Maturitätsausbildung. Im Zentrum steht die Studierfähigkeit beziehungsweise die Hochschulreife. Dabei geht es um die Frage, was von den Studierenden an faktischen Lern- und Denkprozessen beim Lernen und Arbeiten an der Hochschule gefordert wird und über welche fachlichen und überfachlichen Kompetenzen sie dabei verfügen müssen.

Die Nordwestschweizerische Erziehungsdirektorenkonferenz lancierte das Projekt *Benchmarking Sekundarstufe II* (Projektdauer 2004 bis 2010; vgl. Bergmann et al. 2006). Es soll auf regelmässiger Basis ein Vergleich zwischen den Schulen der Sekundarstufe II erstellt werden. Das Projekt definiert wichtige Kennzahlen für schulübergreifende Vergleiche und erhebt vergleichend Kennzahlen, Zufriedenheitsmasse sowie Absolventendaten. Die ermittelten Daten unterstützen die Analyse des Ist-Zustandes und ermöglichen eine konkrete Zielsetzung (z. B. Verbesserung des Indikators um zehn Prozent). Bei einer weiteren Durchführung kann die Zielerreichung überprüft werden.

² Vgl. <http://www.ifes-schuleva.ch>.

2.3 Tertiärstufe

Im Bereich der Hochschulen werden die rechtlichen Grundlagen bezüglich Aufsicht sowie Qualitätssicherung und -entwicklung in den kantonalen Universitäts- und Fachhochschulgesetzen und im Bundesgesetz über die Eidgenössischen Technischen Hochschulen (ETH-Gesetz) von 1991 geschaffen. Durch die Revision der Bildungsbestimmungen in der Bundesverfassung (BV) im Jahre 2006 hält der neue Hochschulartikel (BV Art. 63a) fest, dass Bund und Kantone gemeinsam für die Qualitätssicherung im schweizerischen Hochschulwesen sorgen. Durch das geplante Regelwerk „Hochschullandschaft“ wird die Steuerung des Hochschulbereichs vereinfacht.

Bei der höheren Berufsbildung kommt wie bei der beruflichen Grundbildung bezüglich Qualitätssicherung und -entwicklung das Bundesgesetz über die Berufsbildung von 2002 sowie die Verordnung über die Berufsbildung von 2003 zur Anwendung. Die Kantone üben demgemäss die Aufsicht über die höheren Fachschulen aus, soweit sie eidgenössisch anerkannte Bildungsgänge anbieten.

Evaluationen auf Institutionenebene

Universitäten

Das Organ für Qualitätssicherung (OAQ) hat zwischen 2003 und 2004 eine Bestandesaufnahme und Beurteilung der Qualitätssicherungsmassnahmen der *universitären Hochschulen* mittels Quality Audits vorgenommen. Die Universitäten haben in den 1990er Jahren mit der systematischen Einführung von Qualitätssicherungsmassnahmen für Lehre und Forschung begonnen. Diese Arbeiten sind heute unterschiedlich weit fortgeschritten und richten sich nach den regionalen und universitätsspezifischen Besonderheiten und Bedürfnissen. Die gesetzliche oder reglementarische Verankerung der Qualitätssicherung sowie die Einrichtung einer Qualitätssicherungsstelle zur Koordination der Massnahmen sind in der Regel in jeder universitären Hochschule vorhanden. Einzelne Universitäten, wie etwa die Universität Zürich, sind im Bereich der Selbstevaluation bereits sehr weit fortgeschritten. Die Mehrheit der universitären Hochschulen haben unter anderem folgende Qualitätssicherungsmassnahmen umgesetzt:

- regelmässige Evaluation der Lehrveranstaltungen;
- Qualitätssicherungsmassnahmen für die Forschung;
- Regelung der Berufungsverfahren;
- Didaktikangebote für den Lehrkörper;
- Massnahmen zur Nachwuchsförderung;
- Datenbanksysteme zur Erstellung von Studierendenstatistiken;
- die Evaluation von universitären Untereinheiten (Institute, Abteilungen und ähnliche) ist mehrheitlich noch in Planung.

Insgesamt wurde die strategische und politische Bedeutung von Qualitätsmassnahmen erkannt. Neben den Stärken im Bereich der Qualitätssicherung konnten auch Schwächen ausgemacht werden:

- Zwar findet sich an den universitären Hochschulen eine gute Mischung von Qualitätssicherungsmassnahmen, sie sind aber oft nicht in ein kohärentes Qualitätssicherungssystem eingebunden.
- Die Evaluationsergebnisse werden oft ungenügend für die Planung von Qualitätssicherungsmassnahmen genutzt.
- Eine systematische Evaluation der Studienprogramme und der Curricula fehlt teilweise.

Fachhochschulen

Gemäss ihrem gesetzlichen Auftrag sind die Fachhochschulen daran, ihre Leistungen periodisch zu evaluieren und ein Qualitätsmanagement zu erstellen oder führen bereits ein solches. Sie können Massnahmen der Qualitätssicherung (analog universitäre Hochschulen) ergreifen und eigene Modelle entwickeln oder ein auf dem Markt bestehendes Qualitätsentwicklungsmodell anwenden.

Die Konferenz der Fachhochschulen der Schweiz (KFH) hat eine Fachkommission Qualitätsmanagement (FQM) eingesetzt. Diese verfolgt das Ziel, ein fachhochschulübergreifendes Qualitätsmanagement zu entwickeln, das den gesetzlichen Erfordernissen entspricht. Es soll dazu beitragen, dass fundierte und vergleichbare Aussagen über die Qualität der Leistungen der Fachhochschulen gemacht werden können. Die FQM hat der KFH vorgeschlagen, das Modell der European Foundation for Quality Management (EFQM) zu verwenden und an die Bedürfnisse der Fachhochschulen anzupassen. Dabei hat die KFH einen Bewertungskatalog für Fachhochschulen erstellt, der als Instrument zur Selbstevaluation und Standortbestimmung dienen soll. Dieser erlaubt es, Stärken und Verbesserungspotentiale aufzuzeigen, die in die Planung und Einleitung von Verbesserungsmassnahmen münden (vgl. Klepp 2005).

Höhere Berufsbildung

Im Bereich der *höheren Berufsbildung* können verschiedene der oben genannten Instrumente der Evaluation und der Qualitätssicherung zur Anwendung kommen. Höhere Fachschulen, die einen Bildungsgang anerkennen lassen möchten, müssen in ihrem Beitrittsgesuch Auskunft über ihr Qualitätssicherungs- und Qualitätsentwicklungssystem geben.

Systemebene

Im Zusammenhang mit der Bologna-Deklaration hat sich die Schweiz dafür ausgesprochen, systematisch organisierte Akkreditierungs- und Qualitätssicherungssysteme im Hochschulbereich einzurichten. Das Akkreditierungsverfahren verläuft in drei Stufen. In einem ersten Schritt erarbeitet die Gesuch stellende Hochschule beziehungsweise Hochschuleinheit eine Selbstbeurteilung. Darauf folgt eine externe Begutachtung durch eine unabhängige Gruppe

von Experten und Expertinnen; die Hochschule kann auf den entsprechenden Expertenbericht Stellung nehmen. Aufgrund dieser Ergebnisse fällt für die Universitäten die Schweizerische Universitätskonferenz (SUK), für die Fachhochschulen das Eidgenössische Volkswirtschaftsdepartement oder eine andere Akkreditierungsstelle den Akkreditierungsentscheid. Eine Akkreditierung ist in der Regel sieben Jahre gültig.

Bei den Universitäten erfolgt die Akkreditierung auf freiwilliger Basis. Öffentliche und private Institutionen können sich oder ihre Studiengänge akkreditieren lassen. Im universitären Bereich ist das Organ für Akkreditierung und Qualitätssicherung der Schweizerischen Hochschulen (OAQ) zuständig für die Durchführung von Akkreditierungsverfahren. Es stellt zuhanden der Schweizerischen Universitätskonferenz einen Antrag für den Akkreditierungsentscheid, den die SUK dann fällt.

Im Bereich der Fachhochschulen hat das revidierte Bundesgesetz über die Fachhochschulen die rechtlichen Grundlagen für ein Akkreditierungs- und Qualitätssicherungssystem geschaffen: Die Akkreditierung ist laut Fachhochschulgesetz obligatorisch und an die Genehmigung der Fachhochschulen und ihrer Studiengänge gekoppelt. Das neue Akkreditierungssystem für den Fachhochschulbereich ist seit 2007 eingeführt. Das Eidgenössische Volkswirtschaftsdepartement ist für die Akkreditierung zuständig, es erlässt Richtlinien, um ein Mindestmass an Einheitlichkeit zu garantieren. Alle Fachhochschulen haben mit Unterstützung der Eidgenössischen Fachhochschulkommission umfassende Peer Reviews zur Überprüfung der Qualität des Studienangebots durchgeführt. Gestützt auf die vorgenommenen Überprüfungen der Institutionen und deren Studienangebote hat der Bundesrat Ende 2003 den sieben Fachhochschulen die unbefristete Genehmigung erteilt.

Die Hochschulen werden sich im Rahmen von Rankings mit anderen Schweizer Hochschulen oder mit denjenigen von anderen Ländern vergleichen. Rankings basieren unter anderem auf statistischen Daten der Hochschulen sowie auf Befragungen bei Studierenden und Dozierenden. Dabei werden nicht die Hochschulen sondern einzelne Studiengänge bewertet. Rankings werden in der Schweiz nicht nur für den Hochschulbereich kontrovers diskutiert.

3 Evaluation des Gesamtsystems Bildung und internationale Vergleiche

3.1 Bildungsmonitoring Schweiz

Bis anhin konnten in der Schweiz die verschiedenen Daten zu Bildungsstatistik und Bildungsforschung oftmals nicht genügend genutzt werden, weil sie nicht in einem Gesamtkontext erstellt und nutzbar aufbereitet worden sind. Bund und Kantone haben deshalb 2005 ein nationales Bildungsmonitoring lanciert. Im Rahmen des Bildungsmonitoring Schweiz wird sowohl auf gesamtschweizerischer wie auch auf kantonaler Ebene systematisch und kontinuierlich Wissen für die Steuerung des Bildungssystems gesammelt, analysiert und aufbereitet. Das Bildungsmonitoring dient als Grundlage für die Bildungsplanung sowie für bildungspolitische Entscheide, für die Rechenschaftslegung und öffentliche Dis-

kussion. Vorhandenes Wissen und vorhandene Informationen sollen gesammelt und für die Systemsteuerung nutzbar aufbereitet werden. Zugleich sollen Informationslücken festgestellt und die bestehende Datenlage gezielt ergänzt werden. Das Bildungsmonitoring wird als zyklischer Prozess durchgeführt. Dabei arbeiten Bildungsforschung, Bildungsstatistik und Bildungspolitik respektive Bildungsplanung zusammen. Konkretes Produkt dieses Prozesses ist ein Bildungsbericht Schweiz, der alle vier Jahre vorliegt (SKBF 2007).

3.2 Internationale Vergleichsstudien

Die Schweiz beteiligt sich an internationalen Vergleichsstudien der Organisation for Economic Cooperation and Development (OECD). Die OECD-Staaten vergleichen in verschiedenen Studien die Effektivität ihrer Bildungssysteme. Im Rahmen des Projekts INES (Indicators of Educational Systems) sind seit mehr als zehn Jahren internationale Bildungsindikatoren entwickelt und laufend aktualisiert worden. Diese Indikatoren liefern vergleichende Informationen über die in Bildung investierten menschlichen und finanziellen Ressourcen, über die Funktionsweise und Entwicklung von Bildungs- und Lernsystemen und über die Resultate der Bildungsinvestitionen. Jedes Jahr wird eine breite Palette an Indikatoren in der OECD-Publikation *Education at a Glance* veröffentlicht.

Im Kontext der OECD beteiligt sich die Schweiz im Bereich der Kompetenzmessung am Programme for International Student Assessment (PISA). PISA misst die Grundfähigkeiten bei 15-jährigen Schülerinnen und Schülern in den Bereichen Lesen, Mathematik und Naturwissenschaften sowie fachübergreifende Kompetenzen. Die Verfahren richten sich nach wissenschaftlichen Methoden. Projektplanung, Datenerhebung, Analyse und Berichterstattung erfolgten weitgehend dekontextualisiert vom lokal spezifischen Unterrichtsgegenstand.

Die Schweiz beteiligt sich an verschiedenen OECD-Länderexamen und an thematischen Examen der OECD. Bei letzteren wird ein vergleichender Länderbericht vom teilnehmenden Land als Grundlagebericht erarbeitet. Dabei werden Daten, Forschungsliteratur und Expertenstudien zum untersuchten Thema aufgearbeitet.

Auf internationaler und gesamtschweizerischer Ebene zeigt sich immer mehr, dass Evaluationen im Bildungsbereich mit anderen Bereichen wie Arbeitsmarkt, Ökonomie, Familien-, Sozial- und Migrationspolitik in Verbindung zu bringen sind.

4 Bilanz und Ausblick

Wo die Qualität von Bildungsinstitutionen belegt und entwickelt werden soll, liefern Evaluationen die notwendigen Daten. Sie ermöglichen einen Soll-Ist-Vergleich und weisen auf den Handlungsbedarf hin, der wiederum die Grundlage für zielgerichtete Entwicklungsmassnahmen auf den verschiedenen Ebenen bildet. Evaluationen sind somit ein notwendiges Element jedes Konzepts zur Qualitätssicherung und -entwicklung. Diese Sichtweise hat sich im Bildungsbereich in der Schweiz auf allen Ebenen und zunehmend auch in allen Sprachregionen durchgesetzt.

Obschon ein gewisser Konsens in Bezug auf die Bedeutung von Evaluationen für die Qualitätssicherung und -entwicklung im Bildungsbereich besteht, sind dennoch einige kritische Punkte zu beachten (vgl. EDK 2008), die diskutiert werden:

- *Aufwand und Nutzen:* Die finanziellen und personellen Ressourcen für Evaluationen sind begrenzt. Evaluationen sind daher auf allen Ebenen koordiniert zu planen, um bei den Evaluierten Überdross und dadurch eine fragwürdige Qualität der erhobenen Daten zu vermeiden.
- *Zielkonflikte:* Auftraggeber und Evaluierte verfolgen mit einer Evaluation nicht unbedingt dieselben Ziele. In der Planung von Evaluationen ist daher deren Zweck, das heisst die vorgesehene Nutzung der gewonnenen Daten offen zu legen.
- *Rollenkonflikte:* Im Zusammenwirken der verschiedenen an Evaluationen beteiligten Personen und Gruppen können Konflikte auftreten, wo Rollen nicht sauber auseinander gehalten werden beziehungsweise wo Doppel- oder gar Mehrfachrollen derselben Person nicht offen gelegt werden.
- *Rückkoppelungen:* Eine bevorstehende Evaluation kann Anlass zu kurzfristigen, häufig oberflächlichen Verbesserungsmassnahmen geben oder die Auswahl der zu befragenden Personen oder der zu analysierenden Dokumente beeinflussen.
- *Wirkungen:* Evaluationen können auch im Bildungsbereich unterschiedliche, teilweise nicht beabsichtigte Wirkungen erzielen. Zu nennen sind dabei (vgl. Stamm 2003) „Anstoss zu Verbesserungen und Innovationen“, „Anlass für einzelne spontane Sofortmassnahmen“, „Auslöser von Alibi-Reaktionen“, „Auslöser von Abwehr und Blockade“.
- *Grenzen:* Evaluationen können sehr viele Bereiche und Aspekte von Bildungsinstitutionen erhellen, aber auch an Grenzen stossen. Länger dauernde Prozesse, etwa im Unterricht oder in der Schulführung, sind in zeitlich begrenzten Evaluationen nur schwer zu erfassen. Latente Qualitätsdefizite bei einzelnen Personen oder in der Organisation entziehen sich häufig der Anlage einer punktuellen Evaluation.

Diese Punkte sind zu berücksichtigen, wenn eine Evaluationskultur im schweizerischen Bildungswesen weiterhin auf- und ausgebaut werden soll. Formalisierte und standardisierte Evaluationsverfahren sind hilfreich und notwendig, müssen jedoch regelmässig auch hinterfragt werden, damit sie mit einem sich ständig wandelnden Bildungssystem Schritt halten können.

Evaluation im Bildungsbereich: Ein Ländervergleich zwischen Deutschland, Österreich und der Schweiz

1 Einleitung

Nach der länderspezifischen Schilderung der Situation von Evaluation im Bildungsbereich in Deutschland, Österreich und der Schweiz ist es das Ziel dieses Beitrages, eine ländervergleichende Perspektive einzunehmen und die Situation in den drei Ländern einander gegenüberzustellen.

Zu diesem Zweck werden die historische Entwicklung, die aktuellen Aktivitäten und die institutionellen Verankerungen der Evaluation (Evaluation zunächst verstanden in ihrem weitesten Sinne) nachgezeichnet, bevor eine Bilanz gezogen und ein Ausblick gegeben wird.

2 Wie ist die historische Entwicklung der Evaluation im Bildungsbereich und in welche aktuellen Aktivitäten mündet sie?

In den nationalen Evaluationsaktivitäten im Bildungsbereich in Deutschland, der Schweiz und Österreich lassen sich in der historischen Entwicklung auffallende Parallelen feststellen.

Erste umfangreichere Evaluationsaktivitäten kamen im Bereich der Schule in den 1970er und 1980er Jahren auf. Dabei handelte es sich primär um die Auswertung von einzelnen Schulversuchen und die damit verbundene Systemevaluation auf Makroebene, zum Beispiel die Effektivität von Ganztageschulen (Ipfling 1981) und Gesamtschulen (Fend 1982) in Deutschland, die Neugestaltung der gymnasialen Oberstufe in Österreich (Petri 1984a) sowie verschiedene Innovationen im Primarschulbereich in der Schweiz (Meylan 1983).

Anfangs der 1990er Jahre wandelte sich dieses Bild. Teilweise bedingt durch die wachsende Autonomie und Eigenverantwortlichkeit der Einzelschulen sowie geleitet durch den Gedanken, dass die Arbeit in den Schulen vor Ort richtungsweisend für die Qualität ist, war in allen drei Ländern ein Trend hin zur schulinternen Evaluation festzustellen. In dieser Zeit wurde die (Selbst-)Evaluation an Schulen auch zum Forschungsgegenstand mit international angelegten Projekten (z. B. Czerwanski 2003) sowie vielen praxisbezogenen Publikationen aus allen drei Ländern (z. B. Ackermann 2001; Burkard/Eikenbusch 2000; Burkard/Pfeiffer 1995; Landwehr/Steiner 2003; Müller 2002; Radnitzky/Schratz 1999).

Gegen Ende der 1990er Jahre und mit Beginn dieses Jahrtausends setzte ein Trend ein, der auch heute noch aktuell ist und der die Bildungsszene wie nie zuvor in den Mittelpunkt

der öffentlichen Diskussion gerückt hat: Ergänzend zur Einzelschule werden mit internationalen Schulleistungsstudien wie PISA (PISA-Konsortium Deutschland 2007) und TIMSS (Baumert et al. 2000) systematisch Daten zum allgemeinbildenden Schulsystem gesammelt. Insbesondere in Deutschland wird diese Entwicklung durch nationale und regionale Schulleistungsstudien und Lernstandserhebungen wie zum Beispiel DESI (DESI-Konsortium 2008), ELEMENT (Lehmann/Lenkeit 2008), LAU (Lehmann et al. 1997), MARKUS (Helmke/Jäger 2002), QuaSUM (Lehmann et al. 2001) sowie Vergleichsarbeiten und teilweise zentrale Abschlussprüfungen noch verstärkt. Auch vor der Berufsbildung macht diese Entwicklung nicht Halt, ist doch ein internationales „Berufsbildungs-PISA“ politisch gewollt und organisatorisch auf dem Weg (Baethge et al. 2006).

Neben der Durchführung von Schulleistungsstudien sind aktuell weitere Entwicklungen zu beobachten. In allen drei Ländern werden Bildungsstandards entwickelt beziehungsweise implementiert, die Maßstäbe dafür liefern sollen, was Schülerinnen und Schüler in einem bestimmten Alter können sollen (vgl. Lucyshyn 2006 für Österreich, das Projekt HarmoS¹ für die Schweiz sowie Kultusministerkonferenz Deutschland 2006 für Deutschland). Das gilt mittlerweile auch nicht mehr nur für die schulischen Kernfächer (Tenorth 2008). Zudem hilft eine systematische Bildungsberichterstattung, die Situation des Bildungswesens im jeweiligen Land detailliert zu beschreiben. In der Schweiz (Schweizerische Koordinationsstelle für Bildungsforschung (SKBF) 2007) und in Deutschland (Konsortium Bildungsberichterstattung 2006) sind erste Berichte bereits erschienen, in Österreich sind die Vorarbeiten im Gange².

Darüber hinaus bleibt die Weiterentwicklung der Einzelschule von Interesse. So gibt es in Deutschland Schulinspektionen zur Überprüfung interner Abläufe in den Schulen. In der Schweiz liegen standardisierte Instrumente für die interne Schulevaluation vor, zum Beispiel das Förderorientierte Qualitätsevaluations-System (FQS)³ und das Modell Qualität durch Evaluation und Entwicklung (Q2E) (Landwehr/Steiner 2007). Zudem streben immer mehr Schulen nach einer Zertifizierung im Bereich des Qualitätsmanagements. In Österreich will das Bundesministerium für Unterricht, Kunst und Kultur mit dem Projekt Q.I.S. (siehe <http://www.qis.at/>) die österreichischen Schulen anregen und darin bestärken, selbst ihre Qualität zu hinterfragen, zu kontrollieren und weiter zu entwickeln.

Auch grössere Einzelprojekte prägen die Gegenwart der Evaluationstätigkeit im Schulbereich. Schulversuche und Großprojekte wie zum Beispiel die Evaluation der Maturitätsreform in der Schweiz (EVAMAR) (Ramseier et al. 2005) werden anlassbezogen durchgeführt.

Bei Innovations- und Reformprojekten der Berufsbildung in der Schweiz, für die nach den Artikeln 54 und 55 des Berufsbildungsgesetzes staatliche Förderung beantragt wird, sind deren Evaluationen zwingend vorgeschrieben.

Betrachtet man den Hochschulbereich, so lassen sich systematische Evaluationstätigkeiten im größeren Stil besonders ab den 1990er Jahren mit der Evaluation der Hochschullehre verorten. Allerdings waren und sind diese Aktivitäten wenig koordiniert und entsprangen zumeist der Initiative jeder einzelnen Hochschule auf der Suche nach Qualitätssi-

¹ Vgl. http://www.edk.ch/d/EDK/Geschaefte/framesets/mainHarmoS_d.html [19.5.2008].

² Siehe <http://www.bifie.at/content/view/82/155/> [19.5.2008].

³ Siehe <http://www.erz.be.ch/site/index/fachportal-bildung/biev-index/biev-evaluationen/biev-abgeschlossene-evaluationen/biev-qes/biev-evaluationen-qes-modelle/biev-evaluationen-qes-modelle-fqs.htm> [21.5.2008].

cherung des eigenen Angebotes. Die Autonomie der Hochschulen war und ist ein hohes Gut. Dies gilt vergleichbar für Österreich, die Schweiz und Deutschland. Zusätzlich war in Deutschland nach der Wende die Frage nach der Positionierung von ostdeutschen Studiengängen und Universitäten Gegenstand von Evaluationen.

Im Zuge der Bologna-Deklaration wandelt sich das aktuelle Bild insofern, als systematisch organisierte Akkreditierungs- und Qualitätssicherungssysteme im Hochschulbereich eingerichtet werden, die neben der Evaluation der Lehre auch die Evaluationsaktivitäten der Forschung an Hochschulen verstärken. Der Grad der Verbindlichkeit ist allerdings sehr unterschiedlich. In der Schweiz erfolgt die Akkreditierung der Universitäten und/oder ihrer Studiengänge auf freiwilliger Basis. Öffentliche und private Institutionen können sich oder ihre Studiengänge akkreditieren lassen. Für die Fachhochschulen hingegen ist die Akkreditierung laut Fachhochschulgesetz obligatorisch. In Österreich besteht eine Akkreditierungspflicht nur im privaten Bereich und bei Fachhochschulen. In Deutschland wird die Akkreditierung von Bachelor- und Masterstudiengängen in den ländergemeinsamen Strukturvorgaben der Kultusministerkonferenz verbindlich vorgeschrieben und in den einzelnen Hochschulgesetzen der Bundesländer auf unterschiedliche Weise als Voraussetzung für die staatliche Genehmigung eingefordert.

Zusätzlich ist ein weiterer Trend erkennbar: Mit nationalen und internationalen Hochschulrankings zum Beispiel des Centrums für Hochschulentwicklung, der Österreichischen Qualitätssicherungsagentur AQA, dem Shanghai-Ranking und anderen stellen sich Hochschulen dem nationalen und internationalen, fächerbezogenen wie auch hochschulübergreifenden Vergleich. Von einem Studierenden-PISA liest man zwar schon gelegentlich, doch ist dessen Realisierung noch ungewiss.

Neben der Internationalisierung des Evaluationsgeschehens gibt es zudem anlassbezogene Evaluationen.

3 Wie ist Evaluation im Bildungsbereich heute institutionell verankert?

Mit wachsender Bedeutung des Evaluationsgeschehens im Bildungsbereich ist auch die institutionelle Verankerung von Evaluation im weitesten Sinne stark vorangeschritten.

In Österreich ist Evaluation eine Aufgabe der Schulaufsicht (sowohl im Sinne der Rechtsaufsicht als auch im Sinne der Metaevaluation der Selbstevaluationen von Schulen). Das vormalige Zentrum für Schulentwicklung wird zum Bundesinstitut ausgebaut und soll institutioneller Kern der Evaluation im Schulbereich werden. In Deutschland koordiniert die „Ständige Konferenz der Kultusminister der Länder“ (KMK) Fragen der Evaluation auf Bundesebene; gleichzeitig entstehen auf Bundeslandebene neue Agenturen für die Schulinspektion. In der Schweiz ist die „Arbeitsgemeinschaft Externe Evaluation von Schulen“ (ARGEV) – eine Arbeitsgemeinschaft von Fachstellen aus zwanzig Deutschschweizer Kantonen sowie des Fürstentums Liechtenstein – für die externe Schulevaluation zuständig beziehungsweise befasst sich konzeptionell mit dem Thema.

Der Hochschulbereich in Österreich ist durch Autonomie und eigene Qualitätssicherung geprägt, so dass sich keine weiteren Institutionen gebildet haben. Interne Evaluationseinheiten und internationale Akkreditierungsagenturen prägen das Bild der Qualitätsbemühungen. Zusätzlich koordiniert die „Österreichische Qualitätssicherungsagentur“ (AQA) den fächerbezogenen, hochschulübergreifenden Vergleich. In Deutschland ist rund die Hälfte der Universitäten in die Arbeit von Evaluations- und Akkreditierungsagenturen involviert. Damit einher geht auch eine immer größer werdende Anzahl von hochschulinternen Einheiten, die sich mit Evaluation befassen. Darüber hinaus gibt es weitere institutionelle Akteure auf aggregierter Ebene (vgl. den Beitrag von Stefan Kuhlmann in diesem Band). In der Schweiz hat nahezu jede Hochschule eine eigene Qualitätsabteilung. Darüber hinaus ist im universitären Bereich das „Organ für Akkreditierung und Qualitätssicherung der Schweizerischen Hochschulen“ (OAQ) für die Durchführung von Akkreditierungsverfahren zuständig. Im Bereich der Fachhochschulen hat die „Konferenz der Fachhochschulen der Schweiz“ (KFH) eine „Fachkommission Qualitätsmanagement“ (FQM) gegründet.

4 Wohin geht die Evaluationsreise im Bildungsbereich?

Qualitätssicherung, Schulaufsicht, Akkreditierung, Monitoring, Bildungsberichte, internationale Leistungsvergleichsstudien, Schulranking, Hochschulranking, ... Evaluation scheint zu boomen. Jedoch: Kaum ein anderer Bereich wie die Evaluation sieht sich damit konfrontiert, von sehr viel unterschiedlichen Disziplinen, Bereichen und Politikfeldern so vereinbart werden zu wollen. Schließlich kann jeder evaluieren und alle entsprechenden Aktivitäten sind irgendwie auch Evaluation.

Es würde den Rahmen dieses Artikels sprengen, definitorische Klarheit zu schaffen und Übereinstimmungen, Schnittmengen, aber gegebenenfalls auch Abgrenzungen zwischen dem Begriff Evaluation und den zuvor genannten vornehmen zu wollen. Darüber hinaus ist es unrealistisch zu glauben, diesbezüglich Einigkeit zwischen den verschiedenen Akteuren erzielen zu können. An dieser Stelle wird daher der Anspruch darauf beschränkt, einige Diskussionspunkte aufzuwerfen, die auf dem Weg hin zu einer produktiven Evaluationskultur im Bildungsbereich Beachtung verdienen.

4.1 Professionalisierung der Evaluationsszene

Die starke Zunahme der Tätigkeiten im weit verstandenen Bereich der Evaluation lässt eine Professionalisierung der Evaluationsszene vermuten und erhoffen. Doch stimmt das tatsächlich? Was macht eine Professionalisierung aus? Sicherlich diskutierbare, aber als Grundlage nützliche Kriterien, an denen eine Professionalisierung festgemacht werden kann, sind solche wie die Generierung von spezifischem Expertenwissen mit einer damit einhergehenden Fachsprache, spezifische Ausbildungsgänge mit akademischer Ausrichtung, Richtlinien und Standards, Berufsverbände und -vertretungen et cetera. Wie sieht es diesbezüglich aus mit der Evaluation im Bildungsbereich?

Mit der „Gesellschaft für Evaluation“ (DeGEval) sowie der Schweizerischen Evaluationsgesellschaft“ (SEVAL) existieren zwei Fachgesellschaften, die den deutschsprachigen Raum abdecken. In beiden gibt es mehr oder weniger aktive Arbeitskreise beziehungsweise -gruppen, die sich zumindest in Teilen mit dem Bildungsbereich beschäftigen (DeGEval: AK Aus- und Weiterbildung in der Evaluation, AK Berufliche und betriebliche Bildung, AK Forschungs-, Technologie und Innovationspolitik, AK Hochschulen; SEVAL: AG Forschung und Lehre). Allerdings ist die Gruppe derer, die Evaluationen in der Praxis durchführt oder verantwortet, ungleich größer als diejenige, die sich in den Fachgesellschaften organisiert hat. Hier wird oft neben- und nicht miteinander gearbeitet.

Sowohl die SEVAL als auch die DeGEval haben Evaluationsstandards (Widmer et al. 2000; DeGEval 2002) formuliert, die Qualitätsanforderungen an Evaluationen definieren. Das ist ein guter, wichtiger und richtiger Schritt. Allerdings fehlt es den Evaluationsstandards im Bildungsbereich noch an Verbindlichkeit und Bekanntheit, so dass sie in konkreten Projekten noch zu oft unberücksichtigt bleiben. Gerade im Bildungsbereich sind viele Bildungsforscher tätig, die sogar eigene Evaluationsstandards für spezifische Bereiche entwickeln (Abs/Klieme 2005; Abs et al. 2006; Leutner 1999) und eher disziplinbezogen als evaluationsbezogen zusammenarbeiten. Zudem agieren Lehrkräfte, Vertreterinnen und Vertreter von Schulbehörden, Akkreditierungsfachleute und Qualitätsmanagementbeauftragte, die alle von der Disziplin, aus der sie kommen, geprägt sind. Hier soll nicht deren professionelles Tun in ihrem Sinne in Frage gestellt werden, aber den Besonderheiten von Evaluation wird nicht immer ausreichend Rechnung getragen.

Evaluationsprofessionalität bedarf zudem einer fundierten Ausbildung. In einigen sozialwissenschaftlichen Studiengängen ist Evaluation expliziter Ausbildungsbestandteil, doch ist er oft mit dem Bereich Forschungsmethoden gekoppelt und fristet unter diesem ein Stiefmütterchendasein. Ein Ausweg bietet eine auf Evaluation fokussierte Ausbildung. Hierzu bestehen sowohl in der Schweiz (Bern⁴) als auch in Deutschland (Saarbrücken⁵, Bonn⁶) Angebote auf Masterstufe. Die aktuelle Praxis ist aber nach wie vor die, dass Personen aus anderen Disziplinen ohne eine solche Ausbildung oder äquivalenter Kenntnisse die Evaluationslandschaft im Bildungsbereich prägen.

Die Professionalisierung der Evaluation im Bildungsbereich steht so gesehen erst noch am Anfang. Die Angebote der Evaluations-Community, angereichert mit dem Fachwissen relevanter Disziplinen im jeweiligen Kontext, müssen stärker genutzt werden!

4.2 Koordiniertes Handeln in konkreten Projekten

Da Ressourcen für Evaluationen begrenzt sind und um eine Überevaluierung zu vermeiden – man liest von „Evaluitis“ (Frey 2008), „Von der Inquisition zur Evaluation“ (Hornbostel 2008), „Wir evaluieren uns zu Tode“ (Preußler 2008) –, müssen die konkreten Projektaktivitäten auf allen Ebenen koordiniert werden, um qualitativ hochwertige Evaluationen durch-

⁴ Siehe <http://www.evaluationsstudium.ch/>.

⁵ Siehe <http://www.master-evaluation.de/>.

⁶ Siehe <http://www.zem.uni-bonn.de/www/ZEM/Aufbaustudium.html>.

führen zu können. Nur durch qualitativ hochwertige Arbeit kann dem nach wie vor schlechten Ruf der Evaluation in der Öffentlichkeit begegnet werden.

Im Bereich der Hochschulen bedeutet das zum Beispiel die Integration von Lehrveranstaltungsevaluation in umfassende Rahmenkonzepte (Schmidt 2008) sowie die Verbindung von Evaluation, Akkreditierung und weiterer Steuerungselemente. Dem entspricht, dass die in den vergangenen Jahren dominante Programmakkreditierung durch Verfahren ergänzt werden soll, die eine Integration unterschiedlicher Qualitätssicherungsinstrumente zum Ziel haben (HRK/ACQUIN 2007). Die nach wie vor unzureichende Verknüpfung von Forschungs- und Lehrevaluation spiegelt eine andere Seite der gleichen Medaille wider. Um hektische Überbetriebsamkeit zu vermeiden und um Räder nicht ständig neu zu erfinden ist es notwendig, besonders in konkreten Projekten eine enge, aktive Verbindung der Akteure der Evaluations-Community und ihrer Aktivitäten untereinander, aber auch mit weiteren Aktiven der Evaluation im Bildungsbereich herzustellen!

4.3 Ressourcen

Gerade kleinere Evaluationsprojekte auf der Ebene einer Einzelschule kranken oftmals daran, dass kaum ausreichende Ressourcen für die anspruchsvolle Tätigkeit zur Verfügung gestellt werden. Das hat zur Folge, dass immer wieder interne, in Evaluation kurzfristig weitergebildete Lehrkräfte Evaluation als Teil ihres Pensums durchführen. Auch wenn es viele schöne und gute Beispiele gibt, fehlt es zu vielen Projekten an der notwendigen Evaluationsprofessionalität. Aus diesem Grund gibt es immer mehr extern durchgeführte Schulevaluationen, doch diese begegnen höchstens dem Wissensdefizit. Da sie aber auch Geld kosten, löst sich das Ressourcenproblem damit nicht. Dem gegenüber stehen vielfältige Reforminitiativen, die den Evaluationsbedarf noch weiter erhöhen. Ein fokussierter und professioneller Einsatz der zu Verfügung stehenden Ressourcen ist deswegen unbedingt notwendig!

4.4 Evaluation muss in jedem Fall mehr als Datensammlung sein

Gerade in Großprojekten fällt eine Unmenge von Daten und Ergebnissen an, die manchmal methodisch so kompliziert sind, dass man sie als Nicht-Fachmann kaum versteht und erst recht nicht kritisch hinterfragen kann, da sich die verwendete Methodik (z. B. IRT-Modelle der Schulleistungsstudien) nur wenigen Eingeweihten erschließt. Das macht es für die Verantwortlichen nicht einfacher, Ergebnisse so zu kommunizieren, dass sie die Chance haben, in und für die Praxis handlungsleitend zu werden (und zumindest diese Mindestanforderung sollte Konsens sein, wenn der Begriff Evaluation verwendet wird). Das Ergebnis ist oft, dass Resultate solcher Studien in der Praxis nur sehr eingeschränkt verwendet werden, oder sogar deren Sinn für die Schulpraxis in Frage gestellt wird. Die Rezeptionsforschung (Hosenfeld/Groß Ophoff 2007; Kohler/Schrader 2004) geht hier in die richtige Richtung, indem die Frage untersucht wird, wie Ergebnisse aufbereitet und vermittelt werden müssen, dass sie auch wirklich das Potential haben, genutzt zu werden. Doch eine gute Ergebnisdarstel-

lung führt nicht zwingend zur Ergebnisnutzung (Ditton/Arnoldt 2004). Allzu oft zeigt sich, dass die Nutzung der Ergebnisse in der Praxis – von wem und für welchen Zweck ist dabei noch eine andere Frage – in Projekten, die unter der Überschrift Evaluation laufen, nicht zu Ende (oder überhaupt nicht) gedacht ist. Dann ist ungeklärt, welchen Nutzen diese so genannten Evaluationsprojekte für die Praxis überhaupt entfalten können. Gute Evaluationspraxis muss sich um Fragen der Ergebnisnutzung schon zu Projektbeginn kümmern!

4.5 Auf der Suche nach dem Impact

Nicht nur für die Evaluation im Bildungsbereich gilt, dass sie allzu oft bei der Bestimmung von output oder outcome stehen bleibt. Immer mehr werden aber Fragen nach längerfristigen Wirkungen, nach Impact sowie nach Kausalerklärungen zu bearbeiten sein!

4.6 Ausmerzen von Kinderkrankheiten

Neben den großen Zielen darf es die Evaluation nicht versäumen, sich den eigenen Kinderkrankheiten zu stellen. Für Evaluationen allgemein (Balzer 2005) wie auch für Evaluationsprojekte im Bildungswesen (Balzer 2006) muss festgestellt werden, dass sehr grundlegende Elemente erfolgreicher Evaluationsprojekte – wie die Festlegung klarer, realistischer Evaluationsziele und -fragestellungen, eine klare Definition des Evaluationsgegenstandes, eine klare Definition des Kontextes, in dem der Evaluationsgegenstand angesiedelt ist, sowie die Festlegung klarer Bewertungsgrundlagen und -indikatoren – in der aktuellen Evaluationspraxis noch verbessert werden können. Solides Handwerk in den konkreten Projekten ist gefragt!

4.7 Fazit

Evaluation im Bildungsbereich hat sich auf einen viel versprechenden Weg gemacht, doch ist die Zukunft einer guten Evaluationskultur im Hinblick auf Themen wie Professionalisierung, Koordinierung, Ressourcen- und Ergebnisnutzung, Ausrichtung auf den Impact sowie konkrete Projektdurchführung noch sehr ungewiss. Es gibt weiter viel zu tun.

Energie- und Umweltpolitik

Umweltevaluation in Deutschland – Entwicklungstrends mit Fokus auf dem Energiesektor

1 Einleitung

Umweltpolitik entsteht in Deutschland als eigenständiger Bereich in den späten 1970er Jahren. Willy Brandts Vision vom „Blauen Himmel über der Ruhr“ greift früh das Thema Umweltschutz auf. Ab Mitte der 1970er Jahre wird ein nationales rechtliches Instrumentarium aufgebaut, in dem überwiegend Umweltauflagen festgelegt werden (Bundesimmissionschutzgesetz, Abwasserabgaben, Lärmschutzrichtlinie und Abfallgesetz). Von Prittwitz (1988) mit seinem „Katastrophenparadox“ und Beck (1986) mit seiner „Risikogesellschaft“ bringen erste politik- bzw. sozialwissenschaftliche Erklärungsmodelle einer Umweltpolitik ein.

Mit dem 1987 veröffentlichten Bericht der Weltkommission für Umwelt und Entwicklung unter Vorsitz der ehemaligen norwegischen Ministerpräsidentin und Umweltministerin Gro Harlem Brundtland (Hauff 1987) beginnt eine neue Phase des Umweltdiskurses in Deutschland, der sich besonders in den 1990er Jahren in Richtung der global geführten Nachhaltigkeitsdebatte verlagert (vgl. zur Entwicklung des Ökologie- und Nachhaltigkeitsdiskurses in Deutschland Brand 1997; Brand/Poferl 1997).

Die Vision einer „nachhaltigen Entwicklung“ erreichte auf der Weltkonferenz für Umwelt und Entwicklung in Rio de Janeiro 1992 den Durchbruch (vgl. zur Geschichte der Nachhaltigkeit Luks 2002; siehe auch Meyer 2007).

Für den Energiesektor übernahm die bereits auf dieser Konferenz kontrovers diskutierte Klimarahmenkonvention, welche 1997 in das Kyoto-Protokoll mündete, eine Leitfunktion. Hier ging und geht es um die Frage, wie die führenden Industriestaaten durch geeignete Maßnahmen zur Reduzierung des CO₂-Ausstosses den drohenden Klimawandel abschwächen können. Aufgrund des hohen Beitrags der fossilen Energieerzeugung (besonders durch Kohlekraftwerke) für die Gesamtmenge der CO₂-Emissionen in Deutschland erhält der Energiesektor große Relevanz. Heute haben Umwelt- und Nachhaltigkeitspolitik einen festen Stellenwert innerhalb der Gesamtpolitik. Aktuell findet Klimaschutz, forciert durch Katastrophen und die Vergegenwärtigung entsprechender Risiken auch in Deutschland immer mehr Beachtung.

Vor diesem Hintergrund ist die Umweltevaluation Teilmenge einer Evaluation von Nachhaltigkeit. Nachhaltige Entwicklung geht über Umweltschutz hinaus und umfasst ökonomische und soziale Aspekte, die anhand von Indikatoren und Zielkategorien überprüft respektive evaluiert werden können. Nachhaltigkeit kann dabei eine Zieldimension beziehungsweise eine Bewertungs- oder Kriteriendimension sein, die an umweltrelevante Evaluationsgegenstände angelegt werden kann.

Der Begriff Umweltevaluation ist nach unserem gemeinsamen Verständnis auf alle systematischen, wissenschaftlichen Verfahren zur Analyse und Bewertung ökologischer Sachverhalte anwendbar, welche sachliche, personenunabhängige, problemadäquate und für die praktische Umsetzung verwertbare Ergebnisse zu Ursachen und Wirkungen von Interventionen liefern und dabei sämtliche für die Bewertung relevanten Informationen nutzen.

Dieser Beitrag kann die Vielfalt zu Inhalt und Methodik bisheriger Arbeiten, die sich mit dem Thema von Umwelt und Nachhaltiger Entwicklung in evaluatorischer Absicht befassen, auch nicht in Ansätzen vollständig abbilden. Dies wäre nur durch eine umfassende Studie machbar. Da eine solche nicht existiert, werfen die Autoren dieses Beitrags aus ihren jeweiligen Perspektiven Schlaglichter auf exemplarische Schwerpunkte: Auf die für den Klimaschutz zentrale Politik für Energieeffizienz und erneuerbare Energien sowie auf einige Ansätze der Umweltevaluation im betrieblichen Umfeld.

2 Kontext

Die sozialwissenschaftlich orientierte Evaluationsforschung hat seit den 1990er Jahren einen enormen Aufschwung in Europa genommen. Die 1997 in Deutschland gegründete Evaluationsvereinigung (DeGEval) verfügt seit 1999 auch über einen Arbeitskreis „Evaluationen im Umweltbereich“. Die Umweltpolitik gehört trotz eines frühzeitigen Einsatzes einzelner Evaluationen eher zu den Politikfeldern, in denen die Evaluation bis heute wenig bis gar nicht verankert werden konnte. Dies gilt im Prinzip für alle Ebenen der administrativen Umsetzung von umweltpolitischen Programmen.

Im Umweltbereich hat sich der Evaluationsbegriff zunächst in einer technisch-naturwissenschaftlichen Perspektive etablieren können (vgl. zu den folgenden Ausführungen auch Meyer 2001). Der Begriff Evaluation ersetzte dabei zunehmend den in den 1960er Jahren gebräuchlichen Begriff der Überwachung. Die Einhaltung von Normen und Grenzwerten, die für bestimmte technische Einzelaspekte festgelegt wurden, kontrollieren regelmäßig Institutionen wie zum Beispiel der Technische Überwachungsverein (TÜV) auf Grundlage gesetzlicher Regelungen. Eine Betriebserlaubnis wird darauf aufbauend zeitlich befristet erteilt. Die Umweltdiskussion in Deutschland war lange Zeit durch den Streit über die Festlegung sinnvoller Grenzwerte und die zur Überwachung ihrer Einhaltung zu nutzenden Messverfahren und Maßnahmen geprägt. Die in den 1980er Jahren verabschiedete „Technische Anleitung zur Reinhaltung der Luft“ stellt ein gutes Beispiel dafür dar. Kennzeichnend ist die Konzentration auf messbare technisch-naturwissenschaftliche Teilaspekte, wobei sowohl die Komplexität von Wechselwirkungen innerhalb des Ökosystems als auch ökonomische und soziale Aspekte weitgehend außer Acht gelassen werden.

Die Defizite bezüglich der mangelnden Systemorientierung werden zum Teil durch neuere Verfahren der Umweltbewertungsforschung aufgehoben. Besonders populär ist dabei der Ansatz des „ökologischen Fußabdrucks“ geworden, der verschiedene ökologische Aspekte in eine leicht verständliche, geographische Kennzahl zusammenfasst (vgl. Wackernagel/Rees 1997). Noch weiter verbreitet ist die bereits Anfang der 1970er Jahre als „Environmental Impact Assessment“ in den USA eingeführte „Umweltverträglichkeitsprüfung“ (UVP), die Ende der 1980er Jahre auf Grundlage der europäischen UVP-Richtlinie auch in

Deutschland für alle Planungsverfahren verbindlich vorgeschrieben wurde (vgl. Gassner/Winkelbrandt 2004).

Die mangelnde Einbindung ökonomischer Aspekte in die Bewertung ökologischer Sachverhalte ging einher mit der vor allem von Volkswirten schon frühzeitig diskutierten Problematik einer fehlenden Berücksichtigung ökologischer Fragestellungen bei ökonomischen Kalkulationen sowohl auf betrieblicher als auch auf nationaler Ebene. Aus Sicht der Wirtschaftswissenschaft ergibt sich als Folge ein Versagen marktregulierender Mechanismen, da die Kosten des Erhalts der Natur der Allgemeinheit aufgebürdet und nicht den aus ihrer Nutzung entstehenden individuellen Einnahmen entgegengestellt werden (vgl. Wicke 1993). Dementsprechend können Monetarisierungsversuche der ökologischen Wirkungen von Umweltnutzungen und -schädigungen als Ausgangspunkt ökonomischer Evaluationsmodelle gelten. Berechnet werden dabei auf der Grundlage ökonomischer Indikatoren die zur Wiederherstellung des Ausgangszustands notwendigen Kosten, welche dann den Verursachern von Umweltbelastungen zugewiesen werden sollen. Zu den prominentesten Vertretern dieser Richtung gehören die vom Statistischen Bundesamt unternommenen Versuche zur Integration der Umweltkosten in die Volkswirtschaftliche Gesamtrechnung (vgl. Statistisches Bundesamt 2003). Ende des zwanzigsten Jahrhunderts verlagerte sich der Schwerpunkt auf Ansätze, welche die betriebliche Ebene in den Mittelpunkt stellen und für die Bewertung nicht ausschließlich Geldeinheiten verwenden. Beispiele sind Ökobilanzen, innerbetriebliche Stoffstromanalysen, Produktlinienanalysen, die Berechnung von Material-Intensitäten pro Service-Einheit (MIPS) und ähnliche Konzepte, für die sich „Umweltcontrolling“ als Oberbegriff etabliert hat (vgl. Bundesministerium für Umwelt, Naturschutz und Reaktorsicherheit 2001). Eine erste sozialwissenschaftliche Untersuchung zu deren Leistungsfähigkeit in Bezug auf immanente Kriterien und deren Anschlussfähigkeit liegt vor (Klemisch 2004).

Die Tradition der sozialwissenschaftlich orientierten Evaluationsforschung spielt trotzdem in der Umweltevaluation bisher nur eine untergeordnete Rolle. Sie konnte primär im Rahmen der Umweltbildungs- und -bewußtseinsforschung Einfluß erlangen. Beispiele sind die umfassenden Evaluationsstudien im Auftrag der Deutschen Bundesstiftung Umwelt zur Umweltberatung in Deutschland (Stockmann et al. 2001; Jacoby et al. 2005). Daneben gibt es allerdings noch eine weitere, wesentlich traditionsreichere Richtung sozialwissenschaftlicher Evaluationen im Umweltbereich. Gemeint sind die Konzepte und Verfahren der Technikfolgenabschätzung, welche die Auswirkungen technischer Interventionen auf die menschlichen Lebensbedingungen in den Mittelpunkt stellt (Grunwald 2002). Gemeinsam ist beiden Forschungsrichtungen die humanzentrierte Sichtweise auf Umweltprobleme, wobei sowohl der menschliche Beitrag zur Entstehung von Umweltschädigungen als auch die Implementationsbedingungen und Auswirkungen technischer Lösungen für die beteiligten Menschen im Fokus stehen.

Zur letzten Kategorie der sozialwissenschaftlichen Umweltevaluation zählen als Anbieter unter anderem das Institut für sozialökologische Forschung (ISOE) in Frankfurt, das Institut für ökologische Wirtschaftsforschung (IÖW) in Berlin, das Klaus Novy Institut in Köln und das Centrum für Evaluation der Universität des Saarlandes (CEval) in Saarbrücken.

Die Umweltinstitute kommen traditionell eher aus dem technisch-naturwissenschaftlichen Bereich und haben häufig einen Schwerpunkt bei der Energie- und Klimapolitik. Hierzu zählen das Wuppertal-Institut, das Öko-Institut in Freiburg oder das IFEU-Institut in

Heidelberg. Von ihrem Selbstverständnis würden sich diese Institute jedoch nicht als Umweltevaluatoren bezeichnen und sie sind – von Ausnahmen abgesehen – auch nicht in den entsprechenden Diskussions- und Qualitätsforen der Evaluation wie der Gesellschaft für Evaluation (DeGEval) anzutreffen. Einige Institute, zum Beispiel das IÖW haben eine starke ökonomische Ausrichtung oder sind von den Ansätzen der Managementforschung geprägt wie das Centre for Sustainable Management (CSM) der Universität Lüneburg.

Die politischen Zuständigkeiten im Umweltsektor sind in Deutschland weit verteilt. Dies ist einerseits der Tatsache geschuldet, dass es sich beim Thema Umwelt um ein Querschnittsthema handelt: Fast alle Aktivitäten der Gesellschaftsmitglieder haben einen Einfluss auf die natürliche Umwelt. Das beginnt beim Autofahren, reicht über den Bau- und die Nutzung der Gebäude bis hin zum täglichen Umgang mit Nahrungsmitteln und Konsumgütern. Deshalb gibt es nur wenige Politikbereiche, die nicht auch umweltrelevante Aspekte berühren.

Ein weiterer wichtiger Grund für die breite Streuung der Zuständigkeiten in Deutschland ist das föderale System, welches den 16 Bundesländern weitreichende Zuständigkeiten zuweist: Häufig liegt die Umsetzung und Kontrolle von Gesetzen in der Hand der Bundesländer. Zudem weist die kommunale Ebene eine hohe Teil-Eigenständigkeit auf. Die Kommunen sind in den Bundesländern überwiegend für den Vollzug der Umweltschutzgesetze allein oder im Zusammenwirken mit Behörden auf regionaler oder Landesebene zuständig (Klemisch et al. 1994). Auch dort gibt es politische und planerische Aktivitäten mit erheblichem Umwelteinfluss.

Im Bereich der Energiepolitik sieht die Verteilung der Zuständigkeiten auf Bundesebene wie folgt aus:

- Energieeffizienz: Bauministerium und Wirtschaftsministerium
- Erneuerbare Energien: Umweltministerium sowie Ministerium für Ernährung, Landwirtschaft und Verbraucherschutz
- Konventionelle Energieversorgung: Wirtschaftsministerium

Auf der Bundesebene werden vornehmlich regulative Vorgaben mittels Gesetzgebung oder Verordnungen auf den Weg gebracht. Mittlerweile sind diese Aktivitäten vielfach durch EU-Rahmenvorgaben geprägt, die es auf Bundesebene umzusetzen gilt. Auch größere finanzielle Förderprogramme sowie Informationsmaßnahmen sind zum Teil auf nationaler Ebene angesiedelt. Dabei gibt es vielfache Überschneidungen mit Maßnahmen auf Ebene der Bundesländer, teilweise auch mit den Kommunen.

Seit Ende der 1990er Jahre gibt es eine große Dynamik in der Energiepolitik in Deutschland. Diese wurde ausgelöst unter anderem durch die Liberalisierung des Energieversorgungsmarktes auf der Ebene der EU (1998), mit der ein freier Zugang zum Energieversorgungsmarkt auch für nicht staatlich oder politisch gesetzte Monopole erreicht werden sollte. Ein weiterer wichtiger Anlass ist die zunehmende Bedeutung des Klimawandels (vgl. Rahmstorf/Schellnhuber 2007), den man durch den Einsatz erneuerbarer Energien sowie einem verstärkt effizienten Einsatz der Energie begrenzen möchte (vgl. zur Entwicklung erneuerbarer Energien Hennicke/Fischedick 2007). Deshalb wurden in Deutschland zum Beispiel das Erneuerbare Energiengesetz zur Förderung von Strom aus regenerativen Ener-

gien erlassen (feste, langjährig garantierte Zuschüsse für ins Stromnetz eingespeiste Energie, seit 2000) und das Marktanreizprogramm für die Wärmegewinnung aus erneuerbaren Energien im Jahr 2002 eingeführt (finanzielle Investitionszuschüsse für Anlagen zur Wärmeversorgung). Die Förderung von energiesparenden Maßnahmen bei Gebäuden wird durch verschiedene politisch beschlossene Kreditförderprogramme vorangetrieben (seit 2000). Hinzu kommt der CO₂-Handel auf EU-Ebene, der auch die großen Energieanlagen in Deutschland betrifft. Vielfache Aktivitäten auf Landes- und kommunaler Ebene kommen hinzu. Zu evaluieren gibt es somit reichlich.

Zentrale Akteure auf dem Gebiet der Energieversorgung außerhalb der Politik sind die Energieversorgungsunternehmen, die im Stromsektor weiterhin von vier großen Oligopolen dominiert werden. Viele hundert kommunale Stadtwerke spielen mit ihrer Energiepolitik vor Ort noch eine wichtige Rolle bei Fragen der Energieversorgung. Auf der Seite des Energieeinsatzes beim Endkunden ist die Akteursstruktur deutlich heterogener: Privatpersonen mit ihren heute sehr unterschiedlich geprägten Lebensstilen, die verschiedenen Branchen des Gewerbe-, Handel- und Dienstleistungssektors, die produzierende Industrie sowie öffentliche Einrichtungen machen eine differenzierte Politik zur gezielten Erreichung von Energie- und Umweltzielen notwendig.

Ein übergreifendes, schlüssiges energiepolitisches Konzept liegt in Deutschland bisher nicht vor. Dies liegt unter anderem daran, dass die genannten Ressorts aufgrund der politischen Konstellationen in den letzten Jahrzehnten zumeist in Hand von MinisterInnen verschiedener Parteizugehörigkeit waren und auch aktuell sind. Nur zu Teilbereichen gibt es quer zu den Ressorts vereinbarte Zielsetzungen und damit verbundene politische Instrumente, u. a. zum Klimaschutz oder zum Ausstieg aus der Atomenergie. Aufgrund dieser Ausgangssituation gibt es keine zentrale politische Institution, die für Fragen einer Umwelt- oder Energieevaluation zuständig wäre. Die Ziele, Standards und Methodiken von Evaluationsprojekten werden bei jedem Projekt jeweils mit den Auftraggebern verhandelt und können nur in wenigen Bereichen auf übergreifende Ziele, Wertmaßstäbe oder Standards zurückgreifen.

3 Evaluationsaktivitäten am Beispiel der Energiepolitik

Auf der Bundesebene sind die Hauptauftraggeber für Evaluationen das Umwelt- und das Wirtschaftsministerium. Viele Instrumente wurden in den letzten Jahren ausgewertet. Der Schwerpunkt lag dabei auf finanziellen Förder- oder Steuerprogrammen, wie der „Öko-Steuer“ oder dem Marktanreizprogramm Erneuerbare Energien. Auch Beratungsprogramme (Energieberatung der Verbraucherzentralen) oder Kampagnen („Solar? – Na Klar!“) wurden evaluiert (Dünnhoff et al. 2005; Umweltbundesamt 2002).

Bei den Bundesländern lag der Schwerpunkt der Evaluation (entsprechend dem Schwerpunkt der Maßnahmen) bei Informations- und Beratungskampagnen (wie z.B. das REN-Impulsprogramm Nordrhein-Westfalen¹ und die schleswig-holsteinische Kampagne „Aus. Wirklich Aus?“) (Gruber/Böde 2000; Schötz et al. 2003). Die Förderprogramme auf

¹ REN steht dabei für „Rationelle Energienutzung“.

Landesebene unterlagen zumeist einer Eigenauswertung, seltener wurden sie durch Externe evaluiert (vgl. das Klimaschutz-Plus-Förderprogramm in Baden-Württemberg; Sawillion 2003; Duscha et al. 2007).

Auf der Ebene der Kommunen werden kaum Mittel für die Evaluation von Energieaktivitäten eingesetzt. Häufig fehlt es hier an Evaluationskompetenz und an der Bereitschaft, bei defizitärer Haushaltslage Wirkungsmessungen zu finanzieren. Wenn Evaluationen erfolgten, dann am ehesten bei finanziellen Förderprogrammen, um den Einsatz der knappen Mittel zu rechtfertigen (u. a. in Hannover, München, Münster). Zudem gibt es an manchen Stellen ein instrumentenübergreifendes Controlling, das zum Beispiel die CO₂-Entwicklung der Stadt aufzeigt oder anhand von anderen Indikatoren versucht, die Entwicklungen zu dokumentieren (vgl. European Energy Award, Klimabündnis der Städte). Auch dies wird meist intern bearbeitet. Dabei handelt es sich aber selten um Evaluationen im engeren Sinne.

Generell ist im öffentlich-rechtlichen Bereich in Deutschland davon auszugehen, dass die Evaluationsvorhaben von den jeweils zuständigen Fachreferaten innerhalb der Ministerien und Ämter vergeben und betreut werden. Damit erfolgt die Betreuung überwiegend aus einer Fachoptik, wobei interdisziplinäre und evaluationstheoretische Perspektiven untergeordnet sind.

4 Verbindung mit anderen Evaluationsfeldern: Die betriebliche Ebene als Beispiel

Einen wichtigen Teil der Verantwortung für einen erfolgreichen Umweltschutz trägt neben der politischen auch die betriebliche Ebene. Das Handlungsfeld Energie stellt sich hier als Managementaufgabe dar. Diese Aufgaben werden in den letzten Jahren nicht separat, sondern als Teil von übergreifenden Managementsystemen aufgegriffen. Auf betrieblicher Ebene wurden Umweltschutzmaßnahmen deshalb vor allem in Bezug auf die Einführung von Umweltmanagementsystemen (UMS) evaluiert (u. a. Klemisch 1997; Freimann 1997; BMU/UBA 2000; Dyllick/ Hamschmidt 2000; Klemisch 2004). Dabei standen mehr die ökonomischen bzw. organisatorischen und weniger die ökologischen Effekte im Vordergrund. Trotz einer Vielzahl dokumentierter Maßnahmen im betrieblichen Umweltschutz geben UMS keine gesicherte Auskunft über die tatsächlichen ökologischen Entlastungseffekte. Menge, Art und Qualität dieser Effekte sind stark von den unternehmensintern festgelegten Zielformulierungen abhängig. Insgesamt kann festgestellt werden, dass im Hinblick auf die ökologischen Wirkungen von UMS Verbesserungen der Ökoeffizienz auf betrieblicher Ebene im Vordergrund stehen. Diese beziehen sich in erster Linie bei der Inputseite auf Energie-, Wasser- und Rohstoffeinsatz und bei der häufig noch dominierenden Outputseite auf die Bereiche Emissionen und Abfallwirtschaft.

Es bleibt unklar, ob durch die Einführung von UMS anspruchsvollere Umweltziele als zuvor definiert wurden. Zwischen Environmental management and audit scheme (EMAS) und ISO 14001 lassen sich keine Unterschiede bezüglich der ökologischen Wirksamkeit

feststellen (BMU/UBA 2000: 36).² Die vorliegenden Erkenntnisse lassen keine Aussagen über die absolute Entwicklung von Umweltbelastungen zu und geben daher auch keine Richtungssicherheit. Es wird vermutet, dass trotz Steigerungen der Öko-Effizienz in Teilbereichen die Umweltentlastungen durch Wachstumseffekte kompensiert werden (Dyllick/Ham-schmidt 2000: 61). Ansätze, um im Rahmen einer Umweltleistungsbewertung zu vergleichbaren Daten zu kommen, bestehen einerseits in der Anwendung der ISO Norm 14031 "Leitlinien Umweltleistungsbewertung", andererseits in der Entwicklung von branchenbezogenen Umweltkennzahlen.

Vergleicht man Umweltmanagement und integrierte Produktpolitik als zwei zentrale Konzepte einer betrieblichen Umweltpolitik vor dem Anforderungshintergrund der DeGE-val-Standards für Evaluation (Nützlichkeit, Durchführbarkeit, Fairness und Genauigkeit), so zeigt sich, dass beide Konzepte so angelegt sind, dass sie den Evaluationsstandards weitgehend entsprechen (DeGEval 2002). Für die Analyse beider Konzepte bieten die Standards sowohl nützliche Informationen zum systematischen Vergleich der Instrumente, zu ihren Einsatzfeldern, zu den Analyseaspekten, zu den Beteiligungsstrukturen, zur Kosten-Nutzenstruktur, zur Genauigkeit sowie zur Nachvollziehbarkeit der Kriterien und Ergebnisse. Eine Evaluation betrieblicher Umweltpolitik findet hier eine solide Grundlage, wobei die Instrumente selbst – genauso wie die Instrumente nachhaltigen Wirtschaftens – nach wie vor nur in wenigen Unternehmen Anwendung finden (Klemisch/Potter 2006).

Tabelle 1: Evaluationsstandards und Instrumente betrieblicher Umweltpolitik

| Verfahren/ Evaluationsstandard | Umweltmanagement | Integrierte Produktpolitik |
|--|--|--|
| Zentrale Instrumente | EMAS oder ISO 14000f | Kennzeichnung nach ISO 14020ff, Produktbilanzierung nach ISO 14040ff oder nicht standardisierten Verfahren |
| Gegenstand | Unternehmen/Standort/Organisation | Produkte |
| N: Nützlichkeit: Zentrale Analyseaspekte | Verbesserung der Betriebsökologie und -organisation | Verbesserung der Produktökologie |
| N: Nützlichkeit: Identifikation der Beteiligten | Entwicklung unter Beteiligung be- trieblicher Akteure verbessert die Umsetzungsqualität | Kriterienentwicklung unter Beteili- gung gesellschaftlicher Gruppen erhöht die Akzeptanz |
| D: Durchführbarkeit: | u. a. Klärung von Kosten/Nutzen | u. a. Klärung von Kosten/Nutzen |
| F: Fairness: Transparenz | Verfahren sollte als interner Prozess offen gelegt sein | Kriterien und Verfahren sind in der Regel offen gelegt |
| F: Fairness: Information | Informationsmaterial in Form der Umwelterklärung vorhanden, inter- ner Prozess dokumentiert | Offene Informationspolitik Erfolgsbe- dingung für Bewertungs- und Kenn- zeichnungsverfahren |
| G: Genauigkeit: Nachvollzieh- barkeit der Kriterien | Durch interne Dokumentation und Vorgehen nach ISO Norm; überwie- gend harte Kriterien, die gut nach- vollziehbar sind | Durch festgelegte Verfahren und nach ISO Norm; überwiegend harte Kriterien, die gut nachvollziehbar sind |

² Bei EMAS handelt es sich um eine Verordnung der EU, dagegen regelt die International Organisation for Standardization (ISO) mit ihren Normenreihen weltweit Verfahren der Industrie.

5 Entwicklung und Bilanzierung

5.1 Rückschau und Stand

Bis Anfang der 2000er Jahre sind wenige Studien in Deutschland zu finden, die sich mit den Auswirkungen der Energiepolitik auf die Umwelt im Sinne von Evaluationen beschäftigen. Clausnitzer zieht in einer Studie zur Situation der Energiepolitik in Deutschland im Jahr 2000 das folgende Fazit: „In Deutschland ist Evaluation von Energiepolitik eher selten. Große Bedeutung hat sie leider nicht in dem Sinne, dass neue politische Maßnahmen erst oder nur nach Überprüfung der Wirksamkeit und Effizienz bisheriger Politik beschlossen werden. Standards zum Umfang von [Energie-] Evaluationen existieren nicht.“ (Clausnitzer 2000: 6)

Schwerpunkt waren bis dahin vor allem einfache Auswertungen von regionalen Förderprogrammen, zumeist als wenig veröffentlichte, interne Studien von Stadtwerken, die nur selten den Ansprüchen wissenschaftlicher Evaluationen entsprachen.

Seit dem Jahr 2000 hat die Anzahl wissenschaftlich fundierter Evaluationen zu energiepolitischen Instrumenten deutlich zugenommen. Das Interesse auf Seiten der Politik, ob und wie die Maßnahmen wirken, scheint gewachsen zu sein. Die Ergebnisse werden jedoch nicht nur zur Optimierung und Steuerung der Maßnahmen eingesetzt, sondern auf dem von stark unterschiedlichen Interessen besetzten Akteursfeld oft zur Rechtfertigung verschiedener Positionen. Dies ist um so leichter möglich, weil Standards für die Ermittlung von Effekten und Einsparungen und zur Zuweisung von Effekten zu bestimmten Instrumenten weitgehend fehlen. Das Zusammenwirken unterschiedlicher Instrumente, wie es in der Praxis üblicherweise anzutreffen ist, wird in den Studien wenig bis gar nicht thematisiert. Dies liegt unter anderem darin begründet, dass es, wie oben gezeigt, mindestens vier Bundesministerien gibt, die trotz überschneidender Zielfelder, vorwiegend auf die in ihrer jeweiligen Zuständigkeit liegenden Instrumentarien fokussieren. Eine übergreifende Koordination der Evaluationen und ihrer Standards gibt es in Deutschland auf politischer Ebene nicht. Neben dieser eher systematischen Einschränkung durch die Evaluationsauftraggeber kommen auf Anbieterseite Einschränkungen durch die immer noch vorherrschende disziplinär ausgerichtete Arbeitsweise hinzu, die einer Integration von sozialen, wirtschaftlichen und ökologischen Fragestellungen nicht förderlich ist. Da Energiepolitik zudem weitestgehend als technisches Problem definiert wird, dominieren in diesen Evaluationen technisch-naturwissenschaftliche Ansätze.

5.2 Entwicklungsmöglichkeiten

Entwicklungsmöglichkeiten für umweltrelevante Evaluationen im Energiebereich liegen zunächst in der Qualitätssorge. Das fängt damit an, dass die Evaluationsstandards der DeGEval (DeGEval 2002) Berücksichtigung finden sollten, die zumindest in der eher naturwissenschaftlich geprägten Auftraggeber- und -nehmerschaft wenig bekannt sind. Zudem bleibt zu prüfen, inwieweit materielle Standards für dieses Themenfeld hilfreich sind, um zu einer besseren Vergleichbarkeit und Transparenz der Ergebnisse zu kommen. Damit wür-

den Evaluationen noch wertvoller für die Steuerung und Bewertung von umwelt- und energiepolitischen Maßnahmen und Prozessen werden. Denkbar wären zum Beispiel materielle Standards bezüglich der zu treffenden Annahmen zu technischen Eigenschaften (wie Annahmen zur Lebensdauer elektrischer Geräte, wie sie schon in internationalen Gremien diskutiert wurden). Ebenso ließe sich prüfen, ob nicht auch zielgruppenspezifische Standards (für private Haushalte, Gewerbebetriebe etc.) sowie für spezifische Instrumentenarten (rechtliche Vorgaben, finanzielle Förderung etc.) zur Vergleichbarkeit beitragen könnten, ohne die Aussagemöglichkeiten an anderen Stellen zu stark zu beschneiden.

Zudem erscheint es in einem komplexen energie- und umweltpolitischen Umfeld wichtig, die Wechselwirkungen von Politikinstrumenten verstärkt in den Blick zu nehmen: Was lässt sich über die Wirkungen einzelner Instrumente sagen, wenn doch immer mehrere zugleich wirken. Als Beispiel sei hier die Informationskampagne „Initiative Energieeffizienz“ der deutschen Energieagentur benannt, zu der zeitgleich ein umfassendes Förderprogramm der Bundesregierung zur energetisch verbesserten Sanierung von Gebäuden aufgelegt wurde.

Schließlich ist eine stärkere Integration von sozialen, ökonomischen und ökologischen Fragestellungen in Energie- und Umweltevaluationen wünschenswert, um den Ansprüchen einer nachhaltigen Entwicklung besser Rechnung zu tragen.

6 Schlussfolgerungen

Das Thema Umweltschutz wird deutschland- und weltweit an Bedeutung gewinnen. Insbesondere durch die nicht mehr völlig abzuwendenden Klimaveränderungen werden verstärkt politische Aktivitäten und gesellschaftliche Veränderungen notwendig. Dies betrifft alle Handlungsebenen: Von der Politik, über die Wirtschaft bis hin zu den privaten Haushalten. Evaluationen sollten zur Steuerung dieser Lernprozesse in Richtung Nachhaltigkeit verstärkt genutzt werden.

Zumindest im Bereich der für den Klimaschutz zentralen Energiepolitik sind aktuell zunehmende Aktivitäten in Deutschland zu verzeichnen. Fehlende oder sehr selten eingesetzte Standards (wie die der DeGEval 2002) und zu enge Blickwinkel (technisch-naturwissenschaftlich) schränken die Aussagekraft und somit die Nützlichkeit der Ergebnisse ein.

Es gibt Chancen für konzeptionell bessere und nützliche Evaluationen im Umwelt- und Energiebereich. Zum Beispiel durch Vorgaben der EU, die Effekte nationaler politischer Instrumente nachzuweisen. Der Wunsch nach vergleichbaren Ergebnissen zwischen den EU-Staaten sollte hier Standardisierungen der Evaluationssystematik zur Qualitätssicherung fördern.

Eine weitere große Herausforderung liegt darin, wie in einem komplexen Querschnittsthema wie dem Umweltschutz das (nötige) Zusammenspiel verschiedener Maßnahmen hinreichend abgebildet und bewertet werden kann. Diese Herausforderung wird noch vergrößert, wenn statt einer „Umweltevaluation“ eine „Nachhaltigkeitsevaluation“ im Sinne einer weitgehenden Erfassung von komplexen Wirkungsgefügen angestrebt wird. Dies kann durch die Berücksichtigung der Aspekte Ökonomie, Ökologie, Soziales und die Ent-

wicklung entsprechender Indikatoren geschehen. Voraussetzung ist eine adäquate Methodenentwicklung, welche die hohen Anforderungen durch die Vielfalt der einbezogenen Aspekte und deren Zuordnung und Gewichtung beinhaltet.

Eine Weiterentwicklung der Methodik durch eine Bestandsaufnahme, Überprüfung und Nutzung der in inter- und transdisziplinärer Arbeit entwickelten Verfahren, Methoden und Instrumente einer Umwelt- bzw. Nachhaltigkeitsevaluation sind erstrebenswerte Schritte.

Ein wichtiger Baustein wäre in Deutschland eine erste umfassende Studie, die aktuell die bisherigen Evaluationen im Themenfeld Umwelt und Nachhaltigkeit zusammenträgt, auswertet und daraus methodische, organisatorische und konzeptionelle Vorschläge entwickelt. Hier besteht deutlicher Forschungsbedarf.

Evaluation im Themenfeld Energie und Umwelt – Länderstudie Österreich

1 Die österreichische Umwelt- und Energiepolitik als Kontext von Evaluationen

1.1 Meilensteine, Erfolge und Schwachstellen

Österreich galt viele Jahre hindurch als Umweltmusterland: die Nicht-Inbetriebnahme des fertig gebauten Kernkraftwerks Zwentendorf in Folge einer Protestbewegung und Volksabstimmung im Jahr 1978, die Demonstrationen und der darauf folgende Baustopp des Donaukraftwerks Hainburg im Winter 1984/85 sowie die Auswirkungen der Reaktorkatastrophe von Tschernobyl im Jahr 1986 hatten in der österreichischen Bevölkerung ein breites Umweltbewusstsein verankert. Durch regulative und investitionsintensive Maßnahmen konnten in einigen Feldern beachtliche Erfolge verzeichnet werden¹, so dass sich Österreich auf internationaler Ebene als Umweltvorreiter positionieren konnte. Als Beleg dafür dienen bis heute die strengen Emissionsstandards, die hervorragende Wasserqualität österreichischer Seen, der mit 23 Prozent hohe Anteil erneuerbarer Energie am Bruttoinlandsverbrauch, der mit 15 Prozent im internationalen Vergleich hervorragende Anteil des Bio-landbaus an der gesamten landwirtschaftlichen Fläche, Vorzeigeprojekte von Passiv- und Niedrigenergiehäusern und die Gründung von sechs Nationalparks. In den letzten zehn Jahren kam es nach Meinung kritischer WissenschaftlerInnen und Nichtregierungsorganisationen zu einer Stagnation der umweltpolitischen Anstrengungen. Besonders kritisiert werden die bisher fehlenden Erfolge der Energie- und Klimaschutzpolitik (Österreich liegt derzeit mehr als 30 % über seinem Kyoto-Ziel), die Gesundheitsgefährdungen durch Feinstaubemissionen, die ungebremsste Flächenversiegelung und das Fehlen einer integrierten nachhaltigkeitsbezogenen Verkehrspolitik, die sich in ständig steigenden Transit und Verkehrsaufkommen niederschlägt (vgl. Lauber 1997; OECD 2003; AGEZ/FORUM/ÖKOBÜRO 2006; Pesendorfer 2007).

¹ Zum Beispiel Stopp des sauren Regens und des Waldsterbens durch Reduktion des SO₂-Gehalts von Kraftstoffen und die Umrüstung von Kohlekraftwerken auf Gas; Einführung des Katalysators, bleifreien Benzins und strenger Abgasvorschriften nach kalifornischem Vorbild, deutliche Verbesserung der Wasserqualität von Flüssen und Seen durch den Bau von Kläranlagen.

1.2 Institutionelle Rahmenbedingungen und umweltpolitische Instrumente

Das im Jahre 1972 eingerichtete Umweltministerium hatte lange Zeit hindurch nur wenige Kompetenzen und geringe eigenständige Handlungsmöglichkeiten. Mitte der 1980er Jahre gewann es parallel zur gesellschaftlichen Sensibilisierung für Umweltfragen an politischer Bedeutung und wurde personell ausgebaut. Im Jahr 2000 wurde es mit dem Landwirtschaftsministerium zum „Bundesministerium für Land- und Forstwirtschaft, Umwelt und Wasserwirtschaft“ zusammengeführt und in der Öffentlichkeitsarbeit als „Lebensministerium“ bezeichnet. Diese Fusion vergrößerte einerseits die Handlungsmöglichkeiten der Umweltpolitik (z. B. Förderung von Biolandbau und nachwachsenden Rohstoffen, Verzicht auf den Einsatz gentechnisch veränderter Organismen), hatte andererseits eine Interessenbindung in der Agrarpolitik zur Folge.

Bis heute sind zentrale *Kompetenzen der Umwelt- und Energiepolitik* nicht im Lebensministerium konzentriert, so dass sich komplexe Querschnittskompetenzen und Abstimmungsnotwendigkeiten ergeben, wie beispielsweise:

- mit dem Bundesministerium für Wirtschaft und Arbeit in der Energiepolitik;
- mit dem Bundesministerium für Verkehr, Innovation und Technologie in der Verkehrs- und Technologiepolitik;
- mit den Ländern und Gemeinden in Naturschutz und Flächenwidmung.²

Mit der im Jahr 2002 beschlossenen österreichischen Nachhaltigkeitsstrategie und den darauf folgenden Umsetzungsbemühungen fand eine Ausweitung der Koordinationserfordernisse des Lebensministeriums in andere Politikbereiche statt. Durch die gleichgewichtige Berücksichtigung ökologischer, ökonomischer und sozialer Problemstellungen und der gleichzeitigen Beachtung internationaler Aspekte haben sich neue Schnittstellen zu den für Wirtschafts-, Sozial- und Außenpolitik zuständigen Ressorts ergeben. Diese horizontale Integration soll in den nächsten Jahren um Elemente einer zusätzlichen vertikalen Integration ergänzt werden, indem die Bundesländer verstärkt in die Umsetzung der österreichischen Nachhaltigkeitsstrategie eingebunden werden.

Die *Instrumente der österreichischen Umwelt- und Energiepolitik* der 1980er und frühen 1990er Jahre waren von einem technisch-naturwissenschaftlichen, an Grenzwerten und Verwaltungsverfahren orientierten Zugang geprägt, der zu einer raschen Zunahme der umweltrechtlichen Bestimmungen führte. Seit Mitte der 1990er Jahre ist ein Wandel zu neuen, oft marktorientierten oder freiwilligen Instrumenten der Umweltpolitik festzustellen, indem Managementsysteme (EMAS, ISO14001), Labels (österreichisches Umweltzeichen), Emissionszertifikate im Klimaschutz sowie diverse Forschungs-, Technologie-, Beratungs- und Förderprogramme an Bedeutung gewannen. Der zur gleichen Zeit erstellte nationale Umweltplan (NUP) trug zu einer Kultur transparenter und partizipativer Aushandlungsprozesse bei und unterstützte die Bildung informeller Netzwerke in der österreichischen

² Erschwerend kommt hinzu, dass eine Menge umweltrelevanter Tatbestände in die Gesetzgebungskompetenz der Länder fällt (z.B. Bauordnungen, Naturschutzgesetze) und somit eine historisch bedingte Heterogenität der Rechtsgrundlagen in den verschiedenen Bundesländern vorliegt. Der im Jahr 2005 beendete Verfassungskonvent hat zwar Reformbedürfnisse aufgezeigt, konnte aber zu keiner Vereinheitlichung beitragen.

Umweltpolitik und -verwaltung. Mit der im Jahr 2002 von der Bundesregierung beschlossenen österreichischen Nachhaltigkeitsstrategie und einigen weiteren Strategien (z.B. Biodiversitätsstrategie, Klimastrategie) wurde strategische Steuerung in der österreichischen Umwelt- und Energiepolitik etabliert und mit einer Aufbau- und Ablauforganisation kombiniert, die regelmäßige Evaluationen beinhaltet.

Programme als umwelt- und energiepolitische Steuerungsinstrumente sind in Österreich in vielerlei Varianten anzutreffen: im Agrarbereich (z.B. Österreichisches Programm für die Entwicklung des ländlichen Raums), in Forschung und Entwicklung (z.B. BMVIT-Programm Nachhaltig Wirtschaften, BMBWK-Programm proVision, BMLFUW-Forschungsprogramm PFEIL10), als Maßnahmenförderung (z.B. Umweltförderung, der im Jahr 2007 neu aufgelegte Energie- und Klimaschutzfonds) sowie als geförderte Beratung (nationales Programm klima:aktiv, regionale Beratungsprogramme der Länder). In der Programmperiode 2000-2006 hatten auch die Strukturfondsprogramme der Europäischen Union eine gewisse Bedeutung. Nicht zuletzt aufgrund des Einsatzes beträchtlicher öffentlicher Mittel und den Evaluationsvorgaben der europäischen Ebene werden einige dieser umwelt- und energiepolitisch relevanten Programme in regelmäßigen Abständen evaluiert.

2 Evaluationen im Themenfeld Umwelt und Energie in Österreich

2.1 Die Institutionalisierung von Evaluation

Die Institutionalisierung von Evaluation ist in Österreich bislang gering:

- Die *Nachfrage nach Evaluationen* ist punktuell und hängt zumeist von Initiative und Interesse einzelner Auftraggeber ab. Es gibt keine verfassungsmäßigen und kaum gesetzliche Verpflichtungen zur Evaluation.³ In den wichtigsten Auftraggeber-Institutionen fehlen die institutionellen Kapazitäten, um einen kontinuierlichen Wissensaufbau und Qualifikationserwerb sicherzustellen. So verfügt keines der im Umwelt- und Energiebereich wichtigen Bundesministerien über Stabstellen, die sich kontinuierlich mit der Bündelung des Wissens über Vergabe und Verwertung von Evaluationsergebnissen beschäftigen.
- Auch auf Seiten der *Anbieter von Evaluationen* ist nur eine geringe Vernetzung und Institutionalisierung anzutreffen. Im Gegensatz zur Schweiz und zu Deutschland gibt es in Österreich keine nationale Evaluationsgesellschaft, die sich den Themen Erfahrungsaustausch, Qualitätssicherung und Weiterbildung widmen könnte. Einzig die Plattform Forschungs- und Technologieevaluierung⁴ hat in einem schmalen Themenbereich eine Vernetzungsfunktion übernommen und auch bereits Standards veröffent-

³ Ausnahmen sind die zunehmenden Selbstverpflichtungen zur Evaluation in Strategie-Dokumenten (z.B. österreichische Nachhaltigkeitsstrategie), die regelmäßige Evaluation von Förderprogrammen (z.B. Umweltförderung, Förderung des ökologischen Landbaus) sowie die im Rahmen der Programmplanung vorgesehenen Evaluationen von Strukturfondsprogrammen. Es gibt in Österreich jedoch keine der Schweiz vergleichbare „Nachhaltigkeitsprüfung“ und auch kein verfassungsrechtlich vorgesehenes Evaluationsgebot.

⁴ www.fteval.at [22.8.2007]

licht. Der im Jahr 1995 vom Bundeskanzleramt eingerichtete Checkpoint EVA, eine Koordinierungsstelle für Evaluation von Regionalpolitik und regionalpolitischen Programmen, wurde nur bis zum Jahr 2000 finanziert und daraufhin eingestellt (vgl. Holzinger 2001).

Eine Folge dieses geringen Institutionalisierungsgrades ist eine *geringe Markttransparenz*: Evaluationen werden zwar beauftragt, aber nicht notwendigerweise veröffentlicht und in keiner Weise zentral dokumentiert. Die nachfolgend dargestellte Länderstudie musste daher die erforderliche Datenbasis zur Analyse der Evaluation im Themenfeld Umwelt und Energie erst schaffen. Sie konnte dabei auf einer vergleichbaren Studie aus dem Jahr 2000 (vgl. Kuffner 2000) aufbauen, so dass – mit einigen Einschränkungen – auch Aussagen über die Entwicklungen der letzten Jahre getätigt werden können.

2.2 Methodik und Datenbasis der Länderstudie

Die vorliegende Länderstudie erfolgte in vier Schritten:

1. Recherche von österreichischen Evaluationsberichten in den Themenfeldern Umwelt und Energie aus den Jahren 2000 bis 2007
2. Analyse von Themenstellung, Methodik und Qualität der recherchierten Evaluationsberichte
3. Vergleich der Ergebnisse mit der Studie aus dem Jahr 2000
4. Interpretation der Ergebnisse und Vergleich mit internationalen Entwicklungen

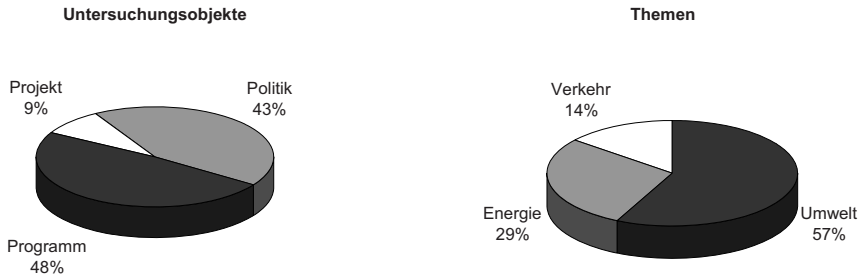
Aufgrund der langjährigen Beschäftigung mit dem Thema Evaluation nachhaltiger Entwicklung (vgl. Martinuzzi 2004; Martinuzzi/Meyer/Kopp 2007) und der Mitarbeit bei der Erstellung und Umsetzung der österreichischen Nachhaltigkeitsstrategie (vgl. Martinuzzi/Steurer 2003; 2005) konnten die Autoren der vorliegenden Studie auf qualifiziertes Adressmaterial von über 500 Personen (Auftraggeber und Ausführende von Evaluationen) zurückgreifen. Diese wurden im Herbst 2006 kontaktiert und um Übermittlung von Evaluationen im Themenfeld Umwelt und Energie bzw. um Hinweise auf derartige Evaluationen ersucht⁵. Der Rücklauf war unerwartet gering und erbrachte trotz mehrfachen Nachfragens nur fünf Evaluationsberichte. Durch eine anschließende Internet- und Bibliotheksrecherche konnte die Datenbasis auf insgesamt 35 Evaluationsberichte erweitert werden (Übersicht siehe unten, Abschnitt 6). Die Berichte wurden anhand eines Kriterienrasters analysiert. Die wichtigsten Ergebnisse sind nachfolgend dargestellt (weiterführende Analysen siehe Leschnik 2007).⁶

⁵ Aufgrund der hohen ökologischen und energetischen Bedeutung wurden auch Studien aus dem Verkehrsbereich einbezogen.

⁶ Die nachfolgenden Aussagen sind dahingehend zu relativieren, dass sie ausschließlich auf der Analyse dieser 35 Evaluationsberichte basieren und dass keine Daten über die Grundgesamtheit aller im Themenfeld Umwelt und Energie durchgeführten Evaluationen verfügbar sind. Den Autoren der vorliegenden Arbeit erscheint es jedoch plausibel, dass die durchgeführte Recherche einen sehr hohen Abbildungsgrad hat.

2.3 Schwerpunkte der Evaluationen

Abbildung 1: Untersuchungsobjekte und Themen österreichischer Evaluationen im Themenfeld Umwelt und Energie



[n=35, Zeitraum 2000-2007]

Von den 35 untersuchten Evaluationsberichten beschäftigen sich 16 mit Programmen, vornehmlich im Themenfeld Umwelt, wohingegen die Mehrzahl der evaluierten Politiken (Gesetze, Strategien und Fördersysteme) im Themenfeld Energie zu finden ist. Projekte als Evaluationsobjekte sind vergleichsweise selten.⁷

Mehr als die Hälfte der untersuchten Evaluationsberichte finden sich im **Themenfeld Umwelt**. Sie befassen sich zumeist mit komplexen Steuerungsinstrumente wie Programmen (Umweltförderung, Cleaner Production Programme, Impulsprogramm Nachhaltig Wirtschaften, ÖkoBusinessPlan Wien, Wirtschaftsinitiative Nachhaltigkeit Steiermark), und Strategien (österreichische Nachhaltigkeitsstrategie). Kaum anzutreffen sind rein naturwissenschaftlich-ökologische Evaluierungen oder Evaluierungen von Ge- und Verboten. Zentraler Auftraggeber im Themenfeld Umwelt ist das Lebensministerium mit zehn von zwanzig untersuchten Evaluationsberichten. Die anderen zehn Berichte verteilen sich auf andere Ministerien (3 Berichte) und Landesbehörden (7 Berichte).

In den **Themenfeldern Energie und Verkehr** ist keine vergleichbare Konzentration auf Seite der Auftraggeber feststellbar. Rund die Hälfte der analysierten Evaluationsberichte wurde von Landesbehörden beauftragt, die andere Hälfte von verschiedenen Bundesbehörden. Gerade im Energiebereich finden sich volkswirtschaftlich orientierte Evaluationen (z. B. volkswirtschaftliche Auswirkungen der Unterstützung von Ökostrom in Österreich, Analyse der volkswirtschaftlichen Bedeutung der energetischen Nutzung von Biomasse für Heizzwecke, ökonomische Evaluierung eines österreichischen CO₂-Pilot-Trading-Systems) und technisch-naturwissenschaftliche Evaluationen (Evaluierung der emissionsseitigen Auswirkungen der Einführung einer durchgehenden 100 km/h Beschränkung, Evaluation der lufthygienischen Auswirkungen der Erweiterung des Nachtfahrverbotes).

⁷ Im Gegensatz zu Projekten, die ein klar definiertes Ziel in einem bestimmten Zeitraum verfolgen, wurden Programme als komplexes Bündel von Angeboten an bestimmte Zielgruppen verstanden, die dadurch zu geänderten Entscheidungen oder Verhalten motiviert werden sollen (z.B. Förderprogramme, Bildungsprogramme, Beratungsprogramme).

Insgesamt betrachtet sind daher Programme im Umweltbereich und volkswirtschaftliche Folgenabschätzungen im Energiebereich die wichtigsten Themenstellungen der untersuchten österreichischen Evaluationen.

2.4 Beteiligte an Evaluationen

Rund die Hälfte der untersuchten Evaluationen wurde von Bundesbehörden beauftragt (v. a. durch das Lebensministerium), rund ein Drittel von Landesbehörden (v. a. in den Themenfeldern Energie und Verkehr). Kommunen und sonstige Auftraggeber (z.B. E-Control GmbH) sind als Auftraggeber von Evaluationen selten.

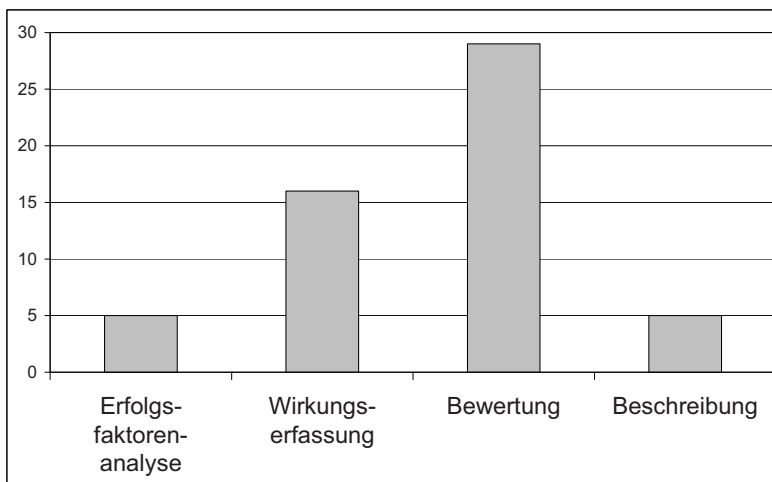
Die Auftragnehmer von Evaluationen sind zu rund je einem Drittel Universitäten, freie Forschungsinstitute und Consulting-Unternehmen. Auffällig ist die regionale Verteilung der Auftragnehmer: Fast sechzig Prozent haben ihren Stammsitz oder zumindest ein Büro in Wien. Mögliche Ursachen könnten in der Konzentration von Universitätsstandorten in der Bundeshauptstadt, in regionaler Nähe oder in informellen Kontakten zum Auftraggeber als Wettbewerbsvorteil bei der Vergabe von Evaluationsaufträgen liegen.

2.5 Merkmale von Evaluationen

Siebzig Prozent der erhobenen Evaluationen wurden begleitend durchgeführt, je fünfzehn Prozent ex-ante und ex-post. Die meisten von ihnen dienten der Unterstützung von Entscheidungen über die Fortsetzung einer Strategie oder eines Programms.

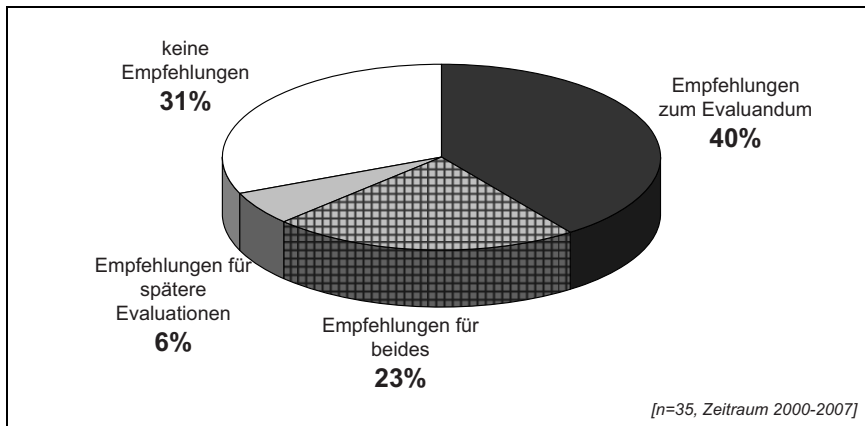
Mehr als neunzig Prozent der analysierten Evaluationen setzten quantitative Erhebungsverfahren ein, rund fünfzig Prozent kombinierten sie jedoch mit qualitativen Methoden. Der ausschließliche Einsatz qualitativer Methoden ist hingegen mit neun Prozent sehr selten.

Abbildung 2: Methodik österreichischer Evaluationen im Themenfeld Umwelt und Energie



Fast alle untersuchten Evaluationen bieten eine Bewertung des Evaluandums, nur die Hälfte enthält aber auch eine Darstellung der ausgelösten Wirkungen. Mit je fünf Fällen sind eine detaillierte Beschreibung des Evaluandums auf der einen Seite und eine Analyse von Erfolgsfaktoren auf der anderen Seite hingegen vergleichsweise selten.

Abbildung 3: Empfehlungen in österreichischen Evaluationen im Themenfeld Umwelt und Energie



Rund ein Drittel der untersuchten Evaluationen enthält überhaupt keine Empfehlungen, rund zwei Drittel bieten Empfehlungen zum Evaluandum und rund ein Drittel zeigt darüber hinaus auch noch Perspektiven für spätere Evaluationen auf. Dies deutet gemeinsam mit dem hohen Anteil begleitender Evaluationen darauf hin, dass im Themenfeld Umwelt und Energie Evaluationen zunehmend als Feedback mit dem Ziel des organisierten Lernens eingesetzt werden.

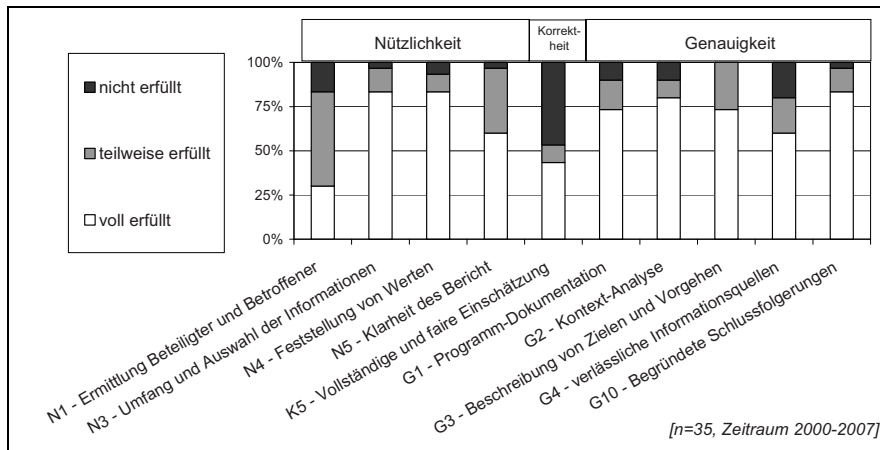
Dieser Befund stimmt mit den persönlichen Erfahrungen der Autoren der vorliegenden Studie überein, die in den letzten sieben Jahren eine spürbare Änderung der Motivlage von Evaluationsauftraggebern feststellen konnten: Stand vor einigen Jahren die Legitimation durch externe Experten im Vordergrund der Beauftragung, nimmt das Interesse der Auftraggeber an neuen Erkenntnissen durch Evaluationen deutlich zu. Sie beauftragen zunehmend partizipative und lernorientierte Evaluationskonzepte und involvieren sich selbst stärker in den Evaluationsprozeß.

2.6 Die Qualität der untersuchten Evaluationen

Um die Qualität der österreichischen Evaluationen im Themenfeld Umwelt und Energie zu untersuchen wurden jene zehn Evaluationsstandards der DeGEval- (DeGEval 2002) und SEVAL-Standards (Widmer et al. 2000) herangezogen, die ausschließlich auf Basis der Evaluationsberichte (und ohne Kenntnis der Details der Evaluationsprozesse) beurteilt werden können. Sie wurden von den Autoren der vorliegenden Studie mittels einer dreiteiligen Skala beurteilt (voll erfüllt, teilweise erfüllt, nicht erfüllt):

- Die **Genauigkeitsstandards („G“)** weisen den höchsten Erfüllungsgrad auf. Die meisten Evaluationsberichte beschreiben die vorliegende Intervention detailliert und führen eine Kontextanalyse durch. Die Schlussfolgerungen sind schlüssig und gut begründet
- Auch die **Nützlichkeitsstandards („N“)** werden in der Regel gut erfüllt, einzig der Nützlichkeitsstandard „Ermittlung der Beteiligten und Betroffenen“ fällt aus der Gesamtansicht: Viele Berichte nennen zwar Auftraggeber und Evaluationsteam, beziehen jedoch keine möglichen Betroffenen ein bzw. verzichten auch auf ihre Erwähnung.
- Der von den **Korrekttheitsstandards („K“)** verwendete Standard „ausgewogene Darstellung von Stärken und Schwächen“, ist am schwächsten ausgeprägt: Viele Berichte tendieren dazu, entweder nur Stärken oder nur Schwächen darzustellen. Hier könnte eine überpointierte Darstellung durch die jeweiligen Evaluatoren vorliegen oder sich ein Einfluss der Auftraggeber (die auf möglichst eindeutige Ergebnisse drängen) manifestieren.

Abbildung 4: Qualität österreichischer Evaluationen im Themenfeld Umwelt und Energie



3 Die Entwicklung seit dem Jahr 2000

Um die Entwicklung der Evaluation im Themenfeld Umwelt und Energie seit dem Jahr 2000 nachzuzeichnen, bietet sich ein Vergleich mit einer im Jahr 2000 veröffentlichten Arbeit über Nachhaltigkeitsevaluationen an (vgl. Kuffner 2000).⁸ Dabei zeigen sich folgende Trends:

- Auf die Bundesebene entfallen heute deutlich mehr Evaluationen als im Jahr 2000, wohingegen Evaluationen im Auftrag der Bundesländer in selben Ausmaß abgenommen haben. Der Anteil von Politikevaluationen hat ebenfalls zugenommen, wohingegen Projektevaluationen heute seltener sind. Dies deutet darauf hin, dass Evaluation an politischem Stellenwert gewinnt.

⁸ Kuffner hatte für den Zeitraum von 1994 bis 1999 insgesamt 43 deutschsprachige Evaluationen recherchiert, die sich mit Aspekten nachhaltiger Entwicklung beschäftigten. Damit deckt ihre Arbeit einen breiteren Untersuchungsraum und ein breiteres Themenfeld ab, als die vorliegende Länderstudie. Für die vorliegende Arbeit wurden einige identische Fragestellungen und Kriterien verwendet, um eine direkte Vergleichbarkeit zu ermöglichen.

- Förderungen und Programme haben als Gegenstand von Evaluationen deutlich zugenommen. Durch ihren Innovationsgrad und die investierten öffentlichen Mittel scheint eine gewisse Selbstverpflichtung zur Evaluation zu entstehen.
- Die Anzahl begleitender Evaluationen nimmt in den letzten Jahren etwas zu, was darauf hin deutet, dass Evaluationen zunehmend als Lernchance gesehen werden.
- Die verwendete Methodik und analytische Tiefe der Darstellung haben sich kaum verändert. Es dominieren quantitativ orientierte Analysen und darauf aufbauende Bewertungen. Evaluationen legitimieren sich zumeist durch repräsentative Stichproben und quantitative Datenanalysen. Qualitativ orientierte Erklärungen von Wirkungsketten und Erfolgsfaktoren sind weiterhin selten. Das auf Steuerungsmöglichkeiten abzielende Erkenntnisinteresse der Auftraggeber spiegelt sich daher noch nicht in den Evaluationsmethoden wider.
- Die Erfüllung des Standards „Feststellung von Werten“ hat sich in den letzten Jahren deutlich verbessert. Dies mag jedoch auch methodisch bedingt sein, da im Jahr 2000 analysierten Evaluationen sich auch mit Fragen der sozialen Nachhaltigkeit beschäftigt hatten und vor der Herausforderung standen, Werte zu thematisieren, wohingegen die nun untersuchten Evaluation im Themenfeld Umwelt und Energie weniger stark gesellschaftspolitische Themen zum Hintergrund haben und daher auch weniger stark Wertefragen tangieren.
- Die Erfüllung des Standards „Vollständige und faire Einschätzung“ hat sich verschlechtert, was nach Ansicht der Autoren der vorliegenden Arbeit als durchaus kritisch zu sehen ist. Hier wären vertiefende Analysen sinnvoll und Mechanismen der Qualitätssicherung zu entwickeln.

4 Verbindungen zu anderen Evaluationsfeldern und zu internationalen Entwicklungen

Wie bereits in Kapitel 2.1. erwähnt, ist die Institutionalisierung von Evaluation in Österreich gering. Dementsprechend gering sind auch die Verbindungen der Evaluationsaktivitäten im Themenfeld Umwelt und Energie mit anderen Evaluationsfeldern sowie mit internationalen Entwicklungen. Daher bietet die durchgeführte Länderstudie wenige Anknüpfungspunkte zur Analyse derartiger Verbindungen. Aufgrund der persönlichen Erfahrungen der Autoren der vorliegenden Studie können einige Trends identifiziert werden:

- *Als umwelt- und energiepolitische Steuerungsinstrumente werden zunehmend Strategien eingesetzt.* Sie enthalten in vielen Fällen Verpflichtungen zur Evaluation als Instrument der Fortschritts- und Wirkungsmessung. Diese Kombination (Steuerung durch Strategien, Lernen durch Evaluationen) könnte als „Strategic Public Management“ (vgl. Steurer/Martinuzzi 2005 sowie Steurer 2007) in den nächsten Jahren europaweit und somit auch in Österreich an Bedeutung gewinnen. Unterstützt wird dieser Trend beispielsweise durch den „governance cycle“ der europäischen Nachhaltigkeitsstrategie, der zweijährige Umsetzungsberichte vorsieht und damit eine (wenn auch schwach strukturierte) regelmäßige Form von Evaluation etabliert.

- *Programme gewinnen als operative Instrumente der österreichischen Umwelt- und Energiepolitik an Bedeutung.* Sie kombinieren einzelne Instrumente der konventionellen Umweltpolitik (z. B. Förderung, Beratung, Öffentlichkeitsarbeit), enthalten komplexe Verhaltensangebote für Betriebe und Einzelpersonen und tragen zum Aufbau von Akteursnetzwerken bei (vgl. Martinuzzi 2006). Aufgrund ihrer Komplexität, ihres Innovationsgrades und der in sie investierten Finanzmittel werden Evaluationen hier zunehmend eingesetzt. Ein Trend, der wahrscheinlich weiter anhalten wird.
- *Strukturfonds werden in den kommenden Jahren in Österreich an Bedeutung verlieren,* da für die Programmperiode 2007-2013 um rund 25 % geringere Finanzmittel für Strukturfondsprojekte in Österreich vorgesehen sind (vgl. ÖKOR 2007). Die von Strukturfonds ausgehenden institutionellen und instrumentellen Effekte werden daher voraussichtlich gering bleiben.
- Theoriebasierte Evaluationsansätze (vgl. Chen/Rossi 1980; Bickman 1987; Donaldson 2003; Chen 2005) haben in den letzten Jahren in Österreich im Themenfeld Umwelt und Energie an Bedeutung gewonnen und bieten Auftraggebern detailliertere Einblicke in die Wirkungsweise von Programmen. Eine weitere Verbreitung dieser Methoden ist zu erwarten
- *Evaluationsstandards* wurden in den letzten Jahren zwar in einigen Ausschreibungen explizit erwähnt⁹, ihre Erfüllung war jedoch kein relevantes Kriterium der Auftragsvergabe und es finden sich in den entsprechenden Evaluationsberichten bisher auch keine Verweise auf Evaluationsstandards.

5 Bilanzierung des bisher Erreichten

Evaluationen im Themenfeld Umwelt und Energie bieten in Österreich derzeit folgendes Gesamtbild:¹⁰

- *Es dominieren punktuell durchgeführte Evaluationen.* Es fehlt eine öffentlich zugängliche zentrale Dokumentation von Evaluationen. Bisher wurden kaum Meta-Evaluationen, Meta-Analysen oder Synthesen von Evaluationen durchgeführt. Eine gesetzliche Verpflichtung zur Veröffentlichung von Evaluationen und eine öffentlich finanzierte Dokumentation von Evaluationen wären hilfreich, um die Nutzenpotenziale von Evaluationen in vollem Umfang zu erschließen.

⁹ Zum Beispiel: „Nach einem Zeitraum von drei Jahren sollen daher die Leitziele, der Umsetzungsprozess und die erzielten Wirkungen der Österreichischen Nachhaltigkeitsstrategie einer unabhängigen Evaluation unterzogen werden. Dabei sollen die von den Europäischen Evaluationsgesellschaften entwickelten Standards für Evaluationen hinsichtlich Nützlichkeit, Durchführbarkeit, Fairness und Genauigkeit angewendet werden.“ (BMLFUW 2002), weitere Erwähnungen von Evaluationsstandards in den Ausschreibungen zur Evaluation des Beratungsprogramms klima:aktiv und des ÖkoBusinessPlan Wien.

¹⁰ Um das bisher Erreichte zu bilanzieren greifen die Autoren der vorliegenden Länderstudie nicht nur auf die durchgeführte Recherche und Analyse von Evaluationsberichten zurück, sondern auch auf ihre langjährige Tätigkeit im Arbeitskreis Umweltevaluation der Deutschen Gesellschaft für Evaluation DeGEval, auf ihre eigenen Erfahrungen als Evaluatoren und die Erkenntnisse aus der Koordinierung der Serie von EU-Konferenzen und Trainings EASY-ECO Evaluation of Sustainability (siehe www.easy-eco.eu).

- *Der hohe methodische Innovationsgrad einiger Evaluationen strahlt bisher kaum auf andere Evaluationen aus.* Die meisten Evaluationen verzichten (vermutlich aufgrund knapper Evaluationsbudgets) auf eine fundierte Analyse der Fachliteratur oder eine Bezugnahme auf andere Evaluationen, sondern starten mit einer Beschreibung des Evaluandums und der Evaluationsmethodik. Damit findet keine Mehrfachnutzung von Evaluationsergebnissen statt und es kommt zu keinem kumulativen Wissensgewinn.
- Evaluationen werden häufig an für die jeweiligen Inhalte stehende Fachexperten vergeben, ein Diskurs über Evaluationsmethoden ist in Österreich kaum vorhanden. Evaluationen werden offensichtlich als Expertengutachten verstanden, bei denen allein die Inhaltsexpertise für Glaubwürdigkeit und Akzeptanz der Evaluationsergebnisse sorgt. Eine transparente und fundierte Darstellung und Analyse von Wirkungen, Kausalketten und Erfolgsfaktoren ist vergleichsweise selten.
- Hohe finanzielle Aufwendungen, innovative umwelt- und energiepolitische Instrumente sowie internationale Anforderungen sind die wichtigsten Nachfrageimpulse. Es fehlen jedoch institutionelle Kapazitäten und fachliche Qualifikationen auf Seite der Auftraggeber, um eine professionelle Vergabe und Nutzung von Evaluationen sicherzustellen. Die ausgelösten Lernprozesse beschränken sich daher zumeist auf das einzelne Evaluationsprojekt und reichen kaum darüber hinaus. Positiv hervorzuheben ist die Tatsache, dass Evaluationen zunehmend als Feedbackschleife eingesetzt und wiederholte Evaluationen häufiger werden. Das Schaffen institutioneller Kapazitäten und die Ausbildung von Auftraggebern stellen nach Ansicht der Autoren der vorliegenden Arbeit zentrale Herausforderungen dar.

Künftige Impulse zur Etablierung von Evaluationen im Themenfeld Umwelt und Energie könnten aus der Verwaltungs- und Bundesstaatsreform, einer Stärkung des Rechnungshofs als „Nachhaltigkeitsprüfer“ und einer Weiterentwicklung der strategischen Umweltprüfung zu einem Instrument einer integrierten Nachhaltigkeitspolitik resultieren. Die dafür erforderliche Lernorientierung ist nach Ansicht der Autoren der vorliegenden Arbeit auf der Ebene der Verwaltung bereits verabredet. Es wird noch Überzeugungsarbeit erforderlich sein, um politische Entscheidungsträger vom Nutzen wirkungsorientierter Politik und faktenbasierter Entscheidungen zu überzeugen.

6 Die analysierten Evaluationsberichte

| Titel | Jahr | Auftraggeber | Auftragnehmer |
|---|------|-----------------------------|---|
| Bericht zur Evaluierung der Vorarlberger Wohnbauförderung. Umweltrelevante Aspekte | 2000 | Vorarlberger Naturschutzrat | Spektrum – Zentrum für Umwelttechnik & -management GmbH |
| Cleaner Production in Österreich – Eine Bilanz | 2000 | BMVIT, BMLFUW | Institut für Industrielle Ökologie, Joanneum, Stenum GmbH |
| Evaluierung der Umsetzung des Österreichischen Umweltzeichens für Tourismusbetriebe | 2000 | BMUJF | Institut für Integrativen Tourismus & Freizeitforschung |
| Ökostadt 2000 – Evaluierung | 2000 | Stadt Graz – Umweltamt | Grazer Ökoteam |

| Titel | Jahr | Auftraggeber | Auftragnehmer |
|---|-------------|--|--|
| Energiebinnenmarkt und Umweltschutz: Evaluierung für Österreich | 2001 | BMLFUW | Energieverwertungsagentur, Institut für Energiewirtschaft [TU Wien] |
| Analyse der volkswirtschaftlichen Bedeutung der energetischen Nutzung von Biomasse für Heizzwecke und Entwicklung von effizienten Förderstrategien für Österreich | 2002 | BMWA | TU Wien |
| Auswirkungen der NÖ ÖKO-Eigenheimförderung auf Baupraxis und CO ₂ -Bilanz | 2002 | NÖ LReg – Wohnbauforschung | Die Umweltberatung, Arbeitsgruppe Energiewirtschaft [TU Wien] |
| Evaluierung der Umsetzung der IPF- und IFF-Aktionsvorschläge in Österreich | 2002 | BMLFUW | Institut für Sozioökonomik der Forst- und Holzwirtschaft [Universität für Bodenkultur] |
| Design und ökonomische Evaluierung eines österreichischen CO ₂ -Pilot-Trading-Systems | 2002 | BMLFUW | Wirtschaftsforschungsinstitut |
| Salzburger Landesverkehrskonzept 1991 – Evaluierung 2001 | 2002 | Sbg. LReg – Fachabt. 6/7 – Verkehrsplanung | DI Dietmar Krammer – Ziviltechnikerbüro für Bauwesen |
| Evaluierung der (heiz-) energierelevanten Förderungen der Stadt Wien | 2003 | Stadt Wien – MA 22 | Energieverwertungsagentur |
| Evaluierung der NÖ Wärmepumpenförderung | 2003 | NÖ LReg – Gruppe Finanzen – Abt. Wohnbauförderung | Haustechnik Planungsgesellschaft für Ver- und Entsorgungsanlagen Ges.m.b.H. |
| Evaluierung der Österreichischen Strategie zur Umsetzung des Übereinkommens über die biologische Vielfalt | 2003 | Umweltbundesamt (UBA) | UBA |
| Evaluierung der Solar-initiative „Sonnenland Kärnten“. Endbericht | 2003 | Selden & Partner | Institut für höhere Studien Kärnten |
| Evaluierung der Umsetzung des erweiterten CBD-Arbeitsprogramms für biologische Vielfalt der Wälder und der IPF- und IFF-Aktionsvorschläge in Österreich | 2003 | BMLFUW | Institut für Sozioökonomik der Forst- und Holzwirtschaft [Universität für Bodenkultur] |
| Evaluierung der Wirksamkeit forstlicher Erschließungsmaßnahmen | 2003 | BMLFUW | Universität für Bodenkultur |
| Forschungsprojekt zur Evaluierung der Akzeptanz von Lärmschutzmaßnahmen an Eisenbahnbestandsstrecken | 2003 | BMVIT; OÖ LReg; Vbg. LReg; ÖBB; OÖ Gemeinden, Vbg. Gemeinden | Institut Retzl, TAS Schreiner GmbH |
| Bewertung der volkswirtschaftlichen Auswirkungen der Unterstützung von Ökostrom in Österreich | 2004 | E-Control GmbH | Institut für höhere Studien Kärnten |
| Evaluierung von Förderungsmaßnahmen der Landwirtschaft im Bundesland Salzburg | 2004 | Sbg. LReg – Abt. 4 – Land- und Forstwirtschaft | Bundesanstalt für Agrarwirtschaft, Bundesanstalt für Bergbauernfragen |
| Impulsprogramm Nachhaltig Wirtschaften. Zwischenevaluierung und Inhaltlich-strategische Stellungnahmen | 2004 | BMVIT | Mandl, Lüthi & Partner |
| Umweltrelevante betriebliche Förderungen in Wien | 2004 | Stadt Wien – MA 22 | Energieverwertungsagentur |

| Titel | Jahr | Auftraggeber | Auftragnehmer |
|---|-------------|--|---|
| Evaluation der luft-hygienischen Auswirkungen der Erweiterung des Nachtfahrverbotes auf der Unterinntal-Autobahn im Winterhalbjahr 2004/05 | 2005 | Tiroler LReg | Oekoscience AG |
| Evaluation der Österreichischen Strategie zur Nachhaltigen Entwicklung | 2005 | BMLFUW | Adelphi Consult GmbH, Hackl & Kienel-Mayer OEG, Forschungsstelle für Umweltpolitik [FU Berlin] |
| Evaluierung der Umweltförderung des Bundes für den Zeitraum 01.01.2002 bis 31.12.2004 | 2005 | BMLFUW | KWI Consultants&Engineers, WIFO, Hydro Ingenieure Umwelttechnik GmbH, Ingeniurgemeinschaft innovative Umwelttechnik |
| ÖkoBusinessPlan Wien – Kurzfassung des siebenten Evaluationsberichts April 2005 | 2005 | Stadt Wien – MA 22 | Forschungsschwerpunkt Nachhaltigkeit und Umweltmanagement [WU Wien] |
| Update-Evaluierung des Österreichischen Programms für die Entwicklung des ländlichen Raums | 2005 | BMLFUW | AGES, AWI, BABF, BAW, BFW, HBLFA Raumberg-Gumpenstein, UBA, WIFO |
| Vertiefende Evaluierung der Wirkung geförderter Maßnahmen zur Verbesserung der Schutzwirkungen des Waldes | 2005 | BMLFUW | BFW, Landesforstdirektion Tirol |
| Evaluierung der emissionsseitigen Auswirkungen der Einführung einer durchgehenden 100km/h Beschränkung auf der A 12 bei Imst anhand gemessener Geschwindigkeitsverteilungen | 2006 | Tiroler LReg | Institut für Verbrennungskraftmaschinen und Thermodynamik [TU Graz] |
| Evaluierungsbericht zur Klimastrategie 2002 | 2006 | BMLFUW | Umweltbundesamt, Österreichische Energieagentur |
| Kyoto-Fortschrittsbericht Österreich 1990-2004 | 2006 | UBA | Umweltbundesamt |
| ÖkoBusinessPlan Wien – Kurzfassung des achten Evaluationsberichts Mai 2006 | 2006 | Stadt Wien – MA 22 | Forschungsschwerpunkt Nachhaltigkeit und Umweltmanagement [WU Wien] |
| UVP-Evaluation | 2006 | BMLFUW | Umweltbundesamt, Haslinger, Nagele&Partner |
| Volkswirtschaftliche Kosten-Nutzen-Analyse der neuen EU-Chemiepolitik (REACH) für Österreich | 2006 | BMLFUW, BMWA, BAK, WKÖ, FB chem. Industrie, Industriellenvereinigung | Alpen-Adria Universität Klagenfurt |
| Wirtschaftsinitiative Nachhaltigkeit – Evaluationsbericht zur Programmperiode 2003 bis 2005 | 2006 | Stmk. LReg – Fachabt. 19D – Abfall- und Stoffflusswirtschaft | Research Institute for Managing Sustainability [WU Wien], Institut für Industrielle Ökologie |
| Zur Effizienz von Schieneninfrastrukturbauvorhaben am Beispiel des Brenner-Basistunnels | 2006 | o. A. | Institut für Transportwirtschaft und Logistik [WU Wien] |

Energie und Umwelt: Stand der Evaluationstätigkeiten in der Schweiz

1 Kontext der Evaluationstätigkeiten

Die Evaluationstätigkeiten in der Schweiz im Bereich Umwelt und Energie sind vor Entwicklungen in den beiden Politikbereichen sowie den Evaluationstätigkeiten in der Schweiz insgesamt zu verstehen.

1.1 Umweltpolitik in der Schweiz

Die Umweltpolitik der Schweiz wurde massgeblich in den 1980er und 1990er Jahren ausgebaut (Knoepfel/Nahrath 2006). In einer Reihe von Gesetzen und Verordnungen erliessen Bund, Kantone und Gemeinden regulative Instrumente für die Einhaltung von Grenzwerten für Immissionen und Emissionen. Das Umweltgesetz von 1983 führte zusätzliche Instrumente wie die Umweltverträglichkeitsprüfung und forderte umfassendere Ansätze wie die Anwendung des Verursacherprinzips ein. Der Umweltschutz konnte vermehrt in Politikbereichen eingeführt werden, in welchen die öffentliche Hand bereits früher aktiv war. Vermutlich international am bekanntesten ist der beachtliche Ausbau der Infrastrukturen für den öffentlichen Verkehr (Alpentransversalen, S-Bahn in Zürich) bei gleichzeitiger Mehrbelastung des grenzüberschreitenden Gütertransports auf der Strasse. Generell verstärkte die öffentliche Hand den Einbezug ökologischer Kriterien in den Bereichen, in welchen sie direkt Geldmittel investierte, namentlich in der Landwirtschaftspolitik mit dem Ersatz der produktionsgebundenen Abgeltungen durch flächenbezogene Direktzahlungen mit ökologischen Auflagen oder der Finanzierung von ökologischen Kompensationsmassnahmen für Infrastrukturen der öffentlichen Hand (z. B. im Nationalstrassenbau). Auf lokaler Ebene werden umweltpolitische Akzente zum Beispiel in der Abfallbewirtschaftung, dem Baurecht oder der Raum- und Verkehrsplanung gesetzt. In vielen Bereichen setzte sich die Einsicht durch, dass sich eine intakte Umwelt und wirtschaftlicher Wohlstand bedingen und dass sich mit Umweltschutz auch Gewinne erwirtschaften lassen, zum Beispiel wenn strenge Umweltauflagen zur Entwicklung neuer international vermarktbarer Technologien führen. Nicht zuletzt dies führte dazu, dass sich ein „Ökobusiness“ als rentabler Wirtschaftszweig etablieren konnte. Wie in anderen Ländern bleiben aber auch in der Schweiz Zielkonflikte zwischen verschiedenen öffentlichen Interessen bestehen. Gegenwärtig sieht sich die Umweltpolitik verstärkt Forderungen nach einer Vereinfachung der Verfahren und einer Senkung der öffentlichen Ausgaben konfrontiert. Die Umweltpolitik wird in den Bereichen

mit bereits institutionalisierten und ausgebauten Regulierungen nun zum Teil wieder zurückgebaut oder konsolidiert.

Das Konzept der nachhaltigen Entwicklung ist gegenüber dem Umweltschutz ambivalent. Es fordert zunächst, dass Umweltanliegen gleichwertig zur wirtschaftlichen Entwicklung berücksichtigt werden. Das unterstützt eine zeitgemässe Umweltpolitik, die zunächst Win-Win-Situationen ausschöpft und die volkswirtschaftlichen Effekte bei der Einführung weiterer Schutzauflagen berücksichtigt. Gleichzeitig setzt dies Grenzen gegenüber einem Umweltschutz, der Ökosysteme oder Arten aufgrund ihrer eigenen Wertigkeit schützen will, auch im Widerspruch zu den Bedürfnissen der Menschen. Nachhaltige Entwicklung erscheint in strategischen Dokumenten auf Bundes-, Kantons- und Gemeindeebene oft sehr prominent, ein eigentliches Politikfeld mit verbindlichen regulativen oder finanziellen Instrumenten gibt es hingegen nicht. Am deutlichsten zutage tritt nachhaltige Entwicklung in Planungsinstrumenten wie den Lokalen Agenda21-Prozessen. Passend dazu ist sie auf Ebene des Bundes im Bundesamt für Raumentwicklung angesiedelt, welches zur nachhaltigen Entwicklung Strategien und Grundlagen erarbeitet sowie Koordinations- und Informationsaufgaben wahrnimmt.

1.2 Energiepolitik

Während die Umweltpolitik als transversale Politik übergreifend auf die unterschiedlichsten Politikbereiche einwirkt, stellt die Energiepolitik eine klassische sektorale Politik dar. Im Zentrum der Aufgaben standen über lange Zeit Fragen der technischen Sicherheit, zum Beispiel bei Stauanlagen, Gasleitungen oder elektrischen Netzen. Die Schweiz importiert zwar Brenn- und Treibstoffe fast vollständig, produziert aber bedeutende Mengen an Elektrizität aus Wasser- und Kernkraftwerken. Produktionsanlagen (inklusive der Kernkraftwerke) sind zu einem grossen Teil im Besitz von Gemeinden und Kantonen, entweder als Aktiengesellschaften (z. B. NOK/Axpo-Gruppe), als öffentlich-rechtliche Anstalten oder direkt als Teile der Verwaltung (z.B. Elektrizitätswerk der Stadt Zürich). Der Bund verfügt im Energiebereich über keine eigenen Infrastrukturen (im Unterschied zum Verkehr), weshalb sein Einfluss im Vergleich zu den Kantonen und Gemeinden weniger direkt ist. Die Aufgaben der verschiedenen Ebenen werden sich in den nächsten Jahren durch die beschlossene schrittweise Liberalisierung des Strommarkts verändern.

Der Bund erhielt 1990 die Kompetenzen zur Förderung der rationellen Energienutzung (d.h. des Energiesparens) und der Produktion von erneuerbaren Energien (BV Art. 24octies). Zu diesem Zeitpunkt zeigte sich, dass ein weiteres geplantes Kernkraftwerk wegen des grossen Widerstands aus der Bevölkerung bis auf weiteres nicht gebaut werden konnte. Die Stimmberechtigten unterstützten zudem einen auf zehn Jahre befristeten Stopp für Planung und Bau von neuen Kernkraftwerken – den vollständigen Ausstieg aus der Kernenergie lehnten die Stimmberechtigten ab. Der Bund nutzte die neuen Kompetenzen und startete umgehend das Aktionsprogramm Energie2000, das später in das Programm EnergieSchweiz überging. Letzteres knüpft stark an die aus dem Kyoto-Prozess formulierte inländische Klimapolitik an. Die Aktionsprogramme beruhen auf freiwilligen Massnahmen mit Instrumenten wie Information und Beratung, Aus- und Weiterbildung, Branchenvereinbarungen oder

dem Aufbau von Marktagenturen. Energie2000 (in beschränktem Masse EnergieSchweiz) verfügte zudem über Mittel zur Unterstützung von Pilot- und Demonstrationsanlagen und Förderprogrammen. Gemeinden und Kantone – zum Teil Elektrizitätswerke – verwenden ähnliche Massnahmen zur Förderung von erneuerbaren Energien und der rationellen Energienutzung. Im Gebäudebereich und der Energieberatung liegt die Federführung bei den Kantonen. Zum Teil werden die Instrumente noch ergänzt, beispielsweise im Kanton Basel-Stadt mit einer Lenkungsabgabe auf Strom (dem „Stromsparfonds“). Die freiwilligen Massnahmen blieben politisch umstritten.

Die Energiepolitik des Bundes ist derzeit im Umbruch begriffen. Bei den erneuerbaren Energien und Energieeffizienz kommen starke Impulse aus der internationalen Klimapolitik, etwa was die Formulierung von Emissions- und Effizienzzielen betrifft. Eine Reihe von finanziellen und regulativen Instrumenten steht erneut auf der politischen Agenda und mehrere Instrumente haben eine intakte Chance, realisiert zu werden. Beispiele für die im politischen Prozess unterschiedlich fortgeschrittenen Instrumente sind die CO₂-Abgabe auf Brennstoffe, die Einspeisevergütung für erneuerbare Energien, die Treibstoffzollbefreiung für biogene Treibstoffe, die Ergänzung der Motorfahrzeugsteuer um ein verbrauchsabhängiges Bonus-Malus-System sowie ergänzende gesetzliche Mindestanforderungen an Geräte und weitere Lenkungsabgaben.

2 Evaluationsaktivitäten

Es gibt in der Umwelt- und Energiepolitik eine ganze Reihe etablierter Verfahren, die auf eine Bewertung von Projekten und politischen Massnahmen abzielen, nicht aber als Ex-Post-Evaluationen im engeren Sinne zu bezeichnen sind (Huber/Müller 2004). Diese Verfahren und Instrumente sollen hier kurz auch wegen der oft geforderten Nutzung von Synergien verschiedener Instrumente mit Ex-Post-Evaluationen diskutiert werden (BJ 2006). Dass zwischen den Instrumenten Ähnlichkeiten bestehen, zeigt sich nur schon darin, dass viele dieser Instrumente von denselben externen Auftragnehmern gleichzeitig abgedeckt werden. Anschliessend werden die Aktivitäten der Ex-Post-Evaluationen im engeren Sinne skizziert.

2.1 Ex-Post-Evaluationen im Kontext ähnlicher Instrumente der Umwelt- und Energiepolitik

Bewertungsfragen sind Gegenstand von politischen Instrumenten, die höchstens am Rande Berührungspunkte zu Ex-Post-Evaluationen haben. Das ist etwa der Fall bei Plangenehmigungsverfahren, technischen Typenprüfungen oder der Umweltverträglichkeitsprüfung. Ebenfalls formalisierte Prüfverfahren verwendet werden bei der Vergabe von Label, die zum Teil auch als Vorstufe zu Normen eingesetzt werden. Letzteres kann der Fall sein für Label im Bereich der biologischen Landwirtschaft, in der Wald- und Holzzertifizierung, den Energieetiketten auf Geräten und Autos oder mit dem Minergie-Label im Gebäude-Neubau. Die öffentliche Hand nutzt ebenfalls Labels, z.B. das Energiestadt-Label. Ein ähnlicher Ansatz wird mit freiwilligen Qualitätssicherungssystemen verfolgt, etwa der ISO 14001-Norm.

Mit der Einführung einer CO₂-Abgabe auf Brennstoffe sowie dem durch eine Abgabe auf Treibstoffe finanzierten Förderprogramm der Stiftung Klimarappen (eine freiwillige Massnahme der Wirtschaft) erhöht sich zudem der Bedarf nach fundierten Beurteilungen der CO₂-Einsparungen von Projekten. Insgesamt teilen diese Prüf- und Zertifizierungssysteme viele Bewertungskriterien und -indikatoren, die auch in Evaluationen verwendet werden.

Amtliche Statistiken und Monitoring-Systeme liefern wichtige Grundlagen für eine wirkungsorientierte Umwelt- und Energiepolitik. Die Gesamtenergiestatistik liefert Angaben zu den fossilen Energieträgern, welche direkt in das Treibhausgasinventar einfließen und den Grad der Zielerreichung der Verpflichtung der Schweiz im Kyoto-Protokoll dokumentieren. Unterschiedlich eng an die politische Steuerung geknüpft sind Monitoring- und Beobachtungssysteme wie beispielsweise das nationale Beobachtungsnetz für Luftfremdstoffe (NABEL), der Kataster der belasteten Standorte (Altlasten), das „Monitoring ländlicher Raum“ oder das Biodiversitäts-Monitoring. Verbreitet sind Indikatorensysteme im Bereich der nachhaltigen Entwicklung. Mit der Liberalisierung des Elektrizitätsmarkts könnte der Bedarf nach einem Monitoring der Elektrizitätsproduktion und -verteilung steigen. Im Unterschied zu den Ex-Post-Evaluationen konzentrieren sich Monitoring-Systeme und deskriptive Statistiken auf Indikatoren und nicht auf die Zusammenhänge zwischen Indikatoren in der Wirkungskette.

Methodisch die grössten Ähnlichkeiten mit den Ex-Post-Evaluationen haben Ex-Ante-Evaluationen der Machbarkeit, Wirkungen oder des Kosten-Nutzenverhältnisses von politischen Massnahmen. Auf Bundesebene werden diese Art der Ex-Ante-Studien in der Regel im Rahmen der Ressortforschung und auf Mandatsbasis betrieben (vor allem koordiniert durch die Sektion Umweltforschung des BAFU bzw. das Forschungsprogramm Energiewirtschaftliche Grundlagen des BFE). Im Vergleich zu den Ex-Post-Evaluationen basiert die Ressortforschung stärker auf ökonomischen Ansätzen und Modellrechnungen, etwa auf ausgedehnten Forschungsstudien zu den externen Kosten und Nutzen des Verkehrs, des volkswirtschaftlichen Nutzens des Umweltschutzes oder der langfristigen Szenarien der Energieversorgung in der Schweiz (Energieperspektiven). Unter dem Begriff der Evaluation werden in der Praxis meistens Ex-Ante-Vorprüfungen der Nutzen und Kosten von einzelnen Projekten verstanden (etwa Infrastrukturanlagen wie die Varianten der neuen Eisenbahn-Alpentransversalen NEAT). Das Bundesamt für Raumentwicklung (ARE) führt mit der Nachhaltigkeitsbeurteilung (NHB) eine Regulierungsfolgenabschätzung von ausgewählten Vorhaben des Bundes durch (ARE 2004). Die Ex-Ante-Evaluationen können sehr unterschiedliche Verwendungszwecke haben, die von der Organisation eines Gesprächsforums über die Berechnungsgrundlagen für Abgaben bis zum konkreten Entscheid über verschiedene Offerten reichen. Ex-Ante-Evaluationen unterscheiden sich in der Praxis von den Ex-Post-Evaluationen vor allem darin, dass Fragen in den Bereichen Organisationsentwicklung, Vollzugsstrukturen und -kulturen sowie der nachträglichen Kontrolle ausgeklammert bleiben.

Schliesslich sind betriebswirtschaftliche Instrumente aufzuführen, welche verwaltungsintern zur Geschäftsführung und Abwicklung von Aufträgen genutzt werden. Operative Controlling-Instrumente sind vorhanden (z. B. Vertrags- und Finanzcontrolling), auf allen politischen Ebenen haben Ämter Ansätze aus der wirkungsorientierten Verwaltungsführung übernommen. Beispielsweise benutzt das Bundesamt für Energie Leistungsvereinba-

rungen mit privat-rechtlichen Agenturen im Rahmen von EnergieSchweiz und hat für diese (sowie für die Energieforschung) Controlling-Handbücher und ein Management-Informationssystem geschaffen. Insgesamt liefern das Controlling und die internen Geschäftsprüfungsverfahren in erster Linie eine Datengrundlage für Evaluationen.

2.2 Ex-Post-Evaluationen in der Umweltpolitik

Ex-Post-Evaluationen verwenden sozialwissenschaftliche Methoden und analysieren und bewerten damit die Konzepte, den Vollzug, die Wirkungen oder die Wirtschaftlichkeit von politischen Massnahmen. Die Umweltpolitik in der Schweiz kann auf eine Fülle an Ex-Post-Evaluationen zurückblicken. Dies ist einer der Befunde aus einer aktuellen Zusammenstellung und Machbarkeitsstudie von Evaluationen in der Umweltpolitik (Mauch/Balthasar 2005). Diese listet 44 Studien auf, die zwischen 1980 und 2005 publiziert wurden. Vor allem für die Sachbereiche „Wald“, „Abfall“, „Gewässerschutz“ sowie „Luftreinhaltung“ lassen sich mehrere Evaluationen finden. Es lässt sich ebenfalls zeigen, dass ein Grossteil auf die Evaluation der Konzeption und des Vollzugs fokussiert. Erst in zweiter Linie untersuchen sie die Wirksamkeit der Massnahmen. Sehr selten sind zudem Studien zur Wirtschaftlichkeit, bzw. der Effizienz und dem Kosten/Nutzenverhältnis der umweltpolitischen Massnahmen. Der Bericht weist auf die methodischen Probleme einer Erfassung des Nutzens von umweltpolitischen Massnahmen hin und empfiehlt die Durchführung von effizienzbezogenen Analysen in erster Linie für Ex-Ante-Evaluationen.

Der grosse Teil der vorliegenden Evaluationen wurde von privaten Forschungs- und Beratungsbüros durchgeführt, die über die für Wirkungs- und Wirtschaftlichkeitsbetrachtungen notwendige wirtschaftliche, technische oder ökologische Expertise verfügen. Diese Büros sind in der Regel im Ökobusiness etabliert, sind aber nicht monothematisch, das heisst erstellen ebenfalls Studien in anderen Themenbereichen. Angesichts des kleinen Evaluationsmarkts in der Schweiz ist es nicht erstaunlich, dass diese Auftragnehmer neben Evaluationen meistens auch in denselben Themenbereichen eine Reihe weiterer Dienstleistungen (wie Beratung) anbieten. Die Mitarbeitenden in diesen Büros haben oft eine Ausbildung in Ökonomie oder im Umweltbereich (z. B. Umweltwissenschaften, Agrarökonomie oder Nachdiplomstudium Raumplanung an der ETH Zürich, Geographie, etc.). Einige bekannte Auftragnehmer sind (in alphabetischer Reihenfolge) BHP – Hanser und Partner AG, Econcept, Ecoplan und Infras. Wichtige Expertisen für Evaluationen kommen aber auch von eher technischen Ingenieurbüros sowie von Selbständigen oder kleineren Büros (vgl. Übersicht im Evaluatoren-Verzeichnis auf der Homepage der SEVAL).

Bei den ältesten im Bericht Mauch/Balthasar (2005) zitierten Studien handelt es sich um Vollzugsevaluationen im Rahmen eines europäischen Forschungsprojekts (Peters/Knoepfel 1980) beziehungsweise in der Form einer ökonomischen Dissertation (Bussmann 1981). Diese ersten Vollzugsstudien zur Umweltpolitik haben – vermutlich zusammen mit Studien zur Raumordnungspolitik an der ETH Zürich (Institut ORL) und den Verwaltungswissenschaften an der Universität Genf – die schweizerische Policy-Analyse mitbegründet. Dass das 1987 lancierte Nationale Forschungsprogramm „Wirksamkeit staatlicher Massnahmen“ (NFP 27) just in einer „umweltpolitischen Hochkonjunktur“ gestartet wurde, dürfte die

Bedeutung dieser akademisch-politikwissenschaftlichen Ausrichtung unterstützt haben. Dabei sind nach den qualitativen Vollzugevaluationen mehr und mehr auch quantitative Wirkungsstudien erstellt worden (Widmer 1991). Es haben sich daraus private Evaluationsbüros gebildet, die auch im Umweltbereich tätig sind, zum Beispiel (als zufällige Auswahl) das Büro Vatter, erasm, evaluanda oder Interface. Im Vergleich zu den genannten Büros und Hochschulen stellen Büros aus der klassischen Unternehmensberatung oder Revisionsstellen als Auftragnehmer von Ex-Post-Evaluationen eine Ausnahme dar.

Der grösste Auftraggeber für Ex-Post-Evaluationen der Umweltpolitik dürfte das Bundesamt für Umwelt sein. Auf Ebene der Kantone und Gemeinden ergehen ebenfalls Aufträge, zum Beispiel an die Evaluation von Massnahmen in der Abfallbewirtschaftung oder Verkehrsplanung. Auf Ebene Bund liessen die Parlamentarische Verwaltungskontrolle und die Eidgenössische Finanzkontrolle einzelne Aspekte der Umweltpolitik evaluieren – zum Beispiel das eher schwache Instrument der Bundesinventare der geschützten Landschaften (Parlamentsdienste 2003) oder forstliche Subventionen (Eidgenössische Finanzkontrolle 2003). Die Eidgenössische Finanzkontrolle arbeitet an Evaluationen im Umweltbereich, die mehrere Ämter betreffen (Umweltschutz im Nationalstrassenbau, Organisation und Wirksamkeit der Lenkungsabgabe auf flüchtigen organischen Verbindungen). Grosse Beachtung fand eine Evaluation durch den Think Tank „AvenirSuisse“, der die immensen volkswirtschaftlichen Kosten des Verbandsbeschwerderechts beklagte (Rentsch 2003). Bereits früher wies eine Studie im Auftrag des Bundesamts für Umwelt auf die hohen Erfolgsquoten der Verbandsbeschwerden hin (Flückiger et al. 2000). Einige der wichtigsten aktuellen Evaluationen des Bundesamts für Umwelt lassen sich vor dem Hintergrund der Forderungen nach Effizienzsteigerungen und Vereinfachung von Verwaltungsverfahren verstehen. Diesbezügliche Verfahrensvereinfachungen werden in einer Evaluation der Umweltverträglichkeitsprüfung vorgeschlagen (Sager/Schenkel 2004) sowie in einer Analyse der Konzeption und der Verständlichkeit der Vollzugshilfen des Amts (Mauch et al. 2006). Zudem evaluierte das Bundesamt ganze Teilbereiche der Umweltpolitik, zum Beispiel die Abfallpolitik des Bundes (Hanser et al. 2005). Weitere aktuelle Evaluationen zur Umweltbildung (Widmer et al. 2007b), zur Umweltaussenpolitik (Mauch et al. 2007) und zur Umweltforschung stehen in Zusammenhang mit möglichen organisatorischen Veränderungen.

2.3 Ex-Post-Evaluationen im Energiebereich

Die Massnahmen zur Förderung erneuerbarer Energien und der rationellen Energienutzung gehören in der Schweiz zu den am besten evaluierten Themen überhaupt. Das Bundesamt für Energie hat dazu seit 1990 rund neunzig Evaluationen durchführen lassen. Bereits das Aktionsprogramm Energie2000 umfasste über sechzig Evaluationen (Balthasar 2000). Im Vergleich zur Situation in der Umweltpolitik ist erwähnenswert, dass diese Evaluationen fast ausschliesslich zentral vom Bundesamt für Energie in Auftrag gegeben wurden. Vereinzelt gibt es solche auch von kantonalen Energiestellen (Hammer et al. 2000) oder von Aufsichtsbehörden des Bundes (Mauch et al. 2005). Die rechtliche Grundlage für die Evaluationen bildet seit 1998 das Energiegesetz (EnG Art. 20), das den Bundesrat dazu verpflichtet, Massnahmen regelmässig auf ihre Wirkungen zu prüfen und dem Parlament zu berichten.

Die Evaluationstätigkeiten konzentrieren sich deshalb auf den Teil der Energiepolitik, der im Energiegesetz beschrieben ist und in welchem der Bund im Vergleich zu anderen Gebieten über einen gewissen Handlungsspielraum und finanzielle Ressourcen verfügt: Subventions- bzw. Förderprogramme für die Forschung, den Technologietransfer und vor allem die Markteinführung von erneuerbaren Energien und Techniken zur rationellen Energienutzung.

Die Ex-Post-Evaluationen sollen die Wirkungen von energiepolitischen Massnahmen aufzeigen und gegenüber der Bevölkerung, dem Parlament und dem Bundesrat Transparenz herstellen. Ein Beispiel dafür sind direkte Förderungsprogramme für Anlagen zur Produktion von erneuerbaren Energien oder Effizienzmassnahmen an Gebäuden, für die Evaluationen neben den Erfolgen auch die zum Teil beträchtlichen Mitnahmeeffekte ausweisen, die je nach Förderprogramm zwischen dreissig und 85 Prozent betragen (Balthasar 2000: 87). Die Ex-Post-Evaluationen sind aber nicht nur ein Instrument zur Rechenschaftslegung und zur Kontrolle von einzelnen Massnahmen, sondern vielmehr auch Lerninstrumente. Sie sollen auf die Probleme und Schwächen von energiepolitischen Massnahmen hinweisen und konstruktive Verbesserungsmöglichkeiten aufzeigen. Dazu untersuchen sie die verschiedensten Etappen in der Wirkungskette, von der Konzeption der Massnahmen, über den Vollzug bis zu den Wirkungen bei den Zielgruppen und dem Verbrauch bzw. der Produktion von Energie. Sämtliche Evaluationsberichte werden publiziert. Seit 2002 sind die Berichte auf der Homepage des Amts zugänglich. Zudem wird jährlich ein kurzer Bericht über die Evaluationen als Anhang zum Jahresbericht von EnergieSchweiz publiziert.

Der im schweizerischen Vergleich hohe Stellenwert der Ex-Post-Evaluationen in diesem Aktionsprogramm ergibt sich in erster Linie aus dem Nutzen für die Steuerung. Das Bundesamt für Energie unterscheidet zwischen strategischen und punktuellen Evaluationen. Strategische Evaluationen werden zu grundsätzlichen energiepolitischen Fragestellungen erstellt, deren Empfehlungen sich in erster Linie an die Leitung des Programms EnergieSchweiz richtet. Beispiele sind die Funktionsweisen und Wirkungen der vom Programm EnergieSchweiz unterstützten „Netzwerkagenturen“ (Rieder et al. 2003) oder die Einbettung von Energiefragen in die Verkehrspolitik des Bundes (Angele et al. 2003). Punktueller Evaluationen werden von der Programmleitung oder von einzelnen Sektionen angeregt und konzentrieren sich auf thematisch engere Gegenstände, wie zum Beispiel eine Informationskampagne für Solarwärme (Gerheuser 2006), Pilotprojekte für Mobilitätsberatung (De Tommasi/Caduff 2006) oder ein Ausbildungs- und Beratungsangebot für Energiesparen in öffentlichen Gebäuden (Dettli et al. 2004). Die Berichte richten sich fallweise entweder stärker an die Leitung des Programms oder an einzelne Fachspezialisten sowie an die Akteure im Feld. Insgesamt lag der Aufwand für die Ex-Post-Evaluationen stets deutlich unter zwei, zumeist unter einem Prozent der Ausgaben des Aktionsprogramms.

Für die Durchführung der Evaluationen im Energiebereich mandatiert das Bundesamt für Energie externe Auftragnehmer. Dabei handelt es sich grösstenteils um dieselben, die Expertise in der Umweltpolitik haben (vgl. oben, Kapitel 2.2). Die Ergebnisse der Ex-Post-Evaluationen des Aktionsprogramms fliessen in das Controlling und die wirkungsorientierte Steuerung des Programms ein. Auf der Seite der Leistungen lassen sich Agenturen an den in Controllinginstrumenten beschriebenen und zum Teil vertraglich vereinbarten Zielindikatoren messen. Quantitative Resultate der Evaluationen zu den Effekten fliessen in die

„Wirkungsanalyse EnergieSchweiz“ ein, die im Sinne eines Benchmarks die Marktentwicklungen einzelner Produkte und ungefähren Wirkungen der Massnahmen illustriert (Iten et al. 2006). Sie werden zudem eingesetzt, um Globalbeiträge des Bundes an Fördermassnahmen der Kantone entlang eines wirkungsorientierten Schlüssels zu verteilen (Schneider et al. 2006).

3 Verbindungen mit anderen Evaluationsfeldern und internationale Entwicklungen

Die Evaluationen in der Umwelt- und Energiepolitik in der Schweiz werden grösstenteils von Beratungs- und Forschungsbüros wahrgenommen, die neben der technischen oder naturwissenschaftlichen auch über sozialwissenschaftliche Expertise verfügen (z. B. zu Marketing, Verwaltung etc.). Diese Unternehmen sind in der Regel sehr gut europäisch vernetzt und führen Aufträge zum Teil mit anderen Unternehmen aus dem deutschsprachigen Raum aus. Die Professionalität und die verwendeten Standards sind im Schnitt hoch und die Unternehmen sind auch in der Schweizerischen Evaluationsgesellschaft (SEVAL) präsent. Auf Seite der Bundesverwaltung sind die zuständigen Ämter mit ihren Fachspezialisten ebenfalls in der SEVAL sowie im Netzwerk Evaluation der Bundesverwaltung vertreten. Schliesslich gibt es im Rahmen der Zusammenarbeit mit internationalen Organisationen Anlässe für einen Wissenstransfer, wobei die Bundesämter von einer systematischen Nutzung weit entfernt sind.

In beiden Politikbereichen ergeben sich Berührungspunkte zu anderen Evaluationsfeldern. So liessen beide Bundesämter Evaluationen im Bereich der Forschung und des Technologietransfers durchführen. Eine Besonderheit ergibt sich für die Evaluationen in der Umwelt- und Energiepolitik dort, wo Zielindikatoren im Rahmen von internationalen Politiken eine Steuerungsfunktion erhalten. Aktuell ist an den Handel mit CO₂-Zertifikaten zu denken, für welche die United Nations Framework Convention on Climate Change (UNFCCC) ein methodisches Regelwerk entwickelte und ständig weiterentwickelt. Diese internationale Entwicklung schlägt sich in Ansätzen der Berechnungsgrundlagen von Brenn- und Treibstoffeinsparungen nieder, besonders auch für Kompensationsmassnahmen (in der Schweiz für die privatrechtliche Stiftung Klimarappen und CO₂-Abgabe auf Brennstoffen). Diese fortlaufende methodische Diskussion fliesst in die Wirkungsberechnungen in den Evaluationen des Bundesamts für Energie ein.

4 Stand der Evaluation

Die Evaluationstätigkeiten im Umweltbereich konnten 2007 im Vergleich zu 2000 von einer Ausweitung und Professionalisierung profitieren, die sich beispielsweise in der Anzahl veröffentlichter Studien oder den konzeptionellen Überlegungen widerspiegeln. Gegenwärtig sind vor allem zwei Herausforderungen hervorzuheben. Zunächst ist an die mögliche Institutionalisierung der Evaluationstätigkeiten zu denken. Diskutiert wird im Bundesamt

für Umwelt (BAFU), ob für die Auswahl (zumindest teilweise) sowie die Abwicklung von Ex-Post-Evaluationen eine Stabsstelle eingerichtet wird. Zudem entwickelt es eine Evaluationsstrategie für das Amt. Diese Institutionalisierung der Evaluationstätigkeiten im BAFU könnte die weitere Professionalisierung in diesem Gebiet begünstigen. Eine zweite Herausforderung besteht darin, dem Thema der nachhaltigen Entwicklung in Evaluationen einen angemessenen Stellenwert einzuräumen. Fachämter sind in der Regel bereits mit der Überprüfung des Zielerreichungsgrads im Umweltbereich sowie den allfälligen Effekten auf BIP und Arbeitslosigkeit ausgelastet.

Im Vergleich zu 2000 verfügt der Bund 2007 über bedeutend weniger finanzielle Mittel für die Förderung von erneuerbaren Energien und die rationelle Energieverwendung. In der Anfangsphase wurde im Parlament sogar die Abschaffung des Aktionsprogramms EnergieSchweiz gefordert. Das jährliche Budget für Evaluationen von EnergieSchweiz wurde entsprechend linear gekürzt. Zusätzlich wurde das institutionelle Arrangement in dem Sinne vereinfacht, dass die externe Begleitgruppe Evaluation und das Mandat eines externen Evaluationsberaters aufgehoben wurden. Zwei wichtige Herausforderungen an die Evaluationstätigkeiten in der Energiepolitik stehen an. Erstens sollte der Anteil an strategischen Evaluationen – gerade auch in anderen Bereichen der Energiepolitik – ausgebaut werden. Zweitens sollte versucht werden, die Ex-Post-Evaluationen noch weiter zu professionalisieren, etwa durch den stärkeren Einbezug von internationalen Standards. Für letzteres liegt ein grosses Potential vor, von internationalen Ergebnissen und Methodiken zu profitieren, etwa in Anlehnung an die Methoden und Ergebnisse aus einem Programm zur Unterstützung der Umsetzung der EG-Richtlinie 2006/32/EG über Endenergieeffizienz und Energiedienstleistungen (AID-EE 2007; Khan et al. 2006).

Ausser den Fachämtern kann keine Institution gefunden werden, die in ähnlichem Masse kontinuierlich Wirkungen von Umwelt- und Energiepolitik evaluiert. Kontrollinstanzen des Bundes führen Evaluationen eher reaktiv oder punktuell durch, wobei die Eidgenössische Finanzkontrolle zunehmend systematisch ämterübergreifende Fragestellungen angeht. Sehr punktuell sind zudem die Evaluationen der wenigen privaten Think Tanks.

5 Bilanzierung

Die erzielten Erfolge der Evaluationen im Umwelt- und Energiebereich in der Schweiz sind als positiv zu würdigen. In beiden Bereichen konnte als Ergänzung zu anderen Instrumenten eine ansehnliche Evaluationstätigkeit aufgebaut werden. Ex-Post-Evaluationen haben als Kontroll- und Lerninstrumente zur effektiven und effizienten Steuerung beigetragen. Nicht verschwiegen werden soll, dass dafür interne Ressourcen sowie Mittel für externe Mandate eingesetzt werden müssen. Die Durchführung interner Evaluationen (zum Beispiel der verschiedenen Geschäftsbereiche eines Amtes) und die Abwicklung eines grösseren Teils der Evaluationstätigkeiten durch internes Personal könnte erwogen werden. Vorteile wären die im Vergleich zu externen Mandaten tieferen Stundenansätze sowie das Schaffen von internen Kompetenzen. Dies ist allerdings gegenüber der geringeren Glaubwürdigkeit und Unabhängigkeit sowie der Machbarkeit abzuwägen.

Neben den beachtlichen Erfolgen sind auch die Grenzen der Evaluationstätigkeiten zu skizzieren. Evaluationen scheitern selten an Widerständen, stehen aber immer in Konkurrenz zu anderen Dossiers. Evaluationen aufgleisen, erstellen und nachbearbeiten braucht Zeit, gerade wenn sie nutzenstiftend in den Kontext der laufenden Verwaltungstätigkeiten eingebettet werden sollen. Wie in anderen Bereichen steigt auch im Umwelt- und Energiebereich der Druck, Staatstätigkeiten abzubauen und Kosten zu verringern. Die bisherigen Evaluationstätigkeiten haben dazu bereits beigetragen, indem die Konzepte und Vollzugsabläufe verbessert und eine Selektion zwischen unterschiedlich wirksamen Massnahmen getroffen werden konnte. Es bestehen aber noch bedeutende Lücken in der Bewertung der Wirtschaftlichkeit und Effizienz, die zum Teil bereits an einer sinnvollen Aufgliederung und Dokumentation der Kostenseite scheitert. Einerseits ist die diesbezügliche Datenlage zu verbessern, andererseits auch jeweils die Bereitschaft zu fordern, die Wirtschaftlichkeit (und nicht nur die positiven Effekte auf Umwelt und Volkswirtschaft) als wichtigen Aspekt in der Beurteilung mit zu berücksichtigen.

Schliesslich wird die Glaubwürdigkeit von Evaluationen auch in Zukunft ein Thema bleiben. Der wichtigste Massstab wird sein, dass mit den für Evaluationen investierten Mitteln auch ausreichende und konkrete Verbesserungen erzielt werden können. Die Voraussetzungen, dass dies den Evaluationen in der Umwelt- und Energiepolitik aufbauend auf der langjährigen Erfahrung mit Evaluationen auch in Zukunft gelingen wird, sind sehr gut.

Ländervergleich Energie und Umwelt

1 Einleitung

Sowohl der Energie- als auch der Umweltbereich haben in den vergangenen Jahren aufgrund anhaltender Umweltbelastungen, der Knappheit fossiler Energieressourcen sowie den befürchteten negativen Auswirkungen der globalen Klimaerwärmung an Bedeutung gewonnen. Diese Entwicklungen lösten in Deutschland, Österreich und der Schweiz einen neuen Energie- und Umweltdiskurs aus, welcher der Politikgestaltung in den beiden Bereichen zunehmend bedeutende Impulse gibt. Auch bei Evaluationen, die in den drei Ländern in diesen Bereichen durchgeführt werden, zeigen sich verschiedene parallel verlaufende Entwicklungen, aber auch einige Unterschiede. Nach einer kurzen Schilderung der rechtlich-institutionellen Kontextbedingungen für die Evaluation in diesen Bereichen in Abschnitt 2, stehen in Abschnitt 3 die gegenstandsbedingten und gegenstandsunabhängigen Gemeinsamkeiten und Unterschiede in den Evaluationsaktivitäten in den drei Ländern im Vordergrund. Abschnitt 4 vergleicht dann die Entwicklungen seit dem Jahr 2000 sowie den aktuellen Stand der Evaluation im Energie- und Umweltbereich. Zum Schluss (Abschnitt 5) ziehe ich eine Bilanz zu diesen Evaluationstätigkeiten und gehe auf einige gemeinsame Herausforderungen für die Energie- und Umweltevaluation in Deutschland, Österreich und der Schweiz ein.

2 Kontext der Evaluationstätigkeit

Energiepolitik – also jenes (vorwiegend) staatliche Handeln, das sich mit der Bereitstellung, Umwandlung, Verteilung und Verwendung von Energie befasst – gehört in Deutschland, Österreich und der Schweiz zu den klassischen Sektoralpolitiken mit verschiedenen Querverbindungen zur Wirtschafts-, Umwelt-, Forschungs- und Technologiepolitik. Zudem ist die Energiepolitik in allen drei Staaten intern aufgrund der verschiedenen Energieträger stark segregiert. Unterschiedlich ist hingegen die föderale Kompetenzregelung in den drei Staaten. Während in Deutschland und Österreich die Energiepolitik vorwiegend Bundesangelegenheit ist (Reiche 2005), liegt in der Schweiz die energiepolitische Kompetenz traditionellerweise bei den Kantonen (Klöti 1993). Auf Bundesebene wurde die Energiepolitik in der Schweiz erst im Jahr 1990 verfassungsrechtlich verankert.

Im Gegensatz dazu ist Umweltpolitik ein relativ junges Politikfeld und hat sich in Deutschland, Österreich und der Schweiz zu Beginn der 1970er Jahre herausgebildet, als die negativen Auswirkungen des Industriewachstums auf die natürliche Umwelt immer deutlicher wurden. In allen drei Staaten haben die Gliedstaaten in der Umweltpolitik relativ weit

reichende Kompetenzen. Allerdings trugen vor allem die umweltpolitischen Kompetenzerweiterungen auf Bundesebene in den 1980er Jahren massgeblich zur Institutionalisierung dieses relativ neuen Politikfeldes bei. Die Umweltpolitik erfuhr seitdem in allen drei Ländern eine stetige Erweiterung und wird heute zunehmend als Querschnittspolitik aufgefasst (Pesendorfer 2007; Jänicke 2006; Knoepfel/Nahrath 2006).

3 Evaluationstätigkeiten

Wie Kurt Bisang in seinem Beitrag aufzeigt, ist es für den Energie- und Umweltbereich sinnvoll, zwischen verschiedenen Bewertungsverfahren zu unterscheiden, die nicht alle als Evaluationen im engeren Sinn zu verstehen sind. Neben betriebswirtschaftlichen Instrumenten gibt es eine Reihe von Bewertungsverfahren vorwiegend naturwissenschaftlich-technischer Art (z. B. technische Prüfverfahren, Zertifizierungen, Umweltverträglichkeitsprüfungen), die zwar oft Gegenstand von politischen Instrumenten sind, aber nur am Rande Berührungspunkte zu Evaluationen aufweisen. Auch die Überprüfung der Einhaltung von Emissionsstandards ist solchen naturwissenschaftlich-technischen Bewertungsverfahren zuzuordnen, die zwar für die Evaluation von politischen Programmen, Projekten und Massnahmen wichtige Daten liefern, aber selbst keine eigentlichen Evaluationen darstellen (vgl. Newburn 2001; Huber/Müller 2004).

Evaluationen im sozialwissenschaftlichen Verständnis zielen üblicherweise darauf ab, die Wirksamkeit von staatlichem Handeln zu überprüfen und aufgrund der Ergebnisse Verbesserungsmöglichkeiten für politische Massnahmen abzuleiten (Weiss 1998a: 4). Beispiele für solche Evaluationen im engeren Sinn sind die Studien zur österreichischen Nachhaltigkeitsstrategie (Martinuzzi/Steuerer 2005), zum Förderprogramm ‚Klimaschutz-Plus‘ in Deutschland (Duscha et al. 2007) oder zu den verschiedenen Massnahmen im Rahmen der nationalen Energieprogramme in der Schweiz (Balthasar/Rieder 2000).

Im Unterschied zu betriebswissenschaftlichen und naturwissenschaftlich-technischen Bewertungsverfahren sind die Evaluationstätigkeiten im eben beschriebenen engeren Sinn in Deutschland, Österreich und der Schweiz zum Teil recht unterschiedlich ausgeprägt. Unterschiede lassen sich vor allem in der thematischen Ausrichtung, den involvierten Akteuren (Auftraggeber und Evaluationsanbieter) sowie bei der Durchführung der Evaluationen feststellen.

3.1 Thematische Ausrichtung

Wie im deutschen Beitrag von Markus Duscha, Herbert Klemisch und Wolfgang Meyer dargestellt, liegt ein bedeutender Schwerpunkt der Evaluation im Energiebereich heute bei der Energieeffizienz und den erneuerbaren Energien. Diese Themenfelder haben mit der Problematik der Klimaveränderung in den letzten Jahren stark an Bedeutung gewonnen und stehen in engem Bezug zu umweltpolitischen Fragestellungen. Aufgrund ihrer wirtschaftlichen Bedeutung und dem Einsatz beträchtlicher öffentlicher Mittel in diesem Bereich werden energiepolitische Programme zur Förderung erneuerbarer Energien und effizienter

Energienutzung heute sowohl in Deutschland als auch in Österreich und der Schweiz in regelmässigen Abständen evaluiert.

Umweltevaluationen sind in den drei Ländern jedoch unterschiedlich ausgeprägt. In Deutschland sind sie heute weitgehend Teil der Evaluation von Nachhaltigkeit, wo neben ökonomischen und sozialen Aspekten auch ökologische Fragen untersucht werden. Generell sind sozialwissenschaftlich orientierte Umweltevaluationen in Deutschland jedoch eher schlecht verankert; am besten konnten sie sich im Bereich der Umweltbildung und -beratung etablieren (z. B. Stockmann et al. 2001). Der Trend zu Nachhaltigkeitsevaluationen lässt sich auch in Österreich und der Schweiz beobachten. Wie die Beiträge zu Österreich und zur Schweiz zeigen, besteht in dieser Entwicklung die Gefahr, dass der Umweltbereich aufgrund von Zielkonflikten mit anderen gesellschaftlichen (vorab wirtschaftlichen) Interessen eine Schwächung erfährt.

3.2 Auftraggeber und Evaluationsanbieter

Dem Energie- und Umweltbereich in den drei Ländern ist gemein, dass sich die Auftraggeber für Evaluationen fast ausschliesslich auf die zuständigen Fachministerien und -ämter beschränken. Für den in allen drei Ländern stark sektoral organisierten Energiebereich überrascht diese Feststellung nicht. Im Umweltbereich wird mit dieser Konzentration der Auftraggeber auf wenige Ministerien und Ämter hingegen deutlich, dass die zunehmend als transversal aufgefasste Umweltpolitik noch zu keiner Ausweitung der Auftraggeber für Umweltevaluationen geführt hat. In Deutschland ist das Umweltministerium der zentrale Auftraggeber für Evaluationen im Umweltbereich auf Bundesebene, in der Schweiz das Bundesamt für Umwelt, in Österreich das Lebensministerium. Ähnlich präsentiert sich die Situation auf den unteren Staatsebenen, wo die jeweiligen Umweltämter die Hauptnachfrager nach Umweltevaluationen sind.

In Österreich scheint die Nachfrage nach Evaluationen hingegen weniger systematisch zu erfolgen als in den anderen beiden Ländern und stärker von der Initiative und Interessen einzelner Dienststellen abzuhängen. André Martinuzzi und Melanie Leschnik erklären diesen Missstand in ihrem Beitrag mit dem Fehlen von institutionellen Kapazitäten, die einen kontinuierlichen Wissensaufbau und Qualifikationserwerb im Evaluationsbereich ermöglichen würden. So verfügt keines der im Umwelt- und Energiebereich wichtigen Bundesministerien über eine Stabstelle für Evaluation. Das Schweizer Bundesamt für Energie (BFE) verfügt hingegen über eine solche Fachstelle für Evaluation. Seit Anfang der 1990er Jahre hat das BFE über neunzig Evaluationen durchführen lassen und ist heute wichtigster Auftraggeber für Evaluationen im Energiebereich auf Bundesebene. Allerdings sind die Bundeskompetenzen im Energiebereich in der Schweiz begrenzt und beschränken sich weitgehend auf die Förderung der Energieeffizienz und von erneuerbaren Energien.

Seitens der Evaluationsanbieter präsentiert sich in allen drei Ländern ein ziemlich heterogenes Bild. Gemeinsam ist allen drei Ländern, dass Evaluationsinstitute und -büros oft Evaluationen sowohl im Energie- als auch im Umweltbereich anbieten – ein klares Indiz für die Kleinheit des jeweiligen Marktes von Evaluationsnachfragern und -anbietern. In Deutschland lässt sich traditionellerweise eine stark technisch-naturwissenschaftliche Aus-

richtung der Evaluationsanbieter ausmachen. Dennoch gibt es einige namhafte sozialwissenschaftliche Forschungseinrichtungen im Umweltbereich. Nach Ansicht der Autoren des deutschen Beitrags verstehen sich diese Evaluationsanbieter jedoch oft nicht als „Umwelt-evaluatoren“ und sind auch selten in der Gesellschaft für Evaluation (DeGEval) vertreten.

Wie der Schweizer Beitrag zeigt, besteht diesbezüglich in der Schweiz eine grössere Vielfalt. Hier hat sich in den letzten zwei Jahrzehnten ein relativ breiter Kreis von Forschungs- und Beratungsbüros herausgebildet, die Evaluationen im Energie- und Umweltbereich anbieten und neben der technisch-naturwissenschaftlichen auch über sozialwissenschaftliche Expertise verfügen. Die meisten dieser Büros und Institute sind auch Mitglied in der Schweizerischen Evaluationsgesellschaft (SEVAL). Bisang schreibt dem Nationalen Forschungsprogramm 27 über die „Wirksamkeit staatlicher Massnahmen“, dessen Lancierung mit der umweltpolitischen Hochkonjunktur in den 1980er Jahren zusammenfiel, hierfür die Initialzündung zu. In Österreich scheint der „Evaluationsmarkt“ im Energie- und Umweltbereich am schwächsten entwickelt zu sein.

4 Stand der Evaluation

Trotz ihres im Vergleich zu naturwissenschaftlich-technisch ausgerichteten Bewertungsinstrumenten nach wie vor eher geringen Stellenwerts, hat die sozialwissenschaftlich orientierte Energie- und Umweltevaluation seit den 1990er Jahren in allen drei Ländern einen Aufschwung erfahren. Diese Entwicklung mag damit zusammenhängen, dass der Energie- und Umweltbereich heute weniger als noch in den 1980er Jahren auf ein rein technisches Problem reduziert wird und statt dessen politische, soziale und wirtschaftliche Faktoren grössere Beachtung finden. Zudem lässt sich in allen drei Ländern eine zunehmende Wirkungsorientierung in der Politikgestaltung feststellen.

Wie Martinuzzi und Leschnik in ihrer systematischen Erhebung von Evaluationsstudien in Österreich feststellen, haben Programmförderungs-massnahmen sowohl im Umwelt- als auch im Energiebereich als Gegenstand von Evaluationen seit dem Jahr 2000 an Bedeutung gewonnen. Zudem hat sich die Zahl der auf Bundesebene durchgeführten Evaluationen in diesem Zeitraum in Österreich deutlich erhöht. Durch ihr Innovationspotential und aufgrund der nicht unbedeutenden Summen öffentlicher Mittel, die in politische Förderprogramme investiert werden, scheint sich in Österreich eine gewisse Selbstverpflichtung zur Evaluation zu entwickeln. Aus der gestiegenen Zahl von begleitenden Evaluationen schliessen die Autoren, dass Evaluationen in den öffentlichen Diensten vermehrt als Lernchance wahrgenommen werden.

In der Schweiz hat sich eine solche Evaluationskultur in der öffentlichen Verwaltung teilweise bereits in den 1990er Jahren etabliert. Vor allem im Energiebereich ist die Evaluation seit der Lancierung des ersten energiepolitischen Programms des Bundes im Jahr 1990 gut verankert. Allerdings verfügt der Bund heute im Vergleich zu 2000 über bedeutend weniger finanzielle Mittel für die Förderung von erneuerbaren Energien und die rationelle Energieverwendung. Diese Budgetreduktion schränkt auch die finanziellen Mittel ein, die für die Evaluation dieser Fördermassnahmen bereitgestellt werden. Es zeichnet sich daher eine stärkere Verlagerung der Evaluation auf strategische Fragen ab. Zudem wird versucht,

die Evaluation weiter zu professionalisieren und auch die internationale Einbindung zu verstärken.

Im Umweltbereich kam es in allen drei Ländern seit 2000 zu einer Ausweitung der Evaluationstätigkeiten. Während sich in der Schweiz über die letzten Jahre ein Forschungs- und Beratungszweig mit verschiedenen Evaluationsaktivitäten entwickelt hat, gehört die Umweltpolitik in Deutschland zu jenen Politikbereichen, in denen die Evaluation bis heute eher wenig verankert ist. Duscha, Klemisch und Meyer gehen jedoch davon aus, dass das Umweltthema in Deutschland und weltweit weiter an Bedeutung gewinnen wird, was sich auch in einer Ausweitung der Evaluationstätigkeiten in diesem Bereich niederschlagen könnte. Vor allem in Verbindung mit klimapolitisch relevanten energie- und verkehrspolitischen Fragen ist zu erwarten, dass die Umweltevaluation in den nächsten Jahren einen höheren Stellenwert einnehmen wird. Eine ähnliche Entwicklung lässt sich grundsätzlich auch für die Schweiz und Österreich erwarten, wenn auch der Energie- und Umweltbereich – wie andere Politikbereiche auch – immer wieder unter dem politischen Druck stehen, Staatstätigkeiten abzubauen und die öffentlichen Ausgaben zu reduzieren.

5 Bilanz und Ausblick

Hatte der Umweltbereich in den 1990er Jahren zwischenzeitlich an Stellenwert verloren, lässt sich heute vor allem auf konzeptioneller und strategischer Ebene wieder eine Zunahme der Evaluationstätigkeiten unter dem Stichwort ‚Nachhaltigkeit‘ feststellen. Inwiefern die Betonung des Nachhaltigkeitsprinzips in der Politik aller drei Länder tatsächlich zu einer Ausweitung von ex-post Evaluationen zu den Wirkungen von politischen Massnahmen führt, bleibt hingegen abzuwarten. Zwar haben sich die Regierungen Deutschlands, Österreichs und der Schweiz dem Grundsatz der nachhaltigen Entwicklung und einer regelmäßigen Überprüfung der Nachhaltigkeit politischer Massnahmen verpflichtet. Eine Konkretisierung und Umsetzung dieser Prinzipien in den verschiedenen Sektoralpolitiken steht jedoch weitgehend noch aus. Zusätzlich dürften sich aber für die Umweltevaluation neue Betätigungsfelder eröffnen, indem etwa raumplanerische und regionalpolitische Instrumente in der Umweltpolitik zunehmend an Bedeutung gewinnen.

Im Energiebereich haben Evaluationen zu energiepolitischen Instrumenten seit dem Jahr 2000 deutlich zugenommen. Mit der derzeitigen Aktualität der Energie- und Klimapolitik wäre eigentlich zu erwarten, dass auch die Evaluationstätigkeiten in diesen Bereichen weiter ausgebaut werden. Dass dies jedoch nicht zwangsläufig der Fall sein muss, zeigt das Beispiel der Schweiz. Hier stellt das derzeit laufende Energieprogramm ‚EnergieSchweiz‘ deutlich weniger finanzielle Mittel für die Evaluation (wie auch für das Programm selbst) bereit als das Vorgängerprogramm ‚Energie 2000‘. So bleibt die Evaluation immer abhängig vom politischen Willen, für die Überprüfung der Wirksamkeit politischer Massnahmen genügend Ressourcen zur Verfügung zu stellen.

Zudem steht die Evaluation in allen drei Ländern – in unterschiedlichem Ausmass – vor der Herausforderung einer weiteren Institutionalisierung und Professionalisierung. In der Schweiz ist die Evaluation im Energiebereich auf Bundesebene institutionell seit Beginn der 1990er Jahre gut verankert. Für den Umweltbereich stellt sich derzeit die Frage, ob die

relativ umfangreiche Evaluationstätigkeit im hauptsächlich zuständigen Bundesamt für Umwelt ebenfalls mit der Schaffung einer Stabstelle für Evaluation organisationell gestärkt werden soll. Während auch in Deutschland eine weitere Institutionalisierung und Professionalisierung der Energie- und Umweltevaluation erwartet werden kann, bleiben die Zeichen aus Österreich widersprüchlich. Eine 1995 im Bundeskanzleramt eingerichtete Koordinationsstelle für Evaluation wurde im Jahr 2000 wieder eingestellt. Hingegen sieht die österreichische Nachhaltigkeitsstrategie aus dem Jahr 2002 für die Umwelt- und Energiepolitik eine Aufbau- und Ablauforganisation vor, die regelmässige Evaluationen beinhaltet.

Entwicklungszusammenarbeit

Evaluierung in der deutschen Entwicklungszusammenarbeit

1 Rahmenbedingungen

1.1 Bereiche und Akteure der Entwicklungszusammenarbeit

Die deutsche Entwicklungszusammenarbeit (EZ) ist – mehr als die der meisten anderen Geberländer – durch eine Vielfalt an Instrumenten und Akteuren gekennzeichnet. Zwar ist die Entwicklungspolitik gebündelt, das heißt das Bundesministeriums für wirtschaftliche Entwicklung und Zusammenarbeit (BMZ) ist zuständig für bi- und multilaterale Entwicklungszusammenarbeit und finanziert durch seinen Haushalt den größten Teil der deutschen Entwicklungshilfeleistungen (2007: rd. 4,5 Mrd Euro)¹; die Durchführung der bilateralen staatlichen Programme (ca. 50 % der jährlichen Entwicklungsausgaben) erfolgt jedoch durch insgesamt acht staatliche (Durchführungs-) Organisationen. Eine noch viel größere Zahl zivilgesellschaftlicher Organisationen (insgesamt über 350, wobei sich die Finanzierungen auf einige wenige große konzentrieren) erhält eine Förderung von Projekten und Programmen. Der multilaterale Teil (rd. 30 % der jährlichen Entwicklungsausgaben) ist konzentriert auf den Entwicklungsfonds der Europäischen Union (EU) und die multilateralen Banken – vorneweg die Weltbank –, daneben sind eine Reihe weiterer Organisationen und deren Programme Gegenstand der Förderung, insbesondere Organisationen der Vereinten Nationen (UN).

Fast alle größeren und eine Reihe von kleineren staatlichen wie nicht-staatlichen Organisationen verfügen – ebenso wie fast durchgängig der multilaterale Bereich – über eigene, zentrale Evaluierungseinheiten oder mindestens Evaluierungsbeauftragte in Controlling- oder Qualitätsmanagement-Einheiten, die „unabhängige Evaluierungen“ im Sinne des Entwicklungsausschusses (Development Assistance Committee; DAC) der Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD) steuern oder durchführen.² Auch wenn die Systeme (d. h. insbesondere der Einsatz von Selbstevaluierungen³ sowie

¹ Official Development Assistance (ODA)/Öffentliche Entwicklungszusammenarbeit anderer Ressorts sowie der Bundesländer bleiben hier unberücksichtigt, da über Evaluierung von deren Finanzleistungen wenig bekannt ist.

² „unabhängig“ bezieht sich demnach zum einen auf die Durchführung von Evaluierungen durch externe Gutachterinnen und Gutachter (oder Mitarbeiter/innen der Evaluierungseinheit), die zuvor nicht nennenswert mit dem Untersuchungsgegenstand befasst waren; zum anderen auf die beauftragende und steuernde Stelle (Evaluierungseinheit), d. h. Arbeitseinheiten, die außerhalb des Einflusses von projektverantwortlichen Arbeitseinheiten (mit operativer Verantwortung) sind (vgl. DAC 1998).

³ Als Selbstevaluierung werden hier Evaluierungen bezeichnet, die von den projektverantwortlichen Stellen gesteuert werden – auch solche, die mit externen Gutachtern durchgeführt werden.

unabhängigen Evaluierungen) im Einzelnen unterschiedlich sind, so ist gerade in der Entwicklungszusammenarbeit die Einigung auf Evaluationsstandards recht weit gediehen.

1.2 Veränderungen im deutschen Kontext

Tiefgehende Veränderungen der entwicklungspolitischen Vorgaben bestimmen das Bild: Die noch bis Ende der 1990er Jahre propagierte Politik der „Pluralität der deutschen EZ“, welche die Vielfalt der Akteure in den Vordergrund stellte, wurde abgelöst durch die Politik einer auf Schwerpunktbereiche konzentrierten „EZ aus einem Guss“ für die staatliche EZ, die Instrumente und Akteure unter das Primat der Politik stellt und eine Verzahnung der bi- und multilateralen EZ anstrebt (vgl. BMZ 2005; Zintl 2006). Dies blieb nicht ohne Auswirkungen auf die Evaluierung. Neue Verwaltungsvorschriften zur Bundeshaushaltsordnung aus den Jahren 2001 bzw. 2006 stellen darüber hinaus hohe Anforderungen an die Erfolgskontrolle als Aufgabe der die Mittel bereitstellenden Bundesbehörde, mit Evaluierungen als einem wesentlichen Bestandteil. Diese werden zum Teil schon lange in der Entwicklungszusammenarbeit erfüllt, wenn auch nicht in gleichem Maße in den diversen Bereichen. Die Neufassung der „Leitlinien für die bilaterale Finanziellen und Technischen Zusammenarbeit“ sieht Schluss- oder Ex-post-Evaluierungen einer angemessenen Stichprobe von Entwicklungsmaßnahmen seitens der beiden größten Durchführungsorganisationen, der Gesellschaft für Technische Zusammenarbeit (GTZ) sowie der KfW Entwicklungsbank (KfW), vor und formalisiert damit die seit Jahren zunehmend angewandte Vorgehensweise (vgl. BMZ 2007).

2 Evaluationsaktivitäten

2.1 Überblick

Grundsätzlich sind alle Maßnahmen in den genannten Bereichen, die aus dem Haushalt des BMZ finanziert werden – staatliche, nicht-staatliche und multilaterale EZ – Gegenstand von Evaluierungen des BMZ. De facto wird nach dem Subsidiaritätsprinzip vorgegangen: Soweit die verantwortlichen Organisationen über eigene Einrichtungen der systematischen Verlaufs- und Ergebnisüberprüfung sowie Evaluierung verfügen, sind BMZ-Evaluierungen nicht oder nur in Ausnahmefällen erforderlich bzw. sinnvoll. Dies gilt insbesondere für die Europäische Kommission, multilaterale Banken sowie UN-Organisationen: Evaluierungen einzelner Mitglieder beschränken sich meist auf zweckgebundene, das heißt für abgrenzbare Projekte und Programme zur Verfügung gestellte Mittel. BMZ-Evaluierungen beschränken sich in aller Regel auf Fragen strategischen Interesses mit Blick auf die politischen Steuerungsaufgaben des BMZ, und werden in den meisten Fällen von der zentralen Facheinheit – dem Referat 120 „Evaluierung der Entwicklungszusammenarbeit“ – geplant und gesteuert.

Die jeweiligen „Systeme“⁴ der Evaluierung der einzelnen Trägerorganisationen sind Gegenstand der periodischen Überprüfung, sei es durch externe Leistungsüberprüfungen oder Peer Reviews (oder einer Kombination von beiden). Dies gilt zunehmend für den multilateralen Bereich (multilaterale Banken und UN-Organisationen) und vor allem auch den bilateralen: Neben der eher kursorischen Überprüfung im Rahmen der alle vier Jahre stattfindenden DAC Peer Reviews, die die Entwicklungszusammenarbeit insgesamt auf den Prüfstand stellen (vgl. DAC 2005a), ist die Qualität, Unabhängigkeit und Nützlichkeit des deutschen Evaluierungssystems nunmehr, wie bereits 1999, erneut Gegenstand einer Überprüfung, die neben der Evaluierung des BMZ auch die von 19 deutschen EZ-Organisationen mit einbezieht. Die Ergebnisse der vom Hamburgischen WeltWirtschaftsinstitut (HWWI) mit dem Centrum für Evaluation (CEval) und zwei Peers (den Leitern der Schweizer und der belgischen Evaluierungsabteilung) durchgeführten Überprüfung werden voraussichtlich im Frühsommer 2008 vorliegen (vgl. Borrmann/Stockmann 2008).

2.2 Evaluierung der bilateralen Entwicklungszusammenarbeit

Evaluierungen durch die staatlichen Durchführungsorganisationen

Dieser größte Handlungsbereich der deutschen Entwicklungspolitik ist in Bezug auf Evaluierung stark arbeitsteilig organisiert: So wurde die Verantwortung für Projektevaluierung bereits Ende der 1990er Jahre an die KfW Entwicklungsbank und die Gesellschaft für Technische Zusammenarbeit (GTZ) delegiert, mit der Maßgabe, DAC-konforme Evaluierungen durchzuführen. Bis Ende 2006 führte die KfW-Evaluierungsabteilung jährlich für alle abgeschlossenen Projekte Ex-post-Evaluierungen unter Einsatz von externen Gutachtern/innen oder KfW-Mitarbeitern/innen, die keine Projektverantwortung haben, durch; das heißt zum Beispiel für 2006 über hundert Projekte. Mit Beginn des Jahres 2007 ist ein Übergang auf eine geschichtete Stichprobe vorgesehen. Die GTZ hat – nach einer langen Tradition von regelmäßigen Selbstevaluierungen – im Jahr 2005 mit systematischer, unabhängiger Evaluation begonnen und beauftragt und steuert zur Zeit rund dreißig Evaluierungen laufender und abgeschlossener Vorhaben aus zwei Themenbereichen pro Jahr, die von externen Instituten oder Firmen durchgeführt werden. Damit wird in dem jeweiligen Themenbereich im Durchschnitt ein hoher Deckungsgrad erreicht (je rd. 50 % Schlussevaluierungen und 10 % der laufenden Vorhaben; allerdings nur 5 % ex-post). Auch der Deutsche Entwicklungsdienst (DED) und die Internationale Weiterbildung und Entwicklung gGmbH (InWEnt) führen Evaluierungen von Projekten und zu Querschnittsfragen durch, überwiegend jedoch im Auftrag der projekt-/ programmverantwortlichen Stellen. Die Ergebnisse dienen zum einen dem eigenen Qualitätsmanagement dieser Organisationen. Zum anderen sind sie eine wichtige Grundlage der Erfolgskontrolle des BMZ. Über den Einzelfall hinaus sind Querschnittsauswertungen von Interesse: So hat zum Beispiel die KfW Entwicklungsbank eine

⁴ Meist gibt es eine Mischung von mehr oder weniger formalisierter Selbst- und interner Evaluation, Fortschrittskontrollen und ähnliches sowie unabhängiger Evaluierung. Nur letztere ist, soweit nicht explizit erwähnt, Gegenstand der Ausführungen. Ex-ante-Evaluierungen wiederum zählen in der deutschen EZ nicht zum Bereich Evaluation, sondern werden als Prüfungen bezeichnet, und sind dementsprechend ebenfalls hier ausgeschlossen.

Auswertung der 2004 und 2005 durchgeführten Evaluierungen von Transportvorhaben vorgenommen (vgl. KfW 2006).

Zusammengefaßt läßt sich sagen, dass pro Jahr über 150 Projekte unabhängig („zentral“) evaluiert werden – überwiegend durch die staatlichen Durchführungsorganisationen selbst und zum Teil durch das BMZ im Rahmen von Querschnittsuntersuchungen (s. u.). Dazu kommt eine viel größere Zahl von Projekten, die auf Veranlassung von Programmverantwortlichen ebenfalls unter Einsatz externer Gutachter/innen evaluiert werden. Anders ausgedrückt: Beinahe jedes Projekt der staatlichen Entwicklungszusammenarbeit wird während seiner Laufzeit mindestens einmal, gelegentlich mehrfach, einer Evaluierung durch projektexterne Gutachter/innen unterzogen – ex ante Evaluierungen (die in der Entwicklungszusammenarbeit als Projektprüfung bezeichnet werden) nicht mit gerechnet.

Nicht-staatliche EZ

Die größeren, auf Entwicklungszusammenarbeit spezialisierten zivilgesellschaftlichen Organisationen, wie die Deutsche Welthungerhilfe oder die kirchlichen Zentralstellen, verfügen seit vielen Jahren über unabhängige Evaluierungseinheiten, die Projektevaluierungen durchführen oder andere Arbeitseinheiten dabei unterstützen und zunehmend auch Querschnittsevaluierungen durchführen. Die Herangehensweise ist im zivilgesellschaftlichen Bereich vielfach eher lern- und partnerorientiert, zumal wenn diese Organisationen ebenfalls nicht selbst Projekte durchführen, sondern Partnerorganisationen längerfristig fördern. So hatte die von einer Gruppe von 32 indischen und 14 deutschen zivilgesellschaftlichen Organisationen veranlasste Evaluierung von Mikrofinanzinstitutionen in Indien vorrangig Empowerment der Nutzergruppen zum Ziel. Insbesondere soweit diese Organisationen im internationalen Wettbewerb um Finanzmittel stehen, kommt der „Kontroll-“Aspekt jedoch gleichgewichtig hinzu. Insgesamt ist das Bild hier noch recht heterogen. Einige, insbesondere kleinere Organisationen – zumal in der humanitären Hilfe (internationale Nichtregierungsorganisationen ausgenommen) oder mit Spezialaufgaben der Aus- und Fortbildung oder politischen Bildung im In- und Ausland – beginnen zum Teil erst, sich mit dem Thema Evaluierung auseinander zu setzen, mit exemplarischen Ausnahmen.

Evaluierungen des BMZ

Seit Ende der 1990er Jahre konzentriert sich das BMZ zunehmend auf komplexe oder „strategische“ Evaluierungen, das heißt Evaluierungen von Themen, Sektoren, Instrumenten oder größeren (insbesondere instrumentenübergreifenden) Programmen, in der Regel mit mehreren Fallstudien. Die Auswahl der Evaluierungen ist überwiegend nachfrageorientiert. Der Bedarf wird alle zwei Jahre im BMZ (und den staatlichen Durchführungsorganisationen) abgefragt, durch eigene Vorschläge des Evaluierungsreferats ergänzt und im „Zentralen Evaluierungsprogramm“ des BMZ durch Staatssekretärsentscheidung festgelegt. Neue Prioritäten führen gelegentlich zu Änderungen und de facto zu einer rollenden Planung. Insbesondere in den letzten Jahren hat die Zahl der Gemeinschaftsevaluierungen mit ande-

ren Gebern zugenommen: Von den 13 komplexen, das heißt projektübergreifenden Evaluierungen mit mehreren Fallstudien, die in den Jahren 2005 bis 2006 abgeschlossen wurden, wurden vier gemeinsam mit anderen Partnern durchgeführt. Im Jahr 2007 werden es fünf von sieben sein. Die Gründe dafür liegen zum einen im Evaluationsgegenstand selbst (wenn es sich z. B. um gemeinschaftsfinanzierte Programme handelt), zum anderen in dem zunehmenden Bestreben nach geberübergreifender Bearbeitung von Themen zwecks gemeinsamen Lernens (z. B. aktuell „Bürgermitsprache und Rechenschaftslegung“, vgl. Foresti et al. 2007; O'Neil et al. 2007).

Anlass, Gegenstand und Fragestellungen der Evaluierungen sind so unterschiedlich wie die Handlungsfelder des BMZ selbst. Eine ganze Reihe konzentriert sich auf Instrumente oder Mechanismen in Deutschland⁵, wie zum Beispiel die Evaluierung der Berücksichtigung von Genderfragen in der deutschen EZ. Untersuchungsgegenstand waren die organisatorischen Vorkehrungen und die (planerische) Praxis im BMZ und den vier Durchführungsorganisationen. Auf weitere Feldstudien wurde verzichtet, da es genug Anhaltspunkte für den Stand und notwendige Verbesserungen gab, der Mehrwert von Feldstudien also nicht gegeben war (BMZ 2005). Eher „klassisch“ verlief zunächst die Evaluierung des Reintegrationsprogramms mit Fallstudien in vier Ländern. Da sich dieses Programm weniger als ursprünglich vom Fachreferat vermutet an den übergeordneten politischen Zielen des BMZ ausgerichtet erwies, wurden im Anschluss, über die üblichen Aufgaben eines Syntheseberichts hinaus, auf Wunsch des Fachreferats von der Gutachterin detailliertere Umsetzungsempfehlungen erarbeitet. Eine Besonderheit in vielerlei Hinsicht stellt die Evaluierung der Nachhaltigkeit ländlicher Regionalentwicklungsvorhaben dar: Zum einen handelt es sich um Ex-post-Evaluierungen von vier sehr großen Vorhaben in vier Ländern, die seit vielen Jahren abgeschlossen waren. Zum anderen wurden die Evaluierungen vollständig von lokalen Gutachtern durchgeführt. In den einzelnen Ländern wurden lokale Steuerungsgruppen eingerichtet und deutsche Gutachter zur methodischen und organisatorischen Begleitung unter Vertrag genommen, die am Ende des mehrjährigen Prozesses schließlich eine Querschnittsauswertung verfassten. In Punkto Komplexität wurde diese durch die internationale Gemeinschaftsevaluierung zu General Budget Support unter der Leitung des britischen Entwicklungsministeriums (UK Department for International Development – DFID) übertroffen: Nicht nur wurden – nach einer länger währenden Designphase – Fallstudien in sieben Ländern durchgeführt. Auch die schiere Zahl der Mitglieder der Steuerungsgruppe – Vertreter und Vertreterinnen von 24 Organisationen sowie Vertreter der Partnerländern selbst – machte dieses Vorhaben zu einem organisatorisch herausfordernden Unterfangen. Derartige Evaluierungsvorhaben können vom BMZ in Federführung nur in Ausnahmefällen geleitet werden, wie bei der Evaluierung der „Enabling Development Policy“ des World Food Programme, die ebenfalls sieben Fallstudien und ebenso viele Mitglieder in der Steuerungsgruppe umfasste. Neu sind Länderprogrammevaluierungen, die sich auf die bilaterale staatliche Entwicklungszusammenarbeit konzentrieren und sich im Sinne eines „*light approach*“ auf Managementfragen (vgl. einleitend: Politik einer EZ aus einem Guss) sowie eine

⁵ Die Beispiele konzentrieren sich auf die 2005 und 2006 abgeschlossenen Evaluierungen. Alle im Folgenden genannten Evaluierungen finden sich auf der Webseite des BMZ unter <http://www.bmz.de/de/service/infothek/evaluierung/index.html> [29.4.2008].

grobe Abschätzung der entwicklungspolitischen Wirksamkeit konzentrieren. Gerade bei Länderprogramm- oder Sektorevaluierungen geht die Tendenz zu Gemeinschaftsevaluierungen, mit möglichst hoher Partnerbeteiligung – wie zum Beispiel bei der laufenden Evaluation des Tansanischen Gesundheitssektors unter der Federführung des dänischen Außenministeriums, an der sich nicht nur das BMZ, sondern auch das tansanische Finanzministerium an der Managementgruppe beteiligt. Neue Instrumente, wie Gemeinschaftsfinanzierungen oder vergleichbare sektorweite Ansätze, lassen eine Zurechnung von Wirkungen auf einzelne Geber kaum mehr zu.

Andere evaluierungsrelevante Aktivitäten

Hier ist vor allem der Arbeitskreis Entwicklungspolitik der DeGEval – Gesellschaft für Evaluation zu nennen, der Auftraggeber, Wissenschaft und Gutachterinnen und Gutachter zu den halbjährlichen Tagungen zusammenbringt. Arbeitsgruppen zu unterschiedlichen Themen ergänzen die Arbeit, wie zum Beispiel zum Thema „Lernen aus Evaluation“ (vgl. AK Evaluation von Entwicklungspolitik 2005). Im Bereich Aus- und Fortbildung sind vermehrte Aktivitäten zu verzeichnen: So wurde in den vergangenen Jahren das Fortbildungsprogramm Evaluation in der Entwicklungszusammenarbeit (FEEZ) vom Centrum für Evaluation der Universität des Saarlandes (CEval) mit der Arbeitsgemeinschaft entwicklungspolitischer Gutachter (AGEG) eingerichtet. CEval hat darüber hinaus einen Masterstudiengang Evaluation eingeführt, der sich auch Entwicklungsfragen widmet. Das Seminar für Ländliche Entwicklung (SLE) an der Humboldt-Universität Berlin befasst sich im Rahmen seiner Ausbildungsgänge in jüngerer Zeit vermehrt mit der Untersuchung von Wirkungen, ebenso wie das Deutsche Institut für Entwicklungspolitik (DIE). Kommerzielle Anbieter von Fortbildungsveranstaltungen, auch aus dem Ausland, runden das Bild ab.

3 Verbindungen mit anderen Politikfeldern und Internationale Entwicklungen

Wie die Entwicklungszusammenarbeit selbst behandelt deren Evaluation fast die gesamte Palette möglicher Politikfelder, selbst über die zehn Themenschwerpunkte dieses Bandes hinaus (wie z. B. Finanzsektorentwicklung, Demokratieförderung u. a.). Experten der verschiedenen Fachrichtungen sind regelmäßig Teil von Teams bei BMZ-Evaluierungen – Evaluierungsexperten der verschiedenen Themenbereiche sind jedoch rar. Ob dies aus einem generellen Mangel an Evaluationsexpertise oder mangelnder Bereitschaft von Experten, sich auf die oft schwierigen Verhältnisse im Entwicklungskontext einzustellen, resultiert, oder umgekehrt aus einer bisher unzureichenden Nachfrage seitens der deutschen Entwicklungsorganisationen, ist schwer auszumachen. Die methodischen Diskussionen – und Erprobungen in der Praxis – werden jedenfalls, mit wenigen Ausnahmen, in anderen Ländern vorangetrieben.

Wie in allen anderen Bereichen der Entwicklungszusammenarbeit, wird der *state of the art* international definiert (auch wenn Prioritäten lokal bestimmt werden). Für die bilateralen Geber ist der Ort der Konsensfindung und gleichzeitig Standard setzende Organisation der Entwicklungsausschuss der OECD (Development Assistance Committee, DAC). Die Beteiligung der multilateralen Banken sowie von Vertretern der UN-Organisationen unterstützt die Verbreitung von Neuerungen und guter Praxis. Gerade in letzter Zeit wurden mehrere Standards und Orientierungslinien verabschiedet: (probeweise) die DAC *Evaluation Quality Standards for Development Evaluation* (vgl. DAC 2006a), die die DAC-Prinzipien von 1991 insbesondere um Qualitätsstandards für Evaluierungsberichte ergänzen (und die zusammen den DeGEval-Standards ähneln) sowie, die Orientierungslinien für die Durchführung von internationalen Gemeinschaftsevaluierungen (vgl. DAC 2006b). Die Evaluierungsabteilung der Weltbank hat im Auftrag des DAC darüber hinaus ein Nachschlagewerk für die Evaluierung von globalen und regionalen Partnerschaftsprogrammen erstellt (vgl. IEG/World Bank 2007), das in nächster Zeit im BMZ erprobt wird. In Vorbereitung sind darüber hinaus Orientierungslinien für die Evaluierung von konfliktpräventiven und friedensfördernden Maßnahmen seitens des DAC.

Die Gruppe der Leiter/innen der EU-Evaluierungsabteilungen ist vergleichsweise lose organisiert. Hier stehen im Zentrum: gemeinsame Evaluierungen von EU-Themen, wie die soeben abgeschlossene „Triple C“-Evaluierung (Koordination, Komplementarität und Kohärenz), die mehrere Jahre in Anspruch nahm und zu recht ernüchternden Ergebnissen kommt (vgl. EUHES 2007) sowie die Integration der neuen EU-Mitgliedsländer. Eher informell ist die vom BMZ gegründete DACH-Gruppe (Deutschland, Österreich, Schweiz), der auch Belgien und die Niederlande angehören; dieser informelle Rahmen erlaubt, auch sensible Fragen offen zu diskutieren, die manchmal entscheidender für die Arbeit sind als methodisch-technische.

Die Evaluierung der Entwicklungszusammenarbeit wird von der Politik stark beeinflusst. Die neuere Ausrichtung der internationalen Entwicklungsdiskussion ist dabei, auch die Evaluierung nachhaltig zu verändern. Die durch die Millenniumserklärung der UN vorgegebenen Ziele und die durch die Pariser Erklärung des Entwicklungsausschusses der OECD (DAC 2006c) von über hundert Entwicklungsorganisationen und Partnerregierungen vereinbarten Verfahrensprinzipien prägen verstärkt das Bild. Die geforderte Harmonisierung seitens der Geber wird zunehmend vom BMZ praktiziert durch vermehrte Gemeinschaftsevaluierungen. Selbst Arbeitsteilung wird gerade mit der Schweizer Entwicklungszusammenarbeit ausprobiert. Eine besondere Herausforderung ist der geforderte Nachweis von Wirkungen, zumal auf der Ebene letztendlicher (Millenniums-)Ziele. Während dies bei klassischen Entwicklungsvorhaben, wie zum Beispiel einem Bewässerungsvorhaben, relativ leicht zu bewerkstelligen ist, sind die methodischen Herausforderungen bei Maßnahmen im Governance-Bereich, der Friedensförderung, oder gar bei Budgethilfe groß. Als Herangehensweise hat sich international die Verwendung von Wirkungsketten – bei komplexeren Entwicklungsmaßnahmen von „Wirkungsbäumen“ – durchgesetzt. Auf der Basis dieser – hypothetischen – Wirkungszusammenhänge beginnt der Versuch, Wirkungen zu belegen, die kurz- und längerfristig, intendiert oder unbeabsichtigt, und damit auch negativ, sein können und gelegentlich Rückkoppelungsschleifen aufweisen. Angesichts vielfältiger externer Einflüsse sind dem „Nachweis“ von Wirkungen der Entwicklungszusammenarbeit oft

enge Grenzen gesetzt, auch wenn das methodische Instrumentarium hier noch nicht in allen Fällen ausgeschöpft ist. Nach vielen Jahren eher prozessorientierter (auch partizipativer) Evaluierung oder Ex-post-Überprüfungen von tendenziell lediglich Outputs, das heißt der erbrachten Leistungen, stehen die Zeichen erneut eher auf quantitativen Methoden, um Wirkungen zu belegen. Der Aufwand dafür ist hoch und der Zeitrahmen unter Umständen lang, jedoch dürfte für bestimmte Fragestellungen kein Weg daran vorbei führen: So führt das Bundesministerium für wirtschaftliche Entwicklung und Zusammenarbeit (BMZ) zurzeit in Kooperation mit der Freien Universität Berlin, Sonderforschungsbereich 700, eine Untersuchung der (kumulativen) Wirkungen von Entwicklungszusammenarbeit auf Regierungsführung und Konfliktregulierung in Nordost-Afghanistan durch, auch mit dem Ziel der Methodenbildung. Die Untersuchung erstreckt sich über drei Jahre und umfasst Befragungen von über 2000 Haushalten in 2007 und erneut in 2009 sowie ergänzende Fallstudien. Erste Zwischenergebnisse, basierend auf der *baseline* Studie, werden demnächst vorliegen (vgl. Zürcher et al. 2007).

4 Die wichtigsten Veränderungen der Evaluierung der deutschen Entwicklungszusammenarbeit

Im BMZ wurde der bereits im Jahr 2000 wirksam werdende Übergang von Projekt- zu „strategischen“ Evaluierungen konsolidiert, das heißt fast alle der seither durchgeführten Evaluierungen befassen sich mit projektübergreifenden Fragestellungen. Neuerungen betreffen Länderprogrammevaluierungen sowie internationale Gemeinschaftsevaluierungen, die in wachsendem Umfang durchgeführt wurden (siehe oben, Abschnitt 2).

Die Schrittfolge von Evaluierungen – von der Gesamtprogrammplanung bis zur Veröffentlichung von Evaluierungsergebnissen im Einzelnen – ist inzwischen weitgehend festgelegt, mit Variationen, die sich aus dem breiten Spektrum der durchgeführten Evaluierungen und Evaluierungsbeteiligten ergeben. Regelmäßig wird mit den betroffenen Fachreferaten im BMZ sowie mit den entsprechenden Stellen in den Durchführungsorganisationen eine begleitende Arbeitsgruppe gebildet, die an der Entwicklung der *Terms of Reference* ebenso beteiligt ist wie an der Kommentierung der Berichte, der Umsetzungsplanung für die Empfehlungen sowie der internen und externen Kommunikation von Evaluierungsergebnissen. Diese lernorientierte Vorgehensweise – die bei Gemeinschaftsevaluierungen noch deutlich mehr Beteiligte mit einschließt – unterstützt die spätere Nutzung der Ergebnisse. Eine besondere Herausforderung ist die Einbeziehung der Partnerländer, die zunehmend erfolgt – nicht nur durch den regelmäßigen Einsatz lokaler Gutachter/innen, sondern auch durch die Einbeziehung verantwortlicher Regierungsstellen in den gesamten Prozess.

Weitere Neuerungen betreffen eine formalisierte Umsetzungsplanung und ein Umsetzungsmonitoring sowie die generelle Publikation von Kurzfassungen der Evaluierungsberichte auf der Webseite des Bundesministeriums für wirtschaftliche Entwicklung und Zusammenarbeit (BMZ). Die vollständigen Berichte werden auf Anfrage zur Verfügung gestellt. Der Verbreitung von Evaluierungsergebnissen insbesondere im BMZ selbst soll ein neu aufgelegter Newsletter dienen.

Die Umsetzung der DAC- und DeGEval Standards mindestens in der Arbeit unabhängiger Evaluierungsabteilungen dürfte heute bei allen größeren deutschen Entwicklungsorganisationen weitgehend gegeben sein. Der Diffusion von *good practice* Standards und von Neuerungen von Verfahren, Methodik oder evaluationsrelevanten entwicklungspolitischen Themen dienen die vom BMZ gegründeten jährlichen Treffen der Evaluierungsabteilungen deutscher Entwicklungsorganisationen ebenso wie gelegentliche („deutsche“) Gemeinschaftsevaluierungen oder Arbeitsgruppen des Arbeitskreises Entwicklungspolitik der DeGEval. Die Grundsätze und zentralen Verfahren werden seitens des BMZ zurzeit in Leitlinien der Evaluierung zusammengefasst. Einzelne Elemente, wie zum Beispiel die an den DAC-Standards für Evaluierung orientierte Checkliste für die Beurteilung der Qualität von Evaluierungsberichten, liegen vor und sind auf der Webseite des BMZ einsehbar.⁶

Von besonderer Bedeutung ist in diesem Zusammenhang auch die im Mai 2005 vom BMZ mit den staatlichen Durchführungsorganisationen eingerichtete Gruppe „Evaluierung aus einem Guss“. Die eingangs erwähnte, weitgehende Delegation der Verantwortung für Projektevaluierung führte in der Praxis zu einem „Auseinanderlaufen“ von Systemen. Die Gruppe hat in einem ersten Schritt gemeinsame Orientierungslinien für Evaluierungskriterien in Anlehnung an die DAC-Evaluierungskriterien beschlossen (BMZ 2006). Auf dieser Basis haben KfW und GTZ ihre Bewertungskriterien vereinheitlicht, so dass erstmalig „*ratings*“ von Projekten grundsätzlich kompatibel sein werden. Eine bessere Verzahnung der Evaluierungsprogramme und Veröffentlichungspraxis sind Gegenstand laufender Maßnahmen, ebenso wie das Thema „Lernen aus Evaluierung“ und Wissensmanagement – auch und gerade institutionenübergreifend.

5 Bilanzierung und Ausblick

Der hohe Deckungsgrad von Projekten und Programmen der deutschen Entwicklungszusammenarbeit durch Selbst- oder unabhängige Evaluierung dürfte im Vergleich zu anderen Politikfeldern einmalig sein. Evaluierungen sind etablierte Verfahren und werden mindestens organisationsintern nachgefragt und hinreichend genutzt, wenn auch in deutlich unterschiedlichem Maße. Besonders auf strategischer Ebene gibt es sicherlich vielfach Ausbaubedarf. Insbesondere für international um Finanzierung oder auch nur Sichtbarkeit konkurrierende Organisationen gehört Evaluation zum Markenzeichen. In Deutschland ist – im Vergleich zu vielen anderen Ländern – die Nachfrage aus dem öffentlichen und parlamentarischen Raum vergleichsweise gering. Schnelle Bilder, impressionistische Erfolgs- (oder Misserfolgs-) Meldungen – bis hin zum Gebrauch kritischer Ergebnisse zur Bestätigung eigener Vorurteile – lassen eine Evaluationskultur nur schwer gedeihen. Umgekehrt muss sich die Evaluierung der Entwicklungszusammenarbeit vielleicht den Vorwurf gefallen lassen, gelegentlich zu spät und zu wenige politisch relevante Bewertungen zu liefern. Auch die Qualität von Evaluierungen lässt gelegentlich zu wünschen übrig.

Auch in Deutschland nimmt die Frage „Was bewirkt die Entwicklungszusammenarbeit?“ an Bedeutung zu. Jeder, der sich mit der Evaluation komplexer Programme mit viel-

⁶ Siehe <http://bmz.de/de/ErfolgKontrolle/ziele/Grundsaeetze.html> [10.3.2008].

fältigen Einflüssen befasst hat, weiß, wie schwierig diese Frage zu beantworten ist. Hier wird eine neue „Qualitätsoffensive“ erforderlich sein, die nur durch ein Zusammenspiel von Evaluierung und empirischer Forschung zu bewältigen sein wird. Auch werden in Zukunft Ressortgrenzen überschritten werden müssen: Bei Friedenseinsätzen, in der humanitären Hilfe oder beim Umweltschutz sind zunehmend mehrere Ressorts am gleichen Ort mit entwicklungspolitischen Zielsetzungen tätig. Kohärenz wird so ein Gebot auch für die Evaluierung.

Entwicklungspolitik ist keine „nationale“ Angelegenheit, sondern eine (von anderen Inputs oft ununterscheidbare) unterstützende Tätigkeit der Entwicklungsanstrengungen der Partnerländer. Was am Ende interessiert ist weniger, was die deutsche Entwicklungszusammenarbeit gebracht hat, sondern was aus den Politiken und Programmen der Partner geworden ist, die wir unterstützt haben. Dies bedeutet nicht weniger, als dass aus geberzentrierten Evaluierungen partnerzentrierte werden müssen. Die Förderung von Transparenz und evidenzbasierter Politik sowie Rechenschaftslegung gegenüber den eigenen Bürgern ist eine der zentralen Aufgabenbereiche der Entwicklungspolitik zur Förderung von *good governance* – und damit eine Herausforderung für die Evaluierung der Entwicklungszusammenarbeit.

Evaluierung in der Österreichischen Ost- und Entwicklungszusammenarbeit

1 Einführung

Bei der Betrachtung der Funktion und Stellung der Evaluierung im System der österreichischen Ost- und Entwicklungszusammenarbeit (im folgenden OEZA) ist es erforderlich, auch Bezug auf die Grundlagen und Rahmenbedingungen zu nehmen, auf denen die Evaluierungsfunktion aufbaut und innerhalb derer sie angesiedelt ist. Die OEZA umfasst alle jene Aktivitäten im Rahmen der Ost- und Entwicklungszusammenarbeit, die beim Bundesministerium für europäische und internationale Angelegenheiten budgetiert und deren Durchführung entweder durch das Bundesministerium selbst (gilt für multilaterale Programme und Projekte) oder durch die Austrian Development Agency (ADA; gilt für sämtliche bilaterale Programme und Projekte) erfolgt. Leistungen anderer Ministerien wie zum Beispiel die Beiträge zum Europäischen Entwicklungsfonds, die vom Bundesministerium für Finanzen bereitgestellt werden, fallen nicht in diese Kategorie.

1.1 Gesetzliche Grundlagen

Das österreichische Parlament hat im Sommer 2002 das Entwicklungszusammenarbeitsgesetz verabschiedet, das seither die öffentliche Entwicklungszusammenarbeit regelt und das bis dahin gültige „Entwicklungshilfegesetz“ aus dem Jahr 1974 ablöste. Bereits ein Jahr später, 2003, kam es zu einer Novellierung des Gesetzes, mit der die Schaffung einer Agentur für Entwicklungszusammenarbeit (ADA) ermöglicht wurde. Damit verbunden war in Folge die bis dahin größte Umstrukturierung in der Geschichte der österreichischen Ost- und Entwicklungszusammenarbeit. Die neu geschaffene ADA übernahm mit 1. Jänner 2004 sämtliche operative Agenden, während die bis dahin für die gesamte Entwicklungszusammenarbeit zuständige Sektion im Außenministerium nunmehr mit der Ausarbeitung und Vorgabe der inhaltlichen Rahmenbedingungen (Politiken, Strategien, Instrumente, geographische Schwerpunkte etc.) sowie mit der Wahrnehmung der Beziehungen zu einschlägigen multilateralen Organisationen wie den Vereinten Nationen geänderte Aufgaben erhielt. Die Zusammenarbeit zwischen den beiden Organisationseinheiten sowie den ebenfalls zur ADA gehörigen Koordinationsbüros in den Partnerländern wurde durch mehrere Vereinbarungen zwischen dem Bundesministerium und der ADA geregelt.

Eine weitere gesetzliche Grundlage stellt das Entwicklungshelfergesetz aus dem Jahr 1983 dar, mit dem die Rahmenbedingungen für Einsätze von Freiwilligen und Experten in

Entwicklungsländern geregelt werden. Dieses Gesetz steht derzeit vor einer Überarbeitung, um seine Bestimmungen den seit 1983 wesentlich veränderten Anforderungen von Personaleinsätzen anzupassen.

Bei der dritten Grundlage, auf der die OEZA aufbaut, handelt es sich zwar um kein Gesetz an sich, sie ist aber im EZA-Gesetz 2002 – wie auch schon im alten Entwicklungshilfegesetz – als ein österreichisches Spezifikum festgeschrieben. Die Rede ist hier vom Dreijahresprogramm der OEZA. Dieses Dokument umfasst jeweils einen Planungszeitraum von drei Jahren, wird aber jährlich unter der Federführung des Außenministeriums und, in Entsprechung des im EZA Gesetz 2002 verankerten Kohärenzgebotes, in Abstimmung zwischen allen mit Agenden der EZA befassten Ministerien fortgeschrieben und dem Ministerrat zur Beschlussfassung vorgelegt. Im Anschluss daran geht es dem Parlament zur Kenntnisnahme zu. Das Dreijahresprogramm ist die Grundlage für das jährliche Arbeitsprogramm der ADA und legt die inhaltlichen und geographischen Prioritäten der OEZA, aber auch die in Aussicht genommenen Veränderungen, konkret fest.

1.2 Geographische und inhaltliche Schwerpunkte der OEZA

Die OEZA ist derzeit in insgesamt sieben Schwerpunktregionen tätig, in denen sich auch die dreizehn Schwerpunktländer befinden. Zu diesen zählen sieben „klassische“ Entwicklungsländer im Süden (Nikaragua, Bhutan, Kap Verde, Burkina Faso, Uganda, Äthiopien, Mosambik) und sechs Länder in Südosteuropa bzw. am Westbalkan (Serbien, Montenegro, Albanien, Kosovo, Mazedonien und Moldau). Daneben ist Österreich noch im Rahmen eines Schwerpunktprogramms in den palästinensischen Gebieten vertreten.

Die österreichischen Konzentrationsbemühungen finden sich auch in den insgesamt sechs inhaltlichen oder thematischen Schwerpunkten wieder. Es sind dies Wasser und Siedlungshygiene, Energie, Bildung und Ausbildung, Ländliche Entwicklung, Good Governance (inklusive Menschenrechte und Friedenssicherung) und Privatsektorentwicklung. Die beiden wesentlichen übergreifenden Themen (Querschnittsmaterien), hinsichtlich derer jedes einzelne Vorhaben bereits in einem sehr frühen Stadium einer Prüfung unterzogen wird, sind die Gleichstellung von Frauen und Männern (Gender) sowie Umwelt und nachhaltige Entwicklung.

1.3 Leistungen und Kennzahlen

Die gesamten öffentlichen Entwicklungshilfeleistungen Österreichs (Official Development Assistance, ODA) haben sich – mit Ausnahme von 2006 – in den vergangenen Jahren kontinuierlich nach oben entwickelt. Betrug der Anteil der ODA am Bruttonationaleinkommen (BNE) Österreichs 2003 noch 0,20 %, stieg er 2004 auf 0,34 % und 2005 auf 0,52 %, fiel jedoch 2006 auf 0,47 % zurück. In absoluten Zahlen bedeutet dies in den Jahren 2005 und 2006 eine ODA-Gesamtleistung von jeweils etwa 1,2 Milliarden Euro.

Dieser „Leistungsbilanz“ muss allerdings hinzugefügt werden, dass die teils massiven Steigerungen auf einen sehr hohen Anteil an Entschuldungen zurückzuführen sind, der in den genannten Jahren jeweils bei rund fünfzig Prozent der ODA-Gesamtleistung liegt.

Ein Blick auf die für bilaterale technische Hilfe zur Verfügung stehenden Mittel relativiert die Gesamtbilanz daher wiederum. Seit 2004 stehen für diesen wichtigen, weil gestaltbaren, Teil der öffentlichen Leistungen jährlich etwa hundert Millionen Euro zur Verfügung, die von der ADA umgesetzt werden. Dazu kommen pro Jahr noch etwa 15 bis 16 Millionen Euro für freiwillige Beiträge an internationale Organisationen.

Insgesamt sind im Jahr 2007 185 Personen für die OEZA tätig. Davon stehen der ADA für die Umsetzung der Programme und Projekte rund 155 Mitarbeiterinnen und Mitarbeiter zur Verfügung, von denen 75 in der Zentrale in Wien und achtzig (21 entsendete sowie 59 lokal rekrutierte Kräfte) in den Koordinationsbüros in den Schwerpunktländern tätig sind. Die mit der Grundsatzarbeit und internationalen Fragen betraute EZA-Sektion im Außenministerium hat derzeit etwa dreißig Mitarbeiterinnen und Mitarbeiter.

2 Evaluierung in der OEZA

In Umsetzung der Empfehlungen des Development Assistance Committee (DAC) der Organisation für Wirtschaftliche Zusammenarbeit und Entwicklung (OECD) wurde 1989 erstmals eine eigene Abteilung für Zwecke der „Evaluierung und Kontrolle“ der österreichischen Entwicklungszusammenarbeit eingerichtet. Damit wurde zwar den Empfehlungen des DAC Rechnung getragen, in der Realität überwogen zunächst jedoch die Aufgaben der finanziellen Kontrolle der bilateralen Technischen Hilfe, die dem kleinen Team dieser Abteilung ebenfalls übertragen waren, und nicht jene der Evaluierung. Umgekehrt führte dies aber auch dazu, dass der Grossteil der Evaluierungen nicht von dieser Abteilung selbst durchgeführt, sondern im Rahmen von Ausschreibungen an spezialisierte Evaluierungseinrichtungen aus dem (vorwiegend deutschsprachigen Raum) vergeben und von diesen durchgeführt wurde. Diese Vorgangsweise ist in der österreichischen Entwicklungszusammenarbeit auch heute noch die Regel und kommt bei allen strategischen Evaluierungen (siehe unten) zum Tragen.

Die bereits erwähnten großen organisatorischen Veränderungen im Zuge der Schaffung der ADA hatten 2004 auch eine Arbeitsteilung im Bereich der Evaluierung zur Folge. Angelehnt an die generelle Trennung in Grundsatzaufgaben (Außenministerium) und operative Tätigkeit (ADA) wurde in der ADA eine „Stabsstelle Evaluierung“ geschaffen, die seit 1. September 2007 organisatorisch der Geschäftsführung der ADA zugeordnet ist. Die bis Ende 2003 mit Evaluierung und Finanzkontrolle befasste Abteilung im Außenministerium ist auch weiterhin mit Agenden der Evaluierung betraut; durch den Wegfall der früheren (operativen) Funktionen wurden ihr aber neue Aufgaben vor allem im Bereich der Politik- und Strategieentwicklung übertragen.

Aufgrund dieser institutionellen Veränderungen innerhalb der OEZA kam es in der jüngeren Vergangenheit bei den für Evaluierung zur Verfügung stehenden personellen Ressourcen zu einer starken Fluktuation. Diese Entwicklung hatte auch eine kurzfristige

Abnahme der Intensität des Evaluierungsgeschehens – zumindest auf der strategischen Ebene – zur Folge.

2.1 Strategische Evaluierungen

Für die vom Außenministerium und der ADA gemeinsam durchgeführten strategischen Evaluierungen sieht die Arbeitsteilung wie folgt aus:

Das Ministerium erstellt ein Zweijahresprogramm, das unter Bedachtnahme auf frühere Zweijahrespläne eine ausgewogene Mischung aus Evaluierungen von Landesprogrammen (alle Regionen sollen über einen mehrjährigen Zyklus hinweg berücksichtigt werden), von Sektoren und von Instrumenten ist. Grundsätzlich können auch wichtige und/oder große Einzelvorhaben, aus denen Erkenntnisse für andere Programmteile ableitbar sind, Gegenstand strategischer Evaluierungen sein. Bei der Durchführung der Evaluierungen ist gemäß den internen „Leitlinien für die Evaluierung in der OEZA“ vorzugehen, die auf entsprechenden Empfehlungen des DAC beruhen und vor fünf Jahren in Kraft gesetzt wurden. Diese Leitlinien werden durch einen sehr konkreten Leitfaden für die Durchführung von Evaluierungen ergänzt. Beide Dokumente befinden sich derzeit in Überarbeitung, um rezent Empfehlungen des DAC und die organisatorischen Veränderungen innerhalb der OEZA entsprechend zu berücksichtigen.

Die ADA-Stabsstelle ist für die operative Umsetzung der strategischen Evaluierungen verantwortlich. Dies beinhaltet, jeweils in enger Abstimmung mit dem Außenministerium und in Übereinstimmung mit den Leitlinien für die Evaluierung in der OEZA,

- die Mitarbeit bei der Erstellung der Terms of Reference (ToR),
- die (europaweite) Ausschreibung,
- die Auswahl des Evaluierungsteams,
- die Vertragsgestaltung,
- die operative Unterstützung des Evaluierungsteams und administrative „Steuerung“ der Evaluierung,
- die Abnahme von Zwischen- und Endberichten,
- die Sicherstellung der Rückkoppelung der Empfehlungen an die zu befassenden Stellen im Außenministerium, in der ADA und in den Koordinationsbüros,
- das Einholen der Umsetzungspläne für die Empfehlungen,
- den Verfolg der Umsetzung der Empfehlungen.

Regelmäßige, mindestens einmal monatlich stattfindende Besprechungen zwischen den Mitarbeitenden der beiden Einheiten stellen sicher, dass alle Verantwortlichen über den Fortgang der aktuellen Evaluierungen im Bilde sind.

Für den Ablauf strategischer Evaluierungen ist in den Leitlinien eine Reihe von Schritten vorgesehen, die unbedingt einzuhalten sind. Dazu gehören die öffentliche (und, je nach Höhe der veranschlagten Kosten, auch europaweite) Ausschreibung der Evaluierung, die Einbeziehung möglichst aller von der Evaluierung betroffenen Stellen und Personen schon in die Erstellung der ToR, die Integration lokaler Fachleute in das Evaluationsteam, die

Präsentation der Ergebnisse für die Beteiligten zu Ende der Feldbesuche, die Abfassung des Berichtes in der Amtssprache des Partnerlandes sowie die Präsentation des Rohberichtes der Evaluierung durch das Team in Wien. Zusammen mit den schriftlichen Stellungnahmen zum Rohbericht werden die Ergebnisse dieser Präsentation vom Evaluierungsteam in den Endbericht eingearbeitet.

Wichtige strategische Evaluierungen der letzten Jahre, die nach dem beschriebenen Procedere durchgeführt wurden, waren:

- Evaluierung des Landesprogramms Albanien (Teile davon gemeinsam mit der Schweiz),
- Evaluierung der Tätigkeit von Nicht-Regierungsorganisationen in Zentralamerika,
- Evaluierung des Landesprogramms Mosambik,
- Evaluierung des Sektors Wasser und Siedlungshygiene,
- Evaluierung des Engagements der OEZA im Bereich Wasserkraft – Kleinkraftwerke,
- Evaluierung des Sektors Bildung und Ausbildung,
- Evaluierung der Relevanz von Kultur und Kulturerbe in der österreichischen Ost- und Entwicklungszusammenarbeit,
- Evaluierung der Bildungs- und Öffentlichkeitsarbeit der OEZA am Beispiel der Millennium Development Goals und
- Evaluierung von Kleinkreditprogrammen.

Durchschnittlich können mit den im Außenministerium und in der ADA zur Verfügung stehenden personellen und finanziellen Ressourcen im Laufe eines Jahres drei bis vier strategische Evaluierungen durchgeführt und abgeschlossen werden.

Zu erwähnen ist an dieser Stelle auch das Bemühen, gemeinsame Evaluierungen mit anderen Gebern durchzuführen. Ein gut gelungenes Beispiel dafür ist die Evaluierung des Landesprogramms Albanien, bei der in Zusammenarbeit mit dem schweizerischen Staatssekretariat für Wirtschaft (Seco) die Rehabilitierung von Wasserkraftwerken, in deren Rahmen die Schweiz und Österreich jeweils unterschiedliche Teilbereiche übernommen hatten, gemeinsam evaluiert wurde.

Derzeit wird eine Evaluierung des Landesprogramms Burkina Faso zusammen mit dem deutschen Bundesministerium für wirtschaftliche Zusammenarbeit (BMZ) bzw. der deutschen Gesellschaft für technische Zusammenarbeit (GTZ) durchgeführt. Die dabei auftretenden Hemmnisse, die nichts mit der auf beiden Seiten vorhandenen großen Bereitschaft, die Evaluierung gemeinsam durchzuführen, sondern mit praktischen Aspekten wie zum Beispiel den unterschiedlichen Schwerpunktsektoren der beiden Landesprogramme zu tun haben, können als stellvertretend für die Problematik gemeinsamer Evaluierungen insgesamt und, in einem weiteren Sinn, bei der Erfüllung der Forderung der Pariser Deklaration nach mehr Geberkoordination und Harmonisierung angesehen werden.

2.2 Programm- und Projektevaluierungen

Evaluierungen auf dieser Ebene fallen ebenfalls unter die Vorgaben der Leitlinien für die Evaluierung in der OEZA und sind nach diesen Regeln durchzuführen. Die beiden wesentlichen Unterschiede zu den strategischen Evaluierungen manifestieren sich zum einen in unterschiedlichen Zuständigkeiten und Verantwortlichkeiten, die für diese Evaluierungen gelten, zum anderen in einer breiteren Palette von Evaluierungstypen, wie externe oder interne Evaluierung, Selbstevaluierung et cetera.

Um dem in den Leitlinien formulierten Ziel „kein Vorhaben ohne Evaluierung“ zunehmend zu entsprechen, sollte die Möglichkeit einer Evaluierung bereits im ursprünglichen Projektdokument vorgesehen sein. Ist dies nicht der Fall ist zu begründen, warum keine Evaluierung erfolgen sollte. Es handelt sich hier um Evaluierungen von bilateralen Programmen und Projekten, die in der Regel von Projektträgern oder Auftragnehmern durchgeführt werden. Verantwortlich für diese Projekte sind häufig Nichtregierungsorganisationen, die in der OEZA sowohl als Projektträger wie auch im entwicklungspolitischen Diskurs nach wie vor eine sehr wichtige Rolle spielen. So werden nach wie vor zwischen vierzig bis fünfzig Prozent der von der ADA verwalteten bilateralen Mittel von Nichtregierungsorganisationen umgesetzt.

Vorbereitet, in Auftrag gegeben und verantwortet werden Programm- und Projektevaluierungen von den Landes- und Sektorzuständigen oder den Koordinationsbüros der ADA in enger Abstimmung mit dem jeweiligen Projektträger bzw. Auftragnehmer. Die Finanzierung erfolgt in der Regel aus dem Projektbudget.

Die OEZA kommt so dem Ziel, Evaluierung zu einem integralen Bestandteil des Projektzyklus zu machen, mit kleinen Schritten näher. Derzeit werden pro Jahr etwa dreißig Projektevaluierungen durchgeführt. Trägerorganisationen budgetieren in Anboten (bei Ausschreibungen) oder Anträgen (bei Kofinanzierungen durch die ADA) zunehmend auch die Kosten für Evaluierungen mit ein. Im Abnehmen begriffen scheint erfreulicher Weise die Praxis zu sein, Evaluierungen vornehmlich als Legitimierung und Begründung für die Weiterführung von Projekten und Programmen einzusetzen.

3 Ausblick und Herausforderungen

In der österreichischen Entwicklungs- und Ostzusammenarbeit gibt es im Feld der Evaluierung nach wie vor einen gewissen Nachholbedarf. Zu oft noch ruft die Erwähnung des Begriffes Evaluierung den Eindruck einer strengen Kontrolle und in Folge eine Abwehrhaltung bei Betroffenen hervor. Im Fall der Entwicklungszusammenarbeit könnte dies nicht zuletzt auch damit zu tun haben, dass Evaluierung und Kontrolle viele Jahre lang als Geschwisterpaar wahrgenommen wurden. So hieß die einschlägige Abteilung im Außenministerium von ihrer Gründung 1989 bis zur Neustrukturierung 2004 „Evaluierung und Kontrolle“, wobei sich Kontrolle ausschließlich auf Finanzkontrolle bezog.

Für die OEZA lässt sich aber auch feststellen, dass das jahrelange Bemühen, Verständnis und Akzeptanz von, sowie die Nachfrage nach Evaluierung zu erhöhen, durchaus

Früchte getragen hat. Dies war und ist nicht nur auf der Ebene der mit der Projektarbeit direkt Befassten feststellbar, sondern auch auf der Management-Ebene.

Hinsichtlich ihrer personellen Ausstattung stehen für operative Aufgaben der Evaluierung in der ADA zwei Mitarbeiter und im Außenministerium ein Mitarbeiter, allerdings mit beschränkter Zeitkapazität, zur Verfügung. Das jährliche Budget für strategische Evaluierungen beträgt 500'000 Euro. Die OEZA trägt der Forderung nach einem hohen Evaluierungsgrad und Evaluierungsstandard aber nicht nur im Bereich strategischer Evaluierungen, sondern, wie im vorhergehenden Abschnitt erwähnt, auch bei Programm- und Projektevaluierungen Rechnung. Da die dafür aufgewendeten Mittel jeweils über die Projektbudgets bereitgestellt werden, ist die Feststellung der dafür insgesamt bereitgestellten Mittel schwierig.

Durchaus interessant ist auch die Erfahrung, dass die Schaffung neuer Strukturen, wie dies 2004 der Fall war, verstärkte Bemühungen braucht, um die bereits erreichte Akzeptanz der Evaluierung halten und nach Möglichkeit ausbauen zu können. In dieser Hinsicht ist anzuerkennen, dass die wichtigsten Partnerorganisationen der OEZA, großteils Nichtregierungsorganisationen, regelmäßige Evaluierungen zu einem integralen Bestandteil ihrer Tätigkeit gemacht haben.

Welche Herausforderungen stellen sich also für die Evaluierung in der österreichischen Entwicklungs- und Ostzusammenarbeit im Anschluß an die strukturellen Veränderungen der letzten Jahre, aber auch im Zusammenhang mit internationalen Entwicklungen wie der Forderung nach stärkerer Harmonisierung und Koordinierung, wie sie vor allem von der Pariser Deklaration verlangt wird?

Die eine Herausforderung liegt in der weiteren Festigung der Stellung und Akzeptanz der Evaluierung sowohl innerhalb der OEZA, aber auch bei den Projektpartnern und den Durchführungsorganisationen. Damit verbunden muss ein möglichst hoher Grad an Unabhängigkeit sein. Die Peer Review 2004 des DAC der OECD hat bereits im ersten Arbeitsjahr der ADA die damalige Zuordnung der Stabsstelle Evaluierung zur Programm- und Projekt- abteilung kritisch hinterfragt (DAC 2004). Dies hat dazu geführt, dass die Evaluierung direkt der Geschäftsführung unterstellt wurde. Eine völlige Loslösung oder Ausgliederung aus der ADA, wie dies in anderen OECD/DAC-Mitgliedsländern der Fall ist, erscheint derzeit unrealistisch.

Die zweite Herausforderung stellt sich im Zusammenhang mit der Pariser Deklaration zur Harmonisierung der Entwicklungszusammenarbeit (PD) aus dem Jahr 2005. Österreich ist als Unterzeichner dieser Erklärung verpflichtet, die Ziele dieser Erklärung in seinem Wirkungsbereich umzusetzen. Dabei ist klar, dass vieles davon in den Partnerländern selbst erreicht werden muss. Entsprechende Aktionspläne der Unterzeichnerstaaten der PD (Geber- wie Partnerländer) sind vorzubereiten und abzuarbeiten. Gerade für den Bereich der Evaluierung galt und gilt die Erfahrung, dass häufig eine Mission der nächsten die Türklinden der besuchten Ministerien in den Partnerländern in die Hand drückt. Mehr gemeinsame Strategien und Programme auf Geberseite, neue Instrumente wie direkte Budgethilfe, bessere Angleichung der bilateralen Programme an die Entwicklungspläne der Partnerländer, die Ausrichtung an Resultaten und nicht an bereitgestellte Mittel sowie die gegenseitige Rechenschaftspflicht, stellen auch an die Evaluierung neue Herausforderungen. Die seit langem diskutierten, aber wegen der (noch) sehr hohen Transaktionskosten bisher nur verein-

zelt durchgeführten gemeinsamen Evaluierungen, müssen Standard werden. Die im Kapitel „Strategische Evaluierungen“ dargestellten praktischen Schwierigkeiten bei der Durchführung gemeinsamer Evaluierungen würden z. B. dann nicht mehr auftreten, wenn die beteiligten Geberländer nicht alle in unterschiedlichen Sektoren, sondern im Rahmen einer gemeinsamen direkten Budgethilfe in dem betreffenden Land tätig wären.

Eng verbunden mit mehr gemeinsamen Evaluierungen muss auch das Ziel bleiben, die Evaluierungskapazitäten in den Partnerländern selbst zu stärken. Was bereits für viele Projekte und Programme, sowie in ganz besonderer Weise für die direkte Budgethilfe gilt, hat auch für die Evaluierung Gültigkeit: nämlich die Bereitschaft, aber auch die konkreten Möglichkeiten der Partnerländer zu fördern, in diesem Bereich Verantwortung zu übernehmen.

Evaluation in der schweizerischen Entwicklungszusammenarbeit¹

1 Die Akteure und Kontextveränderungen

1.1 Einführung

Die Schweizerische Entwicklungszusammenarbeit (EZA) ist geprägt durch die beiden staatlichen Akteure DEZA (Direktion für Entwicklung und Zusammenarbeit) und SECO (Staatssekretariat für Wirtschaft) sowie einer vielfältigen NGO-Landschaft. Die DEZA ist das Kompetenzzentrum des Bundes für die Entwicklungszusammenarbeit mit dem Süden und Osten, die multilaterale Zusammenarbeit sowie für die Humanitäre Hilfe der Schweiz. Das SECO setzt wirtschaftliche und handelspolitische Massnahmen um (inkl. Budgethilfe). Durch Leistungsvereinbarungen werden Mandate und Beiträge an ausgewählte NGO, die Privatwirtschaft und an Universitäten vergeben. Die Abbildung 1 illustriert die Akteurslandschaft schematisch.

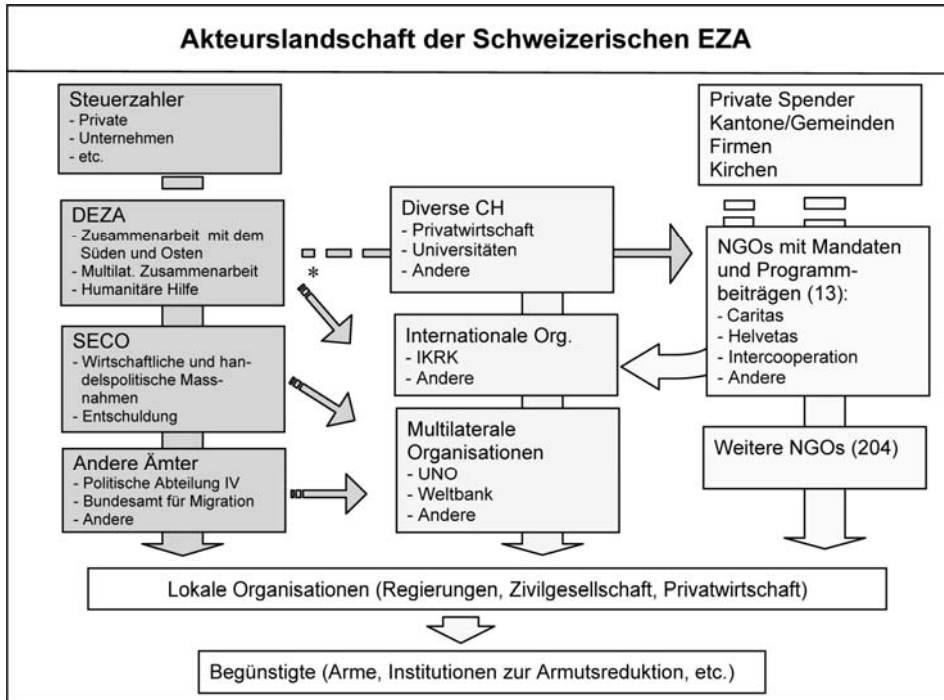
Diese Finanzflüsse und Kooperationsbeziehungen werden regelmässig einer Überprüfung unterzogen. Evaluiert werden vor allem vier Bereiche:

- die Effektivität und Wirksamkeit von einzelnen Massnahmen, die von überschaubaren Projekten bis hin zu komplexen Programmen, welche von mehreren Partnern kofinanziert werden, reichen;
- die Effektivität von Programmen auf Landesebene oder strategischen Themen;
- die Managementleistung (Performance) von Organisationen;
- die Zusammenarbeitsprozesse verschiedener Organisationseinheiten (z. B. im Rahmen der Dezentralisierung) oder zwischen verschiedenen Akteuren.

Grundsätzlich führen alle Akteure systematische Evaluationstätigkeiten aus. Im Rahmen dieses Beitrages wird auf die Evaluationspraxis von DEZA, SECO und die grösseren NGO fokussiert. Grundlage sind Dokumentenstudien und Interviews mit ausgewählten Organisationen.

¹ Der Beitrag beruht auch auf Interviews mit Mitarbeiterinnen und Mitarbeitern im Evaluationsbereich folgender Organisationen im Winter 2006/07: Caritas Schweiz, Luzern (Peter Büchler), DEZA (Peter Meier, Gerhard Siegfried, Samuel Wälty), Helvetas, Zürich (Remo Gesù), Internationales Komitee vom Roten Kreuz – IKRK, Genf (Christoph Luedi), Intercooperation, Bern (Felix von Surry), sowie Staatssekretariat für Wirtschaft, SECO (Odile Keller).

Abbildung 1: Schematische Übersicht über die wichtigsten Akteure der Schweizerischen EZA



* Das Beziehungsgeflecht ist komplex. Die Pfeile stellen nur die wichtigsten Beziehungen qualitativ dar.

1.2 Die wichtigsten Veränderungen im Schweizerischen Kontext

Die wichtigsten Veränderungen sind gekennzeichnet durch eine steigende Zahl von Akteuren in der EZA als Folge der Globalisierung und der thematischen Spezialisierung bei NGO. In diesem Bereich drängen sich neue ausländische Organisationen auf den lukrativen Schweizer Spendenmarkt und bei den staatlichen Akteuren haben andere Bundesämter ihre Tätigkeitsbereiche in der EZA ausgebaut (z. B. die Politische Abteilung IV des EDA zur Konfliktprävention und friedensfördernden Massnahmen). Entsprechend komplex und aufwendig sind die Koordinationsanforderungen im strategischen und operativen Bereich geworden (Gerster 2006).

Die entwicklungspolitische Debatte in den Medien wurde in letzter Zeit geprägt durch die Diskussion um die Wirksamkeit der schweizerischen EZA und fand auch ihren Niederschlag im Bericht der parlamentarischen Geschäftsprüfungskommission zur Kohärenz und strategischen Führung bei der DEZA. Diese konstatiert Lücken in der strategischen Führung und eine mangelhafte thematische und geografische Fokussierung (GPK-S 2006). Die Wirkungsdiskussion ist zwar nicht grundsätzlich neu. Gerade beim Parlament hat die Evaluation seit langem einen hohen Stellenwert (u.a. Basler 1984). Die verstärkte Debatte um die Wirkung der EZA hat in den letzten Jahren den Aufbau von Qualitätsmanagementsystemen und die systematische Einbettung von Evaluationen bei den meisten EZA-Organisationen gefördert.

2 Der Umfang der Evaluationsaktivitäten in der Schweizerischen EZA

Es wird geschätzt, dass pro Jahr zwischen 300 und 400 unabhängige und externe Evaluationen sowie externe Reviews durchgeführt werden. Die Hälfte im Auftrag von Bundesämtern und die andere Hälfte durch NGO sowie Stiftungen und Universitäten. Alle grösseren EZA Vorhaben werden während des Projektzyklus evaluiert. Dabei kann es sich um unabhängige Evaluationen, Mid-Term Reviews, externe Evaluationen, Selbstevaluationen, Wirkungs- oder Impact-Analysen oder andere Evaluationstypen handeln. Eher eine Ausnahme stellen in der Schweiz explizite externe ex-post Analysen von EZA Vorhaben dar (Fischer et al. 2005).

2.1 Evaluationstätigkeit der bilateralen staatlichen EZA

Die bilaterale staatliche EZA wird alle fünf Jahre durch das Development Assistance Committee der OECD einer sogenannten „Peer Review“ unterzogen (DAC 2000; 2005c). Als Antwort auf Empfehlungen der Review im Jahre 2000 haben die DEZA und das SECO ihre Evaluationssysteme überarbeitet. Der Bericht 2005 attestiert bei der DEZA Verbesserungen: Grössere Unabhängigkeit der Sektion Evaluation und Controlling (E + C) vom operationellen Management, alle E + C Mitarbeitenden sind Teil eines Netzwerks (inkl. SECO), welches sich regelmässig über die Erfahrungen austauscht. Die E + C Funktion umfasst jetzt drei Bereiche: Erstens das Controlling, welches relevante Informationen als Entscheidungsgrundlage für das Management sammelt; zweitens die unabhängigen Evaluationen, mit welchen sowohl Rechenschaft gegenüber Parlament und Steuerzahler geleistet wird als auch gemeinsame Lernprozesse gefördert werden; und drittens das Projektzyklusmanagement, mit welchem die gemeinsame Planung, Umsetzung, Monitoring und Evaluation von Vorhaben optimiert wird. Letzteres bestand schon vorher, wurde aber der internationalen Terminologie angepasst.

Die DEZA führte 2002 ein neues Klassifikationssystem für Evaluationen ein:

- Unabhängige Evaluationen (zwei bis vier pro Jahr) fokussieren auf übergeordnete strategische Fragen, werden vom Direktorium genehmigt und von der E + C Sektion in Auftrag gegeben. Diese werden in der Regel auf der Grundlage eines Offertenwettbewerbs an ausländische EvaluatorInnen vergeben. Der Evaluationsprozess wird von einer Kernlerngruppe begleitet, die aus internen und je nach Bedarf externen Mitarbeitenden besteht. Diese kommentiert das Konzept (Approach Paper) der Evaluation sowie den Schlussbericht. Sie verfasst eine eigene Stellungnahme zu den Empfehlungen der Evaluation.
- Externe Evaluationen (ca. 25 pro Jahr) werden von den Programmverantwortlichen an der Zentrale ausgelöst. Ziel ist es, die Projektansätze zu bewerten und mögliche Alternativen aufzuzeigen. Die externen Konsulenten werden bei grossen Vorhaben im Wettbewerbsverfahren rekrutiert. Oft ist jedoch fach- oder kontextspezifisches Wissen wichtig, sodass nur ein kleiner Expertenkreis in Frage kommt.
- Die externen Reviews und Selbstevaluationen (ca. 130 pro Jahr) sind im Verantwortungsbereich des operationellen Managements. Diese Evaluationen werden zuneh-

mend von lokalen Experten durchgeführt. Ziel ist es, im Rahmen des Projektzyklusmanagements die operationelle Tätigkeit zu verbessern.

- Jährlich werden rund zehn interne Audits von Programmen auf Landesebene in den Partnerländern durchgeführt. Diese fokussieren nicht nur auf Fragen der korrekten Buchhaltung und Einhaltung von Richtlinien, sondern versuchen auch die Ergebnisse in Bezug zu den Investitionen zu setzen.

Das SECO kategorisiert nach internen Reviews, externen Evaluationen und unabhängigen Evaluationen und führt rund 25 Evaluationen pro Jahr aus, wovon mehr als die Hälfte als gemeinsame Evaluationen (Joint Evaluations) gelten, da es sich um von mehreren Partnern finanzierte Vorhaben handelt (z. B. Budgethilfe etc.). Die Evaluationsteams sind entsprechend international zusammengesetzt und die fachliche Qualifikation der Konsulenten ist bei der Evaluation von wirtschaftspolitischen Massnahmen zentral. Lokale EvaluatorInnen sind selten. Unabhängige und externe Evaluationen werden ausgeschrieben und gewisse Evaluationen werden durch die Finanzierungspartner ausgeführt (z. B. die Weltbank), mit SECO Beteiligung bei der Definition der Pflichtenhefte und Vernehmlassung der Berichte.

DEZA und SECO haben in den letzten Jahren die resultatorientierte Planung von Projekten und Programmen zum Standard erklärt, unter anderem mittels der Projektplanungsmatrix (Logframe). Diese definiert Ziele sowie Indikatoren auf den Leistungs- und Wirkungsebenen eines Vorhabens (SECO 2005) und ist obligatorisch für alle Vorhaben mit mehr als einer Million Franken Budget. DEZA und SECO haben die diesbezügliche Ausbildung der MitarbeiterInnen forciert und haben neue Controlling-Instrumente für ihre mittelfristigen Strategien eingeführt, um strategisch relevante Informationen für Managemententscheidungen zu generieren. Die Evaluationsergebnisse fliessen systematisch ins strategische Controlling ein und die Erfahrungen werden in jährlichen Monitoring – Berichten zuhanden der Direktionen aufgearbeitet.

2.2 Evaluationstätigkeit der NGO

Die Evaluationstätigkeit wird wesentlich bestimmt von den Ansprüchen der privaten Spender oder vertraglichen Vereinbarungen mit den öffentlichen Auftraggebern. Für die Evaluationspraxis charakteristisch ist, dass offenbar die Spender weniger hohe Ansprüche an die Rechenschaftslegung stellen als institutionelle Geber, mindestens solange der Aufwand für Verwaltung und Mittelbeschaffung vernünftig (10-20 % der Gesamtausgaben) und der Mittelfluss zu den Begünstigten entsprechend gross ist. Die Emotionalität und Professionalität von Kampagnen sind hier oft entscheidender als eine detaillierte und selbstkritische Rechenschaftslegung.

Alle grösseren NGO führen interne und externe Evaluationen durch. Bei grösseren Vorhaben, welche von DEZA, SECO oder der Glückskette² finanziert werden, sind Mid-Term-Reviews und Evaluationen am Phasenende integrierte Bestandteile des Leistungsauftrages. Es wird geschätzt, dass bei circa 2'000 laufenden Vorhaben zwischen fünf und zehn

² Die Stiftung Glückskette ist die humanitäre Solidaritäts- und Sammelplattform der Schweiz.

Prozent derselben pro Jahr evaluiert werden (häufig Selbstevaluationen). Dies resultiert in rund hundert bis 200 Evaluationen pro Jahr. Da einige dieser Vorhaben bereits bei der bilateralen EZA gezählt wurden, dürfte es sich um netto sechzig bis 150 Evaluationen handeln. Auch viele NGO haben die Projektplanungsmatrix in den letzten Jahren zum Standard bei der Projektplanung erklärt.

2.3 Wer sind die EvaluatorInnen?

Die Evaluationsteams bestehen oft aus gemischten Teams, mit internationalen und lokalen Experten, die Expertinnen scheinen noch deutlich untervertreten. Die Expertise wird vorwiegend bei privaten Beratungsunternehmen rekrutiert, sowohl in der Schweiz als auch in den Partnerländern. Vertreter von Hochschulen sind eher selten. Die NGO greifen, unter anderem aus Kostengründen, vermehrt auf lokales Expertenwissen zurück. In Asien und Lateinamerika gibt es erfahrene EvaluatorInnen, wogegen in Afrika stärker auf auswärtige Expertise gesetzt wird.

Akademische Expertise von Hochschulen wird projektbezogen beigezogen, es existieren jedoch keine institutionalisierten Beziehungen zwischen Hochschulen und Schweizerischen EZA-Akteuren im Evaluationsbereich.

3 Verbindungen mit anderen Politikfeldern in der Schweiz und internationalen Entwicklungen

Die Verbindungen mit anderen Politikfeldern sind wenig ausgeprägt. Zwar besteht zur Umsetzung des Verfassungsartikels zur Überprüfung der Massnahmen des Bundes ein Netzwerk innerhalb der Verwaltung, doch der Austausch zwischen der staatlichen EZA mit anderen Politikbereichen ist informeller Art. Ein Austausch findet vor allem dann statt, wenn die EvaluatorInnen auch Gutachten in anderen Politikfeldern durchführen. Die zunehmende thematische Spezialisierung bei Evaluationen dürfte diesen Austausch jedoch limitieren.

Für die EZA-Akteure relevant sind die internationalen Plattformen zur Koordination und Standardisierung von Evaluationsprozessen und -methoden, wie zum Beispiel die Aktivitäten multilateraler Ausschüsse (z. B. des DAC). So wurde kürzlich eine standardisierte Methode zur Erfassung der Wirkung im Armutsbereich entwickelt (DAC 2006d).

Die UN Deklaration zur Erreichung Internationaler Entwicklungsziele im Jahre 2000 hat die Armutsreduktion sowohl bei der Gebergemeinschaft als auch den Entwicklungsländern wieder ins Zentrum der EZA gerückt. So soll der Anteil der Menschen, die Hunger leiden, zwischen 1990 und 2015 halbiert werden. Damit verbunden war die Diskussion um substanziell höhere EZA-Leistungen der Geber, um diese ehrgeizigen Ziele zu erreichen. Politisch war diese Forderung nur durchsetzbar, wenn gleichzeitig auch die entwicklungspolitische Wirksamkeit der Vorhaben signifikant gesteigert wird. Dies führte zur Paris Declaration (DAC 2005b), welche die Strukturen, Prozesse und Instrumente der Zusammenarbeit neu gestalten und harmonisieren will:

- durch Stärkung der Eigenverantwortlichkeit sollen die nationalen Strategien zur Armutsreduktion als zentraler Rahmen für Planung, Monitoring und Evaluation dienen;
- die EZA-Leistungen sollen an den Prioritäten des Partnerlandes ausgerichtet werden;
- die gemeinsame Planung, Durchführung und Evaluation von EZA Massnahmen, eine bessere Abstimmung der Instrumente sowie verbesserte Transparenz soll die Transaktionskosten reduzieren;
- resultatorientiertes Management soll auf die Erreichung der gewünschten Wirkungen fokussieren;
- durch gemeinsame Verantwortung von Geber- und Partnerländern gegenüber ihren jeweiligen Bürgern und Parlamenten soll die Akzeptanz der EZA gestärkt werden.

In diesem Rahmen ist in den letzten Jahren bei den staatlichen wie nicht-staatlichen EZA Akteuren eine stärkere Ausrichtung auf programmorientierte Vorhaben festzustellen, inklusive deren gemeinsame Evaluation. Die Vorgaben sind ehrgeizig. Bis ins Jahre 2010 sollen zwei Drittel der Mittel über programmorientierte Ansätze geleistet werden und vierzig Prozent der Missionen gemeinsam durchgeführt werden. Dies betrifft auch Evaluationsmissionen.

4 Die wichtigsten Veränderungen in der Evaluationspraxis seit 2000

4.1 Methodische Aspekte

Grundsätzlich zeichnet sich die Evaluationstätigkeit in der Schweizer EZA durch eine pragmatische Flexibilität aus und die Pflichtenhefte und Prozesse richten sich stark nach den Bedürfnissen des Kontextes. Die Evaluationskriterien des DAC spielen in den meisten Organisationen eine zentrale Rolle bei der Formulierung der Evaluationsfragen. Die Kriterien sind: Relevanz (Übereinstimmung mit dem Bedarf der Zielgruppe), Effektivität (Ausmass der Zielerreichung), Effizienz (Kosten-Wirksamkeit), langfristige positive und negative Wirkungen (Impact) und Nachhaltigkeit (Andauern der Nutzen) (DAC 1998). Das Kriterium der Kohärenz (aufeinander abgestimmte Zielsetzungen unterschiedlicher Politikbereiche und Geber) wurde in den letzten Jahren wichtiger.

Externe und interne Evaluationen werden als integrierte Bestandteile des Qualitäts- und Wissensmanagements in den Organisationen gesehen. Die Anforderungen, welche im Rahmen von Zertifizierungsbestrebungen notwendig wurden, betrafen die Neugestaltung von organisationsinternen Zuständigkeiten und Prozessen. Die Zertifizierungen hatten überraschend wenig Auswirkungen auf die etablierte Evaluationspraxis von EZA Vorhaben (z. B. beim SECO ISO 9000, Caritas ISO 9000 und das Modell des European Forum for Quality Management) indem die Zahl und Art der Evaluationsvorhaben nicht tangiert wurde.

Methodisch werden vor allem qualitative, eher selten quantitative Methoden der empirischen Sozialforschung angewandt. Die Methoden werden kontextspezifisch konzipiert. Die Zeitknappheit oder der Pilotcharakter von vielen Vorhaben führt zu kleinen Stichproben bei den meisten Evaluationen. Damit erhöht sich die Gefahr von wenig robusten Evidenzen, welches die Generalisierbarkeit von Evaluationsergebnissen erschwert. Der Einsatz

von Evaluationen wird im Rahmen des Wissens- und Qualitätsmanagements situiert. Die Evaluationen versuchen in der Regel die beiden Bedürfnisse nach Rechenschaftslegung einerseits und des gemeinsamen Lernens andererseits durch eine geeignete Methodenwahl und Prozessbegleitung zu optimieren. Selbst-Evaluationen spielen eine wichtige Rolle bei allen Akteuren und werden oft nachfolgenden externen Evaluationen vorgeschaltet. Dies schafft eine bessere Evaluationsbasis und die Projektmitarbeiter müssen sich vorgängig (selbstkritisch) mit der Leistung und Wirkung des Vorhabens befassen.

Substantielle methodische Schwierigkeiten (und empirische Defizite) bestehen nach wie vor bei der systematischen Wirkungserfassung, vor allem auf der Ebene von sektoralen oder länderbezogenen Programmen. Die Bedeutung von Wirkungsanalysen wird allseitig anerkannt, und zum Teil auch in Evaluationsdesigns explizit umgesetzt. Methodisch sind dies jedoch anspruchsvolle und tendenziell aufwendige Studien. Die Durchführbarkeit von sogenannten robusten Analysen mit Kontrollgruppen wird für viele der meist mehrschichtigen Vorhaben der EZA skeptisch beurteilt. Auch die gemeinsame Evaluation der Budgethilfe illustriert die methodischen Probleme bei der Konstruktion einer Vergleichsbasis (IDD 2006: 15).

Durch die systematischere Anwendung von resultat-orientierten Methoden bei der Planung von Vorhaben auf der Projekt-, Programm- und Landesebene wurde die Evaluierbarkeit von Vorhaben in den letzten Jahren stark verbessert. Die Parlamentarische Verwaltungskontrolle sieht jedoch weiterhin Verbesserungsbedarf und fordert eine rigorosere Anwendung von ergebnisbezogenen Zielsetzungen für Kooperationsstrategien der Schweiz (GPK-S 2006).

Grössere Evaluationen werden in der Regel bei allen Organisationen ausgeschrieben, kleinere Vorhaben unter Franken 50'000 werden nach pragmatischen Kriterien freihändig vergeben. Es gibt heute mehr gemeinsame Evaluationen als früher (z. B. TEC 2006), solche sind aber bei NGO die Ausnahme. Nur das SECO mit seinen ko-finanzierten Vorhaben mit internationalen Finanzinstitutionen setzt gemeinsame Evaluationen auf breiter Basis ein. Diese dürften auch für DEZA und NGO wichtiger werden, da die Zahl von gemeinsam finanzierten Vorhaben zunimmt. Allerdings zeigt die bisherige Erfahrung mit gemeinsamen Evaluationen, dass die Steuerung solcher Vorhaben mit vielen Beteiligten entsprechend komplex wird. Gemischte Teams gehören heute zum Standard und in Zukunft werden die Vertreter der Partnerländer eine stärkere Rolle (auch als Teamleiter) in den Evaluationen übernehmen.

4.2 Fallbeispiele innovativer Evaluationstätigkeit

In der Folge soll beispielhaft auf ausgewählte innovative Evaluationstätigkeiten eingegangen werden – ohne Anspruch auf Repräsentativität:

DEZA: Erstens werden für jede unabhängige Evaluation klar definierte Prozessschritte ausgeführt: Evaluationskonzept (Approach Paper), Konstituierung einer begleitenden Kernlerngruppe, Rekrutierung von Experten, Durchführung, Berichterstattung, Debriefing, Beschluss bezüglich Umsetzung durch das Management (Management Response) und Veröf-

fentlichung der Ergebnisse. Ein wichtiges Element ist die Einsetzung und Betreuung der Lerngruppe aus beteiligten Akteuren innerhalb und ausserhalb der DEZA, bestehend aus acht bis zwölf Mitgliedern. Dies erleichtert eine breite Akzeptanz sowohl des Ansatzes als auch der Ergebnisse und vereinfacht deren Umsetzung. Das Management dieses Lernprozesses ist allerdings aufwendig. Zur weiteren Verbesserung der Umsetzung von Evaluationsergebnissen hat die DEZA zehn Leitsätze für Auftraggeber von Evaluationen formuliert (DEZA 2006). Ein zweites Beispiel betrifft die unabhängige Evaluation des Leistungsnachweises (Performance) der DEZA bezüglich der Ermächtigung (Empowerment) der begünstigten dezentralen Verwaltungseinheiten und lokalen Organisationen vor Ort. Das Evaluationsteam kam aus Indien und hat Aktivitäten in Burkina Faso und Bolivien als Fallbeispiele bearbeitet. Das setzte besondere Anforderungen an die sprachlichen und inter-kulturellen Kompetenzen. Durch diesen expliziten Perspektivenwechsel konnte die übliche „Nord-Sicht“ durch Experten aus den Geberländern ausgeschaltet werden (DEZA 2007).

SECO: Je nach zu evaluierendem Instrument kommt eine breite Palette von Evaluationsmethoden in Frage. Im Rahmen des Abschlusses von Projekten fällt den Internen Reviews (Selbstevaluationen) eine wichtige Rolle zu. Im Vergleich zu den anderen EZA-Akteuren spielt der Projektabschluss eine wichtige Rolle bei der Erfahrungsauswertung. Die Abschlussnotizen der Vorhaben werden von der Projektleitung verfasst und sind damit eine ex-post Selbst-Evaluation. Diese dokumentiert das Projekt, den Prozess der Projektumsetzung, die Zielerreichung, die Wirkungen sowie die gemachten Erfahrungen und ist für das interne Wissensmanagement bestimmt. Stellvertretend für die zunehmende Rolle von gemeinsamen Evaluationen kann jene der Budgethilfe in sieben Ländern erwähnt werden (IDD 2006), welche vom SECO unterstützt wurde. Methodologisch wurde die Analyse auf den oben erwähnten DAC Kriterien und dem Konzept der Wirkungskette des Logframe entwickelt: die Erfüllung von Vorbedingungen (Konditionen) löst Inputs aus (hier Finanzflüsse), diese führen zu unmittelbare Effekten oder Aktivitäten (zusätzliche externe Ressourcen für das Staatsbudget), die in Veränderungen bei den Outputs resultieren (z. B. bessere Fiskaldisziplin), was wiederum zu Veränderungen auf der unmittelbaren Wirkungsebene führt (z. B. mehr Mittel für öffentliche Dienstleistungen bei lokalen Behörden). Als langfristige Wirkung resultieren weniger Arme. Die Resultate dieser Wirkungsanalysen zeigten, dass die Wirkungserfassung und Interpretation schwierig ist, da eine Vergleichsbasis fehlt, respektive (theoretisch) konstruiert werden muss.

Caritas: In den letzten Jahren hat Caritas die Qualitätsmanagement- und Evaluationssysteme zusammengeführt und sich mit verschiedenen Gütesiegeln zertifiziert. Rund achtzig Prozent der Evaluationstätigkeit findet im Rahmen der „plan-do-check-act“ Projektzyklusphilosophie statt. Die eigentliche Evaluationstätigkeit in Form von Selbstevaluationen und externen Evaluationen von Vorhaben macht zwanzig Prozent des Ressourceneinsatzes aus (finanziell und personell). Für externe Evaluationen werden vor allem lokale Experten über das lokale Beziehungsnetz rekrutiert. Die Erfahrungen in Lateinamerika und Asien sind sehr gut, in Afrika bestehen gewisse Vorbehalte gegenüber reinen lokalen Evaluationsteams.

Helvetas: Zusätzlich zu den im Projektzyklus eingebetteten Selbst- und externen Evaluationen hat Helvetas in den letzten Jahren versucht, systematisch Wirkungsanalysen in landwirtschaftlichen Projekten in fünf Ländern einzubauen (mit der Methode des Impact Monitoring & Assessments, von Herweg/Steiner 2002). Unterstützt von externen Konsulenten wurden unter Einbezug der Zielbevölkerung Wirkungsindikatoren identifiziert und Hypothesen erarbeitet. Um eine Vergleichsbasis zu erhalten, wurde der Ist-Zustand erfasst und danach wurden Veränderungen in regelmässigen Intervallen gemessen. Die Auswertung dieser Erfahrungen hat gezeigt, dass die Resultate solcher Wirkungsanalysen zwar interessant und aussagekräftig sind, aber auch einen grossen Aufwand bedeuten, den Einsatz von (teuren) Konsulenten erfordern und die Projektmitarbeiter zusätzlich zu ihren Aufgaben stark belasten. Die Erfassung von Wirkungen ist angesichts komplexer Kausalitäten methodisch anspruchsvoll. Eine weitere Schwierigkeit besteht darin, die erfassten positiven und allenfalls negativen Wirkungen richtig zu interpretieren, da Kontrollregionen oder -gruppen oft nicht zur Verfügung stehen (etwa weil diese Gruppen durch Programme anderer Geber tangiert sind).

Intercooperation: Sämtliche Evaluationen sind Teil der Leistungsvereinbarung mit den Auftraggebern (v. a. DEZA). Entsprechend richten sich die Evaluationen an die Vorgaben der Auftraggeber in der Zentrale oder vor Ort. Bei der Auswahl der EvaluatorenInnen besteht eine gewisse Mitsprachemöglichkeit. Zwei von drei Evaluationen werden von einer Selbst-Evaluation eingeleitet. In der Regel werden die Prozesse der Selbst-Evaluation durch einen externen Moderator, einem internationalen oder lokalen Konsulenten, begleitet. Dies garantiert eine gewisse Objektivität des Prozesses, eine ausgewogene Partizipation der Beteiligten und die Mediation in Konfliktfällen. Die Dokumentation ist bei Selbstevaluationen oft eine Schwäche und ein externer Moderator kann dies sicherstellen.

IKRK: Externe Evaluationen sind erst seit 1998 eine offizielle Funktion bei den humanitären Projekten des IKRK. Allerdings wurden bereits vorher und auch heute noch sogenannte Reviews von Operationen von den verschiedenen operationellen Abteilungen an externe Gutachter mandatiert. Die Rekrutierung für externe Evaluationen erfolgt über internationale Ausschreibungen und die Evaluationskriterien folgen dem DAC Standard und einem eigenen Qualitätskompass „urgence, rehabilitation, développement“. Die Berichte werden grundsätzlich nicht veröffentlicht, jedoch gehen die Zusammenfassungen an die jeweiligen IKRK-Geber. Bei einer allfälligen Veröffentlichung der gesamten Berichte wird befürchtet, dass diese weniger kritisch ausfallen. Die Evaluationsthemen werden heute aufgrund grundsätzlicher Überlegungen, auch im politischen Kontext, ausgewählt. Zum Teil wird auch das IKRK zum Gegenstand von separaten Evaluationen grosser Geber (z. B. der EU). Eine wichtige vertrauensbildende Massnahme mit den Geldgebern war die Einführung von gemeinsamen Reviews, bei welchen die wichtigsten Gebervvertreter in die Evaluation eingebunden sind.

5 Bilanz und Schlussfolgerungen

Die Entwicklungszusammenarbeit bleibt ein Politikbereich mit einer vergleichsweise hohen Evaluationsdichte (300 bis 400 pro Jahr). Evaluationen sind integrierte Bestandteile des Qualitätsmanagements vieler Organisationen geworden.

Alle EZA-Akteure stehen unter einem Legitimationsdruck und die Leistungs- und Wirkungsmessung erhält eine entsprechend grosse Bedeutung. Die Auswahl von aussagekräftigen Indikatoren, welche über die unmittelbare Input- und Outputebene hinausgehen, ist sowohl für die staatlichen als auch die nicht-staatlichen Akteure eine Herausforderung. Eine Fokussierung der Portfolios in geografischer und thematischer Hinsicht erleichtert die Herleitung und Aggregation solcher Leistungsindikatoren. Zudem könnte ein Austausch von „Good Practice“ – Erfahrung unter vergleichbaren Organisationen nützlich sein.

Die systematischere Erfassung der langfristigen Wirkung von Vorhaben (vgl. auch Empfehlungen der DAC Peer Reviews; DAC 2000; 2005c) ist zu verstärken. Diese Impact-Analysen sollten demnach systematischer auf Zufallsstichproben bei der Zielbevölkerung *und* einem Vergleich mit Kontrollgruppen basieren. Umstritten scheint, inwiefern es Sinn macht, solche rigorosen Evaluationsansätze auf die unterschiedlichsten Vorhaben anzuwenden. Wichtig für alle Arten von Wirkungsanalysen, welche den Charakter von Forschungsvorhaben haben können, bleibt eine möglichst grosse Unabhängigkeit der EvaluatorenInnen, um die Kollusionsgefahr zwischen ihnen und dem Projektmanagement zu reduzieren (vgl. Michaelowa/Borrmann 2005b).

Angesichts der grossen Zahl von Evaluationsvorhaben in der schweizerischen EZA, und der organisationsspezifischen Auswahl der Vorhaben, könnte eine bessere Koordination bei der Auswahl der Evaluationen einen Mehrwert bringen (z. B. thematische Vertiefung). Von den zahlreichen Evaluationen und Reviews werden zudem nur ein kleiner Prozentsatz systematisch der Öffentlichkeit zugänglich gemacht, zum Beispiel die unabhängigen Evaluationen von DEZA und SECO. Alle anderen Berichte werden nicht systematisch publiziert. Auch wenn die Qualität der Berichte unterschiedlich ist, wäre in einer generellen Offenlegung dieser Evaluationsberichte ein verstärkter methodischer und inhaltlicher Erfahrungsaustausch zu erwarten.

Damit in Zukunft deutlich mehr gemeinsame Evaluationen stattfinden können, benötigt es in den nächsten Jahren eine verstärkte Zusammenarbeit der Geber bei der Konzeption von Vorhaben und eine ausgeprägtere Arbeitsteilung (geografisch und thematisch). Dies ist eine Voraussetzung für programmorientierte Ansätze, welche dann gemeinsam evaluiert werden können. Mit der Umsetzung der Paris Declaration werden vielfältige Entscheidungskompetenzen zu den Partnern und Begünstigten in die Empfängerländer verlagert. Dabei gilt es deren Willen und Möglichkeiten zu stärken, damit Evaluationsresultate zur Formulierung besserer Vorhaben und Politiken effektiv genutzt werden. Dazu müssen die lokalen Kapazitäten für Monitoring und Evaluation von EZA Vorhaben gestärkt werden (u. a. Ausbildung von EvaluatorenInnen, qualitativ bessere statistische Grundlagen).

Ländervergleich: Evaluierung in der Entwicklungszusammenarbeit

1 Struktur und Ausrichtung der Entwicklungszusammenarbeit

Wie aus den vorangegangenen Artikeln ersichtlich wird, sind der strukturelle Aufbau und die Ausrichtung der Entwicklungszusammenarbeit (EZ) in den Ländern Deutschland, Österreich und Schweiz recht ähnlich. Die EZ wird in allen drei Ländern zentral durch den Staat und das zuständige Ministerium für bi- und multilaterale EZ gesteuert und finanziert. Die jeweilige Regierung ist somit verantwortlich für die strategische Ausrichtung und Führung und legt geographische sowie thematische Schwerpunkte in Aktionsprogrammen (Jahres- oder Mehrjahresprogramme) fest. Die letztendliche Umsetzung der EZ wird meist von staatlichen Durchführungsorganisationen (DO) mit Koordinationsbüros in den Partnerländern und zahlreichen NGO übernommen.

Die jeweilige Entwicklungspolitik und somit die strategische Ausrichtung der EZ in allen drei Ländern wird stark durch internationale Gemeinschaftsziele und -vereinbarungen beeinflusst. Anzuführen sind hier vor allem die Millenniumserklärung der UN aus dem Jahre 2000 und die Paris-Deklaration von 2005. Beide Vereinbarungen rücken die Themen „Frieden, Sicherheit und Abrüstung“, „Entwicklung und Armutsbeseitigung“ sowie „Schutz der gemeinsamen Umwelt“ und „Menschenrechte, Demokratie und gute Regierungsführung“ ins Zentrum der EZ, aus denen später acht Millenniumsentwicklungsziele (Millennium Development Goals; MDG)¹ abgeleitet wurden. Zudem wird eine internationale Harmonisierung der Strukturen, Prozesse und Instrumente der EZ gefordert.

2 Internationale Entwicklungen im Bereich EZ-Evaluationen

Die Millenniums-Erklärung, die MDG sowie die Paris-Deklaration sind auch für EZ-Evaluationen von Relevanz, da sie, vor allem die MDG, einen gemeinsamen, überprüfbaren Bezugsrahmen für die internationale Entwicklungszusammenarbeit darstellen. Hervorzu-

¹ Die MDG sollen bis zum Jahr 2015 durch die internationale Gemeinschaft erreicht werden. Sie umfassen die Reduzierung des Anteils der armen Weltbevölkerung (MDG 1), die Sicherung von Grundbildung (MDG 2), die Förderung der Gleichstellung der Geschlechter und der Rechte der Frauen (MDG 3). Außerdem die Verringerung der Kindersterblichkeit (MDG 4), die Verbesserung der Gesundheit von Müttern (MDG 5) und die Bekämpfung von Krankheiten wie HIV/AIDS und Malaria (MDG 6). Schließlich noch die Verbesserung des Umweltschutzes (MDG 7) und der Aufbau von weltweiten Entwicklungspartnerschaften (MDG 8); vgl. VN-Millenniums-Projekt 2005.

heben sind folgende drei Prinzipien, die das Evaluierungsgeschehen international und somit auch in den drei Ländern stark beeinflussen:

1. Orientierung auf Wirkungen der EZ und nicht mehr auf Leistungen: In diesem Kontext ist es die Aufgabe von Evaluationen, zuverlässige Ergebnisse über die Wirkungen von Projekten zur Verfügung zu stellen, die Rechenschaftslegung aber auch Lernen ermöglichen. Die Instrumente wirkungsorientiertes Monitoring (als Steuerungsinstrument zur Ziel- und Wirkungserreichung) und Evaluation (zum Nachweis von Wirkungen) werden zu integralen Bestandteilen der EZ. Besonders das Thema Nachweis von Wirkungen durch „(rigorous) impact evaluations“ (IE) wird intensiv verfolgt.²
2. Harmonisierung und Kohärenzgebot: Durch neue EZ-Konzepte und -Instrumente wie Budget-, Sektor-, Programm- sowie andere Gemeinschaftsfinanzierungen müssen diese auch gemeinschaftlich evaluiert werden. Notwendige internationale Gemeinschaftsevaluationen nehmen entsprechend zu.
3. Gemeinsame Rechenschaftspflicht: Die gemeinsame Verantwortung von Geber- und Partnerländern impliziert, dass nicht nur Gemeinschaftsevaluationen mit anderen Geberorganisationen notwendig sind, sondern die Überprüfung der Wirkungen von Maßnahmen durch Geber- und Partnerländer gemeinschaftlich durchgeführt wird. Dies impliziert das Vorhandensein entsprechender Kapazitäten vor Ort. Da in den Partnerländern das notwendige Wissen und die Erfahrungen mit Programmevaluationen meist fehlen, ist Capacity Building der Partner in Grundlagen von Wirkungsevaluation notwendig.

Diese Prinzipien spiegeln sich in den durch das bilaterale Forum für Geberorganisationen, die Organization for Economic Cooperation and Development (OECD), speziell das Development Assistance Committee (DAC), veröffentlichten Qualitätsstandards, den Prinzipien der Evaluation in der EZ, den Evaluationskriterien für EZ sowie den Orientierungslinien wider (vgl. DAC 1998; 2006a; 2008). Die Berücksichtigung dieser Prinzipien dient im internationalen Vergleich als Bewertungskriterium der Qualität der jeweiligen EZ-Evaluationsaktivitäten.³ In Deutschland ist die Bewertung der Qualität der EZ-Evaluationsaktivitäten, der Institutionalisierungsgrad und notwendige strukturelle Rahmenbedingungen, welche die tatsächliche Nutzung von Evaluationsergebnissen beeinflussen, seit Ende der 1990er Jahre ein immer wiederkehrendes Thema mit dem sich die Wissenschaft beschäftigt. So untersucht das Hamburgische Welt-Wirtschafts Archiv (HWWA) Institut bereits 1999 und 2001 sowie aktuell laufend die Praxis der Erfolgskontrolle in der deutschen EZ (vgl. Borrmann et al. 1999; 2001; Michaelowa/Borrmann 2005a).

² So wurde 2006 ein Zusammenschluss der Netzwerke von EZ-Organisationen sowie multilateraler Entwicklungsbanken gegründet (NONIE – Network of Networks Impact Evaluation Initiative), um die Effektivität der EZ zu verbessern, indem nützliche und relevante, qualitativ hochwertige IE vorangetrieben werden.

³ Vgl. dazu die DAC Peer Reviews (www.oecd.org/dac/peerreviews [4.4.2008]).

3 EZ-Evaluation in den Ländern Deutschland, Österreich und Schweiz

Um diesen Prinzipien so wie auch dem internationalen Qualitätsanspruch an EZ-Evaluation gerecht zu werden, ist die Institutionalisierung von Evaluation in der EZ in den letzten Jahren in den drei Ländern stark ausgebaut worden (Übersicht in Tabelle 1):

- Eine steigende Anzahl von etablierten *Evaluationsabteilungen*, Sektionen oder Einheiten in den zuständigen Ministerien sowie staatlichen und nicht-staatlichen DO ist zu verzeichnen. So haben in Deutschland das BMZ (seit den 1990er Jahren) und die staatlichen DO (nach 2000), in Österreich das BMEIA und die ADA (seit 2004) sowie in der Schweiz die DEZA und das SECO (bereits in den 1980er Jahren) eigene Evaluationseinheiten eingerichtet; dies gilt auch für die Ebene der NGO. Diese Evaluationseinheiten bzw. –beauftragte steuern unabhängige, externe sowie interne Evaluationen, deren Durchführung auf einheitlichen Leitlinien, Standards und Kriterien zur Evaluation beruhen.
- Diese *Leitlinien, Standards und Kriterien zur Evaluation* sind DAC-konform, das heißt sie orientieren sich an den OECD-DAC-Prinzipien (Unparteilichkeit und Unabhängigkeit, Glaubwürdigkeit, Partizipation, Nützlichkeit, Transparenz und Verbreitung; DAC 1998) und DAC-Evaluationskriterien (Relevanz, Effektivität, Wirkung, Effizienz und Nachhaltigkeit; DAC 2008). Resultat ist, dass Evaluationen in der EZ in den drei Ländern vermehrt von unabhängigen Evaluatoren (Unparteilichkeit und Unabhängigkeit) in Zusammenarbeit mit lokalen Experten (Partizipation) durchgeführt werden. Zudem werden an den Kontext angepasste, adäquate qualitative und quantitative Methoden zur Generierung der Evaluationsergebnisse eingesetzt (Glaubwürdigkeit). Die auf diese Weise generierten Evaluationsergebnisse werden zur Steigerung der Nutzung nicht nur mit dem Auftraggeber diskutiert und fließen ins strategische Management ein, sondern werden auch vor Ort präsentiert und zumindest Kurzberichte in der Amtssprache verfaßt (Nützlichkeit). Überdies werden die Evaluationsberichte auch vermehrt der Öffentlichkeit zugänglich gemacht (Transparenz und Verbreitung). Zusätzlich werden weitere, projektübergreifende Fragestellungen, die sich an den MDG orientieren, untersucht. So ist die Gleichstellung von Frauen und Männer ebenso Thema wie Umwelt, nachhaltige Entwicklung als auch Armutsreduzierung. Daneben sind strategische Themen wie Managementleistungen von Organisationen und Zusammenarbeitsprozesse verschiedener Organisationseinheiten Kriterien für die Evaluation in der EZ.
- Ein breites Angebot an *Aus- und Fortbildungen* für die Evaluation im Kontext der EZ demonstriert ebenfalls den Institutionalisierungsgrad. In allen drei Ländern kann in den letzten Jahren eine steigende Nachfrage und auch ein zunehmendes Angebot an Weiterbildungen beobachtet werden. In Deutschland existiert hierzu das am breitesten gefächerte Angebot. Seit 2003 wird eine Fortbildung für Evaluatoren im Bereich der Entwicklungszusammenarbeit (FEEZ) vom Centrum für Evaluation (Ceval), Universität des Saarlandes, in Kooperation mit der Arbeitsgemeinschaft entwicklungspolitischer Gutachter (AGEG) angeboten. Außerdem existiert seit 2004 ein Masterstudiengang Evaluation an der Universität des Saarlandes mit einem Schwerpunkt in EZ. Daneben gibt es weitere EZ-spezifische Studiengänge, die das Thema Evaluation eben-

falls in ihrem Kurrikulum aufgreifen. So thematisiert in Deutschland das Seminar für Ländliche Entwicklung in Berlin das Thema Untersuchung von Wirkungen in der EZ, das Deutsche Institut für Entwicklungspolitik (DIE) in ihrem Postgraduierten-Programm für Entwicklungszusammenarbeit die Planung, Durchführung und Evaluierung von Projekten. In der Schweiz an der ETH im Nachdiplomstudium für Entwicklungsländer (NADEL) wird das Thema Evaluation ebenfalls behandelt.

Tabelle 1: Institutionalierungsgrad der EZ-Evaluation in Deutschland, Österreich und Schweiz

| Institutionalisierung | Deutschland | Österreich | Schweiz |
|---|--|------------|-----------------------|
| Evaluationseinheit/Beauftragte | ja | | |
| Zentrales Evaluationsprogramm | ja | | nein |
| Leitlinien für die EZ-Evaluation | Ja, basierend auf DAC | | |
| Evaluationskriterien | Relevanz, Effektivität, Wirkung, Effizienz, Nachhaltigkeit | | |
| Weitere Projektübergreifende Fragestellungen | Ja, basierend auf MDG | | |
| Aus- und Fortbildungsangebote | Aus- und Fortbildung vorhanden | nein | Fortbildung vorhanden |

Die aufgezeigten Institutionalisierungsbestrebungen der EZ-Evaluation führen zu einer steigenden Anzahl von systematischen Evaluationen in den drei Ländern. So gehören strategische Evaluationen, Projekt-/Programm-Evaluationen und Gemeinsame Evaluationen (jeweils extern und intern) inzwischen zum Standard (vgl. Tabelle 2)⁴.

- *Strategische Evaluationen* umfassen Landesprogramme, Sektoren, Themen und Instrumente und werden durch den jeweiligen Bund koordiniert und an externe Experten per Ausschreibungsverfahren vergeben. Im Ländervergleich wird deutlich, dass Deutschland zwar eine größere Zahl von strategischen Evaluationen durchführt, in Relation zum veranschlagten Budget für EZ wird allerdings ersichtlich, dass die Schweiz und Österreich mit einem wesentlich geringerem EZ-Budget einen deutlich höheren Anteil für strategische Evaluationen verwenden.
- Zusätzlich zu den strategischen Evaluationen werden in den drei Ländern zahlreiche *Projekt/Programmevaluationen* durchgeführt. Koordiniert werden diese durch die Durchführungsorganisationen selbst, die Evaluationsdurchführung wird im Fall von externen Evaluationen meist durch ein Ausschreibungsverfahren an externe EvaluatorInnen oder aber im Fall von Selbstevaluation an interne Projektmitarbeiterinnen vergeben. Im Ländervergleich wird deutlich, dass in Deutschland die Ex-post Evaluation zum Evaluationsrepertoire dazugehört, in der Schweiz und Österreich ist dies noch nicht der Fall. Darüber hinaus werden in allen Ländern Selbst- und interne Evaluierungen durchgeführt.
- Abgesehen von strategischen Evaluationen und Projekt-/Programm-Evaluationen finden immer mehr *gemeinsame Evaluationen* mehrerer Geberländer statt. Diese umfassen

⁴ Im Detail vgl. die Beiträge von Michaela Zintl, Anton Mair und Dieter Zürcher in diesem Band.

meist Länderprogramm- oder Sektorevaluationen. Beispielsweise im Bereich Programmevaluation werden Geber gemeinsam aktiv, so auch Deutschland, Österreich und die Schweiz. Österreich führte zum Beispiel bereits gemeinsam mit der Schweiz die Landesprogrammevaluation Albanien und gemeinsam mit Deutschland die Landesprogrammevaluation Burkina Faso durch. Deutschland führt zurzeit gemeinsam mit der Dänischen Internationalen Entwicklungsagentur (DANIDA) und dem tansanischen Finanzministerium eine Evaluation des tansanischen Gesundheitssektors durch.

Tabelle 2: Systematische EZ-Evaluationen in den Ländern Deutschland, Österreich und Schweiz

| | Deutschland | Österreich | Schweiz |
|--|--|--|------------------------|
| Durchführung von Evaluationen | alle Projekte der DO, viele von größeren NGO | alle Projekte der DO, viele von größeren NGO | alle größeren Projekte |
| Unabhängige, externe Evaluationen | | | |
| <i>Strategische Evaluationen</i> | 7 | Bis zu 4 | 2-4 |
| <i>Projekt/Programm Evaluationen</i> | über 150 von DO, weitere von großen NGO | 30 | 25 DEZA, 25 SECO |
| - laufend | ja | | |
| - Abschluss | ja | | |
| - ex-post | ja | kaum | kaum |
| <i>Projekt/Programm-Audits/Reviews</i> | GTZ: 100 Jährlich | nein | ca. 60 Jährlich |
| Selbst- und interne Evaluationen | | | |
| <i>Projekt/Programm-Evaluationen</i> | alle Projekte | | |
| - laufend | alle | | |
| - Abschluss | alle | | |
| <i>Projekt/Programm-Audits</i> | nein | | ca. 10 in der DEZA |
| Gemeinsame Evaluationen | ja | | |

Das Thema *Nachweis von Wirkungen* durch (rigorous) Impact Evaluations (IE) wird ebenfalls verfolgt. Dabei stecken die meisten EZ-Organisationen jedoch noch in den Kinderschuhen. Die Wissenschaft in Deutschland hat das Thema bereits seit den 1990er Jahre aufgegriffen, unter anderem in Evaluationsstudien zum Thema Wirkungen der deutschen Berufsbildungszusammenarbeit (vgl. Stockmann 1992; Stockmann et al. 2000) und Messung von Wirkungen und Nachhaltigkeit von Projekten der Entwicklungszusammenarbeit (Stockmann 1996; Caspari et al. 2000; Caspari 2004). In diesen Evaluationsstudien gehörten bereits damals die heute diskutierten „anspruchsvollen Methoden zur Wirkungsmessung“, wie zum Beispiel Multimethodenansatz, quasi-experimentelles Design (Vorher/Nachher Messung, Vergleichs- und Kontrollgruppendesign, Zufallsstichproben) sowie hypothesengelei-

tete Ursache-Wirkungsuntersuchung zum Standard. Neue Veröffentlichungen aus der Wissenschaft zu Evaluationsforschung greifen diese Methoden ebenfalls auf (vgl. Stockmann 2006b; 2007). Aktuell ist in Deutschland das BMZ an dieser aus der Wissenschaft kommenden Diskussion aktiv beteiligt und hat zum Beispiel das Ceval mit der Forschungsstudie „Wirkungsevaluierungen – Zum aktuellen internationalen Stand der Diskussion und dessen Relevanz für deutsche EZ-Evaluierungen“ beauftragt, um die in der deutschen EZ-Evaluierung angewandten Methoden auf den Stand des sich international herausbildenden, neuen Anspruchsniveaus zu bringen, Wirkungen fundiert zu belegen. Darüber hinaus plant die GTZ, im Rahmen ihrer jährlich dreißig Evaluationen, dem Thema Wirkungsmessung ein höheres Gewicht zu geben, zunächst einmal in einem Test, in dem eine kleine Anzahl ausgewählter Vorhaben mit anspruchsvolleren Methoden durchgeführt werden sollen.

Tabelle 3: EZ-Wirkungsevaluationen in den Ländern Deutschland, Österreich und Schweiz

| Durchführung von Wirkungsevaluationen | Deutschland | Österreich | Schweiz |
|---------------------------------------|-------------|------------|---------|
| Wissenschaft im Rahmen von EZ | ja | kaum | |
| EZ-Organisationen | ja | | |

4 Schlussfolgerungen und Herausforderungen

In den Tabellen 1 bis 3 sind die ausgeführten Darstellungen zur Evaluationslandschaft der EZ in den drei Ländern überblicksartig zusammengefasst. Die Übersichten verdeutlichen, dass sich Evaluation zu einem integrierten Bestandteil der EZ in allen drei Ländern entwickelt hat. Unabhängige Evaluationseinheiten sind etabliert, national einheitliche und international vergleichbare Leitlinien und Evaluationskriterien sind vorhanden. Strategische, Projekt- und Programmevaluationen sowie auch gemeinsame Evaluationen werden durchgeführt und sind integraler Bestandteil des Projektzyklus'. Insbesondere der Institutionalisierungsgrad in Deutschland ist aufgrund von zahlreichen Aus- und Fortbildungsangeboten zu Evaluation in der EZ und der national angestrebten Standardisierung bei der Evaluationsdurchführung sowie der verbreiteten Durchführung von unabhängigen und Ex-post-Evaluationen am weitesten fortgeschritten. Dies ist nicht zuletzt der ‚DeGEval – Gesellschaft für Evaluation‘ zu verdanken, deren Arbeitskreis ‚Entwicklungspolitik‘ seit über zehn Jahren die Evaluationsbemühungen in der deutschen EZ vorantreibt. Die gesamte Bandbreite der deutschen EZ-Akteure, von NGO, staatlichen Organisationen, Consultingfirmen über die Wissenschaft ist beteiligt. Der Wissenschaft ist ebenfalls zu verdanken, dass die EZ-Evaluation in Deutschland weiter gedeiht. Besonders Hervorzuheben sind hier das CEval und das HWWA, welche die Anwendbarkeit von Evaluationsmethoden und Praktiken in der EZ zum Forschungsthema haben. In der Schweiz ist die Schweizerische Evaluationsgesellschaft (SEVAL) und die dort aktive Arbeitsgruppe Entwicklungszusammenarbeit ebenfalls eine treibende Kraft für die Institutionalisierung der EZ-Aktivitäten.

Trotz des insgesamt hohen Institutionalisierungs- und Professionalisierungsgrads der Evaluation in der EZ können einige Herausforderungen und Verbesserungspotentiale identifiziert werden, der sich die internationale EZ weiterhin stellen muss:

- Obwohl Evaluation bereits jetzt ein integraler Bestandteil des Projektzyklus' in allen drei Ländern darstellt, fehlt es weiträumig noch an ihrer Akzeptanz als Lerninstrument. Innerhalb der Durchführungsorganisationen und bei Projektpartnern muss daher weiterhin kommuniziert und vor allem demonstriert werden, dass Evaluation dem Bereich Qualitätsmanagement und Wissensmanagement zuzuordnen ist und nicht ausschließlich der Kontrolle, sondern der Verbesserung dient. Diese Nutzung von Evaluationsergebnissen für evidenzbasierte Politik sowie Rechenschaftslegung wird in allen drei Ländern noch zu wenig demonstriert und kommuniziert. Auch die Transparenz von Evaluationsergebnissen ist in allen drei Ländern noch verbesserungswürdig. Allzu oft werden Evaluationsergebnisse nicht oder nicht vollständig veröffentlicht. Bisher werden vielfach nur Zusammenfassungen von Evaluationsergebnissen veröffentlicht. In der Regel trifft dies nur auf Ergebnisse von unabhängig durchgeführten Evaluationen zu. Zudem werden Evaluationsergebnisse noch zu wenig für die weitere Planung zurate gezogen.
- Die Harmonisierung der Evaluationsaktivitäten und gemeinsame Evaluationen werden zwar bereits verfolgt, dies sollte allerdings noch weiter vorangetrieben werden zum Beispiel durch Abstimmung der nationalen Evaluationsprogramme. So könnten Synergien nicht nur bei der EZ sondern auch bei der Evaluation genutzt werden und gemeinsame Evaluationen zum Standard werden.
- Die Stärkung der Evaluationskapazitäten in den Partnerländern selbst und der Einsatz lokaler EvaluatorenInnen sollten im Sinne der Paris-Deklaration ebenfalls intensiviert werden. Besonders die Stärkung der Evaluationskapazität in den Partnerländern wird noch zu wenig akzentuiert. In Deutschland hat die staatliche Durchführungsorganisation Internationale Weiterbildung und Entwicklung gGmbH (InWEnt) in Kooperation mit der GTZ bereits erste Anläufe zur Stärkung der Evaluationskapazitäten in den Partnerländern unternommen. So wurden in 2006 und 2007 einige Trainingsmaßnahmen zu Evaluation durchgeführt, unter anderem für den afrikanischen, arabischen und zentralamerikanischen Raum. Eine Vielzahl weiterer Aktivitäten ist allerdings nötig, um die Evaluationskapazitäten in den Partnerländern weiter auszubauen.

Die wohl größte Herausforderung, der sich alle drei Länder stellen müssen, ist die Evaluation von Wirkungen. Um eine systematische Wirkungserfassung vor allem auf der Ebene der sektoralen oder länderbezogenen Programme zu leisten, ist ein intensives Zusammenspiel von Evaluation und empirischer Sozialforschung aufgrund der Anwendung anspruchsvoller Methoden zwingend notwendig (Vorher/Nachher-Messung, Vergleichs- und Kontrollgruppendedesigns, Zufallsstichproben).

Forschung und Technologie

Evaluation von Forschungs- und Innovationspolitik in Deutschland – Stand und Perspektiven¹

Große Forschungsprojekte, übergreifende Forschungsförderungsprogramme und Forschungsinstitutionen wurden zusehends zum Gegenstand von Evaluationsverfahren. Wachsende Autonomiespielräume der Hochschulen gehen einher mit leistungsabhängiger Ressourcenallokation. Das in Gang gesetzte Prinzip des Wettbewerbs zwischen Forschungseinrichtungen erfordert eine verstärkte Profilbildung durch gezieltes Wissenschaftsmanagement. Dies bewirkt in Forschung, Management und Politik einen wachsenden Bedarf an evaluativer Information.

1 Kontext der Evaluationstätigkeit

Unter Forschungs- und Innovationspolitik werden hier alle politischen Initiativen zur Gestaltung des Forschungs- und Innovationssystems gemeint (Larédo/Mustar 2001). Dies schließt Versuche ein, das Innovationsverhalten industrieller Unternehmen zu beeinflussen. Das deutsche System gilt im internationalen Vergleich als relativ gut entwickelt: 2004 waren knapp 473.000 Personen (Vollzeitäquivalente) mit Forschung und Entwicklung (FuE) beschäftigt. Die Gesamtausgaben für FuE beliefen sich auf knapp 55,4 Mrd. Euro, das entspricht 2,49 Prozent des Brutto-Inlandprodukts (Zahlen nach BMBF 2006). Die Forschungsinfrastruktur ist vergleichsweise differenziert:

- Die *Industrie* realisiert den größten Anteil von FuE in Deutschland (298.000 Mitarbeiter); 2004 investierte sie knapp 38,4 Mrd. Euro überwiegend in angewandte Forschung und experimentelle Entwicklung.
- Auf die *Hochschuleinrichtungen* (rd. 370 staatliche und rd. 110 private Hochschulen, darunter 110 Universitäten und vergleichbare Einrichtungen; 97.600 FuE-Mitarbeiter) fällt der zweitgrößte Anteil der Forschungsausgaben (9,1 Mrd. Euro). Sie konzentrieren sich auf Grundlagen- und langfristig anwendungsorientierte Forschung, finanziert durch die Bundesländer sowie durch die Deutsche Forschungsgemeinschaft (DFG, etwa 1,3 Mrd. Euro). Seit den 1980er Jahren ist der Anteil industrieller Forschungsaufträge gewachsen (auf 13,4 % im Jahre 2004).
- Die 15 „Großforschungseinrichtungen“ der Helmholtz-Gemeinschaft (HGF) des Bundes (12.000 FuE-Mitarbeiter; 2,3 Mrd. Euro), leisten vor allem langfristig orientierte

¹ Dieser Beitrag basiert auf überarbeiteten, in Passagen textgleichen Teilen früherer Publikationen, insbesondere Kuhlmann 1998a; 1998b; 2003a; 2003b; 2003c; Kuhlmann/Heinze 2004a; 2004b.

Forschung, die als risikoreich gilt, hohe Kosten verursacht (Anlagen) und große Forschungsteams benötigt. Daneben betreiben verschiedene Ministerien des Bundes sogenannte Ressort-Forschungseinrichtungen zur wissenschaftlich-technischen Unterstützung ihrer Aufgaben.

- Die Institute der Max-Planck-Gesellschaft (MPG; 4.400 Wissenschaftliche Mitarbeiter sowie 12.600 Doktoranden und Gastwissenschaftler; 1 Mrd. Euro) konzentrieren sich auf ausgewählte Felder der Grundlagenforschung in Natur- und Geisteswissenschaften.
- Die Institute der Fraunhofer-Gesellschaft (FhG; 2004: 12.000 Mitarbeiter, nicht nur in der Forschung; Gesamtbudget 1.090 Mio. Euro, davon 413 Mio. staatliche Förderung) sollen die praktische Verwendung wissenschaftlicher Erkenntnisse durch langfristig anwendungsorientierte und angewandte Forschung fördern. Die FhG führt überwiegend Auftragsforschung durch.
- Die Institute der Wissenschaftsgemeinschaft G.W. Leibniz (WGL; früher „Blaue Liste“; 5.700 FuE-Mitarbeiter; 800 Mio. Euro) werden von Bund und Ländern institutionell gefördert.
- Die Arbeitsgemeinschaft industrieller Forschungsvereinigungen (AiF) führt vor allem angewandte Forschung und experimentelle Entwicklung für sektor-spezifische Bedürfnisse von Unternehmen durch.

2004 gaben Bund und Länder knapp 17 Mrd. Euro für FuE aus. Das Spektrum von Instrumenten der staatlichen Forschungs- und Innovationspolitik ist weit ausdifferenziert und reicht von der institutionellen Förderung von Forschungseinrichtungen (mehr als die Hälfte aller Ausgaben) über verschiedene Formen finanzieller Anreize zur Durchführung von Forschung und experimenteller Entwicklung bis zur Gestaltung einer „innovationsorientierten“ Infrastruktur.

Moderne Forschungs- und Innovationssysteme stehen gegenwärtig ganz besonders unter dem Druck der Anpassung an veränderte Umwelten (vgl. zusammenfassend Internationale Kommission 1999):

- Die Entwicklung von Hochtechnologieprodukten hat signifikante Beschleunigung erfahren.
- In vielen Gebieten wächst die „Wissenschaftsbindung“ des Innovationsgeschehens; Grundlagenforschung (etwa molekularbiologische) wird ausgerichtet auf die Entwicklung marktfähiger Produkte (Meyer-Krahmer/Schmoch 1998).
- Komplexe Produkte und Prozesse erfordern eine dichtere Verflechtung und „Fusion“ (Kodama 1995) heterogener Technikentwicklungen, wodurch Grenzziehungen zwischen Wissens- und Technikgebieten verschwinden und interdisziplinäre Kompetenzen an Bedeutung gewinnen.
- Kooperation und internationale Netzwerke (Callon 1992) gehören mittlerweile zum alltäglichen Innovationsgeschehen. Gibbons et al. (1994) fassen diese Entwicklung als Übergang von „Mode 1“ zu „Mode 2“ zusammen. Etzkowitz/Leydesdorff (2000) entwickeln ein „Triple Helix“ Modell, nach dem Wissenschaft ko-evolutionär in die Dy-

namiken des Wirtschafts- und des politischen Systems verflochten ist. In Europa entstehen ‚post-nationale‘ Forschungslandschaften (Kuhlmann 2001).

- Es entstehen besonders dynamische, an einen bestimmten Standort gebundene, thematisch fokussierte Forschungs- und Innovationskapazitäten, die weltweit im Wettbewerb mit nur ein oder zwei anderen Standorten stehen, (z.B. Automobilbau in Deutschland; gentechnologische Pharmazeutik in den USA; Meyer-Krahmer 1999).

Staatliche Politik reagiert auf diese Trends, indem sie mit ihrer Förderpolitik strukturellen Wandel in der „Forschungs- und Innovationslandschaft“ zu bewirken sucht und Modernisierungsdruck auf Forschungsinstitutionen ausübt: So werden seit Mitte der 1990er Jahre viele Förderprogramme als thematische Wettbewerbe ausgeschrieben. Partnerschaften verschiedener Einrichtungen (Forschungseinrichtungen, Universitäten, private Firmen etc.) zielen auf eine Effektivierung ganzer Innovationssysteme.

Ein prominentes Beispiel hierfür bildet die „Exzellenzinitiative“ (begonnen 2005): Mit der Förderung der universitären Spitzenforschung im Rahmen eines Wettbewerbs sollen herausragende Einrichtungen der Wissenschaft in Deutschland entstehen, die auch international ausstrahlen. Die Hochschulen werden durch eine unabhängige Jury (sog. *Gemeinsame Kommission* aus DFG und Wissenschaftsrat) ausgewählt, die zum überwiegenden Teil aus ausländischen Wissenschaftlern besteht. Bis zum Jahr 2011 werden insgesamt 1,9 Mrd. Euro zur Verfügung stehen. Der Bund übernimmt hiervon 75 Prozent.

Zusammengefasst: „Nach jahrzehntelangem Inkrementalismus finden erstmals tief greifende Reformen im System der öffentlich finanzierten Forschung in Deutschland statt. Hierdurch wurde eine Selbsttransformation des Systems mit offenem Ausgang und nicht vorhersehbarer Dynamik in Gang gesetzt“ (Forschergruppe Governance der Forschung 2007: 28). Evaluationsverfahren spielen in diesem Prozess eine zentrale Rolle.

2 Evaluationsaktivitäten

In historischer Perspektive kann man wissenschaftsinterne und -externe Ansätze als zwei heterogene Entwicklungslinien der Evaluation von Wissenschaft sowie der Förderung von Forschung und Innovation unterscheiden. Sie lassen sich als ein Drei-Schalenmodell darstellen:

Erste Schale – Individuelle Forschungsleistungen: Den „Kern“ bilden Peer Review-Verfahren und später zusätzlich Verfahren zur Messung der Forschungsleistung einzelner Forscher und Gruppen (Bibliometrie etc.) als wissenschaftsinterne Instrumente für die Entscheidung über die Allokation von Fördermitteln in Forschungseinrichtungen (Moed et al. 2004; Guston 2003; Hornbostel 1997; Daniel 1993). Peer Reviews sind das vorherrschende Evaluationsinstrument der DFG. Förderanträge werden von Peers bewertet, die im Vier-Jahres-Rhythmus von der wissenschaftlichen Gemeinschaft gewählt werden (Neidhardt 1988). Jeder Gutachter ist angehalten, den Antrag auf der Grundlage seiner wissenschaftlichen Qualität zu beurteilen. Mittlerweile hat die DFG beschlossen, dieses bewährte System („bot-

tom-up“) um strategieorientierte, interdisziplinäre und internationale Komponenten zu ergänzen.

Zweite Schale – Programme: Evaluationsstudien im Sinne von *impact analyses* forschungs- und innovationspolitischer Programme, die im Unterschied zur institutionellen Förderung gezielte Gestaltungsabsichten verfolgen. Zwischen 1985 und 1993 ließ das Bundesforschungsministerium etwa fünfzig größere Evaluationsstudien anfertigen (Kuhlmann/Holland 1995a), in den folgenden Jahren nicht weniger. Solche Studien schließen an die amerikanische Wirkungsforschung und *policy analysis* an (Schön/Rein 1994; Widmer 1996; Kuhlmann 1998a). Dieses Evaluationskonzept kann als *wissenschaftsextern* bezeichnet werden, denn es wird überwiegend vom politisch-administrativen System in Gang gesetzt, um zu prüfen, inwieweit politisch gesetzte Ziele umgesetzt wurden. Als Evaluatoren treten dabei unabhängige Forschungsinstitute im Auftrag der Administration auf. Sie wenden ein relativ breites Spektrum von Evaluationskonzepten und -instrumenten an und sind seit Ende der 1990er Jahre auch professionell in einer „Gesellschaft für Evaluation (DeGEval)“ organisiert.

Dritte Schale – Institutionen: Bewertung der Leistungsfähigkeit von Forschungsinstitutionen oder größeren Wissenschaftsgebieten, wie sie unter anderem der Wissenschaftsrat (WR) durchführt. Nach der deutschen Vereinigung nahmen die WR-Evaluationen gestaltenden Einfluss auf die neue Struktur der Forschungslandschaft Ostdeutschlands (Block/Krull 1990). Seit den 1990er Jahren wurden institutionelle Evaluationen häufiger durchgeführt (Kuhlmann/Holland 1995b). Hierzu zählen auch die „Systemevaluationen“ aller großen Forschungseinrichtungen im Zeitraum 1998 und 2002 (Wissenschaftsrat 2000a; 2001; Internationale Kommission 1999; Evaluierungskommission 1998). Leitfrage war, ob und in welcher Weise die evaluierten Einrichtungen einen wesentlichen Beitrag zur Produktion von Wissen und Innovation leisten. Auch die „Exzellenzinitiative“ kann als „systemorientierte“ (umstrittene; siehe Münch 2007) Evaluationsmassnahme betrachtet werden. Bei der Evaluation von Forschungsinstitutionen werden wissenschaftsinterne und -externe Fragestellungen und Informationsquellen kombiniert.

2.1 Erzeugung evaluativer Information

In den vergangenen Jahrzehnten wurde eine relativ elaborierte Wissenschafts- und Innovationsindikatorik entwickelt, die als Grundlage für eine Matrix von Typen evaluativer Information dienen kann (z. B. Moed et al. 2004). Auf dieser Grundlage können Produktivitäts- bzw. Wirkungsindikatoren erzeugt und in Relation zu Input-Maßen als Referenzgrößen zur Bewertung von Forschungsleistung gestellt werden (Kuhlmann/Heinze 2004a). Im Folgenden wird exemplarisch die Produktion entsprechender Information dargelegt, strukturiert nach (a) Investitionen in Forschung (Input-Maße als Referenzgrößen zur Bewertung von Forschungsleistung) sowie (b) Produktivitäts- bzw. Wirkungsinformationen. Dabei können vertikal die Ebenen von Organisationen, Instituten oder Arbeitsgruppen und Einzelwissenschaftlern sowie horizontal die institutionelle „Landschaft“, innerhalb welcher Forschung

stattfindet, Forschungsprogramme, Fachdisziplinen und Forschungsfelder, unterschieden werden.

(a) Input-Information

Auf den Aggregationsebenen *Programm und Institutionen-“Landschaft“* finden wir eine Reihe regelmäßiger Publikationen (z. B. BMBF 2006; 2007) sowie datenbankgestützter Informationssysteme.² Insbesondere die der ministeriellen und der amtlichen Statistik zugehörigen Berichte enthalten hoch aggregierte Beschreibungen der finanziellen und personellen Ausstattung der deutschen Forschungslandschaft (Hochschulen, außeruniversitäre Forschung, Wirtschaftsforschung), aufgeschlüsselt nach regionalen (Bundesländer), disziplinären und sektoralen Kriterien.

Für die *Aggregationsebene Fachdisziplinen und Forschungsfelder* (auf organisationaler und departmentaler Ebene) liegen Querschnittsbegutachtungen des Wissenschaftsrates in den Feldern Umwelt-, Material- und Energieforschung vor (z. B. Wissenschaftsrat 1994; 1996; 1999). Hier wurden durch Selbstbeschreibungen (schriftliche Befragung der Einrichtungen) und Peer Reviews Informationen zur personellen und materiellen Ausstattung der untersuchten Forschungseinrichtungen erhoben. Weiterhin weist die DFG die seit den 1990er Jahren bewilligten Mittelflüsse an Forschungseinrichtungen aus, gegliedert nach 16 Fachgebieten. Verschiedene Bundesländer haben mit der systematischen Evaluation der universitären Fachdisziplinen begonnen. Eine bundesweite Übersicht der universitären Fachdisziplinen ist bisher nicht verfügbar. Einen ersten Schritt in diese Richtung geht die Studie ‚Forschungslandkarte Hochschulmedizin‘, die vom Fraunhofer-Institut für Systemtechnik und Innovationsforschung durchgeführt wurde (BMBF/Medizinischer Fakultätentag 2002).

Daten über *Universitäten* werden in der Regel von den Wissenschaftsministerien der Bundesländer erfasst. Allerdings fehlen weitgehend Strukturmaße, die Profilunterschiede zwischen einzelnen Universitäten ausweisen. Der Bericht des Wissenschaftsrates (2000b) untersucht relativ ausführlich die eingeworbenen Drittmittel, etwa auf Lehrstuhl- oder An-Instituts-Ebene bzw. im regionalen Vergleich. Er kritisiert zu Recht, dass die für solche Analysen erforderlichen Strukturinformationen bisher weitgehend fehlen. Hingegen weist das DFG-Ranking die bewilligten Forschungsmittel nicht nur nach Regionen und Fachgebieten, sondern auch nach Empfängerorganisationen (Hochschulen; andere) differenziert aus.

Auf der *Aggregationsebene einzelner Forschungsträgerorganisationen* und ihrer Einzelinstitute bzw. Forschungszentren ist die Datenbasis vergleichsweise heterogen. Für die außeruniversitären Einzelinstitute liegen zumeist jährliche Selbstbeschreibungen sowie Fremdbeschreibungen, unter anderem aus Systemevaluationen, vor.

² BMBF und Deutsche Forschungsgemeinschaft (DFG), die Informationen zu Forschungsausgaben nach ihrer Aufgabenstruktur und -entwicklung in Deutschland (DASTAT), hochschulspezifische Datenmengen (ICEbmbf) sowie Angaben zu DFG-geförderten Programmen bereit halten (GEPRIS).

(b) Output- und Wirkungsinformation

Kuhlmann/Heinze (2004b) haben untersucht, welche Produktivitäts- beziehungsweise Wirkungsinformationen auf unterschiedlichen Aggregationsebenen verfügbar sind.

Auf der *Aggregationsebene Institutionen-“Landschaft“* beschreiben Indikatoren die institutionellen Teilbereiche der Forschungslandschaft (Hochschulen, außeruniversitäre Forschung, Wirtschaftsforschung) als Ganzes und schlüsseln sie nach regionalen, disziplinären und sektoralen Kriterien auf. So werden etwa im Bundesbericht Forschung (BMBF 2006) oder im Bericht zur technologischen Leistungsfähigkeit Deutschlands (BMBF 2007) detaillierte Hinweise zur Produktivität der Wissensgenese gegeben. Qualitative oder strukturelle Informationen stehen kaum zur Verfügung.

Das Gewicht der Dokumentation liegt eher auf Output-Größen und weniger auf Wirkungsindikatoren. Zu vermuten ist, dass Outputs leichter zu erheben und auszuwerten sind als Wirkungs-Größen (z. B. Airaghi et al. 1999).

Für die *Aggregationsebene einzelner Forschungsträgerorganisationen* und ihrer Einzelinstitute bzw. Forschungszentren (Universitäten oder Einzelinstitute der außeruniversitären Forschungsverbünde) stellt sich das Informationsangebot sehr heterogen dar. Wie bereits bei den Input-Größen festgestellt sind die außeruniversitären Einzelinstitute durch Jahresberichte und im Zuge von Systemevaluationen umfangreicher beschrieben.

Die von der FlG-Zentrale publizierten Gesamtjahres- und Finanzberichte enthalten nach Instituten aufgeschlüsselte Output- und Wirkungsdaten. Relativ ausführlich sind Informationen zur Produktivität des Forschungshandelns und der Wissensprodukte, und zwar in quantitativer und qualitativer Hinsicht. So werden Projekte auf Institutsebene sehr gut aufgeschlüsselt nach Laufzeit und Finanzierungsgeber, ebenso Patentanmeldungen, erteilte Patente und Lizenzeinnahmen nach Technologiesektoren. Weniger umfangreich sind Daten zu Reputationsaspekten und Aspekte der Sichtbarkeit des generierten Wissens. Insgesamt wird der Darstellung von Forschungswirkungen weniger Raum eingeräumt.

Die MPG veröffentlicht keine Gesamtberichte, welche Output- und Wirkungsdaten institutsbezogenen ausweisen. Wiederum vermögen die Jahresberichte der Einzelinstitute Dies teilweise zu kompensieren. Zwar sind Strukturindikatoren nicht so häufig nachweisbar wie quantitative Maße. Allerdings ergibt sich der Eindruck, dass die Institute der MPG durchaus über ihre Produktivität aussagefähige Selbstbeschreibungen anfertigen. Dies gilt eingeschränkt auch für die Wirkungsseite.

Im Rahmen der Systemevaluation der WGL wurden weder die Gesamtebene, noch die Einzelinstitute hinsichtlich ihrer Forschungsausgaben und -wirkungen beschrieben, zumindest nicht für die breitere Öffentlichkeit. Zwar werden bei einigen Instituten Publikationen und Projekte relativ umfangreich ausgewiesen. Allerdings fällt die Beschreibung sehr heterogen aus.

Im Rahmen der Systemevaluation der HGF wurden die Zentren – im Vergleich zu den Instituten der WGL – relativ ausführlich beschrieben. Der Abschlussbericht hält ein Tableau mit vor allem nicht-wissenschaftlichen „Leistungsdaten“ bereit (z. B. Patent- und Lizenzdaten). Auffällig ist, dass wissenschaftliche Leistungen nicht zentrenbezogen ausgewiesen werden.

Universitäten dokumentieren ihre Forschungsleistungen in sehr heterogener Weise. Einen ersten Ansatzpunkt bietet das erwähnte DFG-Ranking, welches die Bewilligungsempfänger der DFG nach Hochschulen aufschlüsselt. Neben dem absoluten Gesamtvolumen wird das über die Zahl der an einer Hochschule tätigen Professoren und Hochschullehrer relativierte Volumen ausgewiesen. Interpretiert man eine DFG-Bewilligung nicht als bloßen Mittelzufluss und damit als „Input“, sondern als durch den Begutachtungsprozess zertifizierte Qualität geplanter Forschungsarbeit, dann stellt die DFG – zumindest in Näherung – eine Leistungskennziffer für allgemeine Forschungsleistung zur Verfügung. Generell gilt, dass eine für Universitäten übergreifende Dokumentation von Output- und Wirkungsindikatoren weitgehend fehlt. Zwar ist im Hochschulrahmengesetz (HRG) seit 1998 die Qualitätskontrolle von Forschung (und Lehre) rechtlich festgeschrieben worden. Allerdings erscheint dessen Umsetzung, sowohl was die Ländergesetzgebung als auch die gegenwärtig Praxis angeht, relativ uneinheitlich.

2.2 Anbieter evaluativer Information

Eine beachtliche Zahl von Einrichtungen erzeugt evaluative Informationen, jedoch in unterschiedlichen Formaten und nur begrenzt öffentlich zugänglich. Zum einen kann man die Anbieter danach unterscheiden, ob sie Informationen über sich selbst (Selbstbeschreibung) oder über andere erstellen (Fremdbeschreibung). Die Jahresberichte der Institute der MPG und FhG als Selbstbeschreibungen enthalten evaluative Informationen zum Teil für Zwecke der strategischen Forschungsplanung, zum Teil aufgrund bestehender (vor allem Input-)Daten. Die Systemevaluationen der großen Forschungseinrichtungen und die genannten Einzelstudien sind dagegen Fremdbeschreibungen, bei welchen externe Akteure evaluative Informationen in Kooperation mit den und über die zu evaluierenden Einheiten erheben.

Eine weitere Unterscheidung betrifft den institutionellen Ort, an dem evaluative Informationen erzeugt werden. In Deutschland hat die Praxis der Forschungsevaluation keinen gemeinsamen institutionellen Ort und kennt bisher keine gemeinsamen Formate. Im Folgenden werden die relevanten Anbieter evaluativer Informationen skizziert.

In den Zentralverwaltungen der *außeruniversitären „Forschungssäulen“* (MPG, HFG, WGL, FhG) laufen vielfältige Informationen, auch solche zu den Forschungsleistungen der Einzelinstitute, zusammen. Diese werden in unterschiedlichem Masse veröffentlicht.

Die DFG als zentrale Selbstverwaltungseinrichtung der deutschen Wissenschaft zur Förderung der Forschung stellt mit GEPRIS und DFG zwei Tools zur Verfügung mit Strukturinformationen zur Forschungslandschaft. Parallel hat die DFG 2005 ein „Institut für Forschungsinformation und Qualitätssicherung (IFQ)“ errichtet. Das IFQ konzentrierte sich anfangs auf eine verbesserte Informationsbasis über Rahmenbedingungen, Durchführung und Ergebnisse DFG-geförderter Forschung. Dies soll der modellhaften Entwicklung des theoretischen, methodischen und technischen Instrumentariums dienen, das nach Abschluss der Aufbauphase auch über die DFG hinausgehend zu nutzen ist.

Der *Stifterverband* für die deutsche Wissenschaft unterstützt die Forschungsberichterstattung der Bundesregierung und stellt Daten für EUROSTAT, OECD und UNESCO bereit.

Das *Hochschulinformationssystem HIS* führt als von Bund und Ländern finanzierte Einrichtung Erhebungen durch, die vorwiegend die Entwicklung von Verfahren zur Rationalisierung der Hochschulverwaltung, Mitwirkung bei deren Einführung und Anwendung sowie Gutachten zur Schaffung von Entscheidungsgrundlagen im Hochschulwesen zum Zweck haben. Einen dezidiert forschungspolitischen Auftrag hat das HIS nicht.

Der *Wissenschaftsrat (WR)* besitzt als „Fremdbeschreibungsagentur“ wissenschaftlicher Einrichtungen eine hohe Reputation. Er hat neben den Querschnittsevaluationen, die zeitlich und sachlich begrenzte evaluative Informationen dokumentieren, bisher kein systematisches Berichtssystem entwickelt. Auch die Systemevaluationen der HFG und WGL sind – obgleich umfangreich – ihrem Zustandekommen und ihrer Anlage nach ad-hoc-Untersuchungen.

Das *BMBF* erhebt mit seinen Dokumentationen zur strukturellen Situation des deutschen Forschungssystems und der technologischen Leistungsfähigkeit Deutschlands systematisch wichtige evaluative Informationen. Als Bundeseinrichtung verarbeitet das BMBF diese Fremdbeschreibungen auf hoch aggregierter Ebene. Seit vielen Jahren führt das Forschungsministerium Programmevaluationen durch (Kuhlmann/Holland 1995a; Kuhlmann 1998a). Bis heute wird nur ein Teil dieser von externen Evaluationsexperten durchgeführten Studien (häufig ex post-Wirkungsanalysen) der Öffentlichkeit zugänglich gemacht oder regulär publiziert. Wenig evaluative Information liegt bisher bezüglich der vom Bund betriebenen Einrichtungen der problemorientierten Ressortforschung vor.

Einige Wissenschaftsministerien der *Länder* (insbesondere Baden-Württemberg, Bayern, Nordrhein-Westfalen, Niedersachsen) haben eine Reihe von Initiativen zur Evaluation der Hochschulforschung angestoßen.

Einen wichtigen Beitrag zur Generierung und zum Angebot evaluativer Informationen liefern Forschungseinrichtungen mit einem Schwerpunkt im Bereich *Wissenschafts- und Innovationsforschung*. Hier sind vor allem die folgenden zu nennen: Das Institut für Wissenschafts- und Technikforschung (IWT) an der Universität Bielefeld; das Centrum für Hochschulentwicklung (CHE) in Gütersloh; das Wissenschaftszentrum Berlin (WZB); das Fraunhofer-Institut für Systemtechnik und Innovationsforschung (ISI) in Karlsruhe; das Wissenschaftliche Zentrum für Berufs- und Hochschulforschung an der Universität Kassel. Zunehmend tragen auch ausländische Forschungseinrichtungen evaluative Studien zum deutschen Forschungssystem bei.

2.3 Nachfrage nach evaluativer Information

Im Folgenden werden exemplarisch mögliche Nutzungen evaluativer Information behandelt.

Die *Deutsche Forschungsgemeinschaft* hat im Jahr 2000 einen Senatsausschuss „Perspektiven der Forschung“ ins Leben gerufen. Er soll als „Frühwarnsystem“ im Rahmen wissenschaftlicher Bedarfsanalysen innovative Forschungsfelder identifizieren und die forschungsstrategischen Initiativen des Senats vorbereiten. Damit ergänzt die DFG ihre bisherige Praxis, dass

zu jedem Thema zu jeder Zeit ein Antrag gestellt werden kann. Die erforderlichen Bedarfsanalysen können nur auf der Basis evaluativer Information erstellt werden.

Strategische Planung großer Forschungseinrichtungen: Die Forschungsträgerorganisationen wurden im Rahmen der Systemevaluationen aufgefordert, deutlicher als bisher explizite strategische Entwicklungsplanungen zu betreiben. Sie müssen künftig – auf der Grundlage einer regelmäßig aktualisierten, kritischen Analyse ihres eigenen Leistungsspektrums – ihr Forschungsportfolio im Kontext der nationalen und internationalen Forschungslandschaft positionieren.

Bundes- und Landeseinrichtungen der Forschungspolitik (BMBF, BMWi, Bundestag, Wissenschaftsministerien der Länder) sind verantwortlich für die Pflege der regionalen, nationalen und institutionellen Forschungslandschaft im internationalen Kontext. Strategieentwicklung und Planungsarbeiten können bisher nur auf fragmentarische oder ad hoc erstellte Aggregate evaluativer Information zurückgreifen. Weder in der Interaktion mit Forschungsträgerinstitutionen und Universitäten noch gegenüber den Integrationsversuchen der Europäischen Union (Europäischer Forschungsraum; „Mapping of Centres of Excellence“) können Bund und Länder bisher ihre Interessen, Entscheidungen und Vorgehensweisen auf systematische Analysen stützen (abgesehen von – zum Teil guten, aber aufwändigen – ad hoc-Studien durch Einrichtungen der Wissenschafts- und Innovationsforschung). Benötigt werden Input-Informationen (Welche Forschungskapazitäten gibt es bereits, und wer fördert was, national und international?), Output-Informationen auf mittlerem und höherem Aggregationsniveau (Welche Einrichtungen leisten bereits exzellente Forschung in bestimmten Gebieten? Welche Einrichtungen sollten gestärkt werden?) sowie Information über erzielte oder potentielle Wirkungen von Forschung und Förderung (z. B. in der akademischen und sonstigen Bildung; im Bereich Technologie, Innovation und Wirtschaftsentwicklung; zur Lösung sozialer oder ökologischer Probleme, etc.).

Auch der *Wissenschaftsrat* wird bei künftigen Evaluationen von Forschungsinstitutionen oder Themenfeldern und bei weiteren Systemevaluationen Aggregatinformationen heranziehen müssen. Ein besonders dringender Bedarf besteht hinsichtlich Strukturinformationen zur Forschung in Universitäten.

Einrichtungen der *Wissenschafts- und Innovationsforschung* können ihre – teils aus wissenschaftlichem Interesse, teils im Auftrag forschungspolitischer Akteure durchgeführten – Studien zu Stärken, Schwächen und Dynamik des deutschen Forschungssystems bisher nicht auf ein verlässliches, allgemein akzeptiertes System evaluativer Information stützen. Vielfach müssen Grundlagendaten aufwändig im Rahmen von Einzelstudien erhoben werden.

3 Verbindungen mit anderen Evaluationsfeldern und internationalen Entwicklungen

Die fortschreitende Internationalisierung und Europäisierung von Forschung, Technologieentwicklung und Innovation sowie darauf gerichteter politischer Maßnahmen hat Konsequenzen für die Evaluationspraxis. So hat das Generaldirektorat Forschung der *Europäischen Kommission* eine relativ systematische Evaluationspraxis entwickelt, die auch in die Evaluationskultur der Mitgliedländer hineinwirkt (Georghiou 1995). Dazu gehören das Five Year Assessment des Rahmenprogramms durch ein Expertenpanel, jährliche Monitoring-Berichte, Evaluationen der Spezifischen Programme, horizontale Evaluationen übergreifender Fragestellungen sowie Studien der Wirkungen europäischer Programme im nationalen Forschungssystem (z. B. Reger/Kuhlmann 1995). Im Rahmen der Kommissionsinitiative zur Schaffung eines „europäischen Forschungsraums“ (European Commission 2000; Kuhlmann/Edler 2003) gibt es verschiedene Bemühungen, Indikatorenbündel zur Identifikation europäischer wissenschaftlich-technischer Leistungsfähigkeit und entsprechender politischer Fördermaßnahmen zu entwickeln, exemplarisch anzuwenden und in ein öffentlich zugängliches Berichtssystem zu implementieren (z. B. *Trendchart*; *ERAWATCH*). Für die strategische evaluative Analyse ganzer Institutionen-„Landschaften“ können solche Datenaggregationen künftig von großer Bedeutung sein.

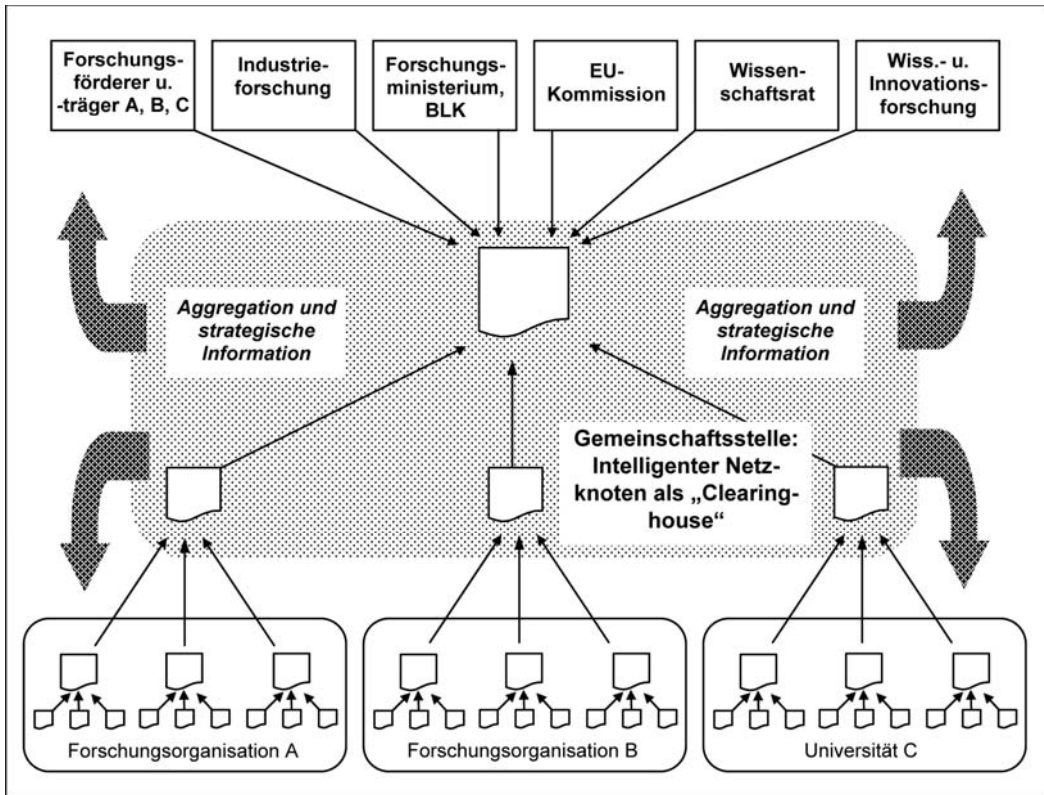
4 Stand und Perspektive der Evaluation

Die deutsche Evaluationspraxis ist einerseits *entwickelt* hinsichtlich des hohen Maßes an Selbstorganisation bei wissenschaftsinternen Verfahren, die Konsens und Verpflichtung zwischen den Forschern fördern, sowie *dynamisch* wegen vielfältiger neuer Evaluationsansätze und Strategieentwicklungen in allen Institutionen. Sie ist andererseits *unsystematisch* und *zersplittert*, weil die verschiedenen institutionell orientierten Evaluationsbemühungen bisher wenig aufeinander abgestimmt sind.

Viele Evaluationsinitiativen (interne und externe Ansätze) beschränken sich nicht mehr allein auf die Feststellung der Qualität einzelner Forschungsleistungen. Zusehends soll evaluative Information auch *strukturbildende* sowie *sozio-ökonomische Effekte* von Forschung und Forschungsförderung aufklären (vor allem aus der Perspektive der Europäischen Union: Airaghi et al. 1999; Polt et al. 2002; Shapira/Kuhlmann 2003). Ein gemeinsames Raster von Indikatoren zur Erfassung und Analyse solcher Effekte wurde bisher nicht verabredet.

Kuhlmann/Heinze (2003; 2004b) empfehlen die Einrichtung einer Gemeinschaftsstelle „Clearinghouse“ zur Unterstützung der Wissenschaftsorganisationen bei der Qualitätssicherung, der Effizienz- und Effektivitätskontrolle, der Selbststeuerung und organisationalem Wandel sowie Maßnahmen zur Strukturverbesserung (siehe Abbildung 1).

Abbildung 1: Gemeinschaftsstelle „Clearinghouse“ für Forschungsstrukturinformationen



(Quelle: Kuhlmann/Heinze 2003: 38)

Eine solche Gemeinschaftsstelle wird gesteuert und überwacht von einem Konsortium (d.h. einer freiwilligen Kooperation) bestehend aus relevanten Einrichtungen der Forschungspolitik, der Forschungsförderung, der Forschungsträgerinstitutionen sowie ausgewählten Einrichtungen der Wissenschafts- und Innovationsforschung. Dieses Konsortium vereinbart Umfang und Formate quantitativer und qualitativer evaluativer Information. Es empfiehlt, welche Informationen Forschungsorganisationen, Forschungsförderer und andere Konsortialmitglieder beitragen, und setzt den Rahmen für Zielsetzungen und Verwendungszwecke. Die Gemeinschaftsstelle wiederum moderiert die erforderlichen Abstimmungsprozesse. Informative Produkte der Gemeinschaftsstelle stehen grundsätzlich allen Konsortialmitgliedern zur Verfügung. Möglichst viele Informationsprodukte sollen der allgemeinen, interessierten Öffentlichkeit zugänglich gemacht werden (z. B. interaktiv gestaltete Internet-Websites).

Evaluative Informationen dürfen nicht „blindlings“ gesammelt, aggregiert, analysiert und verbreitet werden, sondern nur im Rahmen vereinbarter konkreter Evaluationszwecke, Rhythmen und Formate. Die Zusammenarbeit im „Clearinghouse“ schließt ein, dass die Trägereinrichtungen „vertrauliche“ Informationen für vereinbarte strategische Analysen bereit stellen.

Die Evaluation der wissenschaftlichen Qualität und Leistungsfähigkeit von Forschungseinrichtungen gehört in die Hände von qualifizierten *Peers*. Zur Unterstützung ihrer Bewertungstätigkeit sollten sie auf hochwertige „evaluative Information“ zum relativen Stellenwert der bewerteten Einrichtung im internationalen Wettbewerb zurückgreifen können. Sie sollte von eigens dafür qualifizierten wissenschaftlichen Dienstleistungseinrichtungen erbracht werden, die internationalen Standards der Wissenschaftsforschung genügen (Kuhlmann/Heinze 2004b). Das kürzlich von der DFG eingerichtete Institut für Forschungsinformation und Qualitätssicherung (IFQ) stellt einen Schritt in diese Richtung dar.

Die Evaluation von Programmen der Forschungsförderung ebenso wie reformpolitischer Maßnahmen (als Policy-Evaluation) sollte von dafür qualifizierten unabhängigen Einrichtungen der sozialwissenschaftlichen Forschung durchgeführt werden. Dabei sollte die Möglichkeit des Lernens der Entscheider und der Betroffenen hinsichtlich der Gestaltung der Förder- und Reformmaßnahmen im Vordergrund stehen.

Evaluation sollte als „reflexives“ Instrument politischer Entscheidungsfindung in „Multi-Akteur-Arenen“ gesehen werden (Kuhlmann 1998b) und „intelligenter“ Zulieferer von Verhandlungsstrategien der politischen Akteure wie auch der interessierten Öffentlichkeit sein. Unter „intelligent“ wird hier verstanden, die Verknüpfung von Forschung und Technologie, industriellen Anstrengungen, gesellschaftlichen und politischen Bedürfnissen transparenter zu gestalten und so die Zusammenarbeit zwischen den Beteiligten zu erleichtern.

Evaluation von Forschungs- und Technologiepolitik in Österreich¹

Das Feld Forschungs-, Technologie- und Innovationspolitik (FTI-Politik) zeichnet sich durch seine Vielfältigkeit aus. Es weist Überschneidungen mit anderen politischen Themenfeldern wie der Wirtschafts-, Wissenschafts-, Bildungs-, Regional- und Umweltpolitik auf und betrifft sehr heterogene Zielgruppen wie universitäre Forscher/innen, Universitäten als Institutionen sowie außeruniversitäre Forschungseinrichtungen und private Unternehmen. Zudem verfügt das Feld über unterschiedlichste Interventionsmodelle wie etwa Einzelprojektförderungen, Netzwerk- und Clusterförderungen oder aber steuerliche Anreizmechanismen.

Wir versuchen diesem Umstand Rechnung zu tragen, indem wir von einem sehr breiten Forschungs- und Innovationsbegriff ausgehen und Interventionen aus allen oben genannten Bereichen berücksichtigen.

1 FTI-Politik: Kontexte und Ratio

Die FTI-Politik hat in den letzten Jahren in Österreich und auch europaweit stark an Bedeutung gewonnen. Besonders der Lissabon-Prozess und die Barcelona-Ziele waren Ansporn dafür, dass FTI auf der politischen Agenda stark nach oben gerückt sind. Auch die mediale Berichterstattung ist erheblich gestiegen. Printmedien berichten heute über Forschungs- oder Technologiepolitik in eigenen Rubriken, Beilagen und Magazinen (z. B. Die Presse, Der Standard, das Wissensmagazin der Süddeutschen).²

Ein Grund für die Steigerung des politischen Interesses liegt in Österreich wie in Europa wohl darin, dass man dazu neigt anzunehmen, dass Investitionen in FTI langfristig positive Effekte für Wirtschaft und Beschäftigung haben und somit einen Beitrag zu Wohlstand und „Lebensqualität“ leisten. Eine andere, etwas skeptischere Begründung ist, dass die FTI-Politik im Gegensatz zu anderen Instrumenten der Wirtschaftspolitik (geförderte Darlehen, Investitionsförderungen o. ä.) auch deswegen in den Mittelpunkt des politischen Interesses

¹ Dank an Leonhard Jörg, Michael Stampfer, Brigitte Tempelmaier, Alfred Radauer, Wolfgang Neurath, Michaela Topolnik, Sonja Sheikh und Rupert Pichler für wertvolle Kommentare und kontroverse Diskussionen. Die im vorliegenden Artikel ausgedrückten Meinungen sind ausschließlich jene der Autoren und stellen keine offiziellen Positionen der Plattform Forschungs- und Technologieevaluierung, des Wiener Wissenschafts- und Technologiefonds und der Joanneum Research Forschungsgesellschaft mbH dar. Sämtliche zitierten und erwähnten Evaluierungsberichte können, so freigegeben, auf der Homepage der Plattform Forschungs- und Technologieevaluierung www.fteval.at bezogen werden.

² Einige dieser Medien und Publikationen werden freilich von der öffentlichen Hand finanziert bzw. subventioniert.

rückt, weil hier (auch EU-wettbewerbsrechtlich) mehr Gestaltungsspielraum vorzufinden ist.

Mit der Zunahme des politischen Interesses und mit der Steigerung der Ausgaben der öffentlichen Hand wächst natürlich auch der Bedarf an Information, Legitimation, Lernen und Kontrolle. So gibt es in Österreich einerseits Maßnahmen, die versuchen, das öffentliche Verständnis für FTI sowie das allgemeine Interesse an FTI zu erhöhen. Ein Beispiel hierfür (und für die Resonanz, die solche Maßnahmen auch finden) war etwa die ‚Lange Nacht der Forschung‘ 2005, an der an 150 Stationen 48.000 Menschen teilnahmen (Steiner et al. 2007).

Andererseits ist es Aufgabe der Evaluation von FTI-Maßnahmen, über Erfolg und Misserfolg von Maßnahmen Auskunft zu geben, über die Zweckmäßigkeit der Mittelaufwendung für bestimmte Maßnahmen zu urteilen und Politik-Lernen im Sinne eines Verbesserns von Maßnahmen zu gewährleisten. Dass dies kein unbedeutendes Unterfangen ist, zeigen die „nackten“ Zahlen. Die öffentlichen F&E-Ausgaben des Bundes und der Länder betrugen 2006 rund 2,25 Mrd. Euro, das sind mehr als ein Drittel der gesamten Ausgaben für Forschung und Entwicklung, die insgesamt 2,47 Prozent des österreichischen Bruttoinlandsprodukts ausmachten. Da die (öffentlichen) Mittelaufwendungen in den letzten Jahren stark gestiegen sind (um 81 % seit 1998), ist es verständlich, dass im Kontext steigender F&E Ausgaben immer mehr verlangt wird, den Erfolg dieses Einsatzes öffentlicher Mittel nachzuzeichnen.

2 Die Situation in Österreich

Die Evaluation von Maßnahmen der FTI-Politik in Österreich hat sich über die Jahre stark weiterentwickelt: Inhaltlich, indem Methoden wie Logic Charts (Zinöcker et al. 2005b), Logit-Probit-Analysen (Streicher et al. 2004), Matched Pairs (Pointner/Polt 2005) und Focus Groups erstmalig eingesetzt wurden; quantitativ, indem immer mehr Programme und Institutionen auch evaluiert wurden und schließlich qualitativ, weil sich (einige) Entscheidungsträger/innen in den Agenturen und Ministerien ernsthaft mit den Ergebnissen von Evaluation auseinandersetzen und Empfehlungen entweder aufgreifen oder begründet verwerfen. Beispiele hierfür:

- Der Umgang der Politik mit den Evaluationsergebnissen zu den Förderorganisationen Fonds zur Förderung der wissenschaftlichen Forschung (FWF) und des Forschungsförderungsfonds für die gewerbliche Wirtschaft (FFF): Die Empfehlungen in Arnold (2004), etwa bezüglich der Organisation des Kuratoriums, hatten direkten Einfluss auf deren Neugestaltung.
- Der Umgang der Nachfolgeorganisation des FFF, die Forschungsförderungsgesellschaft (FFG), und des Wissenschaftsfonds FWF mit den Ergebnissen der jeweiligen Institutionenevaluierung 2004 (vgl. Binder 2005; Novak 2005).
- Das Interesse in Ministerien und Agenturen an Evaluation: So haben sich ab 2004 über 140 Personen (klassische „Anwender/innen“ von Evaluation) im Rahmen von Workshops, die von der Plattform Forschungs- und Technologieevaluierung angeboten wurden, beteiligt (Zinöcker et al. 2004).

- Die Veranstaltung zweier bedeutender internationaler Konferenzen zu Evaluation im Bereich Forschungs- und Technologiepolitik in den Jahren 2003 und 2006: Eine gut genutzte Möglichkeit für österreichische Akteure, sich über den internationalen Stand der Diskussion in Kenntnis zu setzen.³

Im Zeitraum von 2003 bis 2007 wurden über sechzig Evaluationen im Bereich der FTI-Politik durchgeführt (siehe www.fteval.at). Das ist für ein kleines Land wie Österreich nicht wenig und im Vergleich zu den Jahren zuvor ein bedeutender Anstieg. Die Gründe hierfür sind mannigfaltig; vor allem sollte in Betracht gezogen werden, dass die Programmorientierung der österreichischen FTI-Politik, also das Abwenden der Entscheidungsträger/innen von einer erratischen, projektorientierten Vergabepolitik, erst Ende der 1990er Jahre begonnen und im genannten Zeitraum auch die Anzahl der Programme stark zugenommen hat. Viele Programme sind also erst in den letzten Jahren entstanden und konnten somit auch erst in den letzten Jahren evaluiert werden.

Auch das öffentliche Bewusstsein und Interesse an der Notwendigkeit von Evaluation generell hat erst in diesem Zeitraum zugenommen. Ein Hinweis hierfür sind auch die Gründungsjahre der großen (weit über FTI hinausgehend viele Politikfelder umfassenden) Evaluierungsvereinigungen im deutschsprachigen Raum, SEVAL und DeGEval nämlich 1996 und 1997.

Vor diesem Hintergrund wird im Folgenden der Versuch einer strukturierten Herangehensweise zur Beantwortung der Frage nach dem Status quo von Evaluation in Österreichs FTI-Politik unternommen. Zuerst wird kurz der rechtliche Rahmen für Evaluation skizziert, gefolgt von einer Zusammenstellung der wichtigsten Akteure der Evaluation im Bereich FTI. Dann wird nach Inhalt (Projekt, Programm, Institution, Politik) und nach Zeitpunkt (ex ante, interim, ex post) unterschieden, um eine ausgewogene Darstellung der Situation geben zu können.

2.1 Rechtlicher Rahmen der Evaluation

Über lange Zeit gab es keinen gesetzlichen Rahmen für Evaluation in Österreichs FTI-Politik.

Für Universitäten gab es 1997 eine erste detaillierte Evaluierungsverordnung im Rahmen des Universitätsorganisationsgesetz 1993. Heute wird im Universitätsgesetz 2002 den Universitäten vorgegeben, ein eigenes Qualitätsmanagementsystem aufzubauen, im Rahmen dessen Evaluationen des gesamten Leistungsspektrum der Universität durchzuführen sind (§ 14 Universitätsgesetz; bm:bwk 2002).

Mit den Richtlinien zur Förderung der wirtschaftlich-technischen Forschung und Technologieentwicklung, den sogenannten FTE-Richtlinien⁴, wurde ein weiterer Schritt zur

³ Evaluation of Government Funded Activities in R&D, May 2003. New Frontiers in Evaluation, April 2006. Dokumentationen dieser Konferenzen sind unter www.fteval.at/03conference03 [4.4.2008] und www.fteval.at/conference06 [4.4.2008] zugänglich.

⁴ Richtlinien zur Förderung der wirtschaftlich-technischen Forschung und Technologieentwicklung (FTE-Richtlinien) gemäß § 11 Z 1 bis 5 des Forschungs- und Technologieförderungsgesetzes (FTFG) des Bundesministers für

Verankerung von Evaluationsmaßnahmen gesetzt. Diese Richtlinien sehen vor, dass „für alle auf den FTE-Richtlinien basierenden Förderungsprogrammen und -maßnahmen ein schriftliches Evaluierungskonzept zu erstellen ist, das den Zweck, die Ziele und die Verfahren sowie die Termine zur Überprüfung der Erreichung der Förderungsziele enthält und geeignete Indikatoren definiert. Zum Zweck der Erfassung der erforderlichen Informationen ist ein entsprechendes Monitoring aufzubauen.“ (Abschnitt 2.2., Seite 4). Über die Richtlinien hinausgehend ist Evaluation schon im entsprechenden Gesetz (Forschungs- und Technologieförderungsgesetz, Abschnitt II, § 12) explizit verlangt.

Auch im aktuellen österreichischen Regierungsprogramm (BKA 2007) wird der Evaluation von FTI-Politik und ihren Instrumenten ein großer Stellenwert eingeräumt. Für den Zeitraum 2007–2010 werden folgende Maßnahmen in Angriff genommen:

- Portfolioanalyse der großen Anzahl der FTI-Instrumente in Österreich: Wo sind Bündelungen und Prioritätensetzungen möglich (BKA 2007: 56-57)?
- Evaluation der Universitäten nach europäischen Maßstäben mit dem Ziel der Qualitätssteigerung. Hierzu sei eine Verbesserung der Evaluierungsinstrumente notwendig (BKA 2007: 101).
- Eine Evaluation der Forschungsinstitute der österreichischen Akademie der Wissenschaften (BKA 2007: 102).

Darüber hinaus gibt es zwei Dokumente, die Evaluation zwar nicht normieren oder verbindlich vorschreiben, aber doch beeinflussen. Einerseits hat der Rat für Forschung und Technologieentwicklung⁵ mittels einer Empfehlung die österreichische Bundesregierung dazu angehalten, bewusst mit dem Instrument Evaluation zu arbeiten und es aktiv in die Politikgestaltung einzubeziehen (Ratsempfehlung vom 12. April 2005). Andererseits haben sich die Mitglieder der Plattform Forschungs- und Technologieevaluierung mit stark an den Standards der Evaluation der DeGEval angelehnten „Standards der Evaluation in Österreichs Forschungs- und Technologiepolitik“ selbst einen Rahmen für Evaluation gegeben. In ganz allgemeiner Weise sollen damit die Aktivitäten der Plattform zur Steigerung der „Evaluationskultur“ in Österreich unterstützt werden. Speziell dienen Standards dazu, sowohl den Evaluator/innen als auch beauftragenden Einrichtungen und Evaluierten einen Rahmen und Verhaltensmaßgaben bereitzustellen. Damit wird ein höherer Grad an Verbindlichkeit und Sicherheit für alle Beteiligten erreicht.

2.2 Der Markt für Evaluationen

Auf der Angebotsseite wird die Szene von außeruniversitären, mit öffentlicher Grundfinanzierung ausgestatteten Forschungseinrichtungen geprägt: KMU Forschung Österreich, Jo-

Verkehr, Innovation und Technologie vom 27.9.2006 (GZ 609.986/0013-III/12/2006) und des Bundesministers für Wirtschaft und Arbeit vom 28.9.2006 (GZ 97.005/0012-C1/9/2006).

⁵ Der österreichische Rat für Forschung und Technologieentwicklung hat insofern eine Sonderstellung, als dass er nicht bloß die Bundesregierung „berät“, sondern auch eine aktive Rolle bei der Vergabe von Budgetmitteln einnimmt. Somit haben einige seiner Empfehlungen einen verbindlicheren Charakter.

antrum Research, mit größeren Abstrichen das WIFO sowie die Austrian Research Centers. Bedeutendster privater und zentraler Anbieter von Evaluationen ist das internationale Beratungsunternehmen Technopolis, das auch in Österreich über ein Büro verfügt. Daneben gibt es eine ganze Reihe von privaten beziehungsweise außeruniversitären Einrichtungen, die vereinzelt Evaluationen anbieten. Als ausgesprochen schwach ist die Rolle der Universitäten einzustufen, die es in den letzten Jahren über weite Strecken verabsäumt haben, evaluationspezifisches Know-how und Methodenwissen aufzubauen (ausgeklammert bleibt hier die Evaluationsrolle von Wissenschaftler/innen im Rahmen von Peer Review).

Die Nachfrageseite nach Evaluationen ist dadurch gekennzeichnet, dass die Agenden für FTI auf drei verschiedene Ministerien (die Ministerien für Verkehr, Innovation und Technologie, für Wirtschaft und Arbeit sowie für Wissenschaft und Forschung) aufgeteilt sind, wobei alle auf das Instrument Evaluation setzen. Seitens der Agenturen gibt es sowohl auf Bundes- als auch auf Landesebene starke Nachfrage: Neben den „großen“ Institutionen FWF und FFG für wissenschaftsnahe und unternehmensnahe Forschung, gibt es auf Länderebene in Wien, Oberösterreich und Tirol Agenturen, die Evaluation zum Zwecke des Lernens oder der Legitimation mittlerweile vermehrt und ambitioniert einsetzen.

Festzuhalten ist aber, dass das „Standardmodell“ nach wie vor so ist, dass das für eine Fördermaßnahme verantwortliche Ministerium als Initiator und Auftraggeber der Evaluation auftritt; Beauftragungen von abwickelnden Agenturen direkt sind abgesehen von Initiativen der großen Förderorganisationen FWF und vormals FFF nicht die Regel.

2.3 Projekt, Programm, Institution und Politik

Evaluationen auf Projektebene

Evaluationen auf Projektebene im Sinne einer Auswahl von Projektanträgen sind über weite Strecken avanciert und ausgereift. Der österreichische Wissenschaftsfonds FWF bedient sich seit Jahren eines erprobten Systems von internationalem Peer Review. Hier ist anzumerken, dass gerade diese Evaluierungsmethode in Bezug auf mögliche Befangen- und Voreingenommenheiten einer harschen Kritik ausgesetzt, aber unter Forscher/innen (Arnold 2004), nicht nur in Österreich, weitgehend unumstritten ist (vgl. hierzu Dinges 2006). Die Evaluation des Fonds (Streicher et al. 2004) hat den professionellen Einsatz der Methode zur Projektselektion nachgezeichnet, und auch gezeigt, dass keine Unstimmigkeiten (etwa: „Ältere Einreicher werden jüngeren gegenüber bevorzugt“) nachweisbar sind.

Für den Bereich der unternehmensnahen Projektforschung gilt ähnliches: Hier hat die Evaluation (Jörg/Falk 2004) die Professionalität der abwickelnden Stelle FFF⁶ klar nachgezeichnet, jedoch auch auf Probleme der Risikobewertung in der Projektselektion aufmerksam gemacht.

Daneben gibt es eine Reihe von Programmen, die immer mehr danach trachten, die Fördervergabe von Projekten fair und transparent zu gestalten. Sehr oft werden hierzu For-

⁶ Der FFF ist der Fonds zur Förderung der gewerblichen Wirtschaft, der 2004 Teil der Forschungsförderungsagentur FFG geworden ist.

men von Peer Review (etwa expert panels, extended Peer Review)⁷ herangezogen, Wettbewerbe organisiert, Jurys gebildet. Wichtig ist hierbei die Zusammenarbeit mit der Evaluation auf Programmebene, die eben diese Form der Vergabe von Projekten innerhalb eines solchen Programms klar nachvollziehen und nachzeichnen sollte. Verschiedene Programme (die Sonderforschungsbereiche des FWF, die Kompetenzzentren der FFG, u. ä.) sehen Zwischen- und Endevaluationen ihrer ‚Projekte‘ explizit vor, des Weiteren evaluiert die FFG die Projekte ihres Basisprogramms regelmäßig ex post (Sheikh/Bornett 2005).

Davon gesondert sollte die Auftragsforschung in den Ministerien gesehen werden: Mit der steigenden Bedeutung von FTI-Programmen seit Anfang der 1990er Jahren und den schwierigen budgetären Rahmenbedingungen der Ministerien ist die Auftragsforschung in den Hintergrund getreten. Dennoch gab es noch 2002 über 750 Projekte mit einer Dotierung von etwa 35 Millionen Euro, die auf diesem Wege entstanden sind (Zinöcker/Dinges 2004).

Evaluationen auf Programmebene

In fast allen Programmen wird heute mit Evaluationen gearbeitet, einige Ausnahmen werden weiter unten angeführt. Dies ist maßgeblich auf die Rolle des Rates für Forschung und Technologieentwicklung zurückzuführen, der diese Evaluationen mit Nachdruck eingefordert hat. Als besonders breit diskutierte Beispiele für Programmevaluationen können das Kompetenzzentren-Assessment (Edler et al. 2003) oder die Evaluation des Technologietransfer-Programms (Jörg et al. 2000) gelten, deren Empfehlungen klare Auswirkungen auf die Weiterentwicklung dieser Programme hatten.

Im Bereich der Programmevaluationen hatten zwei Auftraggeber, die sehr früh von der Notwendigkeit von Evaluation überzeugt waren, Vorreiterrollen: Erstens das Ministerium für Verkehr, Innovation und Technologie – in einer Vielzahl von Ausformungen und Akronymen in den letzten sieben Legislaturperioden –, das Evaluationen jener Programme durchführen ließ, die im Rahmen des (heute nicht mehr existenten) Innovations- und Technologiefonds initiiert wurden (neben dem bmvit trat hier auch das BMWA in Erscheinung). Der zweite zentrale Auftraggeber war der ehemalige Fonds zur Förderung der gewerblichen Wirtschaft (FFF)⁸, der systematisch alle in seinem institutionellen Rahmen etablierten Initiativen (Holzforschung, Mikrotechnik, Lebensmittelinitiative etc.) extern evaluieren ließ.

Evaluationen auf Ebene der Institutionen

Im Bereich der Evaluation von Institutionen hat es mit Sicherheit in den letzten Jahren die größte Bewegung gegeben. Die zentralen Institutionen zur Förderung der grundlagennahen und der angewandten Forschung, FWF und FFF, wurden erstmalig in den vierzig Jahren ihres Bestehens 2004 evaluiert (Arnold 2004). Andere Institutionen der Forschungsförde-

⁷ Zu den methodischen Unterschieden siehe Rigby 2002 und 2004.

⁸ Der FFF ist in die österreichische Forschungsförderungsgesellschaft FFG, Bereich Basisprogramme, übergegangen.

rung, etwa die Christian Doppler Forschungsgesellschaft (Schibany et al. 2005) oder die Betreuungsstrukturen für das 6. EU Rahmenprogramm (Sheikh et al. 2004), zogen nach.

Grund für diese zunehmende Bedeutung der Evaluation von Institutionen war auch eine neue, professionellere Arbeitsaufteilung zwischen Ministerien und Agenturen: Die Programmabwicklung weg aus den Ministerien hin zu Förderagenturen auszulagern („Agenturifizierung“, „Rothschild Principle“, Arnold 2004: 96) war eine zentrale Forderung in einer Vielzahl von Evaluationen und Studien, der schließlich auch von der Politik entsprochen wurde.

Das Verhältnis Ministerien zu Agenturen, das sich auch als *Prinzipal-Agent-Problem* darstellen lässt, wird wesentlichen Einfluss auf die Steuerung der österreichischen FTI-Politik haben. All die hier zum Einsatz kommenden Steuerungsmechanismen werden Themen zukünftiger Evaluationen sein.

Im Bereich der Evaluation von Forschungsinstitutionen gibt es auch eine klare Verbindung zu den Universitäten, die mit dem Universitätsgesetz 2002 in die Autonomie entlassen wurden und nunmehr durch Leistungsvereinbarungen an das Wissenschaftsministerium gebunden sind. Zwar steckt die Evaluation von Universitäten nach wie vor in den Kinderschuhen, doch mittlerweile gibt es ein systematisches Erfassen von relevanten Informationen zu den österreichischen Universitäten in Form von Wissensbilanzen (Leitner 2007), und erstmalig ist ein systematisches Herangehen einzelner Universitäten an Evaluation zu verzeichnen.⁹ Zudem gibt es für die Universitäten die Einrichtung der österreichischen Qualitätssicherungsagentur AQA: Die AQA ist eine unabhängige Agentur zur Evaluation und Qualitätssicherung im Hochschulbereich. Sie ist als Verein aufgesetzt und ihre Mitglieder sind die maßgeblichen Akteure aus diesem Bereich. Hierin begründet liegt die schwierige Ausgangslage dieser Agentur, die zur Finanzierung einerseits auf Aufträge aus den Mitgliederorganisationen angewiesen ist, gleichzeitig aber als Qualitätssicherer ein hohes Maß an Unabhängigkeit bewahren muss.¹⁰ Insgesamt ist die Rolle der Universitäten im Evaluationskontext eine ganz besondere, denn hier sind die Stimmen am lautesten, die Evaluation als Instrument stark in Zweifel ziehen und auch vehement gegen eine Professionalisierung von Evaluation argumentieren.

Die wichtigsten institutionellen Vertreter/innen der außeruniversitären Forschung in Österreich (Austrian Research Centers, Joanneum Research), die keine unwesentliche Rolle in Österreichs FTI-Politik spielen, sind als Institutionen noch nie extern evaluiert worden. Dies ist bis zu einem gewissen Grad gerechtfertigt, da diese Forschungseinrichtungen nur zum Teil vom Staat basissubventioniert werden und sich ohnehin im Wettbewerb um ihre Finanzierung bemühen müssen. Andererseits wurde bis dato etwa noch nie an Evaluator/innen die Frage gestellt, ob etwa der (geringe) Anteil an Basissubvention richtig eingesetzt wird.

⁹ Hierbei sind besonders die Wiener Universität für Bodenkultur (www.boku.ac.at) und die Universität Wien (www.univie.ac.at) herauszustreichen.

¹⁰ Im aktuellen Regierungsprogramm der österreichischen Bundesregierung wird auf diese schwierige Ausgangssituation eingegangen und eine Neuausrichtung der Österreichischen Qualitätssicherungsagentur AQA in Aussicht gestellt (Bundeskanzleramt 2007: 101).

Evaluation auf Politik- und Systemebene

Dieser Bereich der Evaluation ist wohl auch international gesehen noch sehr unterentwickelt. Eine Handvoll Studien können dieser Kategorie zugerechnet werden, etwa die Evaluation aller mathematischen Forschungseinrichtungen (Hoffmann/Bourguignon 2005) oder die Wirkungsanalyse frauenpolitischer Maßnahmen im Bereich der FTI-Politik (Wroblewski et al. 2006).

Gleichzeitig ist auf der Politik- und Systemebene auch der Übergang zu anderen Berichten und Studien fließend: Eine ganze Anzahl von Arbeiten, die unter Umständen diesem Bereich zuzuordnen wären, tragen nicht die Bezeichnung „Evaluation“, und haben auch keine explizite Evaluationsfunktion vorgesehen, enthalten aber dennoch Elemente formativer oder summativer Evaluierungen. Hierzu gehören etwa die Forschungs- und Technologieberichte der Bundesregierung (etwa bm:bwk/bmvit/bmwa 2006), die jährlich den Status von FTI(-Politik) abbilden, oder auch das Weißbuch des Wirtschaftsforschungsinstitutes WIFO, das klare Empfehlungen für die Weiterentwicklung der FTI-Politik in Österreich beinhaltet. Die Studie „Instrumente der Technologiepolitik und ihr Mix“ von Schibany und Jörg (2005) hat wiederum zu einer nachhaltigen Diskussion darüber beigetragen, wie die einzelnen Instrumente der FTI-Politik zueinander stehen und wie ihr Einsatz optimiert werden könnte.

Eine wichtige Funktion als Nachfrager solcher Evaluationen könnte hier der Rat für Forschung und Technologieentwicklung übernehmen. Als Berater der Bundesregierung hat er übergreifende Fragestellungen nicht aus den Augen zu verlieren, und sollte diese Funktion auch im Rahmen von Evaluationen stärker ausüben.

2.4 Evaluation im Planungszyklus von Programmen

Generell werden Evaluationen nicht nur nach ihrem Inhalt, sondern auch nach dem Zeitpunkt der Evaluation unterschieden. Auffällig ist, dass ex ante Evaluationen im FTI-Bereich Österreichs (und die auch diese Bezeichnung tragen) sehr selten vorkommen. Eine Reihe von Studien, nennen wir sie Machbarkeitsstudien, die vor oder zu Beginn von neuen Initiativen eingesetzt werden, weisen jedoch Elemente von ex ante Evaluationen auf. Der ex ante Bereich ist ein Feld, in dem noch am ehesten Elemente der Selbstevaluation eingesetzt werden: Intensive Vorbereitungsarbeiten, wie internationale Vergleiche, Recherchen und Interviews stehen oft am Beginn neuer politischer Initiativen.

Das Interesse an Interimevaluationen hat in den letzten Jahren einen starken Aufschwung erlebt. Neben der (weiter oben bereits erwähnten) Vorgabe durch den Rat für Forschungs- und Technologieentwicklung ist hier ein klares Ziel der Auftraggeber/innen das „Lernen“. Man will im Programm verbessern und nachjustieren, und oft setzt man hier auf die Expertise externer Evaluator/innen. So wurden etwa die „großen“ thematischen Programme der letzten Jahre in den Bereichen Informationstechnologie, Genomforschung und Nanotechnologie Interimevaluationen unterzogen, die sehr konkrete, umsetzungsorientierte Empfehlungen für Programmmanagement und Politik beinhalteten (Zinöcker et al. 2005a; 2005b; Jörg/Werner 2006).

Klare ex post Evaluationen, die versuchen langfristig die Wirkung von Programmen und Initiativen nachzuweisen, gibt es in Österreichs FTI-Politik, wie auch im internationalen Vergleich, nur wenige. Zwar werden viele Programme auch zu einem Zeitpunkt evaluiert, an dem man auf eine Reihe von Aktivitäten zurückblicken kann, doch nur in den seltensten Fällen ist genügend Zeit vergangen, um auch nur in Ansätzen von einer methodisch sauberen Wirkungsanalyse ausgehen zu können. Zudem ist die Motivation bei Politik und Programmmanagement ex post Analysen durchzuführen begrenzt, da sich beide eher an Gegenwart und Zukunft orientieren.

2.5 Einflüsse anderer Evaluationsfelder

Die Einflüsse anderer Evaluationsfelder (etwa aus der Arbeitsmarktpolitik oder der Entwicklungszusammenarbeit) sind als relativ gering einzuschätzen. Drei Momentaufnahmen, die Ausnahmen darstellen: Es gibt einzelne Evaluatoren/innen, die etwa auch in der Arbeitsmarktpolitik und/oder der Evaluation von Strukturpolitik tätig sind. Folglich kommt es zu einem ‚Wissenstransfer über Köpfe‘. Die österreichische Beteiligung an Konferenzen der DeGEval ist nicht gering, hier wird man ganz zwangsläufig mit anderen, neueren Herangehensweisen aus anderen Feldern konfrontiert; Schließlich gibt es sehr seltene Fälle, in denen ein methodischer Ansatz aus einem anderen Feld in den Bereich der FTI-Politik „schwappt“. Als Beispiel soll hier die „Matched Pairs Analyse“ dienen, ein Ansatz, der eigentlich aus der Arbeitsmarktpolitik kommt (Heckman et al. 1999), und der dann, über internationale Kooperationen, auch als Instrument der Wirkungsmessung in Österreich Einzug gefunden hat (Pointner/Polt 2005).

3 Bilanz

Eine entwickelte Evaluationskultur stellt einen zentralen Bestandteil einer lernenden und strategisch orientierten FTI-Politik dar. Eine gute Evaluationskultur ist gleichsam Voraussetzung und Folge einer guten, also effizienten, transparenten und fairen Politik.

In Österreichs FTI-Politik wird das Instrument „Evaluation“ von der Mehrzahl der Akteure ernst genommen und mit Sorgfalt eingesetzt. Über Evaluationsberichte wird zum Teil kontrovers diskutiert, und sie werden zunehmend auch in der Öffentlichkeit wahrgenommen.

Der Überblick über die FTI-Politik und ihre Evaluation in Österreich zeugt von einem lebhaften Feld, in dem es trotz der bestehenden Aktivitäten an Herausforderungen für die Zukunft nicht mangelt. Wir sehen sie wie folgt:

1. **Datensituation verbessern:** Die Qualität von Evaluationen steht in unmittelbarem Zusammenhang mit der Qualität der Informationen (Daten), die im Programmverlauf gesammelt werden. Hier heißt es jedoch mit großer Sorgfalt vorzugehen: „Collect all data that is needed and need all data that is collected“ ist ein schöner Leitsatz, der auch an-

gesichts einer möglichen Überforderung der Evaluierten (siehe nächster Punkt) zu berücksichtigen sein wird.

2. Evaluationsmüdigkeit versus „Unter-Evaluation“: Zu leicht kann man mit zu vielen Interviewwünschen, zu vielen (und zu langen) Fragebögen etc. eine „Evaluationsmüdigkeit“ bewirken, die nur kontraproduktiv sein kann. Die Anzahl der Evaluationen auf Institutions- und Programmebene hat in den letzten Jahren massiv zugenommen. Gleichzeitig ist es unzureichend, dass zentrale Fördereinrichtungen – so wie in Österreich mit FFF und FWF geschehen – nur einmal in vierzig Jahren evaluiert werden. Hier muss ein sinnvoller Rhythmus gefunden werden, in welchen Abständen langfristige Initiativen zu evaluieren sind.
3. Realistische Erwartungen: Zu oft werden in FTI-Programmen nachgerade homöopathische Budgetmittel eingesetzt, um Ziele wie ‚Wirtschaftswachstum‘, ‚Hebung des Lebensstandards‘ oder ‚Änderung des österreichischen Mindsets‘ zu erreichen. Hier ist es an der Zeit, realistischere Erwartungen an den Tag zu legen und Zielüberfrachtungen in Programmen zu vermeiden. Ex post Evaluationen vergangener Programme und der gezielte Einsatz von ex ante Evaluationen im Rahmen der Programmgestaltung könnten hier einen wichtigen Beitrag leisten.
4. Raum für „curiosity-driven Evaluation“: Evaluator/innen müssen genügend Freiheiten eingeräumt werden, um ihre Funktion auch sinnvoll ausüben zu können. Im Rahmen von Ausschreibungsverfahren soll und kann nicht bloß der Preis das alleinig entscheidende Kriterium im Vergabeprozess von Evaluationen sein. Evaluator/innen müssen in der Lage sein, sich und das Feld methodisch weiterzuentwickeln und in Ecken zu schauen, über welche Auftraggeber/innen von Evaluationen, sei es bewusst oder unbewusst, in den Ausschreibungsunterlagen schweigen. Hier sollte also eine gewisse Flexibilität der Evaluator/innen gewährleistet sein, Auftraggeber/innen sollten dies im für Evaluierungsvorhaben vorgesehenen Budgetrahmen mitdenken. Eigentliche Methodenentwicklung sollte jedoch andere Finanzierungsquellen finden.
5. Sprache: Im österreichischen Rahmen ist ‚Deutsch‘ die vorherrschende Evaluationsprache: Nicht, dass die Auftraggeber/innen nicht auch englischsprachige Evaluierungsberichte akzeptieren würden (Edler/Rigby 2004; Zinöcker et al. 2005b) die entweder von internationalen Evaluator/innen durchgeführt oder mitbetrieben wurden, aber letztendlich verlangen Dokumenten- und Datenanalyse (und wohl auch Interviews) klarerweise Deutschkenntnisse. Das verringert den Kreis potentieller Evaluator/innen (und somit den Kreis derjenigen, die sich um eine Ausschreibung bewerben) radikal. Evaluation sollte internationaler werden, österreichische Evaluator/innen mehr im Ausland, genauso wie Evaluator/innen aus Großbritannien, Schweiz, Kanada et cetera vermehrt in Österreich tätig sein.

Evaluation von Forschung und Technologie in der Schweiz

Forschungsevaluation ist in der Schweiz nach wie vor selten. Ich behaupte nicht, dass Forschung gar nicht evaluiert werde. Wohl aber, dass sie zu oft weitgehend unter Ausschluss der Öffentlichkeit betrieben wird (die ja im Falle der Hochschulen und der Verwaltung diese Tätigkeit finanziert). Die Folge: Systematisches Wissen über Gegenstände, Vorgehen und Ergebnisse von Evaluationen im Bereich Forschung und Technologie ist rar. Dies gilt namentlich für den privatwirtschaftlichen Bereich. Ich konzentriere mich deshalb im Folgenden auf die mit öffentlichen Mitteln geförderte Forschung und ihre Evaluation. Auch so muss die nachfolgende Darstellung skizzenhaft bleiben.

Bevor die Evaluation von Forschung thematisiert werden kann, soll im ersten Kapitel kurz der Kontext umrissen werden: Umfang und Aufteilung der Mittel für Forschung und Technologie in der Schweiz sowie der rechtliche und institutionelle Rahmen, in dem Forschung stattfindet. Sodann gehe ich auf die Aktivitäten im Bereich der Forschungsevaluation ein, soweit sie bekannt sind: Gegenstände, Nachfrager und Anbieter von Evaluationen. Ein abschliessendes Kapitel diskutiert den derzeitigen Stand und die Perspektiven von Forschungsevaluation in der Schweiz.

1 Kontext

1.1 Aufwendungen für Forschung und Technologie in der Schweiz

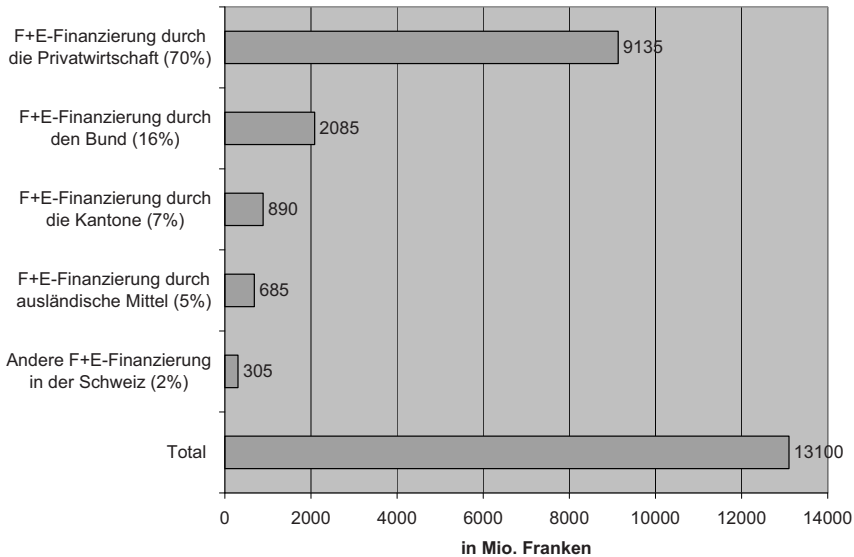
Forschung und Technologie sind in der Schweiz weitgehend Sache der Privatwirtschaft: Gemäss den verfügbaren Statistiken wurden im Jahre 2004 nahezu drei von vier Forschungsfranken von der Privatwirtschaft aufgebracht (70 %, siehe Abbildung 1). Auf Bund und Kantone entfiel nur knapp ein Viertel (16 % bzw. 7 %).

Die öffentlich finanzierte Forschung wird im Wesentlichen in den Hochschulen erbracht und dort wiederum von den Eidgenössischen Technischen Hochschulen (ETH) und den kantonalen Universitäten (69 % der öffentlichen Forschungsmittel, BFS 2006: 10). Auf die Verwaltungen von Bund, Kantonen und Städten (sogenannte Ressortforschung) entfällt demgegenüber nur ein verhältnismässig bescheidener Teil der Mittel.

Schaut man sich die Ressortforschung des Bundes etwas näher an, so findet man eine starke Fokussierung auf einige wenige Politikbereiche: Landwirtschaft, Entwicklungszu-

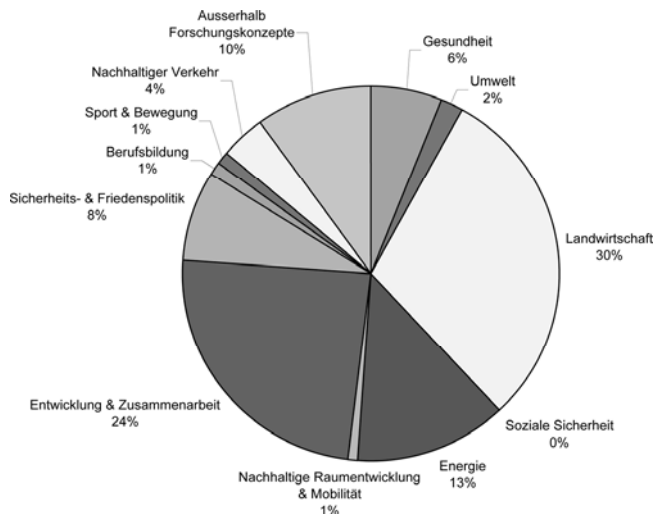
sammenarbeit und Energie machen bereits zwei Drittel (67 %) der Bundesforschung aus (Abbildung 2).¹

Abbildung 1: Finanzierung der durchgeführten Forschung und Entwicklung in der Schweiz nach Quelle, 2004



Quelle: BFS (2006: 7)

Abbildung 2: F+E beim Bund, 2004: Aufteilung nach Politikbereichen



Quelle: Bundesblatt (2006: 810).

¹ Relativierend ist anzumerken, dass der Anteil der Landwirtschaft (30%) zum Grossteil aus den Forschungsanstalten besteht, die in „Agroscope“ zusammengeschlossen sind; dies ist in der Schweiz ein Sonderfall, da die Bundesverwaltung in der Regel nicht über eigene Forschungseinrichtungen verfügt.

Zusammenfassend lässt sich sagen, dass die öffentlichen Investitionen für Forschung und Technologie in der Schweiz schwerpunktmässig in spezifischen Bereichen konzentriert sind: universitäre Hochschulen und ausgewählte thematische Politikfelder. Die Gründe für diese Fokussierung dürften zum Teil mit der Kleinheit des Landes in Verbindung stehen, die eine allseitige Entwicklung von Forschung und Technologie kaum möglich macht, und zum Teil mit der engmaschigen föderalistischen Struktur staatlicher Institutionen zusammenhängen.

1.2 Der rechtliche und institutionelle Rahmen

Die Privatwirtschaft ist weitgehend frei in ihrer Forschungstätigkeit. Im Gegensatz dazu braucht und hat die öffentlich finanzierte Forschung und Entwicklung einen gesetzlichen und institutionellen Rahmen. Die Grundlage dafür bilden das Forschungsgesetz und die dazu gehörende Verordnung (SR 420.1 und SR 420.11). Das Forschungsgesetz definiert als Forschungsorgane des Bundes (Art. 5):

- die Institutionen der Forschungsförderung (Schweizerischer Nationalfonds, wissenschaftliche Akademien, weitere vom Bundesrat anerkannte wissenschaftliche Institutionen)
- die Hochschulen (ETH inkl. Forschungsanstalten, Universitäten)²
- die Bundesverwaltung

Das Gesetz umschreibt recht allgemein die Tätigkeitsbereiche sowie die Grundlagen der Finanzierung dieser Einrichtungen. Dabei wird das Thema der Evaluation der Forschung oder der Forschungseinrichtungen lediglich indirekt angesprochen: Fast etwas beiläufig hält Art. 19 Abs. 1 fest: „Der Bundesrat wacht darüber, dass die Bundesmittel für die Forschung koordiniert *und wirksam* verwendet werden“ (Hervorhebung PF). Daraus liesse sich wohl ein Evaluationsauftrag ableiten. Die Verordnung enthält hierzu allerdings keine weiteren Ausführungen. Soweit ersichtlich hat diese Formulierung bislang kaum praktische Bedeutung erlangt.

Etwas anders verhält es sich mit Art. 5a Abs. 3, welcher dem Schweizerischen Wissenschafts- und Technologierat SWTR als beratendem Organ des Bundesrates „für alle Fragen der Wissenschafts-, Forschungs- und Technologiepolitik“ (Art. 5a Abs. 1) die Technologiefolgenabschätzung und damit eine typisch evaluatorische Tätigkeit als Aufgabe überträgt. Diese Bestimmung hat durchaus praktische Konsequenzen (siehe Kapitel 2.2).

Die wenig präzisen Bestimmungen des Gesetzes und die stark föderalistische Struktur der schweizerischen Hochschullandschaft haben zu einer Vielzahl von forschungsbezogenen Institutionen und zu Finanzflüssen geführt, die auch für Kenner schwer zu überblicken

² Mit der Teilrevision des Forschungsgesetzes gemäss Botschaft „Bildung-Forschung-Innovation 2008-2011“ neu auch: Fachhochschulen.

sind.³ An dieser Stelle und in der gebotenen Kürze können nur die wichtigsten genannt werden:

- Das *Staatssekretariat für Bildung und Forschung SBF* im Eidgenössischen Departement des Inneren ist die zentrale Schaltstelle des Bundes in diesem Bereich. Es ist unter anderem verantwortlich für die Erstellung der vierjährigen Botschaft über „Bildung-Forschung-Innovation“, welche den ganzen, departementsübergreifenden Bereich abdeckt.
- Der *Schweizerische Wissenschafts- und Technologierat SWTR* berät den Bundesrat. Er besteht aus unabhängigen Persönlichkeiten und wird vom Bundesrat gewählt.
- Der *Schweizerische Nationalfonds SNF* ist die wichtigste Institution der Forschungsförderung. Neben der reinen Grundlagenforschung betreut er auch die Nationalen Forschungsprogramme und die Nationalen Forschungsschwerpunkte. Ferner führt er mit DORE („DO REsearch“) ein Programm zur Förderung der angewandten Forschung an den Fachhochschulen in den Bereichen Sozial- und Gesundheitswesen sowie Kultur durch.
- Die *Kommission für Technologie und Innovation KTI* ist für die Förderung der industrienahe angewandten Forschung zuständig (u. a. an den anderen Fachhochschulen).
- Das schweizerische *Hochschulsystem* besteht aus den zwei Eidg. Technischen Hochschulen (ETH) in Lausanne und Zürich, vier ETH-Forschungsanstalten, zehn Universitäten und acht Fachhochschulverbünden, die alle in Forschung und Entwicklung aktiv sind.
- Des weiteren sind *Forschungs- und Infrastruktureinrichtungen* wie zum Beispiel der sozialwissenschaftliche Datenarchivdienst SIDOS oder das Schweizer Haushalt-Panel SHP zu nennen. Die neu gegründete Schweizer Stiftung für sozialwissenschaftliche Forschung FORS beherbergt ab 2008 diese Institutionen und verknüpft sie mit systematischer methodenorientierter Forschung.

2 Aktivitäten

2.1 Evaluationsgegenstände

Was die Gegenstände im Bereich der Evaluation von Forschung und Technologie angeht, so ist es hilfreich, zwischen Projekt-, Programm-, Institutionen- und Politikebene zu unterscheiden.

Projektevaluationen sind vor allem beim SNF und der KTI die Regel. Sie beschränken sich allerdings weitgehend auf die Prüfung von Unterstützungsanträgen. Ausser in den Nationalen Forschungsprogrammen, in denen systematisch nach den Ergebnissen der Projekte gefragt wird und diese an den anfänglichen Zielsetzungen gemessen werden, gibt es kaum

³ Die zum Zeitpunkt der Abfassung dieses Beitrags diskutierte Reorganisation des gesamten Bildungs- und Forschungsbereichs beim Bund (Schaffung eines Bildungsdepartementes) würde zweifellos zu einer Neugestaltung des hier kurz vorgestellten institutionellen Rahmens führen.

Evaluationen nach Abschluss der Projekte. Unter methodischen Gesichtspunkten handelt es sich also vorwiegend um ex ante-Evaluationen, die von Peers durchgeführt werden. Wo Projektevaluationen in der Ressortforschung des Bundes implementiert sind, sind sie nicht selten mit einem betriebswirtschaftlich orientierten Controlling verknüpft.

Programmevaluationen gibt es namentlich in der Ressortforschung, die ihre Forschung oft entlang mehrjähriger Programme organisiert. Dabei handelt es sich in der Regel um die Evaluation von Einzelprogrammen. Die Förderinstrumente des SNF (Nationale Forschungsprogramme und Nationale Forschungsschwerpunkte, früher die Schwerpunktprogramme, aber auch die Personenförderung mittels Beiträgen und Stipendien) waren schon wiederholt Gegenstände von Evaluationen, bei denen aber weniger das Einzelprogramm, sondern das Förderinstrument als solches (Programmart) im Mittelpunkt stand (so z. B. Belser 2006). Beide Arten von Programmevaluationen sind ex post-Evaluationen, die extern durchgeführt werden.

Die Evaluation von *Institutionen* ist insbesondere im Hochschulbereich ein Thema. Die Hochschulen streben an, die Evaluationen so weit wie möglich selber durchzuführen. Dabei haben sich, dem schweizerischen Bildungsföderalismus folgend, unterschiedliche und zum Teil stark voneinander abweichende Praxen herausgebildet und etabliert. Deren Entwicklungsgrad ist im Forschungsbereich markant geringer als im Bereich der Lehre. Meist handelt es sich um expertengestützte Evaluationen (peer evaluation), die sich in unterschiedlichem Ausmass und in wechselnden Kombinationen auf qualitative und quantitative Grundlagen stützen. Forschungseinrichtungen, die eine (Grund-) Finanzierung vom Bund erhalten⁴, werden periodisch evaluiert, oft – aber nicht immer – durch den SWTR oder in dessen Auftrag. Da solche Evaluationen als Entscheidungsgrundlagen für die Weiterfinanzierung dienen, werden sie meist umfassend angelegt (z. B. im Falle des Schweizerischen Sozialarchivs [CEST 2003a] oder der Schweizerischen Friedensstiftung swisspeace [CEST 2003b]). Auch der SNF und die KTI waren schon Gegenstand von Institutionen-Evaluation (Good 2006, SWTR 2002).

Auf der *Politikebene* ist die Evaluation der Steuerungsinstrumente der Ressortforschung (Konzepte, interdepartementaler Steuerungsausschuss) zu erwähnen, die von der Geschäftsprüfungskommission des Nationalrates in Auftrag gegeben und von der Parlamentarischen Verwaltungskontrolle zusammen mit einer externen Stelle durchgeführt worden ist (Bundesblatt 2006: 783-846). Ebenfalls auf dieser Ebene angesiedelt war die Evaluation der Sozialwissenschaften, die von der Schweizerischen Akademie der Geistes- und Sozialwissenschaften (SAGW) initiiert und vom damaligen Schweizerischen Wissenschaftsrat (SWR; heute Schweizerischer Wissenschafts- und Technologierat, SWTR) realisiert wurde (SWR 1993). Generell sind aber forschungspolitisch ausgerichtete Evaluationen in der Schweiz eher die Ausnahme.

⁴ Gestützt auf Art. 16 des Forschungsgesetzes über die Unterstützung von „wissenschaftlichen Hilfseinrichtungen“. Gemeint sind damit vorwiegend Infrastrukturen wie Datenzentren, Bibliotheken etc.

2.2 Nachfrager und Anbieter

Als *Nachfrager* von Evaluationen im Bereich Forschung und Technologie treten zum einen diejenigen Akteure in Erscheinung, die selber Forschung und Entwicklung betreiben: Unternehmungen, die Hochschulen und die forschenden Stellen der öffentlichen Verwaltung. Zum anderen sind jene an Evaluationen interessiert, die Forschung direkt finanzieren oder über die Grundlagen der Finanzierung entscheiden: SNF und KTI sowie die politischen Behörden.

All diese Akteure haben unterschiedliche Interessen, Perspektiven und Verständnisse von Evaluation. Verallgemeinernde Aussagen sind schwierig, zumal sie sich kaum auf greifbare empirische Grundlagen stützen können. So liegen zum Beispiel über den Umfang der Evaluationstätigkeit im Bereich Forschung und Technologie keine verlässlichen Zahlen vor. Unstrittig ist, dass solche Tätigkeiten für die meisten Akteure nicht im Vordergrund stehen, was die methodologische und inhaltliche Entwicklung der Forschungsevaluation in der Schweiz nicht gerade gefördert hat.

Die *Anbieterseite* ist, was externe professionell durchgeführte Evaluationen angeht, etwas weniger stark fragmentiert. Hier treten insbesondere etliche privatwirtschaftlich verfasste Dienstleister auf, die sich auf Evaluationen spezialisiert haben und auch (aber nicht ausschliesslich) Forschung und Technologie evaluieren. Es handelt sich dabei vorwiegend um kleine bis sehr kleine Unternehmen, die meisten mit weniger als zwanzig Beschäftigten. Hochschulen sind ebenfalls auf diesem speziellen Markt aktiv: Einzelne Institute oder Lehrstühle, typischerweise in den Politik- oder den Bildungswissenschaften, führen Evaluationen auf Auftragsbasis durch. Im Unterschied zu den privaten Unternehmungen handelt es sich aber in der Regel nicht um ihr Haupttätigkeitsgebiet (dieses liegt eher auf dem Feld der akademischen Lehre und Forschung), sondern um Aktivitäten, die dem Interesse und dem Engagement einzelner Personen entspringen und auch davon abhängen.

Die einzigen auf die Evaluation von Wissenschaft und Forschung spezialisierten Anbieter in der Schweiz waren bisher der SWTR, TA-Swiss und – stärker interessenpolitisch orientiert – die Akademien der Wissenschaften Schweiz:⁵

- Der SWTR als beratende Instanz des Bundesrates evaluiert vor allem Institutionen im Wissenschafts- und Forschungsbereich, die mit Bundesmitteln unterstützt werden. Dazu zählen namentlich die im Forschungsgesetz erwähnten spezialisierten Organe des Bundes (s. Kapitel 1.2). Zur Erfüllung dieser Aufgabe ist dem SWTR das Zentrum für Wissenschafts- und Technologiestudien CEST angegliedert (Simon/Burla 2004).
- TA-Swiss beschäftigt sich mit Technologiefolgen-Abschätzung und kann demnach als ein Evaluationsanbieter im Bereich Technologie verstanden werden. TA-Swiss ist weitgehend autonom in der Definition der zu bearbeitenden Themen. Methodisch folgt TA-Swiss mehrheitlich einem qualitativ-partizipatorischen Ansatz, der die Mitwirkung der BürgerInnen hoch gewichtet (z. B. Publiforen). TA-Swiss war bis Ende 2007 dem SWTR

⁵ Das Organ für Akkreditierung und Qualitätssicherung der schweizerischen Hochschulen *oag* beschäftigt sich in erster Linie mit der Evaluation und Akkreditierung der akademischen Lehre und wird deshalb an dieser Stelle nicht behandelt.

angegliedert. Auf Beginn 2008 wurde die Stelle zu den Akademien der Wissenschaften Schweiz umplaciert.

- Die *Akademien der Wissenschaften Schweiz*⁶ sind eine interessenpolitische Vertretung der wissenschaftlichen Gemeinschaft, die über ihre Fachgesellschaften in den Akademien Einsitz hat. Als Initiantin von Evaluationen hat sich besonders die Akademie der Geistes- und Sozialwissenschaften hervorgetan (siehe oben, Abschnitt 2.1). Daraus ist unter anderem das sozialwissenschaftliche Schwerpunktprogramm *Demain la Suisse – Zukunft Schweiz* entstanden. Mit der Angliederung von TA-Swiss gewinnen die Akademien zusätzliches evaluatorisches Profil.

Seit März 2007 tritt zusätzlich der SNF als Anbieter von Evaluationen auf. Er kann auch Forschungsvorhaben der öffentlichen Hand evaluieren, die nur teilweise oder überhaupt nicht mit SNF-Mitteln finanziert wurden (SNF 2007). Dabei steht die wissenschaftliche Evaluation im Vordergrund, die als „Kernaufgabe“ verstanden wird. Das Management der Evaluation kann durch eine andere Stelle erfolgen, sofern damit nicht die Qualitätskriterien des SNF in Frage gestellt werden. Für diese Art von Evaluationen erwartet der SNF eine Abgeltung seiner zusätzlich entstehenden Aufwendungen.

3 Stand und Perspektiven

3.1 Forschungsevaluation heute

Will man den aktuellen Stand der Evaluation von Forschung und Technologie in der Schweiz knapp umreissen, so kann man das auf folgende Kurzformel bringen: Sie ist *höchst disparat, meist punktuell, wenig institutionalisiert und kaum öffentlich*.

Die bisherigen Ausführungen haben gezeigt, dass die Aktivitäten in diesem Evaluationsbereich zwar quantitativ überblickbar, in ihrer Anlage, der Wahl des Gegenstandes und in Bezug auf Träger und durchführende Stellen aber *höchst disparat* sind. Sowohl an den Hochschulen wie bei den Forschungsorganen des Bundes wird gelegentlich Forschung evaluiert. Dies geschieht jedoch weitgehend ohne Koordination, ohne ein gemeinsames methodisches Gerüst, ohne inhaltliche Absprachen und ohne Weitergabe von Know-how. Es existiert auch kein Kontakt- oder Austausch-Forum, das zur Qualitätssicherung von Evaluationen in diesem Bereich beitragen könnte. Damit ist nicht gesagt, dass die realisierten Evaluationen im Allgemeinen qualitativ ungenügend seien. Bloss entsteht Qualität eben nicht aufgrund eines gezielten Prozesses, sondern eher situativ dank günstiger Umstände (z. B. gelungene Kooperation zwischen Nachfrager und durchführender Stelle).

Der Forschungsevaluation in der Schweiz mangelt es des Weiteren an Kontinuität. Selbst dort, wo (Beispiel SWTR/CEST) regelmässig evaluiert wird, ist ein eigentliches Evaluationsprogramm und dessen Umsetzung kaum erkennbar. Forschungsevaluation ist *meist punktuell* auf den jeweils aktuellen Bedarf ausgerichtet.

⁶ In diesem Verbund haben sich die Akademien der Geistes- und Sozialwissenschaften, der Naturwissenschaften, der medizinischen und der technischen Wissenschaften zusammengeschlossen.

Ausserhalb des Bereichs SWTR/TA-Swiss gibt es keinen definierten institutionellen Ort für Forschungsevaluation. Sowohl auf der Nachfrager- wie auf der Anbieterseite ist Forschungsevaluation eines von mehreren Tätigkeitsfeldern, dem man sich widmet, wenn Bedarf danach besteht. Ist ein Gebiet so *wenig institutionalisiert*, so stellen sich unter anderem Fragen der Sichtbarkeit und Transparenz.

In der Tat ist Forschungsevaluation zu weiten Teilen *kaum öffentlich* zugänglich oder dokumentiert. Dort, wo sie stattfindet, ist sie fast immer den direkt betroffenen Kreisen vorbehalten. Ausnahmen stellen die Evaluationen auf der Politikebene dar, die von der Natur ihrer Auftraggeber her (Geschäftsprüfungskommission Nationalrat) oder aus einer gegebenen Interessenkonstellation heraus (SAGW) publiziert wurden und zum Teil beträchtliche Resonanz erzeugt haben. In den meisten Fällen jedoch legen die Nachfrager der Evaluation wenig bis keinen Wert auf Öffentlichkeit. Das ist dort irritierend, wo die Forschung – und die Evaluation – mit öffentlichen Geldern finanziert wird. Die evaluatorischen Kernaspekte von Wirksamkeit, Effektivität und Effizienz werden hier mit Bezug auf den Einsatz von Steuergeldern untersucht und sollten deshalb öffentlicher Sichtbarkeit nicht entzogen werden.

3.2 Forschungsevaluation: Die Zukunft

Die Evaluation von Forschung und Technologie in der Schweiz hat zweifellos Entwicklungspotential. Geht man davon aus, dass eine Entwicklung wünschbar ist, so sind folgende Schritte ins Auge zu fassen:

- *Qualitätskriterien und Verfahrensstandards* sind eines der wichtigsten Instrumente zur Professionalisierung des Feldes. Einen Ausgangspunkt dafür bilden die Evaluationsstandards der SEVAL (Widmer et al. 2000), die aber möglicherweise für den spezifischen Bereich präzisiert oder angepasst werden müssen.
- Damit sich die Forschungsevaluation entwickeln kann, muss sie aus der Grauzone der Halb-Privatheit heraustreten oder herausgeholt werden. Fragestellungen, Verfahren und Ergebnisse sollten in der interessierten *Öffentlichkeit* diskutiert werden können.
- Die Diskussion sollte nicht an den Landesgrenzen Halt machen. *Internationale Vernetzung* ist ein wesentliches Element der Qualitätssicherung und der Entwicklung allgemein gültiger und weitherum verwendbarer Standards.
- Stabilisierung und Kontinuität sind nicht ohne eine gewisse *Institutionalisierung* erreichbar. Die Einrichtung auf Evaluation spezialisierter Institute oder Lehrstühle an den Universitäten und/oder Fachhochschulen ist ein erster wichtiger Schritt in diese Richtung.
- Schliesslich ist auch Forschungsevaluation nicht umsonst zu haben. Hier wie generell bei Evaluationen gilt: Soll sie seriös gemacht werden können, ist ein angemessenes *Budget* im Rahmen der zu evaluierenden Gegenstände einzuplanen, seien dies Projekte, Programme, Institutionen oder Politiken.

Evaluation in den Bereichen Forschung, Technologie und Innovation in Deutschland, Österreich und der Schweiz

Die Beiträge, welche die Situationen bezüglich des Zustandes und aktuellen Trends der Evaluation von Forschung und technologischer Leistungsfähigkeit sowie von Forschungs-, Technologie- und Innovationspolitik (FTI-Politik) in Deutschland, Österreich und der Schweiz nachzeichnen (Stefan Kuhlmann, Zinöcker/Dinges und Farago, alle in diesem Band), geben einen sehr guten Blick auf die Gemeinsamkeiten wie auch die nicht unbeträchtlichen Unterschiede zwischen den Ländern. Sowohl erstere als auch letztere spiegeln die jeweiligen institutionellen Verfaßtheiten (Grad und Ausmaß föderaler Politikorganisation im Bereich der FTI-Politik, Einbindung und Ausmaß der Orientierung an der Politik der Europäischen Union etc.), den Grad der Ausdifferenzierung und Arbeitsteilung zwischen den Akteuren im gegenständlichen Politikbereich sowie den Stand der ‚Evaluationskultur‘ (Ausmaß der Öffentlichkeit, der Internationalität und des methodischen und fachlichen Austausches zwischen den verschiedenen mit Evaluationen befaßten Akteursgruppen) wider und sind damit auch ein Abbild für und Ausdruck von jeweiligen politischen Kulturen.

1 Ausgangslagen

Die im folgenden gegenübergestellten Beobachtungen sind vor dem Hintergrund zu sehen, daß den Autoren eine unterschiedlich gut ausgeprägte empirische Basis zur Verfügung stand: Während in Deutschland und Österreich ein relativ guter Überblick über die im letzten Jahrzehnt im Bereich FTI und FTI-Politik durchgeführten Evaluationen besteht, der nicht zuletzt auch durch die entsprechenden Aktivitäten der Fachvereinigungen (DeGEval in Deutschland, Plattform FTEval in Österreich) befördert wurde, fehlt ein solcher Überblick offenbar für die Schweiz. Dieser wäre nicht nur eine lohnende selbstreflexive Aufgabe für die Schweizer Evaluationsforschung, sondern auch für die Schweizer FTI-Politik eine gute Handhabe zur Identifikation von Schwachstellen und zur Steuerung zukünftiger eigener Anstrengungen.

In allen entwickelten Industrieländern geht das Wachstum und der Bedeutungszuwachs der FTI-Politik mit einer Ausdifferenzierung der Akteure und Politikinstrumente und mit einem *wachsenden Bedarf nach ‚evaluativer Information‘* (Stefan Kuhlmann in diesem Band) einher. Die vielfach (in diesem Band von Farago) beklagte Vielzahl von forschungsbezogenen Institutionen als Schwierigkeit für die Evaluation ist gleichzeitig eine wesentliche Bedingung der wachsenden Evaluationsnachfrage! Allerdings unterscheidet sich die Schweiz in einigen Punkten von Deutschland und Österreich, die wiederum eher repräsen-

tativ für allgemeine Entwicklungen im Bereich Evaluation scheinen. Gemeinsamkeiten wie Unterschiede zwischen den drei Ländern sollen im Folgenden entlang verschiedener Dimensionen der Evaluationspraxen betrachtet werden.

2 Evaluationspraxen

In den drei Ländern ist der *rechtliche Rahmen* für Evaluationen unterschiedlich stark (bzw. schwach) ausgeprägt. Überall existieren zumindest allgemeine öffentliche Evaluationsaufträge, was generelle Vorgaben für Mittelverwendungen, aber auch was die wissenschaftliche Forschung und die Universitäten angeht. In Bezug auf einzelne Materien (Technologiefolgenabschätzung, Stammzellenforschung) existieren zum Teil auch konkretere gesetzliche Aufträge. In Deutschland und Österreich gibt es auch im Bereich der angewandten Forschung und Innovation einen guten normativen Rahmen und zwar sowohl in Form von entsprechenden Gesetzesabschnitten als auch einschlägigen Empfehlungen und Initiativen (in Österreich etwa des Rates für Forschung und Technologieentwicklung [RFTE]). Zudem existieren durch die Evaluationsgesellschaften definierte ‚Standards‘ als ‚Code of Conduct‘, die ebenfalls einige Bindungswirkung aufweisen.

Art und Häufigkeit von Evaluationen scheinen zwischen den Ländern deutlich zu differieren, wenn auch in den verschiedenen Ebenen von Evaluation in unterschiedlichem Ausmaß:

So ist etwa die *Evaluation von Forschungsprojekten* in allen betrachteten Ländern nach Einschätzung der Berichterstatter sehr gut ausgebaut: in der Begutachtung von wissenschaftlichen Forschungsprojekten in aller Regel durch *ex-ante* Peer Review, zum Teil mit hoher internationaler Komponente (auch wenn die Grenzen dieser Evaluationsform durchaus gesehen werden, vgl. etwa Zinöcker/Dinges in diesem Band). Was die Förderung von unternehmensnaher und angewandter Forschung angeht (eine Förderungsart, die in Deutschland und Österreich eine längere Tradition hat, in der Schweiz aus ordnungspolitischen Gründen kaum vorkommt), sind die Formen der *ex-ante* Projektselektion in den letzten Jahren stärker ausdifferenziert worden und greifen auf eine wachsende Zahl von Indikatoren zurück, die Projektdurchführbarkeit und Erfolgchancen abschätzbar machen sollen.

Die *ex-post* Evaluation von Projekten hat in den letzten Jahren international wie auch in den betrachteten Ländern zugenommen. Dabei werden zunehmend (etwa beim Schweizerischen Nationalfonds [SNF] in der Schweiz, bei der Deutschen Forschungsgemeinschaft [DFG] in Deutschland oder beim Fonds zur Förderung der wissenschaftlichen Forschung [FWF] in Österreich) auch andere Dimensionen als die Qualität des Forschungsausgangs betrachtet: zum Beispiel werden die Auswahlprozeduren selbst auf systematische Verzerrungen – etwa bezüglich Geschlecht und Alter – hin untersucht und bewertet.

Institutionen und ihre Leistungsfähigkeit waren in den letzten Jahren offenbar in allen Ländern immer wieder zentraler Ansatzpunkt für Evaluationen: Dies ist insbesondere an den Hochschulen stark ausgebaut und zum Teil auch institutionalisiert worden. Im Unterschied zu anderen Institutionen, wurde an den Hochschulen sehr stark auf die Fähigkeit zur Selbstevaluation gesetzt, allerdings stark auf die Lehre und weniger auf die Forschungsleistung bezogen. Andere Institutionen-Evaluationen (etwa von außeruniversitären For-

schungseinrichtungen oder Finanzierungsagenturen für Forschung und technologische Entwicklung) werden dagegen meist sehr umfassend angelegt, weil sie vom Auftraggeber auch häufig mit der Frage nach strategischer Ausrichtung der Institution beziehungsweise ihrer Weiterführung respektive Schließung verbunden werden (etwa die großen deutschen Institutionen-Evaluationen in den 1990er Jahren).

Zumindest in Deutschland und Österreich sind *Programmevaluierungen* sehr häufig, fast regulär geworden. In wesentlich geringerem Ausmaß scheint dies für die Schweiz zu gelten, wo Farago (in diesem Band) beobachtet, daß oft eine ganze Programmart statt des Einzelprogramms evaluiert wird – ein wenig verbreiteter Zugang in Deutschland und Österreich.

Auf der Ebene der *FTI-Politik* finden sich in allen Ländern zwar einzelne Beispiele und Versuche, entsprechende Evaluationen sind aber eher die Ausnahme und sehen sich zudem Schwierigkeiten gegenüber, auf die unten noch eingegangen wird.

Auf den ‚Märkten‘ für Evaluationen gibt es Unterschiede in der *Anbieterstruktur* und Institutionalisierung: In der Schweiz sind es meist kleine private Institutionen, die sich auf Evaluation (wenn auch nicht ausschließlich in FTI) spezialisiert haben oder Universitätsinstitute, für die Evaluation nur einen kleinen Teil der Tätigkeiten darstellt. In Deutschland finden sich unter den Anbietern sowohl Institutionen mit universitärer Anbindung als auch außeruniversitäre Forschungseinrichtungen. In Österreich ist wiederum eine nahezu vollständige Absenz der Universitäten (mit der Ausnahme von sie selbst untersuchenden Einrichtungen) bei einer Dominanz von privaten Anbietern und außeruniversitärer Forschung festzustellen. In allen Fällen (überraschenderweise auch in Deutschland) ist die Zahl der Anbieter relativ klein. Implikationen für die Wettbewerbsintensität der „Evaluationsmärkte“ sind wohl am besten durch transparente und konsequent wettbewerbliche Ausschreibungen sowie durch eine „Internationalisierung“ der Evaluationsmärkte zu kontrollieren (siehe auch weiter unten).

Institutionalisierungen von Evaluationen sind ungeachtet der Unterschiede in den rechtlichen Rahmenbedingungen in allen drei Ländern schwach ausgeprägt. Sie finden sich jeweils im Bereich der Hochschulen angesiedelt (Organ für Akkreditierung und Qualitätssicherung der Schweizerischen Hochschulen [OAQ] in der Schweiz, Österreichische Qualitätssicherungsagentur [AQA] in Österreich, Institut für Forschungsinformation und Qualitätssicherung [iFQ] in Deutschland), auch wenn die konkreten Aufgabenzuschreibungen leicht differieren. Abgesehen von Institutionalisierungen von Technology Assessment (das Farago in diesem Band zu Recht im weiteren Sinn zu den Evaluationsinstitutionen zählt), das ebenfalls in allen drei Ländern existiert, kennt nur die Schweiz mit dem beim Schweizerischen Wissenschafts- und Technologierat (SWTR) angesiedelten Zentrum für Wissenschafts- und Technologiestudien (CEST) eine weitere Institutionalisierung von Evaluation in Form einer öffentlichen Stelle. Eine zusätzliche institutionelle Eigenheit ist, daß in der Schweiz auch der SNF als Anbieter von Auftragsstudien im Bereich der wissenschaftlichen Evaluation auftritt, zumindest solange kein Interessenskonflikt auftritt.

3 Resümee des Vergleichs der drei Länder

Der Betrag von Farago bringt die Situation in der Schweiz auf die Kurzformel: ‚höchst disparat‘, ‚meist punktuell‘, ‚wenig institutionalisiert‘ und ‚kaum öffentlich‘. Von einer solchen Beschreibung unterscheiden sich die anderen Länder insgesamt doch deutlich: Stefan Kuhlmann etwa beschreibt die Situation in Deutschland als ‚entwickelt‘ und ‚dynamisch‘, gleichzeitig aber auch ‚unsystematisch‘ und ‚zersplittert‘, womit er die Situation in Deutschland, zumindest was die Häufigkeit, die methodische Entwicklung und wohl auch was die Öffentlichkeit angeht, in einer Weise beschreibt, die sich von der Schweizer Zustandsbeschreibung doch deutlich positiv abhebt. Aber auch die österreichische Praxis unterscheidet sich in der Häufigkeit und der Öffentlichkeit von der der Schweiz offenbar doch recht stark: Zum einen ist Evaluation in Österreich wie in Deutschland eine häufiger geübte Praxis, und zwar sowohl auf der Programm- als auch auf der Politikebene. Zum anderen ist auch der Grad der Öffentlichkeit von Evaluationen in Österreich im Vergleich mit den anderen zwei Ländern etwas höher.

Was den Grad der Öffentlichkeit, der Sichtbarkeit und der Transparenz von Evaluationen angeht, der nicht nur bei Farago für die Schweiz, sondern auch im Beitrag von Stefan Kuhlmann für Deutschland als zu gering beklagt wird, wäre zu fragen, was denn verschiedene Ausprägungen des Öffentlichkeitsbegriffs sein könnten. Obwohl etwa Österreich nicht das Ausmaß der Öffentlichkeit von Evaluationsstudien erreicht, das in den nordischen Ländern vorherrscht, finden sich in öffentlichen Dokumentationen von Evaluationsstudien eine nicht unbeträchtliche Zahl, die auch öffentlich zugänglich gemacht wurde. Wie in Deutschland gibt es aber auch in Österreich eine nicht unbeträchtliche Zahl von Studien, die in der alleinigen Verfügung des Auftraggebers verbleiben. Auch diesbezüglich ergäbe sich eine mögliche Rolle für die Fachvereinigungen in der Sensibilisierung der politischen und fachlichen Öffentlichkeit für die Bedeutung einer umfassenden Verfügbarmachung von Evaluationsstudien. Diese allein ist eine notwendige aber keine hinreichende Bedingung für eine öffentliche Diskussion. Dazu bedarf es weiters entsprechender Foren und Formen öffentlichen Diskurses. Ein solcher ist meiner Ansicht nach nur in Bezug auf sehr wenige Evaluationen im Bereich der FTI-Politik in den letzten Jahren zustande gekommen.

4 Aktuelle Trends und Herausforderungen

Der Trend in allen Ländern (über die hier betrachteten drei hinaus) geht eindeutig in Richtung der ‚*Internationalisierung*‘ von Evaluationen durch ausschließliche oder teilweise Vergabe des Evaluationsauftrages an internationale ExpertInnen. Dies ergibt sich als Notwendigkeit in der Schweiz und in Österreich schon aus der geringen Größe der Länder und der damit oft verbundenen Rollenkonflikte und Unvereinbarkeiten für die beschränkte Zahl der Akteure: im Peer Review Verfahren etwa durch Konkurrenzsituationen, oft auch durch den Rollenkonflikt des potentiellen Evaluators, wenn die entsprechende Einrichtung an der Formulierung und/oder Umsetzung der zu evaluierenden Programme oder Politiken beteiligt war. Aber auch in Deutschland ist offenbar die Zahl der Anbieter von Evaluationen

nicht sehr groß und eine Internationalisierung schon aus Gründen der Aufrechterhaltung von hinreichender Wettbewerbsintensität geboten. Hier käme – neben der Funktion des regelmäßigen fachlichen Austausches – den Evaluationsgesellschaften beziehungsweise den einschlägigen Kapiteln dieser Organisationen auch eine positive Rolle in der Bekanntmachung und Verbreitung entsprechender Ausschreibungen und in der Unterstützung bei der Partnersuche für gemeinsame Evaluationsstudien zu.

Etwas Skepsis muß bezüglich der Forderung nach stärkerer *Institutionalisierung* von Evaluation angemeldet werden. Wenn damit die stärkere normativ-legistische Verankerung oder einfach die häufigere Durchführung von Evaluationen gemeint sein sollte, kann man dieser Forderung für alle drei Länder etwas abgewinnen – auch wenn beides in Österreich und Deutschland weniger ein Problem scheint als in der Schweiz. Wenn damit die Einrichtung von ‚hauptverantwortlichen‘ öffentlichen Institutionen (ähnlich der Institutionen, wie sie in allen drei Ländern für die Hochschulen geschaffen wurden) gemeint sein sollte, dann ist der Vorteil der Schaffung von kritischer Masse und ausreichender Kapazität sowie die eindeutige Benennung von Verantwortlichkeiten für Evaluationen abzuwägen gegen die Vorteile eines Settings mit mehreren unabhängigen Anbietern von Evaluationen was Konkurrenz und Methodenvielfalt angeht.

Die Art der Nachfrage geht – nicht zuletzt unter dem Eindruck des New Public Management (NPM) – immer stärker auch in Richtung *ergebnis(output/outcome/impact) orientierter Evaluation*. Dies etwas stärker ausgeprägt in Deutschland und Österreich: nicht weil dort NPM stärker das Verwaltungshandeln bestimmen würde als in der Schweiz, sondern weil in diesen zwei Ländern die Förderung von Forschung und Entwicklung viel stärker auf die Unterstützung der Unternehmen abstellt. In Bezug auf deren Förderung wird natürlich viel stärker substantieller Impact in Form von gesteigerter Innovationsleistung und der damit verbundenen privaten und sozialen ökonomischen Erträge angestrebt (siehe Polt/Vonortas 2006). Dieser Nachfragetrend hat die methodischen Anforderungen an Evaluationen stark erhöht, bislang existieren aber in allen drei Ländern nur wenige Beispiele für umfassende Impact-Studien.

Allen drei Ländern gemeinsam ist die Beobachtung, daß *Indikatoren und Daten für Evaluationen* sehr häufig ad hoc erhoben werden (müssen). Während sowohl die Forschungseinrichtungen als auch die Forschungs- und Innovationsfinanzierungsagenturen in den letzten Jahren ihre internen Projektbewertungs- und Monitoringsysteme systematisch ausgebaut haben, fehlt bislang in allen Ländern ein harmonisierter Rahmen dieser Datengenerierung, so daß in aller Regel schon der Vergleich zwischen verschiedenen Institutionen scheitert. Auch eine von Stefan Kuhlmann zu Recht eingemahnte weitgehende Aggregierbarkeit von evaluativer Information, die nötig ist, um von der Betrachtung des jeweiligen Einzelfalles zu einer Systembetrachtung (oder gar zu einem internationalen Vergleich) kommen zu können, fehlt weitgehend. Hier ist der Vergleich der drei Länder nicht nur instruktiv sondern auch repräsentativ: Erst Fortschritte auf diesem Weg werden ‚Politiklernen durch Evaluationen‘ aus dem gegenwärtigen Zustand des vielfach anekdotischen heben können.

In einer Vielzahl von Ländern ist in den letzten zehn Jahren zudem ein *Trend Richtung ‚systemischer‘ Evaluation* festzustellen. Die Fragestellungen gehen bei dieser Art von Evaluation dahin

- eine Institution, ein Programm, eine Maßnahme oder eine bestimmte Politik nicht nur isoliert, sondern im Kontext mit und im Vergleich zu anderen Institutionen, Programmen, Maßnahmen und Politiken zu bewerten;
- die Leistung von (regionalen, nationalen, sektoralen, wissenschaftsdisziplinären) Systemen auf einer höheren Aggregationsebene zu bewerten und dazu geeignete Metriken und Indikatorensysteme zu entwickeln.

In beiden Fällen kamen dazu von der Europäischen Union und den dortigen Evaluationspraktiken beziehungsweise Methodendiskussionen wichtige Anstöße (vgl. Polt et al. 2002 und Polt/Vonortas 2006).

Gesundheit

Evaluation im Bereich Gesundheit – Beispiel Deutschland

1 Einleitung: Zum Kontext von Evaluation im Gesundheitsbereich

Einer der wichtigsten Grundsätze guter Evaluation ist, dass sie gegenstandsangemessen konzipiert und durchgeführt werden sollte. In den Evaluationsstandards, wie sie im internationalen Bereich entwickelt und von der Deutschen Gesellschaft für Evaluation für Deutschland und Österreich adaptiert wurden (vgl. DeGEval 2002) kommt dies in den Kriterien der Durchführbarkeit und Fairness zum Ausdruck: Die Verfahren sollen angemessen sein (Standard D1), von den Beteiligten und Betroffenen akzeptierbar sein (Standard D2) und deren unterschiedliche Sichtweisen in Rechnung gestellt werden (Standard F4). Evaluation kann also nicht nach einem Einheitsschema, in methodischem Monismus, über alle Praxisbereiche gestülpt werden, sie muss sich auf den konkreten Kontext einlassen.

Was sind nun die konkreten Kontextbedingungen im Gesundheitsbereich? Ich möchte hier sechs Punkte anführen, die entscheidend für die Planung von Evaluationsprojekten sind:

- Im Gesundheitssektor wird mit und am Menschen gearbeitet. Dies unterscheidet ihn beispielsweise vom Wirtschaftsbereich. Wir haben es hier vorwiegend nicht mit Gütern, Waren, Produkten zu tun. Deshalb ist dieser Bereich besonders sensibel.
- Aus diesem Grunde wird Gesundheitsversorgung heute vorwiegend als öffentliche Aufgabe definiert, ähnlich wie Bildung. Die im Gesundheitsbereich Tätigen tragen öffentliche Verantwortung; es wird mit Steuergeldern agiert.
- Die Folgen von Fehlentscheidungen sind im Gesundheitsbereich meist deutlich und drastisch, sie äußern sich in subjektivem Leiden. In der Regel zeigen sich diese negativen Folgen unmittelbar, nicht wie im Bildungsbereich (die Folgen einer Ohrfeige für den Bildungsverlauf einer Person zu evaluieren stellt sich dort als äußerst schwierig dar).
- Das Feld der im Gesundheitsbereich Arbeitenden ist interdisziplinär gestaltet. Wir haben es hier mit ÄrztInnen, Pflegepersonen, PsychologInnen, TherapeutInnen, SozialarbeiterInnen und weiteren Berufsständen mit unterschiedlicher Ausbildung und Berufsstatus zu tun. Dies führt dann oft zu Abstimmungsproblemen. Beispielsweise bei Wissenschaftsevaluation ist die Situation einfacher, da es eine relativ standardisierte Ausbildung zum Wissenschaftler/zur Wissenschaftlerin gibt.
- Im Gesundheitsbereich existiert ein Bestand hoch spezialisierten professionellen Wissens, das nach streng methodischen Prozeduren laufend erweitert und überprüft wird

(vgl. unten Cochrane-Library), ähnlich wie im Wissenschaftsbereich. Daneben jedoch steht das Laiensystem, das Alltagswissen über Gesundheit, die Alltagskompetenzen, beispielsweise in der Familie, die Gesundheitsversorgung selbst in die Hand zu nehmen. Jedes Handeln im Alltag ist auch Gesundheitshandeln, hat Auswirkungen auf das Befinden.

- Die in der Gesundheitsversorgung Tätigen agieren, auch aus der Geschichte verständlich, zwischen zwei Polen: einerseits wird der Gesundheitsbereich als „Hightech“-Bereich begriffen, in dem Apparatemedizin und hoch spezialisiertes Krankenhaus den Kern bilden; andererseits existiert das Bild der ärztlichen Tätigkeit als Kunstlehre, die therapeutische Beziehung als einfühlsamer, persönlicher Bereich mit gewissen Ähnlichkeiten zum Bildungsbereich, jedoch extremen Unterschieden zum Wissenschafts- oder Wirtschaftsbereich.

2 Zur Begriffsbestimmung von Gesundheit

Zum Kontext von Evaluation im Gesundheitsbereich gehört auch eine Begriffsbestimmung von Gesundheit. Erst wenn die zentrale Zielgröße der verschiedenen Maßnahmen genau definiert ist, lassen sich Evaluationskonzepte entwickeln.

Es zeigt sich dabei, dass dies gar nicht so einfach ist (vgl. zum Folgenden Faltermaier 2005). Der traditionelle, in Mittelalter und Neuzeit entwickelte Gesundheitsbegriff definiert sie als Restkategorie (Abwesenheit von Krankheit). Die Weltgesundheitsorganisation WHO hat demgegenüber bereits bei ihrer Gründung (1946) ein positives Konzept gesetzt: Gesundheit ist ein Zustand des vollkommenen körperlichen, geistigen und sozialen Wohlbefindens und nicht nur die Abwesenheit von Krankheit und Gebrechen (WHO-Präambel; vgl. www.who.int). Diese Definition ist aber in der Folge als utopisch kritisiert worden. In der sich entwickelnden interdisziplinären Gesundheitswissenschaft wurde der Begriff weiter ausgeweitet und nicht nur gesundheitliches Befinden, sondern auch die kognitive Ebene (Gesundheitseinschätzungen) und Handlungsebene (Krankheitsbewältigung, Gesundheitserhaltungskompetenzen) thematisiert (vgl. Faltermaier 2005). Damit ist Gesundheit ein schillernder Begriff geworden, der es in Evaluationsstudien erschwert, das Bewertungskriterium festzulegen.

3 Die verschiedenen Felder im Gesundheitsbereich

Weiter erschwerend kommt hinzu, dass die Aktivitäten im Gesundheitsbereich, die es zu evaluieren gilt, völlig unterschiedlicher Natur sind. Man kann hier mindestens 14 Felder unterscheiden, die zum Gegenstand von Evaluationen gemacht werden.

Tabelle 1: Evaluationsgegenstände in 14 Feldern

| | |
|---------------------------------|---|
| Ärztliche Versorgung | 1. Medikamente |
| | 2. Heilverfahren |
| | 3. Pflege |
| | 4. Krankenhäuser |
| | 5. Arztpraxen |
| Psychotherapeutische Versorgung | 6. Therapiezulassung |
| | 7. Therapieerfolg |
| | 8. Freier Therapiemarkt |
| Psychosoziale Versorgung | 9. Gemeindenahe Versorgung |
| | 10. Beratung |
| Prävention | 11. Betrieblicher Bereich |
| | 12. Schulischer Bereich, Gesundheitserziehung |
| | 13. Sport |
| | 14. Übergreifend („Public Health“) |

Bei der nun folgenden Analyse des Standes der gesundheitsbezogenen Evaluationen in Deutschland müssen wir uns jeweils auf ein konkretes Praxisfeld beziehen, da die Verfahren und Kriterien hier höchst unterschiedlich sind. Wir werden uns dabei auf die zentralen Felder beschränken müssen.

3.1 Evaluation der Wirkung von Medikamenten

Hier werden nicht nur in Deutschland die härtesten Evaluationskriterien und Evaluationsverfahren zu Grunde gelegt. Das „Gesetz über den Verkehr mit Arzneimitteln“¹, erlassen 1961 auch in Reaktion auf den Conterganskandal (drastische Nebenwirkungen eines in der Schwangerschaft eingenommenen Schlaf- und Beruhigungsmittels auf das ungeborene Leben), regelt die Zulassung von Medikamenten. Es legt fest, dass nur solche Medikamente in Deutschland in den Verkehr gebracht werden dürfen, die eine Zulassung einer speziellen Bundesbehörde besitzen bzw. über eine Zulassung der Europäischen Kommission bzw. des Rats der Europäischen Union verfügen. Die deutsche Bundesbehörde ist derzeit das Bundesinstitut für Arzneimittel und Medizinprodukte BfArM in Berlin (www.bfarm.de). Die Zulassung wird von pharmazeutischen Unternehmen beantragt, die eine behördliche Zulassung zur Herstellung von Medikamenten besitzen. Dem Antrag sind Unterlagen über die Bestandteile des Arzneimittels, die Darreichungsform, die Wirkungen, die Anwendungsgebiete, die Gegenanzeigen, die Nebenwirkungen, die Wechselwirkungen mit anderen Mitteln, die Dosierung, die Anwendbarkeit, die Haltbarkeit, die Umweltrisiken von Produktion und Lagerung sowie die Methoden zur Kontrolle der Qualität (Kontrollmethoden) beizufü-

¹ Siehe http://bundesrecht.juris.de/amg_1976/index.html

gen. Darüber hinaus müssen analytische, pharmakologische und (zentral) klinische Prüfungen vorgenommen sein und mit Sachverständigengutachten belegt werden (§ 21 AMG). Für diese klinischen Studien in der Medikamentenevaluation ist ein mehrstufiges Verfahren vorgeschrieben:

- An gesunden Freiwilligen wird die Wirkung (Nebenwirkung) des Präparats im menschlichen Organismus untersucht.
- An einer Patientengruppe wird die Wirkung verschiedener Darreichungsformen verglichen.
- In Doppelblindstudien nach experimenteller Logik wird die Wirkung im Vergleich mit Kontrollgruppen überprüft.
- Unter Praxisbedingungen wird die Markteinführung des Präparates über eine längere Zeit hinweg beobachtet.

Das gesamte Verfahren muss von einer Ethikkommission genehmigt werden.

Für die dritte Phase, die Doppelblindstudien, sind genauere Erläuterungen notwendig. Es handelt sich hier um ein experimentelles Design, in dem Personen durch Zufall (Randomisierung) einer Untersuchungsgruppe beziehungsweise einer Kontrollgruppe zugeordnet werden (Randomized Controlled Trial RCT). Dadurch soll erreicht werden, dass die Gruppen vergleichbar sind, keine systematischen Unterschiede aufweisen und somit eine gegebenenfalls beobachtete Gesundheitsverbesserung ausschließlich auf das Medikament zurückzuführen ist (interne Validität). Placebogruppen sind Kontrollgruppen von Personen, die ein Präparat ohne Wirkung bekommen, ohne dies zu wissen. Auch der Versuchsleiter (Arzt) weiß nicht, ob er ein Placebo oder das zu testende Medikament („Verum“) verabreicht (Doppelblindversuch). Die Anzahl der Versuchspersonen in den einzelnen Gruppen sollte nach neuesten statistischen Verfahren (Poweranalyse, Effektstärkenschätzung), am besten von biostatistischen Experten, vorausberechnet werden. Solche randomisierten Experimente sollten repliziert werden; nur wenn genügend Studien auf die beabsichtigte Gesundheitswirkung hinweisen, ist das Medikament akzeptabel. In eigens gegründeten Instituten werden weltweit Studien, die diesen Kriterien genügen, gesammelt und verglichen (Cochrane-Institute, z. B. Deutsches Cochrane Zentrum am Universitätsklinikum Freiburg²).

Auf den ersten Blick erscheint das Evaluationsverfahren extrem aufwändig. Man stelle sich ein ähnliches Vorgehen bei der Einführung neuer Unterrichtsmethoden vor! Experimentelle Designs werden allerdings von manchen Evaluationsexperten als einzig verlässlicher Wirkungsnachweis und damit beste Evaluationsmethode bezeichnet (z. B. Bortz/Döring 2006). Jedoch häuft sich die Kritik an diesem Verfahren:

- Das Verfahren ist zu lang (rund 18 Monate).
- Die Studien werden in der Regel von der Pharmaindustrie durchgeführt. Eine eigene staatliche Agentur, welche die Studien durchführt (wie z.B. in Österreich gegründet, in Deutschland in der Vorbereitung), könnte hier unabhängiger sein.

² Vgl. www.cochrane.de

- Oftmals sind die Standards nicht umsetzbar (Randomisierung scheitert oft am Einverständnis der Patienten; die notwendige Probandenanzahl ist nicht zu rekrutieren; ethische Einwände). Werden die Standards aber nicht voll eingehalten, sind die Ergebnisse eigentlich wertlos.
- Das experimentelle Design kann hier nur die rein stoffliche Wirkung eines Präparates unter Standardbedingungen überprüfen. Kontextbedingungen, situative oder personbezogene Besonderheiten sowie die Art der Arzt-Patient-Beziehung werden als Störvariablen abgetan (vgl. Elkeles 2006).
- Es wird nur einmalig die Zulassung geprüft; die weiteren Praxiserfahrungen sind nicht Gegenstand einer Evaluation.

3.2 Evaluation von Heilverfahren (inkl. Psychotherapie)

Das Evaluationsverfahren ist hier ähnlich, am experimentellen Design und der staatlich kontrollierten Zulassung orientiert, nur die Kriterien sind nicht ganz so hart wie bei der Medikamentenzulassung. Medizinische Heilverfahren werden in den Leistungskatalog der Krankenkassen aufgenommen. Richtungweisend sind hier die gesetzlichen Krankenkassen, die einen „Gemeinsamen Bundesausschuss der Ärzte, Zahnärzte, Psychotherapeuten, Krankenhäuser und Krankenkassen in Deutschland“³ unter Aufsicht des Bundesgesundheitsministeriums gebildet haben, wie es im V. Buch des Sozialgesetzbuches (§ 135)⁴ festgelegt wurde. Der diagnostische und therapeutische Nutzen, die medizinische Notwendigkeit und Wirtschaftlichkeit eines Heilverfahrens wird auf Antrag der Krankenkassen überprüft. Der Bundesausschuss hat eine Verfahrensordnung festgelegt, die zwar am experimentellen Design randomisierter kontrollierter Studien orientiert ist, aber auch schwächere „Evidenzstufen“ zulässt (Gemeinsamer Bundesausschuss 2005: 16):

- I a Systematische Übersichtsarbeiten von Studien der Evidenzstufe I b
- I b Randomisierte klinische Studien
- II a Systematische Übersichtsarbeiten von Studien der Evidenzstufe II b
- II b Prospektive vergleichende Kohortenstudien
- III Retrospektive vergleichende Studien
- IV Fallserien und andere nicht-vergleichende Studien
- V Assoziationsbeobachtungen, pathophysiologische Überlegungen, deskriptive Darstellungen, Einzelfallberichte, u. ä.; nicht mit Studien belegte Meinungen anerkannter Experten, Berichte von Expertenkomitees und Konsensuskonferenzen

Bei drängenden Problemen der Gesundheitsversorgung oder Schwierigkeiten in der Umsetzung experimenteller Designs sind auch niedrigere Evidenzstufen als die Idealforderung Ia zulässig.

³ Siehe www.g-ba.de

⁴ Siehe www.sozialgesetzbuch.de/gesetze/05

Exemplarisch war hier die Diskussion um die Kassenzulassung von Akupunktur zur Behandlung chronischer Schmerzen. Mehrere randomisierte Studien überprüften die Effizienz der Akupunktur, indem sie auch Placebogruppen (die PatientInnen wurden an „falschen“ Punkten akupunktiert) einbauten. Als weitere Kontrollgruppen wurden Patienten mit traditioneller medikamentöser Schmerzbehandlung eingeführt. Die Ergebnisse gingen in die Richtung, dass die Untersuchungsgruppen (Akupunktur), aber auch die Placebogruppen (falsche Akupunktur) Schmerzreduktionen zeigten (vgl. z. B. Endres et al. 2007). Nach experimenteller Logik wäre dies ein gescheiterter Wirkungsnachweis. Trotzdem hat der Gemeinsame Bundesausschuss 2006 Akupunktur bei chronischen Rücken- und Knie-schmerzen zugelassen und sich damit mit einer niedrigeren Evidenzstufe zufrieden gegeben, um chronischen Schmerzpatienten eine Hilfe zu gewähren, auch wenn sie nur auf dem Placeboeffekt basiert. An diesem Beispiel zeigt sich die ganze Problematik experimenteller Evaluationskonzepte in der Gesundheitsförderung.

Das Evaluationsverfahren bei Psychotherapie ist ähnlich. In dem seit 1998 gültigen Psychotherapiegesetz wurde festgelegt (§ 11), dass ein Wissenschaftlicher Beirat Psychotherapie⁵ die Wissenschaftlichkeit eines Verfahrens prüft. Dieser hat sich als Mindestanforderungen gestellt, dass empirische Studien zur Wirksamkeit vorliegen müssen, mit klar beschriebener Stichprobe, reliablen Methoden und Kontrollbedingungen. Danach sind bisher anerkannt: psychoanalytisch begründete Verfahren, verhaltenstherapeutische Verfahren sowie in bestimmten Bereichen Gesprächspsychotherapie. Systemische Therapie, Neuropsychologische Therapie und Psychodrama beispielsweise wurden abgelehnt, da zu wenig Wirkungsevidenz in systematischen Studien vorliege.

Hier ist die Kritik an der Art der experimentell orientierten Evidenzbasierung der Evaluation am stärksten, da es bei Psychotherapie noch stärker auf die (im experimentellen Paradigma als Placeboeffekt eingestufte) Therapeut-Patient-Beziehung ankommt. So wirft Kriz (2003) in seinem Gutachten zur Wissenschaftlichkeit in der Psychotherapie dem Wissenschaftlichen Beirat unzulässige Reduzierung der Wissenschaftlichkeit auf kontrollierte randomisierte Studien, ein einseitiges Labor-Wirksamkeitsmodell sowie eine Taylorisierung („Fließbandarbeitsteilung“) der Therapie vor. Legewie (2000) hat als Alternative ein Verfahren qualitativ-quantitativer Therapieevaluation (regionale Prozess-Erfolgsstudien) vorgeschlagen (nach Kriz 2003). Andere empfehlen gar eine Abkehr vom Konzept der Evidence Based Medicine, hin zu einer „Narrative Based Medicine“, die eine systematische Sammlung von Fallexpertisen und klinischen Berichten umfasst (Greenhalgh 1999).

3.3 Evaluation von Krankenhäusern und Arztpraxen

Für bundesdeutsche Krankenhäuser ist laut Sozialgesetzbuch (§ 137 SGB V) ein Qualitätsbericht zumindest alle zwei Jahre vorgeschrieben, der Struktur- und Leistungsdaten sowie externe Qualitätssicherungsmaßnahmen inklusive einer Qualitätspolitik der Klinik vorschreibt. Eine Bundesgeschäftsstelle Qualitätssicherung wurde zur Unterstützung eingerichtet.

⁵ Vgl. www.wbpsychotherapie.de

tet.⁶ Die Situation ist hier besonders komplex, da im Krankenhaus auch organisatorische und ökonomische Aspekte in den Vordergrund rücken (vgl. Lauterbach/Schrapppe 2004). So wurden hier schon früh komplexe Evaluationsmodelle entwickelt (z. B. Struktur-, Prozess- und Ergebnisevaluation nach Donabedian, vgl. Badura/Siegrist 1999).

Auch Arztpraxen sind seit 1999 zur Einführung von Qualitätsmanagementsystemen verpflichtet (§ 135a SGB V). Allerdings gibt es erst seit 2006 eine Richtlinie, welche die Einführung eines internen Qualitätsmanagementsystems verlangt, dabei jedoch viel Spielraum lässt, nur die möglichen Ansätze und Ziele beschreibt (Ausrichtung an fachlichen Standards, Patientenorientierung, Mitarbeiterorientierung im Team, Fehleranalysen und Beschwerdemanagement, Dokumentation, Kooperation in der Versorgung⁷).

Verschiedene qualitätssichernde Maßnahmen wurden in diesen Bereichen eingeführt (vgl. Lauterbach/Schrapppe 2004; Hurrelmann/Laser 2003; Loidl-Keil/Laskowski 2006):

- Systematische Patientendokumentationssysteme sollen alle Behandlungsschritte erfassen.
- Normierte Qualitätsmanagementsysteme mit Zertifizierung wurden auf Krankenhausbedingungen und Arztpraxisbedingungen angepasst (z.B. DinEnIso, EFQM). Viele Kliniken verfügen bereits über eine solche Qualitätszertifizierung.
- Es wurden für spezifische Diagnosegruppen feste Behandlungsleitlinien entwickelt, an die sich der Arzt halten muss (Die Krankenkasse bezahlt nur nach Leitlinie). Dadurch soll ärztliches Handeln standardisiert werden. Leitlinien werden systematisch gesammelt und evaluiert (Deutsches Instrument zur methodischen Leitlinienbewertung DELBI⁸).
- Um die Vernetzung der Patientenversorgung, vor allem zwischen ärztlichen und pflegerischen Maßnahmen und zwischen verschiedenen Institutionen zu verbessern, wurde das Konzept der Versorgungspfade (Clinical Pathways) eingeführt, das im interdisziplinären Team Behandlungen plant und überwacht (vgl. Greiling 2004; Oberender 2005). Gerade das Ernstnehmen von Pflegemaßnahmen, das Einbinden in den Versorgungsprozess ist hier für Deutschland innovativ und geht einher mit einer Akademisierung der Pflege (vgl. Behrens/Langer 2004).
- Aus dem Wirtschaftsbereich werden Qualitätssicherungsmaßnahmen übernommen, wie die „Balanced Scorecard“, durch die Leistungen und Maßnahmen konzipiert werden als Balance zwischen finanzieller Perspektive, Kundenperspektive, interner Prozessperspektive und Innovationsperspektive.
- Ebenso aus dem Wirtschaftsbereich stammt die Konzeption der Qualitätszirkel, bei der die MitarbeiterInnen eines Krankenhauses oder einer Arztpraxis sich zusammensetzen und systematisch Stärken und Schwächen der eigenen Arbeit diskutieren.
- Schließlich ist die Entwicklung eines Leitbildes der Institution, meist professionell angeleitet und moderiert, zu nennen. Hier sollen die Ziele und Arbeitskonzepte der In-

⁶ Siehe www.bqs-online.com

⁷ Vgl. www.q-m-a.de

⁸ Vgl. www.versorgungsleitlinien.de

stitution gemeinschaftlich erarbeitet, festgeschrieben und laufend überarbeitet werden (vgl. Schubert 2005).

Aber auch hier ist die Diskussion um die geeigneten Evaluationsmittel groß. Patientendokumentationssysteme werden als die Basis einer Prozess- und Ergebnisevaluation bezeichnet, in der Praxis aber oft der enorme (Zusatz-)Aufwand beklagt. Leitlinien werden als der Schlüssel der Qualitätsentwicklung im Gesundheitswesen gelobt, von anderen als die Einführung einer „Knopfdruckhaltung“ statt einer ärztlichen Beziehung kritisiert (Loidl-Keil 2004). Insgesamt scheinen die gesetzlichen Auflagen zur Qualitätssicherung und Evaluation im Krankenhausbereich stärker befolgt zu werden als im Bereich freier Arztpraxen. Es zeigt sich eine Parallele zum Hochschulbereich, wo auch die gesetzlichen Evaluationsauflagen in den Hochschulgesetzen erst spät und zögerlich umgesetzt wurden und werden.

3.4 Der freie Markt der Gesundheitsversorgung

Natürlich können Patientinnen und Patienten auch Angebote zur Gesundheitsversorgung auf dem freien Markt (gegen Bezahlung) wahrnehmen, und hier gelten die Auflagen der Zulassung und Evaluation nicht in diesem Maße. Hier zeigt sich die Kehrseite der Gesundheitsevaluation. Beklagen sich viele über die strengen, einseitigen Auflagen bei der Zulassung, über den enormen Aufwand bei der Dokumentation und dem Verfassen von Qualitätshandbüchern und Berichten, so wird im Alternativmedizinbereich oft ganz auf Evaluation und Qualitätsmanagement verzichtet. Es werden Fallberichte vorgelegt (Frau X und Herr Y haben nach der Behandlung wunderbare Heilerfolge gezeigt!), ohne dass dabei die Kriterien guter Fallanalysen (vgl. Mayring 2002) eingehalten werden.

3.5 Psychosoziale Versorgung und Gesundheitsprävention

Neben der medizinischen Versorgung im stationären und ambulanten Bereich gibt es eine große Zahl interventiver und präventiver Gesundheitsmaßnahmen, vor allem auf kommunaler und staatlicher Ebene. Hier spielen Selbstevaluationsmodelle (König 2000) und Modelle kommunaler Qualitätssicherung eine wichtige Rolle. Im Kapitel zur Sozialen Arbeit in diesem Bande wird darauf näher eingegangen.

Gesundheitspräventionsprogramme stellen besondere Anforderungen an Evaluation, da es sich hier um meist sehr langfristige, schwer überprüfbare Effekte handelt. Elkeles (2006) hat an Beispielen von Herz-Kreislauf-erkrankungsprävention, betrieblicher Rückenbeschwerdenprävention und kommunaler Erwerbslosen-Gesundheitsförderung herausgearbeitet, dass die klassischen experimentell orientierten Evaluationskonzepte hier versagen und stärker auf prozessbezogene, formative und partizipative Konzepte zurückgegriffen werden sollte. Kaluza/Lohaus (2006) haben in einer Analyse von zehn empirisch evaluierten Gesundheitspräventionsprogrammen im Bereich Kindheit und Jugend festgestellt, dass nur zwei mit einem randomisierten Verfahren evaluiert wurden, meist ein quasi-experimentelles Design verwendet wird, bei dem zu einer unsystematisch ausgewählten Interventions-

gruppe eine Vergleichsgruppe gesucht wird. Die 1967 gegründete Bundeszentrale für Gesundheitliche Aufklärung⁹, die Gesundheitsprävention als zentrales Ziel hat, sammelt Evaluationsstudien, führt eigene Studien durch und gibt Empfehlungen zur Qualitätssicherung (Entwicklung von Leitlinien und Qualitätskriterien, Mindestanforderungen, Evaluation der didaktisch-methodischen Ansätze sowie Qualifizierung der Multiplikatoren), gibt aber kein Verfahren vor.

3.6 Gesundheitssystemevaluationen

Es gibt wenig übergeordnete Gesundheitssystemforschung, die ganze Länder vergleicht. Auch sind hier die methodischen Probleme noch größer. In einer Expertise der Hans-Böckler-Stiftung (Kreß et al. 2006) werden zwei internationale Studien verglichen. Ein WHO-Report aus dem Jahre 2000 hat fünf Kriterien angelegt: Die (invaliditätsbereinigte) Lebenserwartung, Gesundheitsunterschiede in der Bevölkerung, Patientensouveränität, gesundheitsbezogene soziale Gerechtigkeit sowie Fairness der Finanzierung. Danach landet die Bundesrepublik Deutschland nur auf einem mittelmäßigen Platz (Platz 14 von 191 untersuchten Ländern), nach dem (härtesten) ersten Kriterium allein gar nur auf dem 22. Rangplatz. Ein OECD-Projekt, das bis heute Berichte liefert, arbeitet ebenfalls an der Entwicklung von Gesundheitssystemindikatoren und vergleicht Länder. Beispielsweise in der 5-Jahres-Überlebensrate nach Brustkrebsoperationen landet Deutschland nur auf Rang 14 (von 19 Ländern), bei der Raucherquote auf Rang 9 (von 22), bei den Grippeimpfungen bei über 65-Jährigen auf Rang 17 (von 20).

So problematisch solche Gesundheitssystemindikatoren auch sein mögen, so bestürzend sind die Ergebnisse. Denn Deutschland liegt bei den Ausgaben für den Gesundheitsbereich in der absoluten Spitzengruppe (Platz 3 hinter Schweiz und USA). Gesundheitspolitiker leiten daraus die Forderung nach mehr Evaluation und Qualitätsmanagement ab.

4 Schlussfolgerungen

Es wurde gezeigt, wie unterschiedlich die verschiedenen Felder und Praxiskontexte im Bereich Gesundheitsförderung sind, wie unterschiedlich damit die Anforderungen an Evaluation sich gestalten, wenn man hier kontextsensibel nach den Standards der DeGEval (2002) vorgehen will.

Legt man bei der Analyse der gefundenen Evaluationsmodelle die Differenzierung von Kromrey (2003) in Forschungs-, Kontroll- und Entwicklungsparadigmata zugrunde, so finden wir alle drei Modelle: Das Forschungsparadigma ist grundlegend bei den experimentellen Wirksamkeitsstudien im medizinischen Bereich. Hier wird auch viel mit wissenschaftlichen Zeitschriftenveröffentlichungen gearbeitet. Das Kontrollparadigma liegt den Zulassungsverfahren für Medikamente und Heilverfahren zu Grunde, aber auch ökonomisch orientierte Ansätze in Krankenhaus- und Gesundheitssystemevaluation fallen darunter. Das

⁹ Siehe www.bzga.bund.de

Entwicklungsparadigma steht oft bei Gesundheitspräventionsprogrammen im Vordergrund.

Insgesamt lässt sich, unterstützt vom Konzept der Evidence Based Medicine, eine Dominanz des experimentellen Evaluationsdesigns feststellen, wie in wohl keinem anderen Praxisfeld. Allerdings ist hier die Kritik am Experiment auch besonders stark, wobei hier unterschiedlich gut fundierte Argumente und Alternativen zu beobachten sind.

Ganz allgemein ist im Gesundheitssystem von heute in Deutschland wegen der immer stärker werdenden Forderung nach Evaluation eine Tendenz zur Vereinheitlichung, zur Kodifizierung, zur Standardisierung festzustellen, die das Bild des Arztes verändern. In vielen Bereichen verschmelzen Alltags- und Evaluationstätigkeiten.

Defizite der Evaluation im Gesundheitsbereich in Deutschland sind vor allem festzustellen in übergeordneter Gesundheitssystemforschung, in der Beachtung des Pflegebereiches und der Vernetzung von Maßnahmen, in der Vernachlässigung von Prozessevaluation und schließlich in der Entwicklung tragfähiger, methodisch abgesicherter Alternativen zum experimentellen Design (vgl. hierzu auch Oevreiteit 2002). Eine Lösung kann hier nur in der Kombination und Integration verschiedener Modelle und Ansätze im Sinne einer „Mixed Methodology“ (vgl. Mayring et al. 2007) liegen, derzufolge wir statt von einer Evidenzhierarchie besser von einer Evidenztriangulation sprechen sollten.

Evaluation im österreichischen Gesundheitswesen – ein Überblick

1 Vorbemerkung

Der Versuch einer Übersicht über die Evaluationstätigkeit im österreichischen Gesundheitswesen muss in Anbetracht des hohen Differenzierungsgrades innerhalb des Gesundheitswesens allgemein bleiben und sich auf generelle Entwicklungstendenzen und manche Besonderheiten beschränken. Die Diskussion einzelner Evaluationen tritt daher in den Hintergrund, zu Gunsten einer Betrachtung der Makro- und Mesoebene, auf welchen Rahmenbedingungen für die Evaluationstätigkeit gesetzt werden.

Die föderale Organisation des Gesundheitswesens bedingt eine Verschränkung bundesstaatlicher, gliedstaatlicher und kommunaler Kompetenzen, wodurch Regelungs- und Handhabungsunterschiede in den Bundesländern grundgelegt sind.

Neben den unterschiedlichen Gesetzesebenen erschwert die sektorale Gliederung des österreichischen Gesundheitswesens eine ganzheitliche Betrachtung. So erfolgt beispielsweise im primären Sektor die ambulante Versorgung zum guten Teil durch die niedergelassenen Ärzte, die im Rahmen der Ärztekammer Selbstverwaltung ausüben. Im sekundären Sektor der stationären Versorgung hingegen sind überwiegend landeseigene Krankenanstalten tätig, die wiederum unvergleichliche Organisationsstrukturen aufweisen (Hofmarcher/Rack 2006).

Eine weitere Schwierigkeit ist die unklare Differenzierung der Begriffsfelder „Evaluation“ und „Qualitätsmanagement“. Evaluation wird im Gesundheitswesen primär als „Tool“ des Qualitätsmanagements verstanden und begrifflich von diesem überlagert. Ein eigenständiger Evaluationsdiskurs, geschweige denn eine ausgebildete Evaluationskultur, ist bislang nicht zu verzeichnen. Wenn von Evaluation und Evaluierung gesprochen wird, so bezieht sich dies überwiegend auf die Bewertung von legislativen Vorgaben in der konkreten Anwendungs- und Umsetzungspraxis. Weiters werden Evaluationen punktuell im Bezug auf konkrete Projekte thematisiert und in der Regel an externe Auftragnehmer vergeben, die wiederum individuell sehr unterschiedliche und nicht allgemein nachvollziehbare Maßstäbe oder Standards anlegen.

2 Gliederung

Um die Entwicklung der Evaluationstätigkeiten im österreichischen Gesundheitssystem nachvollziehbar zu machen, muss die landesspezifische Systemarchitektur ein Hauptpunkt

der Betrachtung bilden. Im ersten Teil werden daher die föderale Organisation, die Leistungserbringer, die formalrechtlichen Rahmenbedingungen und die „große“ Gesundheitsreform 2005 unter dem Blickwinkel der Qualitätsentwicklung und Evaluationsbemühungen betrachtet.

Dazu sollen im zweiten Teil die Ergebnisse einer Umfrage unter den Landesgesundheitsträgern den Stand der Evaluation, insbesondere die Kenntnisse und Umsetzung der Standards der DeGEval (2002), illustrieren.

In einem dritten Teil werden Entwicklungen in der Evaluationskultur an verschiedenen Beispielen dargelegt. Zum einen anhand der gesetzlich verankerten Selbstevaluation der niedergelassenen Ärzteschaft, weiters am Beispiel der freien Berufsverbände. Nachdem neben gesetzlichen und institutionellen Rahmenbedingungen vor allem auch das atmosphärische Umfeld die Evaluationstätigkeit beeinflusst, wird dieses exemplarisch anhand einer skandalisierten Evaluationsstudie aufgezeigt.

3 Strukturen im österreichischen Gesundheitswesen

Die **Verwaltung des österreichischen Gesundheitswesens**, insbesondere der gewichtige Krankenanstaltenbereich, ist von der Kompetenzteilung zwischen Bund und Ländern gem. Art. 12 der Bundesverfassung (B-VG) geprägt. Demnach hat der Bund eine Kompetenz zur Grundsatzgesetzgebung während Ausführungsgesetze Angelegenheit von neun (!) Landesregierungen sind. Die Koordination zwischen Bund und Ländern, besonders in Fragen der Finanzierung, erfolgt mittels Gliedstaatsverträgen (gem. Art. 15a B-VG), die der Bund mit den Ländern abschließt und die unter dem Begriff „15a-Vereinbarungen“ in Österreich geläufig sind. Als wichtiger Financier im österreichischen Gesundheitswesen agieren die Sozialversicherungsträger. Regional und berufsgruppenspezifisch gegliedert gibt es in Österreich insgesamt 24 eigenständige Sozialversicherungsträger, die mitunter nur wenige tausend Versicherte einschließen.

Der Krankenanstaltensektor in Österreich ist geprägt von öffentlichen Krankenhäusern. In den letzten dreissig Jahren wurden die einzelnen Krankenanstalten bundesländerweise in Betriebsgesellschaften zusammengeführt und aus der unmittelbaren Landesverwaltung ausgegliedert. Auf diese Weise sollten auch Schwächen der staatlichen Verwaltung überwunden werden (z. B. einjährige Budgets), eine betriebswirtschaftlich angemessene Betriebsführung mit längerfristigen Planungs- und Finanzierungshorizonten und insgesamt einer höheren Flexibilität und Effizienz ermöglicht werden. In den einzelnen Ländern wurden diese ausgegliederten Bereiche in Betriebsgesellschaften mit unterschiedlichen Rechtsformen organisiert. Diese arbeiten unter Zielvorgaben der jeweiligen Landesregierung. Unter zunehmendem Kostendruck erfolgt die Betriebsführung verstärkt auf Grundlage betriebswirtschaftlicher Prämissen. Diese im Hinblick auf Effizienz wünschenswerte Entwicklung birgt die Gefahr gegenstandsverfehlenden Managements. Solches liegt etwa vor, wenn Steuerungsinstrumente der erwerbswirtschaftlichen Betriebslehre ohne Bedachtnahme auf spezifische Rahmenbedingung öffentlicher Krankenhäuser in diesen implementiert und entscheidungsrelevant werden.

Im niedergelassenen Bereich sind vor allem ÄrztInnen, PsychologInnen, PsychotherapeutInnen und funktionelle TherapeutInnen zu nennen. Für sie bestehen unterschiedliche, rechtliche Rahmenbedingungen. Die ÄrztInnen sind in der Ärztekammer Pflichtmitglied. Für PsychologInnen, PsychotherapeutInnen und MTD (medizinisch-technische Dienste) bestehen seit Beginn der 1990er Jahre eigene Berufsgesetze, jedoch keine Kammern.

4 Qualitätsmanagement im Bereich der Krankenanstalten

Ab den 1990er Jahren wurde die Qualitätsentwicklung im österreichischen Gesundheitswesen forciert, zunächst im Bereich der Krankenanstalten. Ablesbar ist dies besonders an der rechtlichen Entwicklung, so zum Beispiel am Krankenanstalten- und Kuranstaltengesetz 1993 (BGBl 801/1993), in dem die Qualitätssicherung in den Dimensionen Struktur-, Prozess- und Ergebnisqualität ausdrücklich verankert wurde. Die zu treffenden Qualitätsmaßnahmen waren auf die Intention ausgerichtet, vergleichende Prüfungen mit anderen Krankenanstalten zu ermöglichen. Freilich waren die Beweggründe für die verstärkte Qualitätsarbeit weniger in einem inhaltlichen, gar wirkungsorientierten Evaluationsinteresse verankert, als vielmehr in der wachsenden Notwendigkeit kostendämpfender und effizienzsteigernder Steuerungsmaßnahmen. Das Bestreben vereinheitlichter Bewertungsgrundlagen zu vergleichenden Prüfungen setzte sich in den darauffolgenden Jahren fort und ist an den 15a-Vereinbarungen für die Jahre 1997-2000 (BGBl 111/1997) bzw. für 2001-2004 (BGBl 60/2002) abzulesen. Ungeachtet der Finanzierungsteile sahen beide Vereinbarungen in ihrem inhaltlichen Teil qualitätsorientierte Änderungen vor. Erstere brachte 1997 das leistungsorientierte Finanzierungssystem (LKF-Scoring), letzteres eine Vereinheitlichung der Dokumentation mittels ICD¹-Codierschlüssel, verpflichtend ab 2001 und ebenso eine Verankerung einer übergreifenden Qualitätsarbeit, freilich wiederum mit dem expliziten Ziel einer Effizienzsteigerung. Die 15a-Vereinbarung für die Jahre 2005 bis 2008 (BGBl 73/2005) stand bereits ganz im Zeichen der Gesundheitsreform 2005 und bildet einen wesentlichen Bestandteil derselben. Neben dieser ist das Gesundheitsreformgesetz (BGBl 179/2005) mit den Teilbereichen des Gesundheitsqualitätsgesetzes (GQG) und dem Gesundheitstelematikgesetz (GtelG) Kernstück der Gesundheitsreform.

4.1 Gesundheitsqualitätsgesetz (GQG)

Mit dem GQG wurde die rechtliche Klammer für eine bundesweite und sektorenübergreifende, das heißt nicht auf Krankenanstalten beschränkte, Qualitätsentwicklung gelegt. Explizit wird darauf verwiesen, dass die Bestimmungen des Gesetzes unabhängig von der Organisationsform der Leistungserbringer sektorenübergreifend in Geltung gesetzt werden. Ebenfalls wurden im Rahmen des GQG qualitätsrelevante Begriffe definiert, so zum Beispiel Begriffe wie: Qualitätsstandards, Struktur-, Prozess- und Ergebnisqualität, Effizienz, Effektivität und viele mehr. Evaluation spielt in diesem Gesetz keine explizite Rolle. Das GQG ist

¹ ICD- International Statistical Classification of Diseases

eigentlich ein zukunftsweisendes Gesetz geworden, das ausreichend Grundlage bieten würde für eine rege, auf Outcome bzw. Wirksamkeit und Nachhaltigkeit gerichtete Evaluationstätigkeit.

4.2 Institutionelle Umsetzung der Gesundheitsreform 2005

Mit der Gesundheitsreform 2005 wurden auch die organisatorisch-institutionellen Rahmenbedingungen für die inhaltliche Qualitätsarbeit neu definiert. Mit der Bundesgesundheitsagentur und den Landesgesundheitsfonds wurden Fonds als Rechtspersonlichkeiten geschaffen, die nicht nur über Finanzierungsfragen entscheiden, sondern auch einheitliche Qualitätsvorgaben in allen Sektoren des Gesundheitswesens umsetzen sollen, namentlich auch eine einheitliche und umfassende Dokumentation als Analyse- und Vergleichsbasis sicherstellen sollen. Tabelle 1 stellt diese Struktur dar.

Tabelle 1: Institutionelle Struktur der Gesundheitsreform 2005

| 15-a Vereinbarung (BGBl 73/2005) | Bundesebene | Landesebene |
|---------------------------------------|-------------------------------|--------------------------|
| Fonds (=Rechtsperson) | 1 Bundesgesundheitsagentur | 9 Landesgesundheitsfonds |
| Fondsorgan (= personelles Gremium) | 1 Bundesgesundheitskommission | 9 Gesundheitsplattformen |

Für die Planungs- und Forschungstätigkeit (traditionell eine Bundesaufgabe) wurde mit der Gesundheit Österreich GmbH (GÖG) eine Kapitalgesellschaft im hundertprozentigem Eigentum des Bundes gegründet. Unter diesem Dach wurden die bislang selbständigen Rechtsträger ÖBIG (Österreichisches Bundesinstitut für Gesundheitswesen) und FGÖ (Fonds gesundes Österreich) zusammengeführt. Die GÖG agiert im wissenschaftlichen Bereich weisungsfrei. Sie übernimmt im Auftrag des Gesundheitsministeriums Aufgaben der Gesundheitsplanung, der Entwicklung struktur- und prozessbezogener Richtlinien und führt wissenschaftliche Analysen zu spezifischen Themenstellungen durch. Darüber hinaus soll unter dem Dach der GÖG auch das BIQG (Bundesinstitut für Qualität im Gesundheitswesen) entstehen. Das BIQG soll ein bundeseinheitliches Qualitätssystem umsetzen, wobei Patientenorientierung, Effektivität und Transparenz die leitenden Kriterien sind. Aufgabe des BIQG wird auch die Evaluierung des Qualitätssystems sein. Unter Qualitätssystem wird dabei ein „Koordinierungs-, Förderungs-, Unterstützungs- und Überwachungssystem des Bundes verstanden, mit dem Ziel, die Qualität der Gesundheitsleistungen kontinuierlich zu verbessern.“ (§ 2 Abs. 1 BQG) Qualität selbst wird als Grad der Erfüllung der Merkmale von patientenorientierter, transparenter, effektiver und effizienter Erbringung der Gesundheitsleistungen verstanden. Als zentrales Anliegen wird in diesem Zusammenhang die Optimierung von Struktur-, Prozess- und Ergebnisqualität angestrebt.

Insgesamt lässt sich im Rahmen der Gesundheitsreform 2005 eine differenzierte Ausformung von Institutionen feststellen, die jeweils einen spezifischen Beitrag zu einem bundesweiten, einheitlichen Qualitätssystem erbringen sollen. Ob dieser komplexen und viel-

gliedrigen Systemarchitektur drängt sich die Frage nach gemeinsamen und kooperationsleitenden Vorstellungen der Beteiligten auf, unter denen sich die wechselseitigen Bemühungen zu tatkräftigen und letztlich wirksamen Entwicklungen verbinden lassen. Abgesehen vom begrifflichen Zierwerk der Politik, welches die Gesundheitsreform begleitet und in den Gesetzen üppig textiert wurde, lassen sich kaum weitergehende inhaltliche, methodische oder prozessleitende Ansagen feststellen. So gibt es etwa zum zentralen Thema der Evaluation im Gesundheitswesen keinen feststellbaren Diskurs hinsichtlich der Standards, der Anwendbarkeit, der methodischen Designs, ihrer Abgrenzung zum Qualitätsmanagement, oder ähnlichen, wichtigen Fragen.

4.3 Umfrage zur Kenntnis der Evaluationsstandards

Den Mangel an übergreifenden und verbindenden, inhaltlichen und methodischen Entwicklungen haben wir zum Ausgangspunkt unserer Umfrage gemacht. Dabei fokussierten wir auf die Frage nach der Kenntnis der Evaluationsstandards bei wichtigen Akteuren der Gesundheitsreform 2005, namentlich den neun Landesgesundheitsfonds. Am 28.12.2006 haben wir daher per E-Mail folgende drei Fragen ausgesandt:

1. Werden im Hinblick auf die Ergebnisqualität von Gesundheitsleistungen in ihrem Bundesland Evaluationen durchgeführt oder sind solche geplant?
2. Sind die Standards der Evaluation (www.degeval.de) in ihrem Bundesland bekannt, wenn ja werden diese in der Planung von allfälligen Evaluationsvorhaben berücksichtigt?
3. Gibt es alternative Evaluierungsstandards, an welchen Sie sich orientieren?

Von neun angefragten Landesgesundheitsfonds haben sieben schriftlich geantwortet (siehe Tabelle 2).

Wie die Befragungsergebnisse zeigen, sind die Standards der Evaluation, wie sie im deutschsprachigen Raum von der DeGEval (2002) gepflegt werden, durchgängig unbekannt. Dies ist umso bedauerlicher, als die Anwendung und Kenntnis der Evaluationsstandards zentrale, evaluationsspezifische Kompetenzen darstellen (Joint Committee 2006). Die punktuell berichteten Aktivitäten sind umso schwerer zu beurteilen, als offensichtlich auch kein gemeinsames Begriffsverständnis gegeben ist und ein wenig nachvollziehbarer Qualitätsbegriff als Sammelbecken für unterschiedlichste Aktivitäten dient, die mehr oder weniger mit geleiteter Evaluationstätigkeit zu tun haben. Wo Evaluationen durchgeführt wurden, waren diese vielfach an externe Beratungsfirmen vergeben, wobei die Auftraggeber nicht anzugeben im Stande waren, auf welche Standards sich die externen Evaluatoren stützen würden!

Tabelle 2: Antworten zu den Fragen zu den Evaluationsstandards

| Bundesland Rückmeldedatum | Frage 1 | Frage 2 | Frage 3 |
|-------------------------------|---|---------|---|
| Steiermark 12.1.2007 | Vereinzelte Initiativen vor allem zur Patienten- zufriedenheit, kein genauer Überblick | Nein | Keine Angabe |
| Burgenland 12.1.2007 | Ja, regelmäßig, anhand von Checklisten und Fragebögen im Hinblick auf bestimmte Kriterien, z. B. Schmerzempfinden, Kundenzufriedenheit, Pflegeinterventionen, etc. | Nein | Keine Angabe |
| Tirol 23.1.2007 | Ja, laufende Projekte vor allem im Rahmen des Institutes für klinische Epidemiologie ² | Nein | Keine Angabe |
| Oberösterreich 31.1.2007 | Quality Indicator Project (QIP) | - | Keine Angabe |
| Wien 5.2.2007 | Nein, allenfalls punktuell in einzelnen Anstalten, keine Übersicht über die Aktivitäten | - | Keine Angabe |
| Kärnten 8.2.2007 | Patientenorientierte Messung der Ergebnisqua- lität in Kärntner Spitälern 1998-2000 | Nein | Allgemeine Richtlinien des Qualitätsmanagements |
| Niederösterreich 15.2.2007 | Nein, Schwerpunkt auf Struktur- und Prozess- qualität gemessen an Behandlungspfaden, Stich- probenüberprüfung von Krankengeschichten | Nein | keine standardisierten Eva- luationen von Ergebnisqua- lität, daher werden keine Standards angewandt |
| Vorarlberg Σαλζβουργ | Keine Rückmeldung | | |

4.4 Zum Aspekt der Ergebnisorientierung

Das Qualitätsmanagement im österreichischen Gesundheitswesen lässt sich in drei historische Entwicklungsphasen einteilen (Hauke et al. 2001), in deren Verlauf eine allmähliche Hinwendung zur Ergebnisorientierung feststellbar ist.

Ausgehend von einer traditionellen Dominanz sanitätsbehördlicher Kontrollverfahren (fokussierend auf Strukturqualität) begann in den frühen 1990er Jahren die kunden- bzw. patientenorientierte Phase (Prozessqualität), begleitet von der Erstellung von Leitfäden und Zertifizierungsbemühungen, vielfach auf freiwilliger Basis.

Ende der 1990er Jahre ist eine verstärkte Ergebnisorientierung zu verzeichnen, die in der Gesundheitsreform 2005 unter dem Schlagwort der Ergebnisqualität ausdrücklich formuliert wurde. Während struktur- und prozessbezogene Ansätze zum Qualitätsmanagement bereits längere Entwicklungsphasen durchlaufen haben, steckt die Ergebnisqualität – so die Auskunft eines Ministeriumssprechers – noch in den „Kinderpatscherln“ (Kinderschuhe). Im Hinblick auf die fortlaufende Erstellung einer reliablen Datengrundlage zur (vergleichenden) Bewertung der Ergebnisqualität wurde jedoch bereits die Übernahme nationaler und internationaler Outcome- und Indikatoren-Projekte im medizinischen Be-

² Siehe www.iet.at

reich forciert, andererseits erfolgte eine starke Konzentration hin auf die Erhebung der subjektiven Lebensqualität und Patientenzufriedenheit als Maßstab und Indikator der medizinischen Leistungen (BMSG 2001).

Ein gutes Beispiel ist das „International Quality Indicator Project“ (QIP), welchem seit 1986 weltweit über 1900 Spitäler angeschlossen sind. Ab 1998 beteiligten sich zunehmend auch österreichische Kliniken an diesem Projekt, wobei mittlerweile über dreißig Krankenanstalten teilnehmen (BMSG 2003). Ziel des QIP ist die Qualitätsverbesserung durch vergleichende Selbsteinschätzung, anhand ergebnisbezogener Indikatoren, wozu den teilnehmenden Institutionen vierteljährlich anonymisierte Vergleichswerte zur Verfügung gestellt werden. Es handelt sich somit um ein internationales Benchmark-Verfahren, wobei der Vergleich mit anderen Organisationen und im Zeitablauf die Grundlage für Ergebnisanalysen und struktur- wie prozessorientierte Korrekturmaßnahmen bildet. Die Teilnahme am QIP wurde erst im Bundesland Oberösterreich im Jahr 2000 per Gesetz verbindlich für alle privaten und öffentlichen Krankenanstalten eingeführt (LGBL 77/2000). Einschränkend muss angemerkt werden, dass Indikatorenprojekte unterschiedliche Differenzierungen in den einzelnen medizinischen Fächern aufweisen, relativ stark nebenwirkungs- und komplikationsorientiert sind und ungeplante Versorgungspfade im Einzelfall darin schlecht abgebildet werden.

Neben den Indikatorenprojekten ist auch eine wachsende Zahl von Registerprojekten zu verzeichnen, wobei die Initiativen dazu überwiegend in einzelnen Fachdisziplinen und Regionen oder gar Krankenanstalten ihren Ausgangspunkt nehmen. Indikatorenprojekte wie auch Register werden zunehmend als wichtige Voraussetzung zur reliablen Messung von Ergebnisqualität gesehen.

Im Zusammenhang mit Ergebnisorientierung muss auch auf eine Entwicklung der letzten Jahre kurz eingegangen werden, nämlich den breiten Einsatz von Balanced Scorecards (BSC) in Krankenhäusern (Borges/Schmidt 2002), so beispielsweise auch am LKH Klagenfurt. Die BSC ist ein Instrument um strategische Ziele in die operative Fläche zu bringen und die Zielerreichung über Kennzahlen zu kontrollieren. Insofern sich diese Ziele auf Leistungs-, Qualitäts- und Kundenaspekte beziehen, ergibt sich auch eine gewisse Nähe zu formativen Evaluationsdesigns, für die sie selbstverständlich keinen Ersatz darstellen. Die am LKH Klagenfurt in einem top-down-Prozess einzelnen Organisationseinheiten verordnete BSC erwies sich sehr struktur- und prozessorientiert und eher ungeeignet für ergebnis- bzw. wirkungsorientierte Aussagen. Neben den BSC-Dimensionen Mitarbeiter, Finanzen und Prozesse enthält die BSC am LKH die Dimension Patienten/Kunden, in welche ausschließlich Zufriedenheitskriterien eingehen. Eine solche reduktionistische Systematik ist jedoch ungeeignet, um die eigentliche Aufgabe und das anzustrebende Ergebnis eines öffentlichen Krankenhauses abzubilden, nämlich die Erreichung des Gesundheitsziels!

5 Evaluation im niedergelassenen Bereich

5.1 Selbstevaluation der Ärzteschaft

Grundlage für die Evaluationstätigkeit bildet eine Novelle des Ärztegesetzes aus dem Jahr 1998 (§ 118a BGBL 169/1998). Darin wurde die Ärztekammer verpflichtet eine Gesellschaft mbH zu gründen und mit der Entwicklung und Umsetzung einheitlicher Qualitäts- und Evaluierungsstandards zu betrauen. Diese Gesellschaft besteht nun seit dem Jahr 2004 unter dem Namen Öqmed und hat folgende Aufgaben:

- die Ausarbeitung von Qualitätskriterien einschließlich Kriterien für die Struktur- und Prozessqualität
- die Qualitätsevaluierung mittels Evaluierungsbögen
- die Führung eines Qualitätsregisters, in welches die Ergebnisse der Evaluierung und Kontrolle aufzunehmen sind

Im Rahmen dieser Aufgaben wurde ein Konzept zur Selbstevaluierung der Praxen entwickelt und unter dem Titel „Qualitätssicherungsverordnung“ 2006 in Wirkung gesetzt. Inhalt der Verordnung ist unter anderem ein Fragebogen zur Selbstevaluierung der Praxen, die jedoch wiederum nur auf Struktur- und Prozessqualitätskriterien Bezug nehmen. Der Prozess der Selbstevaluierung läuft in folgenden Teilschritten ab:

- fragebogengeleitete Selbstbeurteilung der Ärzte, ob die Ordination den vorgeschriebenen Kriterien für Struktur- und Prozessqualität entspricht
- Auswertung der Selbstevaluierung
- Überprüfung der Validität der Selbstevaluierung (des Zutreffens der einzelnen Parameter) durch stichprobenartige Ordinationsbesuche
- allenfalls Definition der Verbesserungsmaßnahmen und Kontrolle der Mängelbehebung

Ergibt die Evaluierung oder die Kontrolle vor Ort eine unmittelbare Gefährdung der Gesundheit oder unterbleibt die Selbstevaluierung gänzlich, gilt dies als schwerwiegende Berufspflichtverletzung und stellt einen Vertragsauflösungsgrund für die Sozialversicherungsträger dar.

5.2 PsychologInnen und PsychotherapeutInnen

1990 wurde für PsychologInnen und PsychotherapeutInnen jeweils ein Berufsgesetz realisiert (BGBL 360/1990; 361/1990), das jedoch primär auf Richtlinien zur Ausübung der fachlichen Tätigkeit ausgerichtet ist, die verbindliche Fortbildung festschreibt, jedoch keine weiterreichenden Bestimmungen zu Qualitätssicherung oder gar Evaluierung enthält. Es fehlt also im Unterschied zur Ärzteschaft an einer gesetzlichen Verpflichtung zur Evaluation.

5.3 MTD – medizinisch-technische Dienste

Die Gruppe der medizinisch-technischen Berufe (MTD) ist ähnlich wie die PsychologInnen und PsychotherapeutInnen seit 1992 in einem Berufsgesetz geregelt (MTD-Gesetz). Vom Geltungsbereich sind folgende Berufe abschließend erfasst (§ 1):

1. physiotherapeutischer Dienst;
2. medizinisch-technischer Laboratoriumsdienst;
3. radiologisch-technischer Dienst;
4. Diätdienst und ernährungsmedizinischer Beratungsdienst;
5. ergotherapeutischer Dienst,
6. logopädisch-phoniatrisch-audiologischer Dienst;
7. orthoptischer Dienst.

Das Gesetz regelt die Ausbildung sowie Art und Weise der Berufsausübung der MTD. Bezeichnenderweise finden sich in diesem Gesetz keine direkten oder indirekten Evaluationsempfehlungen oder -verpflichtungen. Allein eine allgemeine Dokumentationspflicht der Behandlung ist im § 11 kurz geregelt. Der Begriff „Evaluation“ findet sich im gut zwanzigseitigen Gesetzestext nicht, der Begriff „Qualität“ hingegen zumindest einmal im Zusammenhang mit der Ausbildung im radiologisch-technischen Bereich (§ 20, übrigens als letzter Punkt einer 28 Ziffern umfassenden Aufzählung!). Die MTD sind in freien Berufsvereinigungen, diese wiederum im Dachverband der MTD-Austria organisiert. Nach Auskunft der MTD-Austria ist Qualitätssicherung ein wichtiges Anliegen, das jedoch vorrangig durch Weiterbildungsangebote umgesetzt wird. 2006 wurde im Bereich der MTD eine eigene Ausbildungsverordnung für Fachhochschulen erlassen (BGBl 2/2006). Darin finden sich neben der Qualitätssicherung auch die Dokumentation und Evaluation als Ausbildungsinhalt aller MTD explizit angeführt, womit eine wichtige Schiene zur verstärkten Evaluationstätigkeit gelegt wurde.

6 Evaluationskultur

Das Thema der Evaluation ist nicht nur ausgehend von den rechtlichen und institutionellen Rahmenbedingungen zu diskutieren, sondern auch von der Atmosphäre, in der sie stattfinden soll. Anlass dazu gibt es in Österreich genug: Vor einigen Jahren gab es in Österreich eine aufsehenerregende Buchveröffentlichung (Waldhauser et al. 2003), in der österreichische Ärzte anhand statistischer Datenauswertungen Unterschiede in der postkardiochirurgischen Mortalität an Österreichs Kinderherzzentren schlüssig darlegten. Der Publikation vorangegangen waren langwierige Bemühungen der beteiligten Autoren, anhand der Evaluationsergebnisse notwendige Strukturveränderungen anzustoßen. Allein da dieses Bemühen erfolglos blieb und von den zuständigen Stellen teilweise auch ignoriert wurde, wählten die Autoren den Weg einer Publikation ihrer Evaluationsergebnisse. Diesem Schritt folgten Unterlassungs- und Schadensersatzklagen und auch disziplinarrechtliche Maßnahmen. Einer der Autoren, der als Professor an einer österreichischen Universitätsklinik über viele

Jahre erfolgreich tätig war, wurde auf diese Weise existenziell massiv bedroht. Dieses Beispiel soll verdeutlichen, dass es neben gesetzlichen Rahmenbedingungen vor allem auch einer Entwicklung der gelebten Evaluationskultur bedarf.

7 Zusammenfassung und Ausblick

Im österreichischen Gesundheitswesen entwickelte sich in Rahmen der Gesundheitsreform 2005 begrifflich modernes Bundesrecht, das einen Rahmen zur Förderung wirkungsorientierter Evaluationen bieten würde. Die Struktur- und Prozessorientierung im Rahmen von Qualitätsmanagement könnte damit überwunden werden. Dazu gibt es bislang kaum nennenswerte Ansätze. Die Ausschöpfung gesetzlicher Möglichkeiten scheitert bislang auch an der föderalen Organisation des Gesundheitswesens bzw. an erheblichen Reibungsverlusten zwischen den Bundes- und Landesinstitutionen. Feststellbare Evaluationstätigkeit geht primär von einzelnen medizinischen Disziplinen aus. Daneben sind vereinzelte Aktivitäten auf Landesebene zu verzeichnen, wobei Evaluationen vielfach an extern beauftragte Firmen vergeben werden ohne dass die Beauftragter sich mit den Evaluationsstandards dieser Dienstleister auseinandersetzen würden. Insgesamt sprechen unsere Befunde für eine unterentwickelte Evaluationskultur, die nicht zuletzt auch in feindseligen Haltungen gegenüber veröffentlichten Evaluationsstudien und deren Autoren deutlich wird.

Wichtige Maßnahmen zur Verbesserung der derzeitigen Situation liegen nahe: So zunächst die vollständige Umsetzung der Gesundheitsreform 2005, namentlich die vollständige Realisierung des BIQG, damit Gesetze nicht nur zukunftsweisend sind, sondern auch praktisch wirksam werden können. Bemerkenswert in diesem Zusammenhang ist etwa der Artikel 170 der Schweizer Bundesverfassung, der eine verpflichtende Wirksamkeitsüberprüfung mit dürren Worten festschreibt: „Die Bundesversammlung sorgt dafür, dass die Maßnahmen des Bundes auf ihre Wirksamkeit überprüft werden.“ (SR 101).

Weiter wäre ein Paradigmenwechsel wünschenswert, speziell die Abkehr von der Dominanz struktur- und prozessbezogener Betrachtungsweisen zu Gunsten einer wachsenden Outcome- und Nachhaltigkeitsorientierung, die sich beispielsweise in einer systematischen Förderung von Outcome- und Replikationsforschung manifestieren könnte. Für einen solchen Paradigmenwechsel bedarf es wohl eines von Fachgesellschaften getragenen Evaluationsdiskurses, der sehr grundsätzlich und nicht einzelfallbezogen die Entwicklung und Anwendung von Evaluationsmethoden und -standards zum Gegenstand macht und damit auch bestehenden Basisinitiativen in einzelnen Handlungsfeldern einen Anknüpfungspunkt bietet. Dabei könnte auf Entwicklungen in Deutschland und in der Schweiz Bezug genommen werden, wo mit der Einrichtung von Evaluationsfachgesellschaften starke Kristallisationspunkte für entsprechende Bemühungen gegeben sind.

Evaluation im Gesundheitsbereich: Schweiz

1 Begriffs-Umschreibungen: Evaluation – Gesundheit – Krankheit

Eine Umschreibung relevanter Begriffe im Bereich Evaluation wurde auf nationaler Ebene vorgenommen durch das Bundesamt für Gesundheit (BAG). Nachstehend die generelle Begriffsumschreibung von *Evaluation* (nicht spezifisch für den Gesundheitsbereich, aber auch für diesen gültig) (BAG 2005a: 5):

„Evaluation ist der Prozess der Beurteilung des Wertes (Beitrag zum gesellschaftlichen Wohlergehen), der Qualität und/oder der Berechtigung eines Evaluationsgegenstandes (z. B. einer staatlichen Massnahme). Evaluation bedient sich (meist) sozialwissenschaftlicher Methoden der systematischen Sammlung und Analyse von Daten, die oft nicht „de routine“ verfügbar sind, um zu ihren Urteilen über den Evaluationsgegenstand zu kommen.“

Eine Abgrenzung wird vorgenommen zwischen Evaluation einerseits, Monitoring und Controlling andererseits. *Monitoring* meint die systematische Sammlung von vergleichbaren Daten zu einem Phänomen mit dem Ziel der Identifikation von zeitlichen Trends und Veränderungen. *Controlling* ist eine Managementfunktion zur Steuerung von Planungs- und Umsetzungsprozessen, wozu Daten aus der Evaluation benützt werden können. Bei der Begriffsumschreibung von *Evaluation* steht die Public Health Perspektive im Vordergrund, das heisst der Wert einer Massnahme im Dienste der Gesellschaft.

Auf staatlicher Ebene wird im Weiteren unterschieden zwischen Evaluation und Ressortforschung. Verwaltungstechnisch sind Evaluationen ein Teil der Ressortforschung; diese umfasst auch andere Forschungsaufträge als Wirksamkeitsüberprüfungen. Trennscharfe Kriterien zwischen Forschung, Begleitforschung, Wirksamkeitsüberprüfungen/Evaluation gibt es allerdings nicht. Kontext und jeweilige Fragestellung bestimmen den Umgang mit der Begrifflichkeit. Auf jeden Fall zielen Evaluationen auf das staatliche Handeln ab.

Für diesen Text wird Evaluation in einem erweiterten Sinne verwendet, als übergeordneter Begriff für die Bewertung von Prozessen, Auswirkungen und Wirtschaftlichkeit von Angeboten und Interventionen im Gesundheitsbereich. Aus naheliegenden Gründen wird dabei der Evaluation derjenigen Aspekte besondere Aufmerksamkeit geschenkt, die für die Gesundheitspolitik wichtig sind.

Was derzeit als *Krankheit* gilt, ist medizinisch in den internationalen, deskriptiv diagnostischen Standardwerken ICD-10 und DSM-IV aufgelistet, und für die Schweiz juristisch im Allgemeinen Teil des Sozialhilfegesetzes umschrieben (ATSG, Art. 3 Abs. 1): Krankheit ist jede Beeinträchtigung der körperlichen oder geistigen Gesundheit, die nicht Folge eines Unfalles ist und die eine medizinische Untersuchung oder Behandlung erfordert oder eine

Arbeitsunfähigkeit zur Folge hat. Dem medizinischen Krankheitsverständnis liegt das biopsycho-soziale Modell (Engel 1977) zugrunde, einschliesslich subjektiver und sozialer Kriterien, die kulturspezifisch und deshalb weder allgemeingültig noch unveränderlich sind. Die rechtliche Würdigung stellt einschränkend vor allem auf nachprüfbare medizinische Sachverhalte ab.

Für die Evaluation im Gesundheitsbereich stehen, im Sinne der Public Health Perspektive, im Vordergrund:

- Massnahmen zur Erhaltung und Förderung der Gesundheit in einem umfassenden Sinne
- Massnahmen zur Verminderung von Krankheitszuständen; Krankheitszustände mit gesamtgesellschaftlicher Relevanz haben den Vorrang (besonders verhütbare Krankheiten wie Suchtkrankheiten)
- Massnahmen zur wirksamen, wirtschaftlichen und qualitativ guten Behandlung von Krankheiten
- Evaluation der Ausgestaltung der Krankenversicherung (Programm-Systemevaluation).

2 Ziele von Evaluation

2.1 Evidenzbasierte Gesundheitspolitik

Die Evaluation im Gesundheitsbereich auf nationaler Ebene steht im Dienste der in der Verfassung festgeschriebenen systematischen Wirksamkeitsüberprüfung staatlicher Massnahmen. Dabei geht heute die Gesundheitspolitik über die traditionelle Krankheitsbekämpfung hinaus und umfasst auch Massnahmen gegen neue Herausforderungen wie Gewalt, Stress, Bio-Terrorismus, Versorgungs-Ungleichheiten, Gesundheitsverhalten. Interventionen sind dementsprechend interdisziplinär konzipiert und stellen auf Partnerschaften zwischen staatlichen und nichtstaatlichen Akteuren ab (BAG 2005b).

2.2 Effiziente und kosten-effektive Krankenversorgung

Angeichts der Diversifikation therapeutischer Verfahren und der kontinuierlich steigenden Versorgungskosten kommt dem Ziel einer Überprüfung der Effektivität, der Effizienz und der Kosten-Effektivität von Behandlungsmethoden und Heilmitteln eine erhebliche Bedeutung zu. Daran interessiert sind einerseits die Anbieter, um sich im Markt zu behaupten, andererseits die Kostenträger und für die Behandlungsqualität Verantwortlichen, und schliesslich die Patienten und Patientenorganisationen. Dazu kommt die Evaluation der Ausgestaltung der Krankenversicherung (z. B. der Spitalplanung, der Prämienverbilligung, der Tarifbildung (vgl. Widmer et al. 2004), des Risikoausgleichs, des Zulassungsstopps etc.).

2.3 Abgeltung durch Krankenversicherer

Was in der Schweiz als Krankheitsbehandlung durch die Krankenkassen und Versicherungsgesellschaften abgegolten wird, ist im sogenannten Leistungskatalog zusammengefasst. Die Eidgenössische Kommission für Allgemeine Leistungen ELK prüft und schlägt als Pflichtleistung vor, was die Voraussetzungen dafür erfüllt, die Eidgenössische Arzneimittelkommission EAK tut dasselbe für Medikamente (beide beim BAG). Die Grundsätze und Kriterien dafür werden von der Eidgenössischen Kommission für Grundsatzfragen der Krankenversicherung GLK (beim BAG) erarbeitet.¹ Für die Entscheide der Kommissionen werden Unterlagen zum Wirksamkeitsnachweis gefordert.

2.4 Zulassung Heilmittel und Medizinprodukte

Eine eigene Behörde (Swissmedic) ist zuständig für die Prüfung und die Qualität der in der Schweiz zugelassenen Heilmittel und Medizinprodukte einschliesslich Diagnostika. Die Voraussetzungen für eine Zulassung auferlegen den Herstellern beziehungsweise Importeuren unter anderem die Pflicht zum Nachweis der Wirksamkeit, aber auch eine Risikobewertung eines Präparates. Swissmedic bietet Unterlagen für die Vorbereitung und Antragstellung von klinischen Versuchen.²

2.5 Patientenzufriedenheit

Die Zufriedenheit von Patienten mit den angebotenen Leistungen (eines Spitals, einer Abteilung, einer Arztpraxis etc.) wird von privaten Organisationen gemessen, wobei neben der Wirksamkeit der erhaltenen Leistungen auch die Kundenfreundlichkeit (Abläufe, Information etc.) eine Rolle spielen.

3 Rahmenbedingungen

3.1 Gesetzliche Grundlagen

Grundlage für Evaluation im Public Health Bereich ist Art. 170 der Schweizerischen Bundesverfassung. Danach sind Massnahmen des Bundes auf ihre Wirksamkeit hin zu überprüfen. Weitere Grundlagen sind die das BAG betreffenden Gesetze und Erlasse (siehe BAG 2006a).

Das Forschungsgesetz (FG) ist die allgemeine Grundlage der Ressortforschung. Die Bundesämter gelten laut FG als Forschungsorgane, soweit sie für die Erfüllung ihrer Aufgaben selbst Forschung durchführen, in Auftrag geben oder unterstützen.

¹ Siehe www.bag.admin.ch/themen/krankenversicherung

² Siehe www.swissmedic.ch

Die Organisationsverordnung für das Eidgenössische Department des Innern (EDI) sieht vor, dass das BAG die Forschung auf dem Gebiet der Gesundheit steuert.

Grundlage für die Prüfung und Zulassung von Heilmitteln ist das Bundesgesetz über Arzneimittel und Medizinprodukte (Heilmittelgesetz HMG).

Laut Krankenversicherungsgesetz (KVG) haben Leistungen der Krankenversicherung wirksam, zweckmässig und wirtschaftlich zu sein. Die Wirksamkeit muss wissenschaftlich nachgewiesen sein. Das KVG enthält ausserdem Bestimmungen zur Gesundheitsförderung. KVG-Verordnungen sind weitere Grundlagen für Evaluationen (zum Beispiel Verordnung zum Risikoausgleich).

Die Verordnung über den Tabakpräventionsfonds (TPFV) hält fest, dass die aus einer Lenkungsabgabe auf Zigaretten stammenden Mittel für Prävention und Gesundheitsförderung bestimmt sind, mit Fokus auf Forschung. Damit wird die Evaluation geeigneter Massnahmen unterstützt.

Das Bundesstatistikgesetz ist die Grundlage für die Beschaffung relevanter Daten im Bundesamt für Statistik und besonders im Gesundheitsobservatorium OBSAN.

Auf kantonaler Ebene sind für Evaluationen im Gesundheitsbereich besonders die kantonalen Gesundheitsgesetze massgeblich.

3.2 Ethikverträglichkeit

Die Evaluierung von Interventionen im Gesundheitsbereich bedarf im Grundsatz der Bewilligung durch die zuständige Ethikkommission. Seit der Auflösung der überregionalen Ethikkommission UREK sind ausschliesslich die Kantone dafür verantwortlich. Daraus entstehen Ungleichheiten in Arbeitsweise, Zuständigkeitsbereich, Zusammensetzung, Finanzierung et cetera der Ethikkommissionen. Auch die gegenseitige Anerkennung (etwa bei mehreren Kantonen umfassenden Evaluationsprojekten) ist nicht einheitlich geregelt.

3.3 Standards

Die Evaluations-Standards der Schweizerischen Evaluationsgesellschaft (SEVAL-Standards, Widmer et al. 2000) definieren generell die Anforderungen an Evaluatoren, Auftraggeber, Adressaten und weitere Beteiligte. Sie sind ein Instrument zur Qualitätssicherung und -förderung, im Sinne von Empfehlungen. Die SEVAL überwacht deren Einhaltung nicht. Sanktionen bei Nichteinhaltung sind nicht vorgesehen. Die SEVAL-Standards eignen sich nicht für Zwecke der Akkreditierung oder Zertifizierung.

Das BAG stellt die SEVAL-Standards ebenfalls als Arbeitshilfen im Gesundheitsbereich zur Verfügung³ und publiziert Erläuterungen (Läubli Loud 2004).

Für die Evaluation bei der Zulassung von Heilmitteln und Medizinprodukten gibt es eigene Standards beim Schweizerischen Heilmittelinstitut Swissmedic.⁴

³ Siehe www.bag.admin.ch/evaluation

⁴ Siehe www.swissmedic.ch

3.4 Finanzierung von Evaluation im Gesundheitsbereich

Die wichtigsten Finanzierungsquellen sind hier (ohne Anspruch auf Vollständigkeit) aufgeführt.

Wirksamkeitsstudien im Sinne von Ressortforschung werden vor allem durch eidgenössische und kantonale Behörden finanziert. Auch kommunale Behörden kommen in Frage. Für die Auftragserteilung sowie für die Ausschreibung und Offertstellung gibt es Vorgaben (BAG 1997; BAG 2006a).

Der Schweizerische Nationalfonds (SNF) ist hauptsächlich für die Finanzierung von Grundlagenforschung zuständig. Evaluationen im Sinne klinischer Studien zur Wirksamkeitsüberprüfung sind eher Ausnahmen; zum Beispiel im Rahmen der Co-Finanzierung von Nationalen Kompetenz-Zentren für wissenschaftliche Forschung.⁵ Auch im Rahmen Nationaler Forschungsprogramme (NFP) werden Evaluationsstudien finanziert (z. B. Studien zur Krankenversicherung im NFP45 „Probleme des Sozialstaats“). Zudem verfügen die Universitäten über Forschungsfonds, aus welchen Projekte unterstützt werden können.

„Gesundheitsförderung Schweiz“ ist eine Stiftung zur Durchführung, Unterstützung und Evaluierung von gesundheitsfördernden und präventiven Massnahmen. Seit 2007 fokussiert sie neu auf drei strategische Bereiche: „Gesundheitsförderung und Prävention stärken“, „Gesundes Körpergewicht“ und „Psychische Gesundheit – Stress“. Ein besonderes Gewicht liegt auf der ökonomischen Evaluation.⁶

In der Schweiz gibt es eine grosse Zahl privatrechtlich organisierter Fonds und Stiftungen. Deren Mehrzahl finanziert oder unterstützt Aktionen und Interventionen, seltener auch Forschungsvorhaben.

Kommerzielle Anbieter von Leistungen im Gesundheitswesen kommen dann für die Finanzierung von Wirksamkeitsstudien in Frage, wenn es sich um die Vorbereitung von Anerkennungs- oder Zulassungsdossiers oder um Zwecke des Marketings handelt.

Konsumentenschutz-Organisationen und Medien lassen Wirksamkeitsstudien durchführen, zum Beispiel im Sinne von vergleichenden Analysen von Wirkungen oder Nebenwirkungen von Produkten im Gesundheitsmarkt.

Einen Überblick über die für Evaluation im Gesundheitsbereich aufgewendeten Mittel gibt es nur aus dem BAG: Danach reduzierten sich die Ausgaben für Evaluationen von rund 2.5 Millionen Franken im Jahr 2000 auf 865'000 Franken per 2006, und diejenigen für Ressortforschung (exkl. Evaluation, aber inkl. Forschungsbeiträgen an Externe) im gleichen Zeitraum von rund 8.86 Millionen Franken auf rund 7.9 Millionen.

⁵ Siehe www.snf.ch/E/targetedresearch

⁶ Siehe www.gesundheitsfoerderung.ch

4 Anwendungsfelder

4.1 Prävention und Gesundheitsförderung

Medienkampagnen im Dienste der Prävention und Gesundheitsförderung sind besonders auf nationaler Ebene im Auftrag des BAG evaluiert worden. Bekanntestes Beispiel ist die Aidsprävention. Ausserdem gibt es eine generelle Studie zu Medienkampagnen und deren Evaluation (Bonfadelli/Friemel 2006).

Evaluations-Studien im *schulischen Bereich* wurden beispielsweise im Rahmen Europäischer Mehrländerstudien durchgeführt, zum Beispiel „euro-peers“ und „euro-net“, (Guggenbühl 1999a; 1999b) und „Gsundi Schuel“ (Uchtenhagen/Dombrowski 2001). Auch die Einführung präventiver Massnahmen im Sinne des Krankenversicherungsgesetzes wurde evaluiert (z. B. Battaglia/Junker 2001).

Evaluation von Prävention und Gesundheitsförderung am *Arbeitsplatz*: ein Beispiel ist die Studie über Lärm und Raumluft in Bürogebäuden mit Evaluierung von Schutzmassnahmen (Leutenegger 2006). Für den Gesundheitsschutz von Arbeitnehmern gibt es gesetzliche Vorschriften und den Interkantonalen Verband für Arbeitnehmerschutz IVA; das Staatssekretariat für Wirtschaft (Seco) stellt unter anderem Materialien zum Gesundheitsschutz am Arbeitsplatz zur Verfügung, ebenso Arbeitgeber- und Arbeitnehmer-Organisationen. Deren Anwendung wird überprüft, namentlich durch die Schweizerische Unfallversicherungsanstalt (SUVA).⁷ Auch Weiterbildung für betriebliches Gesundheitsmanagement wurde evaluiert (BAG 2005c).

Prävention auf *Gemeinde-Ebene* wurde z.B. im Auftrag des BAG evaluiert in Bezug auf Sucht- und Drogenprobleme (Peters/Wapf 2006).

Was den Bereich *Freizeit und Sport* angeht, ist das Bundesamt für Sport ein Partner für die Vergabe von Evaluationsstudien. Das Sportwissenschaftliche Institut (SWI) nimmt als Hauptaufgaben anwendungsorientierte Forschung und Entwicklung, Dienstleistung sowie Lehre in verschiedenen Bereichen der Sportwissenschaften wahr: Über Studien aus dem BAG zum Thema „Prävention im Sport“ gibt es einen zusammenfassenden Bericht (BAG 2004).

Prävention der *Pflegebedürftigkeit älterer Menschen*, beispielsweise durch Hausbesuche, ist ebenfalls ein Thema von Wirksamkeitsprüfungen (Stuck et al. 2002).

Eine gezielte gesundheitliche Prävention bei *Risikogruppen* gibt es im Bereich etwa der Strafgefangenen (Berger/Uchtenhagen 2000; Karger et al. 2000) und der Migranten (Ledermann et al. 2006); hier liegt bisher der Schwerpunkt auf Problemen der Umsetzung.

Im Bereich der *Sekundärprävention* sind zwei konkrete Beispiele zu nennen. Zum einen die Einführung und Förderung von Methoden zur Früherfassung und -intervention bei riskantem und schädlichem Substanzkonsum (im Rahmen der BAG-Kampagne „Alles im Griff“; Peters et al. 2002), zum anderen Massnahmen bei drogengefährdeten Jugendlichen, in einem BAG-Interventionsprogramm 1999-2005 systematisch evaluiert (Hüsler et al. 2005; Hüsler/Werlen 2006).

⁷ Siehe www.suva.ch

Vorsorgeuntersuchungen, vorbeugendes Screening und generelle oder gezielte Gesundheitsberatung gibt es in vielfältiger Form und durch vielerlei Akteure. Ein Überblick über deren Ergebnisse liegt nicht vor.

Eine Evaluation einer professionellen Suchtpräventionsstelle wurde im Kanton Zürich durchgeführt (Eberhard/Lanz 1997), sowie eine Prozessevaluation der Suchtpräventionsstellen im Kanton Aargau (Guggenbühl 2000). Eine vertiefte Auseinandersetzung mit evaluierter Suchtprävention bei Risikogruppen folgte (Fabian/Guggenbühl 2000).

4.2 Behandlungsmethoden und -settings

Evaluationen im Bereich therapeutischer Interventionen sind wesentlich vielfältiger und umfangreicher. Sie situieren sich in einer dynamischen Entwicklung sowohl bei der wissenschaftlichen Forschung wie auch beim kompetitiven Kampf um Marktanteile für Heilmittel und medizinische Technologien. Ein Überblick ist mangels entsprechender Zusammenführung der Angaben nicht möglich. Einige Beispiele von Evaluationen mit gesundheitspolitischer Priorität mögen genügen.

Pharmakotherapien werden mit seltenen Ausnahmen durch die Pharmaindustrie entwickelt und nach den geltenden Vorschriften getestet, wobei hier die letzte Phase der klinischen Prüfung mit vergleichendem Wirksamkeitsnachweis interessiert.

„Goldstandard“ dieser Versuche ist das randomisierte Design mit geeigneter Kontrollgruppe. Aber auch prospektive Kohortengruppen liefern wichtige ergänzende Kenntnisse. Ein Beispiel ist die HIV Kohortenstudie⁸, ein anderes die Kohortenstudie zur heroingestützten Behandlung Opiatabhängiger (Prozess-, Ergebnis- und Kosten-Nutzen-Evaluation und Evaluation bezüglich Langzeitwirkungen (Uchtenhagen et al. 1999a; Güttinger et al. 2002).

Nur ausnahmsweise wird die *Nachhaltigkeit* von therapeutischen Anwendungen untersucht. Ein frühes Beispiel waren die katamnästischen Ergebnisse stationärer Therapie bei Alkoholproblemen (Maffli et al. 1995).

Ein Evaluationsbericht zum Programm Komplementärmedizin wurde vom BAG publiziert (2005d).

Weitere aktuelle Fragestellungen betreffen z.B. die *Therapie von problematischem Cannabiskonsum bei Jugendlichen*.⁹ Ein randomisierter klinischer Versuch mit Cannabisextrakt betraf die Behandlung spastischer Muskelkrämpfe bei Multipler Sklerose (Vaney et al. 2006).

Anders sieht die Situation aus bezüglich Evaluation von *Psychotherapie*. Der Kreis der Interessierten ist wesentlich kleiner, und die methodischen Probleme sind ungleich grösser wegen der Vielzahl intervenierender Faktoren. Trotzdem gibt es wichtige Ansätze, sowohl was die Wirksamkeitsforschung (Lutz/Grawe 2001; 2005), wie auch was die psychotherapeutische Versorgung betrifft (z. B. Lutz et al. 2004). Eine umfassende Studie des OBSAN wird demnächst publiziert (Margraf 2007).

Evaluationsstudien zur *Arbeitsrehabilitation* gab es ebenfalls im Bereich der Suchtkrankenhilfe (Uchtenhagen et al. 1999b) sowie aktuell in der Psychiatrie („supported employ-

⁸ www.shcs.ch

⁹ www.bag.admin.ch/themen

ment“, Rüst/Debrunner 2005), ausserdem zur *Integration Behinderter* (Baumgartner et al. 2004).

Evaluationen bei *speziellen Zielgruppen* betreffen speziell den Umgang mit Gesundheitsproblemen bei Migranten (Effionayi-Mäder et al. 2001; Ledermann et al. 2006; Achermann/Chimienti 2006; BAG 2006b) sowie bei Strafgefangenen (z. B. Mehrländerstudie zur Evaluation von gerichtlich angeordneten Therapien bei Drogenabhängigen; Stevens et al. 2006).

Evaluationen zu Massnahmen der *Schadensminderung* bei Substanzmissbrauch (Spritzenumtausch, Injektionsräume u. a.) sind durch das BAG im Rahmen der Evaluation MaPaDro vorgenommen und publiziert worden (siehe unten, Abschnitt 4.4).

Eine Evaluation von Medizinprodukten und -technologien gibt es im Hinblick auf deren Zulassung durch Swissmedic.

Die Nutzung der Versorgungseinrichtungen und die dafür relevanten Faktoren wurden evaluiert im Rahmen des NFP45 (Bisig/Gutzwiller 2004) und des Krankenversicherungsgesetzes KVG (Kägi et al. 2004).

Evaluationsstudien zur *Förderung von Behandlungsqualität in Systemen* gibt es seitens von Krankenhäusern oder Fachgesellschaften und von Seiten der Krankenversicherung (z. B. Lüthi et al. 2002). Patientenzufriedenheit und Behandlungsqualität in Arztpraxen ist das Thema von Untersuchungen (Eisenring/Hess 2004; Beule et al. 2004). Ein weiteres Beispiel auf nationaler Ebene fokussierte auf die Therapie Drogenabhängiger (Schaaf 2003; Wettach 2005).

4.3 Gesundheitsökonomie

Eine ökonomische Evaluation von Massnahmen der Gesundheitsförderung und gesundheitlichen Prävention ist ein besonderes Anliegen der Stiftung Gesundheitsförderung Schweiz. Sie publiziert zusammenfassende Berichte dazu (zuletzt Schug et al. 2006).

Wirtschaftlichen Aspekte und Auswirkungen des Krankenversicherungsgesetzes KVG wurden eingehend evaluiert (z. B. Haari et al. 2001; Hammer et al. 2001; Peters et al. 2001; Widmer et al. 2004).

Kosten-Nutzen (cost-benefit), Kosten-Wirksamkeit (cost-effectiveness) und Kosten-Nützlichkeit (cost-utility) spielen eine zunehmende Rolle in der Evaluation im Gesundheitsbereich, speziell bei kontrovers diskutierten Massnahmen in der Gesundheitspolitik. Ein Beispiel ist die Kosten-Nutzen-Analyse der heroingestützten Behandlung Heroinabhängiger (Rossier-Affolter 2000; Frei et al. 2000).

Viele Studien betreffen die Einführung neuer Medikamente oder Verfahren im Vergleich zu bestehenden Medikamenten oder Verfahren – oder auch im Vergleich zu Placebo, was allerdings zunehmend auf ethische Bedenken stösst. (siehe zum Beispiel: Szucs et al. 2004; Cornuz et al. 2006; Neeser et al. 2007).

Ebenso bedeutsam sind wirtschaftliche Studien zur Optimierung von Abläufen und zur Verbesserung der Patientenzufriedenheit (Schwappach et al. 2003; 2004).

4.4 Übersicht Evaluationen beim BAG

Für den Schwerpunkt Evaluationen im Suchtbereich (zur Auswertung des Massnahmenpakets Drogen MaPaDro) gibt es einen Sammelband für die Jahre 1990-2000 (BAG 2002). Weitere Studien finden sich in den drei Bänden „Suchtforschung im BAG“ (Grundlagenforschung 2002-2003, Prävention 2002-2003, Behandlung und Betreuung 1999-2001). Einen aktualisierten Überblick über Evaluationen im Gesundheitsbereich bietet das BAG auf seiner Homepage.¹⁰ Inhaltlich erstrecken sie sich auf HIV/Aids, Alkohol, Tabak, Drogen, Migration, Bildung und Gesundheit, Strahlung, Ernährung, Infektionskrankheiten, Transplantation und weitere Themen. Weitere Arbeiten wurden durch die jeweils zuständigen Bundesämter veranlasst. Einen Überblick gibt Balthasar (2007).

5 Methodik

Es kann hier nicht um eine detaillierte Darstellung der verwendeten Methoden gehen. Ausführliche Darstellungen zur Planung und zu Instrumenten sind vorhanden, die als Leitfaden benützt werden (BAG 1997; Widmer et al. 2000). Eine vermehrte Kombination verschiedener Methoden wäre je nach Fragestellung für die Zukunft sinnvoll.

5.1 Ergebnisevaluation

Für eine Bewertung der Wirksamkeit von gesundheitspolitischen, präventiven und therapeutischen Massnahmen werden sowohl quantitative wie qualitative Methoden eingesetzt. Bei den quantitativen Evaluationen werden strukturierte Erhebungsinstrumente verwendet und soweit möglich vergleichende Analysen durchgeführt, bei den qualitativen Evaluationen geht es um Interviews sowie um Fokusgruppen als ergänzende Informationsquellen (z. B. zur Generierung relevanter Fragestellungen, zur Interpretation quantitativer Resultate, zu Qualitätsfaktoren etc).

Für die Wirksamkeitsprüfung von medizinischen Verfahren und Heilmitteln kommen vorzugsweise randomisierte kontrollierte Studien, allenfalls Kohortenstudien zum Zug. Für die Datenerhebung werden standardisierte Instrumente bevorzugt.

In zunehmendem Masse werden im Gesundheitsbereich Meta-Analysen von Evaluationsstudien durchgeführt, international vor allem als Cochrane Reviews. Ein nationales Beispiel ist die Evaluation aller Studien zu Substitutionsbehandlungen in der Schweiz (Meta-Analyse der Literatur: Rehm/Uchtenhagen 2001; Evaluation durch ein Expert Panel: Varder et al. 2003).

Eine Sonderform von Meta-Analyse ist die Vorbereitung, Durchführung und Auswertung von Konsensus Konferenzen zu einer bestimmten Fragestellung. Konsensus Konferenzen drängen sich besonders auf, wenn die vorhandenen Erkenntnisse zu Nützlichkeit,

¹⁰ Siehe www.bag.admin.ch/evaluation

Wirksamkeit und Wirtschaftlichkeit einer Massnahme oder eines Verfahrens kontrovers sind.

5.2 Prozessevaluation

Hier stehen ebenfalls quantitative und qualitative Methoden zur Verfügung. Das Vorgehen zur Stichprobengewinnung nach der Methode des Theoretical Sampling wird von Sozialwissenschaftlern bevorzugt. Ausserdem gibt es zum Beispiel Schneeballverfahren und Stichprobensuche via Medien. Als kostengünstigste Methode sind Fokusgruppen beliebt.

5.3 Ökonomische Evaluation

Methodische Probleme bestehen in Bezug auf die Berechnung von wirtschaftlichem Nutzen und nicht-monetären Effekten. Zu diesem Bereich gibt es eine neuere Studie (Vaterlaus et al. 2004); er ist aber in der Schweiz noch nicht ausreichend vertreten. Eine systematische Abschätzung der Folgen bei Neueinführung von Verfahren (im Sinne eines Health Technology Appraisal) gibt es noch nicht.

5.4 Formative Evaluation

Evaluationen als Lernprozesse mit dem Ziel einer Leistungs- und Qualitätsverbesserung sind in verschiedenen medizinischen Bereichen gut eingeführt. Sie setzen eine Bereitschaft zu Transparenz und Lernprozessen voraus und benötigen deshalb einen frühzeitigen Einbezug aller an der Evaluation Beteiligten. Im Unterschied zu Ergebnisevaluationen kann dies auch als Selbstevaluation durchgeführt werden.

5.5 Meta-Evaluation

Eine Bewertung von Evaluationsstudien im Gesundheitsbereich gibt es auf internationaler Ebene: rigorose methodische Bewertung für die systematischen Reviews (Cochrane, Campbell libraries).

6 Evaluations-Netzwerke

Die nachstehenden Institutionen und Organisationen haben sich untereinander und mit relevanten Expertengruppen vernetzt (nicht vollzählig): SEVAL, BAG, Medizinische Fachdisziplinen und -gesellschaften, Schweizerischer Nationalfonds, Schweizerisches Heilmittelinstitut, Schweizerisches Gesundheitsobservatorium. Querverbindungen bestehen zu den Evaluationsgesellschaften in Deutschland, Frankreich und England sowie zur Europäischen

Evaluationsgesellschaft. Die Kontakte betreffen vorab den Austausch von Informationen und Erfahrungen. Wichtige Beziehungen betreffen etwa Forschungsgruppen und -institute in den verschiedenen medizinischen Disziplinen, behördliche und private Auftraggeber, Institutionen und Stiftungen der Forschungsförderung, politische Gremien (Sanitätsdirektoren-Konferenz, Gesundheits-Kommissionen des Parlaments), zentrale Institutionen mit Gesundheits- und Forschungsdaten (OBSAN, SIDOS).

7 Stellenwert der Evaluation im Gesundheitsbereich

7.1 Rückblick

Eine eigentliche Evaluationskultur entstand in der Schweiz relativ spät während der 1990er Jahre.

Eine Meta-Evaluation von 15 Evaluationsstudien wurde 1993-95 durchgeführt (Widmer 1996; Widmer 2000). Keine dieser Studien betraf den Gesundheitsbereich. Die Studie bestätigte die Anwendbarkeit der *Program Evaluation Standards* von 1994 (Joint Committee 2006; vgl. auch Hansen 1994) für die Schweiz, hob aber auch kritische Punkte hervor.

Konzeptionell ist die Ausweitung der Gesundheitspolitik auf den Bereich der Gesundheitsförderung (Salutogenese) mit einer entsprechenden Anpassung des Krankenversicherungsgesetzes (KVG) hervorzuheben.

Das Bundesamt für Sozialversicherung führte zahlreiche Studien zur Wirkungsanalyse des Krankenversicherungsgesetzes KVG durch (BSV 2001).¹¹

2001 wurde im BAG das Kompetenzzentrum für Evaluation CCE eingerichtet. Ein Rahmenkonzept beschreibt die Modalitäten und die Organisation (BAG 2005b).

1996 wurde die Schweizerische Evaluationsgesellschaft SEVAL gegründet, die 2002 eine Mitglieder-Liste von 174 Evaluatoren und Evaluatorinnen auswies (13 für den Fachbereich Gesundheit).¹²

Trotz dieser Schritte musste noch vor wenigen Jahren festgestellt werden, dass die Schweiz zu den Ländern mit einer wenig entwickelten *Evaluation Community* gehört. Ausbildung in Evaluation nimmt in den sozialwissenschaftlichen Universitäts-Lehrgängen bisher einen geringen Stellenwert ein (Widmer 2000).

7.2 Ausblick

Es ist damit zu rechnen, dass in allen Bereichen staatlichen Handelns der Stellenwert punktueller und systematischer Evaluation zunehmen wird. Anstrengungen zu einer Weiterentwicklung der vorhandenen Ansätze sind erkennbar, namentlich zur Professionalisierung und Vernetzung sowie zu einer vermehrten Ausbildung von Evaluatoren. Ungewiss ist, ob Ressortforschung weiterhin in relevantem Ausmass finanziert werden kann.

¹¹ Siehe auch www.bsv.admin.ch

¹² Siehe www.seval.ch/de/evaluator

Das Kompetenzzentrum für Evaluation CEE des BAG erhielt einen detaillierten Aufgabenkatalog.¹³ Es arbeitet für viele dieser Aufgaben mit der Schweizerischen Evaluationsgesellschaft SEVAL zusammen. Die SEVAL verfügt über einen rasch angewachsenen Mitgliederbestand (derzeit rund 400 Evaluatoren und Evaluatorinnen, Auftraggeber, Ausbilder, Berater) und ein Aktivitätsprogramm in Qualitätsentwicklung, Wissensvermittlung, eigenen Veranstaltungen et cetera.

Das 1999 beschlossene und 2001 konzeptionell konkretisierte und eingerichtete Schweizerische Gesundheits-Observatorium (OBSAN; eine Organisationseinheit des Bundesamts für Statistik) ist ein Instrument der nationalen Gesundheitspolitik, im Interesse von Evaluation und Verbesserung der Gesundheitsversorgung (www.obsan.ch).

Abschliessend seien noch einige wenige Probleme für die Zukunft genannt:

Eine vermehrte Zusammenarbeit des Bundes mit den Kantonen ist zur Förderung der Evaluationskultur erforderlich. Dazu gehört die Einrichtung von Datenbanken zur Evaluation.

Instrumente und Infrastruktur für ökonomische Evaluationen zuhanden der Gesundheitspolitik bedürfen einer Weiterentwicklung.

Von den deklarierten Zielen der Evaluation bleibt dasjenige der Nachhaltigkeit von Interventionen vermehrt zu berücksichtigen.

Ebenso bleibt die systematische Evaluation bei Neueinführung einer Intervention auf der System-Ebene (Auswirkungen auf das betreffende Versorgungssystem) eine Zukunftsaufgabe.

Ein eklatantes Defizit bildet die mangelnde Evaluierung von Weiterbildungen im Gesundheitsbereich; von den Auswertungsebenen nach Donald Kirkpatrick werden in aller Regel nur gerade diejenige der Teilnehmerzufriedenheit berücksichtigt, während die Auswirkungen auf deren Behandlungspraxis, auf der Institutsebene und letztlich auf der Patienten/Klienten-Ebene unberücksichtigt bleiben (Kirkpatrick 1998). Eine Evaluation der Anwendbarkeit und Praxisrelevanz der in zunehmender Zahl erarbeiteten Richtlinien und Manuale für präventive und therapeutische Interventionen steht noch weitgehend am Anfang.

¹³ Siehe www.bag.admin.ch/evaluation

Mehr Hetero- als Homogenität. Zum Stand der Evaluation im Bereich Gesundheit in Deutschland, Österreich und der Schweiz

1 Einleitung

Im Evaluationsfeld der Gesundheit fallen nach der Lektüre der drei sehr unterschiedlichen Länderdarstellungen vorerst einmal die zahlreichen Unterschiede auf, die Gemeinsamkeiten sind etwas schwieriger zu finden. Entlang der Stichworte „Evaluationskultur“, „Evaluationsgegenstand“ sowie „Stand der Evaluation“ sollen verschiedene Aspekte der drei Länderbeiträge verglichen werden.

2 Evaluationskultur

„Evaluationskultur“ zu definieren und zu beschreiben, stellt eine eigene und spannende Forschungsaufgabe dar. Im vorliegenden Kontext vereinfachen wir die Dinge und meinen damit den Grad der Selbstverständlichkeit, mit dem Evaluationen durchgeführt, mit dem Evaluationen methodisch korrekt erstellt, mit dem Evaluationsergebnisse transparent dargestellt und diskutiert, sowie die gelernten Lektionen umgesetzt werden.

In der Schweiz gibt es mit Artikel 170 der Bundesverfassung seit dem Jahr 2000 eine explizite Grundlage, auf die sich Evaluationsaktivitäten des Bundes stützen können. In verschiedenen Spezialgesetzen finden sich weitere Grundlagen. In Deutschland fehlt zwar eine verfassungsmässige Verankerung, aber auch hier schreiben verschiedene Gesetze Evaluationstätigkeiten im Gesundheitsbereich vor und regeln deren Ablauf. Auch in Österreich gibt es gesetzliche Grundlagen, die beispielsweise eine weitgehende Evaluationstätigkeit in Krankenhäusern ermöglichen würden. Ganz offensichtlich unterscheidet sich zwischen den drei Ländern aber der Vollzug der gesetzlichen Basis: Während in der Schweiz und in Deutschland tatsächlich rege evaluiert wird – in der Schweiz nicht nur beim Bund, sondern auch in den Kantonen, in den Gemeinden und bei Nichtregierungsorganisationen –, führen die Gesetzesgrundlagen in Österreich zu keiner ausgeprägten beziehungsweise befriedigenden Evaluationstätigkeit. Als Ursache wird angegeben, dass in Österreich die Bereitschaft, Evaluation durchzuführen, nicht stark verbreitet ist. Offensichtlich reicht also eine verfassungsmässige beziehungsweise eine gesetzliche Verankerung alleine nicht aus, um eine Evaluationskultur zu entwickeln. Dazu braucht es auch einen von den wichtigsten Akteur/innen getragenen Konsens, dass Evaluationen wichtige Feedback-Elemente des politischen Prozesses sind. Für die Schweiz könnte man sogar die Meinung vertreten, dass die

Verankerung der Evaluation in Artikel 170 der Bundesverfassung nicht den Auftakt, sondern vielmehr den Abschluss eines Prozesses der Bewusstwerdung darstellt: Erste Evaluationsaktivitäten gehen nämlich bereits auf die späten 1980er Jahre zurück.

Hilfreich für eine positive Entwicklung der Evaluationskultur kann auch eine (zentrale) Bundesstelle sein, die selbst Evaluationsaufträge vergibt und damit die Kultur (mit-)prägen kann. In der Schweiz nimmt das Kompetenzzentrum für Evaluation im Bundesamt für Gesundheit seit vielen Jahren eine solche Rolle ein, obschon bei weitem nicht alle Bundesevaluationen im Gesundheitsbereich über diese Stelle vergeben worden sind. In Deutschland könnte die Bundeszentrale für gesundheitliche Aufklärung eine analoge Rolle einnehmen. Evaluationen sind bisher aber nur ein Geschäft unter anderen und somit zu wenig prominent.

Interessant ist der Unterschied zwischen der Schweiz und Österreich mit Blick auf die Auswirkung des stark föderal organisierten Gesundheitswesens. Während dies in der Schweiz eher dazu geführt hat, dass einzelne Kantone beziehungsweise Nichtregierungsorganisationen mit innovativen Ansätzen vorangegangen sind (bspw. Bern und Genf), scheinen die föderalen Strukturen in Österreich eher gezielt zur Abwehr von Evaluationen eingesetzt worden zu sein.

Zur Evaluationskultur gehören auch die Verankerung der Aus-, Weiter- und Fortbildung in der Evaluationswissenschaft sowie die Tätigkeiten der Fachgesellschaften (inkl. der Evaluationsstandards). Im Vergleich zu anderen Wissenschaften gibt es in allen drei Ländern und in beiden Bereichen Rückstände zu diagnostizieren. In der Schweiz und Deutschland sind die Evaluationsfachgesellschaften recht gut etabliert, es gibt Weiterbildungen an Universitäten und die Standards scheinen sich zunehmend zu verbreiten. In der Schweiz scheint die Organisation der Evaluator/innen aus dem Gesundheitsbereich in der SEVAL weitreichender zu sein als in Deutschland in der DeGEval. In Deutschland wird oft nicht von „Evaluation“, sondern von „Wirkungsforschung“ oder „Begleitforschung“ gesprochen. Anders sieht es wiederum in Österreich aus, wo noch nicht vom gleichen Stand der Entwicklung berichtet werden kann.

3 Evaluationsgegenstand

Für Deutschland und die Schweiz wird eindrücklich aufgezeigt, wie breit gefächert der Evaluationsgegenstand im Bereich der Gesundheit ist. Er reicht von der Evaluation von Medikamenten im Zulassungsverfahren bis hin zur Evaluation von lokalen Präventionsprogrammen und zur Evaluation von Gesundheitssystemen. Ganz offensichtlich geht mit dieser Breite des Gegenstandes in allen Ländern eine grosse Heterogenität der Evaluationsformen und -methoden einher. Dargestellt werden in den Länderbeiträgen vor allem die extern durchgeführten Evaluationen. Sehr zahlreich sind aber auch die intern wie extern durchgeführten Selbstevaluationen, die naturgemäss viel schwieriger zugänglich sind.

Kritisch sei angemerkt, dass der Evaluationsbegriff im Bereich der Gesundheit sehr breit verwendet wird und es sich daher lohnen würde, präzisere Gegenstandsbeschreibungen herauszuarbeiten. Beispielsweise werden sowohl im deutschen wie auch im schweizerischen Beitrag die Studien im Bereich der Medikamentenzulassung unter den Begriff der Evaluation subsumiert. Selten werden im Rahmen solcher Untersuchungen aber der Kon-

text, mögliche Alternativen ausser den beiden verglichenen Medikamenten oder die Auswirkungen der untersuchten Medikamente auf die volkswirtschaftlichen (indirekten und intangiblen) Kosten untersucht. Dies widerspricht verschiedenen Standards, die durch die Evaluationsgesellschaften aufgestellt worden sind.

Von einer dem Gesundheitsbereich eigenen Evaluationsperspektive kann in keinem der drei Länder die Rede sein. Die angesprochene Breite und Heterogenität führt dazu, dass der Gesundheitsbereich fraktioniert ist. Die Fachdiskussionen im Bereich der Gesundheitssystemforschung sind bspw. nur lose – oder in der Schweiz gar nicht – mit denjenigen im Bereich der Präventionsevaluation verbunden. Ein gezielter und systematischer Austausch könnte zu gegenseitigen Lerneffekten und somit mittelfristig auch zur Qualitätssteigerung der Evaluationen führen.

4 Stand der Evaluation

In Deutschland und der Schweiz wird von einer breiten Evaluationstätigkeit berichtet, in Österreich von einer stark eingeschränkten. In Deutschland werden zunehmend mehr Evaluationen gefordert, in der Schweiz kann beobachtet werden, dass die Aktivitäten in den letzten Jahren zugenommen haben. Allerdings kommen alle Autoren zum Schluss, dass nach wie vor gewichtige Defizite bestehen. Diese zeigen sich in verschiedener Hinsicht: Zum einen müssen die methodischen Grundlagen, vor allem im Bereich der gesundheitsökonomischen Evaluation, der Evaluation von Präventionsmassnahmen oder auch bei der Evaluation der Qualität von Leistungserbringer/innen, weiter entwickelt werden. Zum anderen gilt es aber auch, wichtige Themen aufzugreifen, die bisher eher brach lagen (bspw. die Evaluation von Prozessketten).

In der Ökonomie wird zwischen einer Makro- und einer Mikroeffizienz unterschieden. Auf der Makroebene stellt man sich die Frage nach dem richtig ausgestalteten Design des Gesundheitswesens und nach der Ressourcenverteilung zwischen dem Bereich Gesundheit und den anderen Bereichen. Auf der Mikroebene stehen eher einzelne Massnahmen im Fokus. In der Evaluationsforschung aller drei Länder liegt der Schwerpunkt heute auf der Mikroebene. Bei aller Notwendigkeit der Mikroevaluationen bringt dies die Gefahr mit sich, dass die „grossen“ Fragen nicht mehr gestellt werden und die Feedbackleistungen der Evaluation in Richtung Politik sehr partiell bleiben. Die Evaluationsforschung sollte sich daher in Zukunft viel stärker mit den (methodisch sehr anspruchsvollen) Gesundheitssystemanalysen und -vergleichen befassen.

Ein kritischer Hinweis sei mit Blick auf den Einsatz der Selbstevaluationen gemacht. In Österreich kommen sie beispielsweise bei der Ärzteschaft zum Einsatz. Ungenügende Evaluationsergebnisse sind dabei mit scharfen Sanktionen verbunden. Vor dem Hintergrund einer insgesamt nur mangelhaft ausgebildeten Evaluationskultur scheint ein solches Arrangement problematisch zu sein und kaum zur Stärkung der Evaluation als sinnvolles Instrument beizutragen.

Nur untergeordnete Rollen spielen für die Evaluation zurzeit die Krankenversicherer. In der Schweiz gibt es nur einzelne Evaluationen, meistens von besonderen Versicherungs- und Leistungserbringungsformen (bspw. Managed Care Organisationen und -instrumenten,

Telefontriage-Anbietern etc.). In Deutschland ist die Situation vergleichbar. Besondere Erwähnung verdient das Wissenschaftliche Institut der Allgemeinen Ortskrankenkassen, das eine aktivere Rolle spielt. Ähnlich liegt die Situation für die Verbände der Leistungserbringer/innen. In der Schweiz ist einzig der Verband der Ärzt/innen (FMH) mit der Evaluation einzelner Modellvorhaben aktiv.

Die Zulassung neuer diagnostischer und therapeutischer Verfahren bzw. neuer Medikamente durchlaufen in allen drei Ländern spezifische Verfahren. Bereits lange werden sog. unabhängige Health Technology Assessment Zentren gefordert, die nicht nur in den genannten Prozessen eine Schlüsselrolle einnehmen, sondern auch alle wissenschaftlichen Evidenzen zu bereits vorhandenen Leistungen zusammentragen und bewerten sollen. In Deutschland (Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen) und Österreich (Ludwig Boltzmann Institut für Health Technology Assessment) wurden bereits grössere Schritte in diese Richtung unternommen. In der Schweiz hat sich diesbezüglich noch sehr wenig entwickelt (im Zentrum für Technologiefolgen-Abschätzung).

5 Fazit und Ausblick

Ausgehend von einem unterschiedlichen Stand der Evaluationskultur und der Evaluationsaktivität besteht in allen drei untersuchten Ländern ein beträchtliches Potenzial und auch eine Notwendigkeit zur Verbesserung der Situation. Einen Königsweg gibt es dabei nicht. Vielmehr müssen die nächsten Schritte auf die länderspezifischen Bedingungen abgestimmt werden. Gefordert sind dabei verschiedene Akteur/innen, namentlich natürlich die Auftraggeber/innen von Evaluationen. In allen drei Ländern sind dies vor allem der Bund, die Gliedstaaten und die Kommunen. Aber auch die Leistungserbringer/innen, ihre Verbände sowie die Versicherer und ihre Verbände spielen eine wichtige Rolle. Eine unterstützende Wirkung könnte dabei vom Trend zur Evidence Based Medicine ausgehen, weil diese von dokumentierten Wirkungszusammenhängen ausgehen muss.

Zentral sind aber auch die Aus-, Weiter- und Fortbildung der Evaluatoren/innen sowie die vernetzenden Aktivitäten der Fachgesellschaften SEVAL und DeGEval. Für die Schweiz lässt sich für die letzten zwanzig Jahre eindrücklich zeigen, wie sich die Evaluationskultur durch das Zusammenspiel der genannten Faktoren sukzessive aufgebaut und verbreitet hat.

Eine Gefahr für die Zukunft der Evaluation geht ganz generell von den knapper werdenden staatlichen Finanzmitteln aus. Im Kanton Bern führte dies beispielsweise bereits dazu, dass eine über zehn Jahre aufgebaute Evaluationstradition kurzerhand gestoppt wurde. Die Gefahr kann zweifach angegangen werden. Zum einen sollte man Sorge dafür tragen, dass die finanziellen Mittel für Evaluationen eher aufgestockt als gekürzt werden. Zum anderen sollten aber auch Evaluatoren/innen auf die Ressourcenknappheit reagieren und weniger aufwändige und daher kostengünstigere Methoden entwickeln. Dabei die wichtigsten Qualitätsaspekte nicht zu vernachlässigen und gleichwohl ein schlankes Design zu konzipieren, dürfte die grösste Herausforderung darstellen. Besonders anforderungsreich ist die Evaluation von kleinen Massnahmen und Programmen. Hier müssen zwangsläufig verstärkt qualitativ-explorative Methoden zum Einsatz kommen. Dies mag für Evaluatoren/innen

aus berufsethischen und wissenschaftlichen Gründen nicht immer befriedigend sein, stärkt aber insgesamt die Bedeutung der Evaluation.

Institutionelle Politik

Die Evaluation institutioneller Politik in der Schweiz

1 Einführung und Begrifflichkeit

Der vorliegende Beitrag befasst sich mit der Evaluation institutioneller Politik in der Schweiz. Die institutionelle Politik ist kein einheitlich definierter Begriff, sondern umfasst ganz unterschiedliche Phänomene wie Verwaltungsreformen, Gebietsreformen, Regierungs- und Parlamentsreformen ebenso wie die generellen Strukturen eines politischen Systems wie der Föderalismus und die politischen Mitspracherechte. Um dieser Vielfalt von Inhalten gerecht zu werden, soll der Begriff hier einleitend definiert und geklärt werden. Institutionelle Politik ist in einem ersten Schritt von der substanziellen Politik abzugrenzen. Die substanzielle Politik sucht mit inhaltlichen Vorgaben oder Policy-Instrumenten direkt auf Adressatenverhalten Einfluss zu nehmen, hat also Policy-Ziele. Die institutionelle Politik dagegen sucht über die Veränderung des organisatorischen Institutionengefüges einer politischen Einheit die Machtverteilung zwischen Akteuren und Prozesse zu ändern, hat also primär Politics-Ziele, die erst sekundär zur Erreichung allfälliger Policy-Ziele beitragen (Knoepfel/Bussmann 1997: 59). Institutionelle Politik umfasst sowohl die politischen Institutionen als auch die Institutionenpolitik, welche die Gegenstände des vorliegenden Beitrages darstellen.

Der Begriff ‚Politische Institution‘ bezeichnet eine feste Untersuchungsgrösse der Politikwissenschaft und bezieht sich auf die Meta-Ebene politischer Systeme. Politische Institutionen können übergreifend als anerkannte Regelsysteme politischer Willensbildung und Willensumsetzung definiert werden (March/Olsen 1989). Sie regeln den Zugang zu Entscheidungen und generieren damit individuelle Erwartungssicherheit und Stabilität. Beispiele sind unterschiedliche Wahl- und Regierungssysteme, Mitspracherechte wie die direkte Demokratie und die gliedstaatliche Kompetenzverteilung des Föderalismus.

Institutionenpolitik bezieht sich im Gegenzug auf die Meso-Ebene von Verwaltungsstrukturen und Organisationsmerkmalen korporativer Akteure. In Anlehnung an Ritz (2003) kann der Zweck von Institutionenpolitik im Wandel der Steuerungsmechanismen, der Strukturen sowie Prozesse und der Kultur als Voraussetzung veränderten administrativen Handelns gesehen und so in einen Bezug zur „Verwaltungspolitik“ (Bogumil/Jann 2005) gebracht werden. Institutionenpolitik entfaltet einerseits direkte Wirkungen innerhalb einer Institution und „ist daher vom politischen Handeln der Verwaltung zu unterscheiden“ (Bogumil/Jann 2005: 251), andererseits hat sie indirekte Wirkungen auf die Anspruchsgruppen ausserhalb der Institution und auf substanzielle Politiken. Wir definieren den Begriff hier entsprechend über seine Funktion: Institutionenpolitik schafft die Voraussetzung dafür, dass materielle Politik wirken kann. Diese Definition umfasst sowohl Reformen als auch Zustände, wobei der Fokus im Folgenden auf Ersteres gelegt wird. Beispiele von Institutio-

nenpolitik sind die Vollzugsstrukturen öffentlicher Politiken, Verwaltungsreformen und -reorganisationen oder Gebietsreformen innerhalb einer politischen Einheit wie zum Beispiel Gemeindefusionen in einem Kanton.

Der Beitrag geht im nächsten Teil auf die Evaluation von Institutionenpolitik ein und handelt anschliessend die Evaluation von politischen Institutionen in der Schweiz ab. Im letzten Teil wird ein Ausblick auf die Zukunft der Evaluation institutioneller Politik in der Schweiz skizziert.

2 Die Evaluation von Institutionenpolitik

In diesem Kapitel wird in einem ersten Abschnitt eine Übersicht über die Institutionenpolitik in der Schweiz gegeben, bevor in den folgenden beiden Abschnitten der Stand der entsprechenden Evaluationstätigkeiten sowie ihrer Erkenntnisse und Nutzung diskutiert wird. Im letzten Abschnitt werden Überlegungen zu den Möglichkeiten und Grenzen von Evaluationen von Institutionenpolitik angestellt.

2.1 Institutionenpolitik in der Schweiz

Institutionenpolitik findet auf allen drei föderalen Ebenen der Schweiz statt. Da die drei Staatsebenen Bund, Kantone und Gemeinden über ausgeprägte Kompetenzen und insbesondere die Kantone über die verfassungsmässig garantierte Organisationsautonomie verfügen, sind sowohl die bestehenden Strukturen als auch Reformtätigkeiten in mehrerer Hinsicht ausgesprochen heterogen. Die Bundesebene verfügt über eine professionelle und ausdifferenzierte Verwaltung, die mit circa 37'000 Mitarbeitenden in Relation zur Grösse der Schweiz im internationalen Vergleich hinsichtlich Grösse verhältnismässig klein ist (Widmer 2007). Auf der kantonalen Ebene finden sich grosse Unterschiede bezüglich Personalbestand, Professionalisierungsgrad sowie organisatorischer Ausdifferenzierung und Spezialisierung (Bochsler et al. 2004; Germann/Weis 1995; Vatter 2006a). Bereits in den 1970er Jahren machten Geser und Höpflinger (1977: 112; auch Geser 1981) die unterschiedlichen Kantonsstrukturen zum Untersuchungsgegenstand und konstatierten zur Frage nach der Leistungskapazität einer kantonalen Verwaltung: „Je kleiner ein Kanton, desto geringer ist tendenziell seine Kapazität, den Verwaltungsapparat strukturell auszdifferenzieren. Der Mangel an struktureller Differenzierung ist mit Einbussen im Umfang und in der Qualität der Verwaltungsleistungen verknüpft.“ Noch ausgeprägter sind die Unterschiede auf kommunaler Ebene, wo das Spektrum von reinen Milizstrukturen (wie in der Schweiz die nicht-professionelle, nebenamtliche Erfüllung öffentlicher Aufgaben genannt wird, vgl. auch den Begriff der Milizarmee für eine Nicht-Berufsarmee) in kleinen Landgemeinden bis hin zu vollständig ausdifferenzierten und professionellen Verwaltungen in den Städten reicht, die von ihrer Leistungsfähigkeit kantonalen Verwaltungen gleichzustellen sind.

Diese unterschiedlichen Strukturen auf den drei föderalen Ebenen wurden verschiedenen Reformen unterzogen. Die wichtigste Art von Reformen waren dabei Verwaltungsreformen. Widmer und Rieder (2003) zählen für den Zeitraum zwischen 1990 und 2000 250

Projekte. Diese Reformen folgen vor allem dem Modell des New Public Management (NPM), dessen Schweizer Version sich durch eine ausgeprägt politische Prägung in Form des Anspruchs stärkerer politischer Steuerung auszeichnet.

Auf Bundesebene wurden Instrumente wie Leistungskontrakte, Globalbudgetierung, Produktgruppengruppierungen und Controllingprozesse im Jahr 1997 durch die spezifische Experimentierklausel Art. 44 im Regierungs- und Verwaltungsorganisationsgesetz (RVOG) zur Führung mit Leistungsauftrag und Globalbudget eingeführt. Auf der kantonalen Ebene wurde NPM in rund der Hälfte aller Kantonsverwaltungen flächendeckend eingeführt (Ritz 2005; 2007; Rieder/Widmer 2007). Weniger Verbreitung fand NPM in den Schweizer Gemeinden. Im Jahr 2005 haben nur fünf Prozent der Gemeinden ein NPM-Projekt voll implementiert (Ritz 2007).

Weitere Reformen betreffen Regierungs- (insgesamt 23) und Parlamentsreformen (insgesamt 39), wobei wir ab Mitte der 1990er Jahre einen Trend zu Verkleinerungen feststellen können. Widmer und Rieder (2003) beobachteten des Weiteren 47 Reformen des Finanzausgleichs und 38 Justizreformen im Zeitraum zwischen 1990 und 2000, denen lediglich 21 Gebietsreformen auf der Ebene Kantone¹ gegenüber stehen. Ab 2000 lässt sich eine grosse Zunahme von Gemeindefusionen konstatieren (Horber-Papazian 2006).

2.2 Evaluationstätigkeit

Die genannten Strukturen und ihre Reformen wurden in unterschiedlichem Ausmass evaluiert. Tabelle 1 gibt einen Überblick über die Evaluationen von Institutionenpolitik auf Bundes- und Kantonsebene. Aufgeführt sind dabei nicht die Reformen selber, sondern lediglich ihre Evaluationen.² Die Tabelle gibt entsprechend keinen Überblick über die Gesamtheit der erfolgten institutionellen Reformen. Die grösste Zahl von Evaluationen betrifft NPM-Reformen.

Wie erwähnt, weist die schweizerische Form des NPM eine stark politische Komponente aus. Vorab die Parlamente sollen zur wirkungsbasierten Steuerung befähigt werden, weshalb der Wirkungsorientierung dieser Reformen grosse Bedeutung zukommt. Widmer (2002: 102-104) unterscheidet zwei Formen von Evaluationen im Zusammenhang mit Reformprojekten in der Schweiz, nämlich Evaluationen *innerhalb* der Reformen und Evaluationen *von* Reformen. Diese idealtypische Unterscheidung wird im Folgenden übernommen.

Die erste Form, also die Evaluation *in* Reformen, liefert im Falle von NPM die Grundlagen, um den Anspruch der verbesserten Steuerungsfähigkeit einzulösen. Von der Grundidee von NPM ausgehend, sollte sich Verwaltungshandeln an Wirkungen orientieren, da die staatliche Leistungserbringung keinen Selbstzweck darstellt, sondern auf die Veränderung eines gesellschaftlichen Zustandes ausgerichtet ist. Die Aufgabe des Staates ist erst erfüllt,

¹ Eine kantonal gesteuerte Gebietsreform kann mehrere Gemeindefusionen umfassen. Fusionen aus Eigeninitiative wurden dabei nicht berücksichtigt.

² Gemeint sind Evaluationen *von* Reformen, vgl. die nachfolgenden Ausführungen. Die Übersicht beruht auf einer Internet-Recherche auf den kantonalen Internetseiten und einer Anfrage per E-Mail an die Staatskanzleien der 26 Kantone, von denen elf eine Rückmeldung schickten. Die Tabelle erhebt deshalb keinen Anspruch auf Vollständigkeit. Insbesondere die Zahl interner Evaluationen kann grösser sein als aufgeführt.

wenn die erwünschte Wirkung seiner Politik eingetreten ist. Evaluationen *in* Reformen stehen aufgrund ihrer materiellen Wirkungsorientierung Evaluationen substanzieller Politik nahe. Die Güte der Umsetzung dieser Wirkungsorientierung ist ihrerseits Gegenstand von Evaluationen *von* NPM-Reformen (Ladner et al. 2007; Rieder 2005; Zürcher 1999).

Tabelle 1: Evaluationen von Institutionenpolitik 1990-2007

| | Evaluationen institutioneller Reformen 1990-2007 | | | |
|-----------|--|-------------------------------------|-------------------------------------|---------------------------------|
| | <i>Verwaltung*</i> | <i>Regierung</i> | <i>Justiz</i> | <i>Gebietseinteilung</i> |
| AG | NPM extern | neue Departementsstrukturen, intern | - | Gebietsreform, Evaluation läuft |
| BE | NPM extern | - | - | Gemeindereform, extern |
| BS | NPM intern | - | - | - |
| FR | NPM extern | - | - | - |
| GE | NPM extern | - | - | - |
| GR | NPM extern | - | - | - |
| JU | - | - | - | - |
| LU | NPM extern | - | - | Grundlagenstudie extern |
| OW | Zwischenevaluation Zusammenarbeitsprojekte OW-NW: Sicht von OW, regierungsintern | - | - | - |
| SG | - | - | Reorganisation Rechtspflege, intern | - |
| SO | NPM extern | - | - | - |
| SZ | NPM intern | - | - | - |
| TG | NPM intern | - | - | - |
| VS | NPM extern | - | - | - |
| ZH | NPM extern | - | - | Pilotstudien intern |
| CH | NPM extern | - | - | - |

* Die verschiedenen Reformen zu einer wirkungsorientierten Verwaltungsführung werden je nach Kanton verschieden bezeichnet, zum Beispiel WOV, NEF, NPM, wif! etc.; in der obigen Tabelle werden diese Reformen jeweils mit NPM bezeichnet.

Die Evaluation *von* NPM-Reformen leistet „eine Erfolgsüberprüfung der Reformprojekte zuhänden Dritter“ (Widmer 2002: 104). Bei NPM-Reformen spielen Vorabklärungen und Machbarkeitsstudien in Form von Ex ante-Evaluationen eine wesentliche Rolle bei der Planung. Ebenfalls lassen sich begleitende und/oder formative Evaluationen beobachten, die auf eine laufende Verbesserung der Reformprozesse zielen. Formative Evaluationen beziehen sich dabei häufig auf Pilotprojekte von NPM-Reformen. Bei summativen Evaluationen überwiegt ebenfalls der Fokus auf den Reformprozess selber. Dies ist zum einen zeitlich bedingt, indem Evaluationsergebnisse bei Pilotprojekten zu einem frühen Zeitpunkt vorlie-

gen müssen. Zum andern hängt dies auch mit der teils symbolischen oder taktischen Nutzung (Johnson 1998: 94) von Reformevaluationen zusammen. So kann die einvernehmliche Durchführung von institutionellen Reformen als Erfolg gelten und damit der Legitimation des Vorhabens dienen, ohne dass damit Fragen der Wirksamkeit der Reformen angesprochen sind.

Wie in Tabelle 2 aufgeführt, werden Evaluationen *von* NPM-Projekten vor allem durch externe Stellen durchgeführt. Dagegen wurden die beiden einzigen identifizierten Evaluationen von Regierungsreorganisationen (im Kanton Aargau) und Justizreformen (im Kanton St. Gallen) durch interne Stellen erarbeitet. Von vier evaluierten kantonalen Gebietsreformen wurden drei durch Externe untersucht.

Tabelle 2: Status der evaluierenden Stellen

| | Evaluationen institutioneller Reformen in den Kantonen 1990-2007 | | | |
|---------------|--|------------------|---------------|--------------------------|
| | <i>Verwaltung</i> | <i>Regierung</i> | <i>Justiz</i> | <i>Gebietseinteilung</i> |
| intern | 4 | 1 | 1 | 1 |
| extern | 9 | 0 | 0 | 3 |

2.3 Erkenntnisse und Nutzung

Die Evaluationen von Institutionenpolitik betreffen vor allem NPM-Reformen, auf die sich die folgenden Ausführungen beziehen. Während diese Evaluationen unterschiedliche Aufgabenstellungen und Ausrichtungen haben, lassen sich gleichwohl Tendenzen über die Gesamtheit hinweg festhalten. Auf der Basis einer Befragung von Verwaltungskadern und Parlamentsmitgliedern in den Kantonen Basel-Landschaft, Solothurn, Thurgau und Zürich kommt Ritz (2005: 57-58) zum Schluss, dass die Reformvorhaben erfolgreich waren, indem sie zu erhöhter Kostentransparenz, zu einer Steigerung des Qualitätsbewusstseins bei der Aufgabenerfüllung, zu einer stärkeren Bürger- bzw. Kundenorientierung und zur Steigerung sowohl der Effizienz als auch der Transparenz von Verwaltungshandeln beitrugen.

Andere Reformziele wurden dagegen nicht erreicht. Diese betreffen primär die politische Seite des schweizerischen NPM-Modells, namentlich die politische Steuerung. Im Vordergrund steht dabei die Messung von Wirksamkeit via Output-Indikatoren. Ein erstes Problem liegt bei der Zielformulierung. Häufig werden als Ziele schlicht die Aufgaben bzw. die gesetzlichen Pflichten der Dienststelle umschrieben, womit aber kein anzustrebender Zustand deklariert ist. Des Weiteren treten oft typische Fehler wie zu vielschichtige, zu ambitionierte oder zu schwammige Ziele auf (Rieder 2005: 153). Das zweite zentrale Problem ist die Verknüpfung von Zielen und Indikatoren. Aufgrund der besseren Messbarkeit von Leistungen beziehen sich Wirkungsindikatoren häufig auf diese Ebene. In diesem Zusammenhang halten Schedler und Proeller (2000: 61-62) fest: „Auch wenn daher Wirkungen die konzeptionell besseren Ziele wären, fehlen heute vielfach noch die nötigen Analyseinstrumente. Aus diesem Grund beschränkt sich die NPM-Praxis oftmals noch auf die Produktorientierung. Um das eigentliche Ziel der Wirkungsorientierung zu erreichen, muss die Verwaltung Erfahrungen sammeln und Anstrengungen in diese Richtung lenken. Denn

die Produkteorientierung ist konzeptionell gesehen lediglich eine Art 'Zwischenstufe' auf dem Weg zur Wirkungsorientierung. (...) Eine wirkungsorientierte Steuerung der Verwaltung verlangt nach neuen Instrumenten und Mitteln. Eine Steuerung über Wirkungen und Leistungen ist erst möglich, wenn diese messbar gemacht wurden. Das Bedürfnis nach derartigen Informationen kann von den traditionellen Kontroll- und Finanzinstrumenten nicht erbracht werden.“ So stellt die Evaluation der NPM-Reform der Stadt Bern (Ladner et al. 2007; Ritz/Sager 2007) eine gesamthafte Nicht-Übereinstimmung von Zielen und Indikatoren von 34 Prozent fest, das heisst 34 Prozent von insgesamt 341 analysierten Indikatoren messen nicht auf der Zielebene, auf der ihr Ziel angesiedelt ist. Wenn Leistungsziele, also Ziele auf der Output-Ebene, ausgeklammert werden, steigt die Quote drastisch: Nur 21 Prozent von 52 Indikatoren für Outcome-Ziele messen tatsächlich Outcomes³, und nur noch 9 Prozent von 77 Indikatoren für Impact-Ziele messen Impacts⁴ von Verwaltungshandeln. Die gesamthafte Fehlzuordnung von Indikatoren auf der Wirkungsebene beträgt beeindruckende 86 Prozent. Die analytische Verkürzung durch die Gleichsetzung von politischem Output und seiner Wirkung wurde in zahlreichen NPM-Projekten als Problem erkannt (Rieder/Furrer 2000; Koch et al. 2002). Eine weitere Erkenntnis, die damit zusammenhängt, besteht im konstatierten Bedeutungsverlust statt –gewinn von Parlamenten aufgrund des durch die Berichterstattung durch die Verwaltung generierten Informationsüberflusses (Ladner et al. 2007; Ritz/Sager 2007).

Ritz (2005: 58) hält zusammenfassend fest: „Aus einem Vergleich der Zielerreichung mit den ursprünglich intendierten Zielen wird ersichtlich, dass die Zielerreichung aus heutiger Sicht der Befragten durchwegs tiefer als anfänglich erhofft ist. Am grössten ist diese Differenz bei den Aspekten einer besseren Aufgabenteilung zwischen Politik und Verwaltung (v. a. aus Sicht der Verwaltungskader), der Einführung von mehr Wettbewerb sowie der Erhöhung der Steuerbarkeit durch das Parlament. Im Gegensatz dazu stimmen Reformziel und Zielerreichung bei der Steigerung des Bewusstseins als strategisch agierender Staat, einer erhöhten Kostentransparenz sowie einer Steigerung der Qualitätsorientierung am stärksten überein.“ Dennoch besteht ein grundsätzlich hohes Interesse an Evaluationsergebnissen, die sowohl instrumentell im Sinne der Umsetzung von Empfehlungen (Weiss 1998b: 24) als auch symbolisch im Sinne der Legitimation der eigenen Position (Johnson 1998: 94) genutzt werden.

2.4 Möglichkeiten und Grenzen

Die obigen Ausführungen zeigen, dass Evaluation *in* Reformen potenziell durchaus einen Beitrag zu einer verbesserten politischen Steuerung leisten kann, dass zu diesem Zweck aber noch wesentliche Verbesserungen in der Wirkungsmessung geleistet werden müssen. Eine stärkere Anlehnung an die Programmevaluation substanzieller Politik kann dabei ein möglicher Lösungsweg sein (vgl. Ladner et al. 2007: 183-189; Sager et al. 2009).

³ Unter Outcome wird hier die Verhaltensänderung der Politikadressaten einer Massnahme verstanden.

⁴ Unter Impact wird hier der Problemlösungsbeitrag der erzielten Verhaltensänderung bei Politikadressaten verstanden.

Die bisherigen Erfahrungen zeigen des Weiteren, dass die Evaluation *von* Institutionenpolitik grosse Anforderungen bezüglich der zu testenden Wirkungsmodelle stellt. Widmer (2002: 102) bezeichnet dabei zwei spezifische Charakteristika von Institutionenpolitik: „(1.) Die zu untersuchenden Wirkungsketten sind im Vergleich zu substanziellen Politiken komplexer. Die beabsichtigten Wirkungen treten – falls überhaupt – vermittelt über mehr Stationen auf, was deren Evaluation massgeblich erschwert. (2.) Die Evaluation institutioneller Politiken erfolgt in der Regel losgelöst von materiellen Politiken, was wiederum den Abstraktionsgrad der Evaluationen erhöht.“ Um die Begrifflichkeit von Knoepfel et al. (1997) zu bemühen, kann gesagt werden, dass sich die untersuchten Wirkungsketten von Institutionenpolitik gänzlich innerhalb des Evaluationsgegenstands ‚Behördenarrangement‘ bewegen, also innerhalb der Gesamtheit der Umsetzungsstrukturen öffentlicher Politik. Das analytische Problem dieser Sichtweise ist, dass ein solches Modell nur bis zu den produzierten Leistungen des Behördenarrangements reicht, also bis zu den Outputs. Eine Erklärung für diese Beschränkung dürfte bei seiner Herkunft aus den Betriebswirtschaften zu suchen sein: Das Produkt eines Marktbetriebes ist gut, wenn es der Kundschaft gefällt und diese es kauft. Die Beschränkung auf die Output-Ebene macht also Sinn. Das Produkt von Politik dagegen genügt sich selber nicht. Politik ist erst dann gut, wenn sie wirkt und damit zur Lösung gesellschaftlicher Probleme beiträgt. Dieser Sichtweise wird die Evaluation von Institutionenpolitik bislang nicht oder nur begrenzt gerecht.

Schliesslich ist auch ein Problem bei der Qualität der Evaluationen selber zu suchen. Widmer (2002: 111) äussert sich kritisch gegenüber einer Auswahl von Studien und sieht den Grund in den wenig günstigen Rahmenbedingungen, „haben doch die verantwortlichen Stellen kaum evaluationsspezifisches Know-how. (...) Auch die oftmals beigezogenen externen Berater haben in den meisten Fällen einen anderen Hintergrund und kaum Erfahrungen in der Konzeption, Begleitung oder Durchführung von Evaluationen.“ Diese Situation hat sich seit dem zitierten Text nicht wesentlich gebessert. Die mangelnde Erfahrung der Auftraggeber hat nicht nur falsche Erwartungen an die Evaluationen zur Folge, sondern auch zu knapp bemessene zeitliche und finanzielle Mittel für die Durchführung von fundierten Untersuchungen.

3 Die Evaluation von politischen Institutionen

Nach der Diskussion von Institutionenpolitik wird in den folgenden Ausführungen auf die Evaluation von politischen Institutionen eingegangen. Im ersten Abschnitt werden die zentralen politischen Institutionen der Schweiz kurz vorgestellt. Anschliessend wird der Stand der Evaluationstätigkeiten referiert, bevor wiederum auf Erkenntnisse und Lehren sowie die Möglichkeiten und Grenzen von Evaluationen politischer Institutionen eingegangen wird.

3.1 Politische Institutionen in der Schweiz

Politische Institutionen bezeichnen nicht allein formale Regeln, d.h. rechtlich verankerte Verfassungs- und Gesetzesnormen eines politischen Systems, sondern ebenso im Verlauf

der Zeit standardisierte, informelle Regelungstechniken und Organisationsmuster. Die zentralen formalen Institutionen des politischen Systems der Schweiz sind der ausgeprägte Föderalismus, die Partizipationsrechte der halbdirekten Demokratie auf allen föderalen Ebenen sowie die Kombination der beiden unterschiedlichen Wahlrechte der Verhältniswahl (Proporz) und der Mehrheitswahl (Majorz). Die wichtigsten informellen Institutionen sind das machtheilige Regierungssystem der Konkordanz, das über rund fünfzig Jahre hinweg stabile Parteiensystem, das erst in jüngster Zeit grösseren Umwälzungen unterliegt, sowie im Bereich der öffentlichen Politik der Vollzugsföderalismus, der die eingespielte Form bei der Formulierung und Umsetzung von Bundespolitik bezeichnet. Für umfassende Darstellungen der formalen wie informellen politischen Institutionen der Schweiz sei auf das Übersichtswerk von Linder (2005) sowie auf die ausführlichen Einzeldarstellungen in Klöti et al. (2006) verwiesen.

3.2 Evaluationstätigkeiten

Es besteht ein breiter Forschungsapparat zur Wirkungsweise politischer Institutionen in der Schweiz. Diese Untersuchungen haben jedoch keinen explizit evaluatorischen Fokus, wie Vatter (2006b: 15) festhält: „Während in den letzten Jahren zahlreiche Untersuchungen in der Schweiz zur Evaluation substanzieller Politiken durchgeführt wurden, stellen Beiträge, die unter einer explizit evaluativen Perspektive politische Institutionen analysieren, nach wie vor die grosse Ausnahme dar.“ Dieser Befund kann vor allem für konkrete Soll-Ist-Vergleiche bestätigt werden, während ein Grossteil der politikwissenschaftlichen Institutionenforschung in der Schweiz als Wirkungsanalysen im Sinne einer „goal-free evaluation“ (Scriven 1973) bezeichnet werden kann, die für die Bewertung nicht allein die Zielsetzungen des Evaluationsgegenstandes beziehen.

Von einer Soll-Ist-Evaluation kann gesprochen werden, wenn einer politischen Institution eine bestimmte Absicht auf der Politics- und/oder auf der Policy-Ebene zugrunde liegt, die in der Untersuchung als zu erreichende Zielvorgabe behandelt werden kann. Entsprechende Studien liegen eher spärlich vor und betreffen vor allem die Institutionen des Föderalismus (Balthasar et al. 1995; Gerheuser et al. 1997; Vatter/Sager 1996; Vatter 2006b). Die generelle Zielsetzung des Föderalismus ist die Integration territorial definierbarer und politisch institutionalisierter Minderheiten durch ihre Gleichbehandlung unabhängig von der Bevölkerungsgrösse. Die genannten Studien bewerten unterschiedliche Elemente des Föderalismus aufgrund ihres Erfolgs, dieses Ziel zu erreichen.

Der Grossteil der politikwissenschaftlichen Institutionenforschung kann allerdings als allgemein wirkungsorientierte Evaluationsforschung bezeichnet werden, indem alle Wirkungen einer politischen Institution berücksichtigt werden. Die schweizerische Politikwissenschaft nutzt hierbei häufig das ‚Labor Schweiz‘, in dem sich unter dem Dach eines einheitlichen übergeordneten Systems 26 autonome und dennoch vergleichbare Politiksysteme finden, die sich für systematische Vergleiche anbieten. Entsprechende Untersuchungen betreffen eine Vielzahl von politischen Institutionen, so die direkte Demokratie, die Parteiensysteme, die Regierungszusammensetzung und weitere politische Systemmerkmale (vgl. Feld/Savioz 1997; Freitag/Vatter 2000; Freitag et al. 2003; Stutzer/Frey 2000; Vatter 2002).

3.3 Erkenntnisse und Lehren

Die Erkenntnisse aus den genannten Studien sind der Vielfalt des Forschungsbereichs entsprechend breit. Als Vorteile des föderalen Staatsaufbaus erkennt die Forschung die damit einhergehende Entlastung für den Bund sowie die resultierende milieugerechte Politik, wodurch sowohl eine integrative Wirkung als auch eine Akzeptanzsteigerung von Bundespolitik erzielt werden (Vatter 2006b). Gleichzeitig erkennt die Föderalismusforschung in der Schweiz aber auch erhebliche Nachteile dieses Systems. Zu diesen Nachteilen gehören auf der Inputseite des politischen Systems das wachsende Ungleichgewicht zwischen Föderalismus und Demokratieprinzip und die zunehmende Vetokraft deutschsprachiger, konservativer Landkantone bei bundespolitischen Entscheidungen. Diese Mängel zeigen sich beispielhaft beim so genannten Ständemehr, d.h. der Erfordernis der Zustimmung durch die Mehrheit von Volk und Kantonen bei Verfassungsabstimmungen, das aufgrund demografischer Veränderungen immer weniger dem integrativen Ziel des Föderalismus gerecht wird (Vatter/Sager 1996; 2006). Ein zweites Problem des Föderalismus betrifft die Output-Seite und besteht darin, dass die unterschiedliche Ausstattung der Kantone die Gefahr von Vollzugsdefiziten beinhaltet und so teils das Postulat der Verteilungsgerechtigkeit öffentlicher Politik verletzt (Kissling-Näf/Knoepfel 1992; Kissling-Näf/Wälti 2006; Sager 2003). Die direkte Demokratie erweist sich in den genannten Untersuchungen unter anderem als disziplinierendes Rückbindungsinstrument für Staatsausgaben und Steuerbelastung (Freitag et al. 2003; Stutzer/Frey 2000).

3.4 Möglichkeiten und Grenzen

Die Gestaltungsmöglichkeiten durch die Evaluation von politischen Institutionen sind begrenzt, da die Reformfähigkeit politischer Institutionen selber beschränkt ist. Politische Institutionen erweisen sich als sehr dauerhaft, da sie vielfach mit sich selbst erhaltenden Mechanismen ausgestattet sind wie zum Beispiel das bereits genannte Ständemehr, dessen Änderung oder Abschaffung selber ein doppeltes Mehr erfordern würde. Analog sind die meisten politischen Institutionen mit einer Vielzahl von Vetopunkten ausgestattet, binden also mannigfaltige Interessen, die sich einer Änderung aus Eigennutzüberlegungen entgegenstellen müssen.

Die Evaluation politischer Institutionen findet denn auch primär mit nicht-instrumentellen Zwecken statt, in Form von Grundlagenforschung⁵ im Sinne einer begleitenden Reflexion über die Wechselwirkung von politischem System und Gesellschaft, und nur in den wenigsten Fällen als gezielte Auftragsforschung mit dem instrumentellen Ziel, Änderungsempfehlungen abzuleiten. Ein Beispiel für Letzteres ist die Evaluation des Vernehmlassungsverfahrens mit Blick auf die Stellung der Kantone von Gerheuser et al. (1997).

Im besten Falle erhalten Evaluationen politische Relevanz über den Umweg der kognitiven Nutzung (Weiss 1998a), wenn sie zu einer veränderten Wahrnehmung politischer

⁵ Für die Abgrenzung von Grundlagenforschung und angewandter Forschung im Evaluationskontext vgl. Ritz 2003: 31-35.

Institutionen aufgrund von Forschungsergebnissen führen. Ein aktuelles Beispiel ist der politische Umgang mit Metropolitanräumen, für die mehrere Untersuchungen die negativen Folgen der eklatanten Deckungsungleichheit zwischen politischen und funktionalen Räumen nachgewiesen haben (Kübler et al. 2003; 2005; Sager 2002). Diese Sichtweise wurde bei der politischen Lösungsfindung zumindest teilweise aufgenommen (TAK 2004).

4 Ausblick

Der vorgenommene Vergleich der Evaluation von Institutionenpolitik und von politischen Institutionen in der Schweiz hat die Unterschiede zwischen diesen beiden Evaluationsformen aufgezeigt. Insgesamt spricht die Reformträgheit politischer Institutionen für die Beibehaltung derer Evaluation als nicht-instrumentell angelegte Grundlagenforschung. Hingegen weisen die beschriebenen Grenzen der Evaluation von Institutionenpolitik auf die Notwendigkeit einer stärkeren Nutzungsorientierung dieser Form von Evaluationen hin. Dies beinhaltet angesichts der Wechselwirkungen der verschiedenen Arten institutioneller Politik nicht zuletzt eine bessere Berücksichtigung der Kontextabhängigkeit von Institutionenpolitik. Die NPM-Reform der Basler Vorortsgemeinde Riehen ist nicht dasselbe wie die NPM-Reform der Stadt Bern. Der institutionelle Kontext spielt eine wesentliche Rolle für die Funktions- und Wirkungsweise von Institutionenpolitik, was ein Abkommen von so genannten *best practice*-Modellen und einen stärkeren Fallbezug von Evaluationen nahe legt (Ladner et al. 2007).

Das Überwinden der rein symbolischen Nutzung von Evaluationen von Institutionenpolitik bedingt, wie ebenfalls gesehen, eine Verbesserung der Rahmenbedingungen ihrer Durchführung. Dies betrifft namentlich die Professionalisierung der Auftraggeberseite und die ausreichende Ressourcenausstattung der evaluierenden Stellen. Unter solchen favorablen Bedingungen ist eine wachsende Bedeutung von informierten Begleitevaluationen von Institutionenpolitik im hier definierten Sinne, die Lernprozesse auslösen und zu Verbesserungen führen können, denk- und wünschbar.

Die Evaluation von Institutionenpolitik in Deutschland: Verwaltungsmodernisierung und Wirkungsanalyse im föderalen System

1 Einleitung

Der folgende Beitrag befasst sich mit der Evaluation institutioneller Reformpolitik in Deutschland. Der Fokus liegt dabei in zeitlicher Hinsicht auf den Entwicklungen der vergangenen zehn Jahre und inhaltlich auf Maßnahmen der Binnenreform öffentlicher Verwaltung, also auf interner Verwaltungspolitik, wohingegen weitergehende Institutionenreformen (Territorialreformen, Funktionalreformen, Privatisierungspolitik und partizipative Institutionenreformen) an diese Stelle ausgeklammert bleiben (vgl. dazu den Beitrag von Sager in diesem Band; ferner Kuhlmann 2008). Ziel ist es, einen Überblick über den inzwischen erreichten Entwicklungsstand institutionell interessierter Evaluation auf den verschiedenen Ebenen des deutschen föderalen Systems zu geben, wohingegen die substanziellen Erträge der verschiedenen Evaluationsuntersuchungen hier nur sehr knapp und eher exemplarisch zur Sprache kommen können. „Evaluationstypologisch“ soll es im Folgenden vor allem um die Bilanzierung der „Evaluation *von* Institutionenpolitik“ (evaluation *of*) in Deutschland gehen, womit eine Meta-Perspektive eingenommen wird. Dagegen wird auf Maßnahmen zur fortlaufenden internen Selbstevaluation und „Erfolgskontrolle“, wie Controlling, Berichtswesen, Benchmarking, Leistungsmessung et cetera, also „Evaluation *in* Institutionen“ (evaluation *in*) hier nicht systematisch eingegangen (siehe dazu Kuhlmann 2005)¹. Es sollen zum einen jene Evaluationsaktivitäten betrachtet werden, die von einer unabhängigen Forschungseinrichtung (sei es als Auftragsforschung, sei es als Eigenprojekt) durchgeführt worden sind beziehungsweise in denen die Initiierung, Durchführung oder Beauftragung der Evaluation durch Institutionen „außerhalb des Bannkreises der zu evaluierenden Verwaltungseinheit“ (Wollmann 2002: 82-84), also etwa durch Parlament, Rechnungshof oder unabhängige Reformkommissionen etc., erfolgt ist („externe Evaluationen“). Zum anderen werden solche Untersuchungen herangezogen, bei denen das Evaluationsverfahren zu einer institutionellen Reformmaßnahme entweder direkt in der operativen Verwaltungseinheit oder zumindest im unmittelbaren Kontext von Verwaltung und Exekutive initiiert,

¹ Generell zur Unterscheidung von Evaluation „in“ und „von“ siehe Widmer 2002. In Anlehnung an Wollmann (2002) werden Controlling, Monitoring, Benchmarking und andere Instrumente aus dem performance-orientierten NPM-Konzeptrepertoire hier als Formen der internen Selbstevaluation in Institutionen („evaluation in“) aufgefasst, womit dem vorliegenden Beitrag ein eher weites Begriffsverständnis von „evaluation in“ zu Grunde liegt.

institutionalisiert und durchgeführt wurde („interne Evaluation“)². Kurz: es geht um die Bilanzierung *externer* und *interner* Evaluation von Verwaltungsmodernisierung in Deutschland.

2 Evaluation von Verwaltungs(reform)politik: konzeptionelle, methodische und institutionelle „Fallstricke“

Um den „Objektbereich“ (Evaluandum) der Evaluation institutioneller Politik zu umreißen, sei noch einmal an die in der Evaluationsforschung gebräuchliche Unterscheidung zwischen „normaler“ sektoraler (oder auch „substanzieller“) Politik (policy) und Verwaltungs- oder Institutionenpolitik erinnert (siehe auch Sager in diesem Band). Institutionen- oder Verwaltungspolitik wird als das zu evaluierende Handlungsprogramm (policy) und als der Versuch politisch-administrativer Akteure aufgefasst, die institutionelle Ordnung, innerhalb derer sie Entscheidungen treffen (polity), zu verändern. Sie wird daher auch als „polity-policy“ bezeichnet (Wollmann 2000: 199-200; Ritz 2003: 180). Da sich Verwaltungspolitik somit in der Regel als Verwaltungsreform darstellt, ist der Gegenstandsbereich der Evaluation die jeweilige Verwaltungsreformmaßnahme beziehungsweise die durch sie ausgelöste Wirkung. Die Evaluation von Verwaltungspolitik bedeutet also normalerweise, dass Verwaltungsreformen evaluiert werden, so dass die in anderen (substanziellen) Politikfeldern gebräuchliche analytische Differenzierung der Evaluation von *Politiken* einerseits und der Evaluation von *Reformen* andererseits hier wenig Sinn macht.

Dieser Typus von Policy ist mit spezifischen Steuerungsproblemen verbunden (Einheit von Subjekt und Objekt der Intervention; Veränderungen innerhalb des politisch-administrativen Systems als Interventionsziel, Wichtigkeit von Leitideen etc.; vgl. Jann 2001: 330-332; Benz 2004: 22). Für die Wirkungsanalyse hat dies zur Folge, dass die Evaluation von Institutionenpolitik – im Unterschied zur Evaluation von Sektoralpolitiken – durch eine komplexere analytische Architektur gekennzeichnet ist. Denn der Gegenstand institutionenpolitischer Wirkungsanalysen sind zunächst bestimmte Veränderungen innerhalb des politisch-administrativen Systems (Institutionenevaluation), sodann die Auswirkungen dieser auf das Handeln der Akteure und die Leistungsfähigkeit der Institution (Performanzevaluation) und schließlich in einem dritten Analyseschritt die weiteren Wirkungen im Umfeld des politisch-administrativen Systems (Outcome-Evaluation). Wie zu zeigen sein wird, beziehen sich die verfügbaren Studien zur Wirkungsanalyse von Verwaltungsreformen in Deutschland überwiegend auf den ersten Analyseschritt. Deutlich weniger empirische Evidenz herrscht hinsichtlich der Performanzevaluation und zu den Outcome-Effekten liegen nur vereinzelte Hinweise, aber keine systematischen Untersuchungen vor.

Neben methodisch-konzeptionellen „Fallstricken“ (siehe Wollmann 2000: 201-203) treffen institutionelle Evaluationen auf erhebliche Beschränkungen dadurch, dass der Gegenstandsbereich einer Verwaltungsreformevaluation im deutschen föderalen System hochgra-

² In der Praxis ist die Zuordnung freilich nicht immer trennscharf vorzunehmen. So bewegen sich etwa Berichte der Ministerien zum Umsetzungsstand von Verwaltungsreformmaßnahmen zwischen Monitoring/„Selbstevaluation“ (als „evaluation in“) und „interner Evaluation“ (im Sinne von „evaluation of“).

dig vertikal und horizontal fragmentiert und außerordentlich vielgestaltig ist. Während in unitarischen Ländern sowohl im Reformprozess als auch bei dessen Evaluation oft ein zentraler Akteur eine bestimmende Rolle spielt (z. B. in England die nationale Audit Commission), erweist sich bei der Initiierung und Durchführung von verwaltungspolitischen Evaluationen die das politisch-administrative System der Bundesrepublik insgesamt prägende dezentral-fragmentierte Struktur als erschwerend (Wollmann 2001: 14; siehe auch den Beitrag von Pleschberger in diesem Band). Dass bislang in der Bundesrepublik – ebenfalls im Unterschied zu anderen Ländern – ein allenfalls geringes Interesse der Parlamente und Rechnungshöfe an einer Evaluation der Verwaltungsmodernisierung zu konstatieren ist, dürfte aber auch daraus resultieren, dass die Ergebnisse solcher Studien, zumal wenn kritische bis schlechte „Erfolgsbilanzen“ in Aussicht stehen, den politischen Profilierungs- und Selbstdarstellungsabsichten der Akteure wenig entgegenkommen (hierzu weiter unten).

3 Evaluation der Verwaltungsmodernisierung in Deutschland

Einerseits hat der Aufschwung, den die Politikevaluation in der Bundesrepublik in den 1960er und 1970er erlebte, dazu beigetragen, dass sich die Evaluationsforschung (im Bereich substanzieller Politik) als Analyse- und Informationsverfahren bei der Politikvorbereitung und Wirkungskontrolle fest etablieren konnte (Wollmann 2001: 14; Derlien 1997). Andererseits weist Deutschland nach wie vor ein markantes Evaluationsdefizit im Hinblick auf verwaltungspolitische Reformprogramme auf, welches angesichts dessen erstaunen muss, dass die öffentlichen Akteure zum Teil erhebliche Ressourcen in die Konzipierung und Implementation entsprechender Projekte investieren, deren Erfolgsbilanz mangels korrespondierender Wirkungsanalysen weitgehend im Dunkeln bleibt. Auch mit Blick darauf, dass die Bedeutung von institutionellen Arrangements und von Institutionenpolitik gerade in der neueren politik- und verwaltungswissenschaftlichen Theoriediskussion (wieder) in den Vordergrund getreten ist, erscheint die Vernachlässigung der Evaluationsfunktion in diesem Bereich öffentlicher Politik als eine auffällige Forschungslücke. Wie im vorangehenden Abschnitt dargelegt, lässt sich jedoch für dieses Defizit eine Reihe von Gründen benennen, die bislang entsprechende Evaluationsaktivitäten in diesem Bereich öffentlicher Politik erschwert, wenn nicht teils gänzlich verhindert haben. Insgesamt kann konstatiert werden, dass die Transparenzforderung, die den neueren Ansätzen der Verwaltungsmodernisierung in Deutschland zugrunde liegt, auf die Wirkungsanalyse und Evaluation eben dieser Maßnahmen nicht oder nur sehr begrenzt „durchgeschlagen“ hat. Während Maßnahmen der Leistungsmessung und Erfolgskontrolle als Verfahren der fortlaufenden internen Selbstevaluation im Zuge des betriebswirtschaftlichen Managerialismus enorm an Bedeutung gewonnen und sich inzwischen in vielen Verwaltungen ausgebreitet haben (vgl. Kuhlmann 2005), sind externe Evaluationen von Verwaltungsreformprojekten, nicht zuletzt jener, welche die Installierung der Evaluationsfunktion in der Verwaltung zum Gegenstand haben (z. B. Aufbau von Controlling-Systemen), nach wie vor nur vereinzelt anzutreffen. Die zu Beginn dieses Jahrzehnts konstatierte „Evaluationslücke“ in der deutschen Verwaltungspolitik (siehe Wollmann 2002) ist damit bislang nicht geschlossen worden.

Wirft man einen Blick auf die verschiedenen föderalen Ebenen des deutschen Politik- und Verwaltungssystems, so lässt sich eine grundlegende Feststellung machen. Die Modernisierungsaktivität im Bereich der administrativen Binnenstrukturen und -prozesse nimmt ebenso wie die Intensität und Breite ihrer Evaluation ab, je weiter man im föderalen System „nach oben“ steigt. Greift man auf die Theorie der „heuchelnden Organisation“ von Bruns-son (1989) zurück, so ist die Kluft zwischen „Reden“ (talk) und tatsächlichem Reformhandeln (action) und damit empirisch messbaren Reformeffekten (results/impacts) auf der Bundesebene am größten, während sie in den Kommunen am geringsten ausgeprägt ist und die Länder eine mittlere Position einnehmen. Diese Tatsache stellt natürlich auch die Evaluation von Verwaltungsreformpolitik vor Probleme, da es ihr auf der Ebene der Bundesbürokratie in weiten Teilen schlicht an Substanz und empirischem „Anschauungsmaterial“ fehlt. Hinzu kommen politisch-institutionelle Erklärungsfaktoren, die den unterschiedlichen Reform- wie auch Evaluationsstand auf den verschiedenen föderalen Ebenen erklären und auf die im Folgenden näher einzugehen sein wird.

3.1 Bundesebene

Die oben konstatierte Evaluationslücke institutioneller Politik klappt auf der Bundesebene am weitesten. Lediglich für den Umbau der Bundesagentur für Arbeit im Gefolge der Hartz IV-Reform, die nach dem hier verfolgten Verständnis von Institutionenpolitik eine verwal- tungspolitische Reformmaßnahme in einem substanziellen Politikfeld darstellt, wurden umfassende externe Evaluationen in Auftrag gegeben (vgl. Jann 2006; ISO 2006). Zudem gibt es regelmäßige und umfangreiche Dokumentationen zur Umsetzung von Reformelementen, wie Kontraktmanagement, Kosten- und Leistungsrechnung, Controlling, dezentraler Res- sourcenvverantwortung und Budgetierung. Diese basieren jedoch ausschließlich auf interner Berichterstattung der Bundesregierung (siehe Jann 2006), die – angefangen beim „General- kalender Schlanker Staat“ über die verschiedenen Berichte der Stabsstelle „Moderner Staat – Moderne Verwaltung“ (vgl. z. B. BMI 2002) bis hin zu den aktuellen Veröffentlichungen zur „Initiative Bürokratieabbau“ (siehe BMI 2004) – allesamt exekutive Selbstdarstellungen sind, denen man nur mit Not eine interne Evaluationsfunktion zuschreiben kann. Denn sie bein- halten im Wesentlichen Zählungen instrumenteller Umsetzungen (Zahl der Kontrakte, Zahl der Projekte zur Einführung der Kosten- und Leistungsrechnung etc.) und haben unüber- sehbar die Funktion institutioneller „Außenwerbung“. Eine zaghafte, aber immerhin er- wähnenswerte Neuorientierung in Richtung verstärkter externer Evaluation von Verwal- tungsreformmaßnahmen auf Bundesebene wird in den jüngsten Aktivitäten des Bundes- rechnungshofes (wie auch der Landesrechnungshöfe; siehe weiter unten) sichtbar. Auf Bitte des Haushaltsausschusses des Deutschen Bundestages führte er eine Begleitforschung zur modellhaften Erprobung flexibler Haushaltsverfahren sowie zur Einführung von Produkt- haushalten, Kosten- und Leistungsrechnung und zur erweiterten Kameralistik in der Bun- desverwaltung durch und erstattete hierzu Bericht (vgl. Bundesrechnungshof 2006). Diese Aktivitäten könnten einen ersten Schritt in Richtung einer systematischen externen Instituti- onenevaluation auf Bundesebene markieren. Sie könnten zudem insgesamt zur Aufwertung

der evaluativen Funktion des deutschen Bundesrechnungshofes in Fragen der Verwaltungsmodernisierung, die dieser bisher kaum wahrgenommen hatte, führen.

Zu den „Performanzeffekten“ dieser institutionellen Veränderungen (Input- und Outputeffekte, Outcomes) gibt es bislang keinerlei verlässliche Untersuchungen (Jann 2006: 110). Das Defizit an Performanzevaluationen hängt – neben den generellen Problemen institutionenorientierter Evaluation (vgl. Wollmann 2000) – mit der besonderen Stellung der Bundesebene im deutschen föderalen System zusammen (siehe Jann 2006: 100-102). Dadurch greifen, im Unterschied zur Landes- und vor allem zur Kommunalebene, eine Reihe von „Reformtreibern“ nur wenig und muss sich somit auch die Evaluation der Verwaltungsmodernisierung in Grenzen halten. Im deutschen Verwaltungsföderalismus fehlt für verwaltungspolitisch interessierte externe wie interne Evaluation auf Bundesebene schlicht der Untersuchungsgegenstand. So ist der Bund im Wesentlichen für Regulierung, Regelsetzung und Programmformulierung zuständig. Er verfügt nur begrenzt über administrative Funktionen und kaum über einen eigenen Behördenunterbau, da die Verwaltungszuständigkeiten weitgehend bei den Ländern und innerhalb dieser zuallererst bei den Kommunen liegen. Der Bund ist damit qua Verfassung „bürgerfern“. Eine Reihe von Reforminstrumenten des NPM taugt vor diesem Hintergrund nur bedingt zur Implementation auf Bundesebene. Sie lassen sich daher weder im Hinblick auf ihre Umsetzung evaluieren (Institutionenevaluation) noch hinsichtlich ihrer weiteren Effekte (Performanz- und Outcome-Evaluation). In Fragen der Entkoppelung von politischer Zielformulierung und administrativer Durchführung, wie sie in Agency-Modellen verfolgt wird, weist der deutsche Verwaltungsföderalismus aufgrund der Abschichtung der Verwaltungszuständigkeit auf die Länder- und Kommunalebene ohnedies einen Reformvorsprung auf, so dass auch hier ein „Reformdruck“ auf Bundesebene nur bedingt gegeben ist. Hinzu kommt, dass wesentliche verwaltungspolitische Themen in Deutschland nicht mit Fragen der Verwaltungsmodernisierung verbunden worden sind, wozu lediglich die Bundesagentur für Arbeit eine Ausnahme bildet, und zudem der „Verwaltungsaufbau Ost“ eine „Sternstunde“ der klassisch-Weberianischen Bürokratie und nicht der (manageriellen) Verwaltungsmodernisierung war. Die Beharrungskraft des Bundes in Fragen der Verwaltungsmodernisierung (im Vergleich zu anderen Ebenen wie auch international) wird vor diesem Hintergrund verständlich und erklärt zumindest auch teilweise das Fehlen entsprechender Evaluationen.

3.2 Landesebene

Etwas optimistischer fällt die Bilanz für die Landesebene aus. Zwar ist auch hier das Fehlen einer systematischen, flächendeckenden externen Evaluation der Verwaltungsmodernisierung zu beklagen (siehe Reichard 2006b: 87). Jedoch haben die Länder inzwischen eine Reihe vorzeigbarer Modernisierungsaktivitäten verfolgt und liegen hierzu auch einige interne und externe Evaluationen vor, auf die sich die hier verfolgte Meta-Bilanz stützen kann. Hinsichtlich der internen Evaluationen ist vor allem auf Berichte der Landesregierungen sowie einzelner Landesbehörden zu verweisen, die teils mit einem expliziten Selbstevaluierungsanspruch ihrer Reformprojekte auftreten (etwa in Baden-Württemberg, Berlin oder NRW; siehe beispielhaft Senatsverwaltung für Inneres 1999). Basierend auf diesem Daten-

material wurden inzwischen auch zusammenfassende und vergleichende Darstellungen zum institutionellen Umsetzungsstand der Verwaltungsmodernisierung in den deutschen Bundesländern erstellt (vgl. Untersuchungsausschuss „Allgemeine Verwaltungsorganisation“ 2002; BMI 2000; Konzendorf 1998; Reichard 2006a sowie Datensammlung des WiDuT³). Darüber hinaus finden sich für einzelne Bundesländer externe wissenschaftliche Evaluationen, in denen allerdings ebenfalls überwiegend der institutionelle Umsetzungsstand der Reformmaßnahmen in den jeweiligen Landesverwaltungen untersucht worden ist. So liegen inzwischen Studien zur Bilanz der Verwaltungsmodernisierung in Berlin (siehe Engelniederhammer et al. 1999) und einige länderübergreifende Arbeiten vor (Kißler et al. 2000; Bürsch/Müller 1999). Kennzeichnend für diese Studien ist – im Unterschied zu vorliegenden Untersuchungen auf kommunaler Ebene (siehe weiter unten) –, dass sie sich (noch) weniger auf eine einheitliche „Zieldimension“ der Institutionenpolitik (NSM-Modell) stützen können, da die Länder eher mit einzelnen betriebswirtschaftlichen Instrumenten, vor allem im Bereich des Haushaltsmanagements und der Ressourcenbewirtschaftung, experimentieren und kein in sich geschlossenes (und als solches „evaluierbares“) Reformkonzept verfolgen. Dies stellt an die Wirkungsforschung die bekannten Zielbildungs- und Konzipierungsprobleme, was ihre bislang „magere“ Ausprägung mit erklärt.

Unbeschadet dieser Schwierigkeiten haben inzwischen einige Landesrechnungshöfe bemerkenswerte Vorstöße in Richtung einer externen Evaluation der Verwaltungsmodernisierung auf Landesebene unternommen, womit auch diese sich – noch ausgeprägter als der Bundesrechnungshof – zunehmend in institutionenevaluativen Fragen engagieren, was bislang für den deutschen Reformkontext ausgesprochen untypisch war⁴. Besonders hervorzuheben ist dabei der im März 2007 publizierte Bericht des Rechnungshofes Baden-Württemberg, der die bisherige Umsetzung Neuer Steuerungsinstrumente (NSI) in den Landesbehörden kritisch bilanziert und zu dem ernüchternden Ergebnis kommt, dass sich die „bis Ende 2005 angefallenen Projektkosten für die flächendeckende Einführung der NSI in Höhe von 220 Mio. Euro (...) bisher nicht annähernd amortisiert“ hätten (Rechnungshof Baden-Württemberg 2007: 1). Ferner wird festgestellt, dass die „NSI keines ihrer Ziele umfassend erreicht“ hätten sowie, dass sie abgesehen vom Haushaltsmanagementsystem bisher „kaum positive Wirkungen in Bezug auf die Effizienz und die Effektivität der Landesverwaltung entfaltet“ hätten (Rechnungshof Baden-Württemberg 2007: 2). In eine ähnliche Richtung weist der im Jahre 2006 vom Berliner Rechnungshof vorgelegte Bericht zur Umsetzung und zu den Wirkungen Neuer Steuerungsinstrumente in den Senatsverwaltungen (insbesondere dezentrale Ressourcenverantwortung, Ziel- und Servicevereinbarungen, Kosten- und Leistungsrechnung, Steuerungsdienste, Qualitätsmanagement, Personalmanagement). Auch hier ist davon die Rede, dass „die von der Verwaltungsreform erwarteten Erfolge und Wirkungen (...) bisher nur in geringem Umfang eingetreten (sind)“ (vgl. Rechnungshof von

³ Das WiDuT ist eine Transferstelle am Forschungsinstitut der Deutschen Hochschule für Verwaltungswissenschaften in Speyer; vgl. www.foev-speyer.de/widut [10.4.2008].

⁴ Zwischenbilanzierungen binnenorientierter Verwaltungsmodernisierung, insbesondere „Neuer Steuerungsinstrumente“ und teils unter Einbeziehung weitergehender Wirkungsfragen (Performanz), liegen inzwischen vor vom Sächsischen Rechnungshof, vom Landesrechnungshof Schleswig-Holstein, vom Rechnungshof von Berlin, vom Rechnungshof Baden-Württemberg und vom Bayerischen Obersten Rechnungshof. Sie sind nach dem hier verwendeten Verständnis als externe Institutionenevaluationen einzuordnen.

Berlin 2006: 6). Diese beiden Berichte können allerdings bislang eher als Ausnahmebeispiele evaluatorischer Tätigkeit von Rechnungshöfen in Fragen der administrativen Binnenmodernisierung angesehen werden. Dies gilt namentlich für ihre explizite Bezugnahme auf Performanzeffekte der Modernisierung (Effizienz, Effektivität, Einsparungen, Qualität etc.), die ansonsten in diesem „Prüfbereich“ nur selten anzutreffen ist.

Ähnlich wie auf der Bundesebene liegen jedoch auch für die Länder Untersuchungen zu den weiterreichenden Effekten der Verwaltungsmodernisierung (Performanzevaluation) kaum vor, geschweige denn sind diese in flächendeckender oder auch nur länderübergreifender Form verfügbar. Dagegen gibt es für andere institutionelle Reformfelder der Landesebene, wie Verwaltungsstruktur-, Funktionalreformen und Gebietsreformen, inzwischen externe Wirkungsstudien, teils mit dem Anspruch einer ersten ex post Bilanz (siehe Bogumil/Kottmann 2006; Bogumil/Ebinger 2005; Bauer et al. 2007), teils – aus naheliegenden zeitlichen Gründen – in Form von ex ante-Wirkungsabschätzungen (siehe bspw. Seitz 2005). Schließlich sei noch auf ein laufendes externes Evaluationsprojekt hingewiesen, welches sich der Wirkungsanalyse staatlicher Aufgabendezentralisierung in international vergleichender Perspektive widmet und damit evaluative Befunde zur Dezentralisierung und Funktionalreform in Westeuropa (darunter auch in den deutschen Bundesländern) erwarten läßt.⁵

3.3 Kommunen

Auf kommunaler Ebene ist inzwischen ein durchaus bemerkenswerter Stand institutionell orientierter Evaluation zu konstatieren. Dieser hängt zweifelsohne vor allem mit dem fortgeschrittenen Reformstand und damit zusammen, dass die Kommunen die Vorreiter der Verwaltungsmodernisierung in Deutschland waren und sind, was wiederum institutionell in ihrer Rolle als den Hauptträgern des staatlichen Verwaltungsvollzugs im deutschen Verwaltungsföderalismus begründet liegt. Vor diesem Hintergrund hat sich inzwischen eine Reihe von Wissenschaftlern und externen Forschungseinrichtungen mit Wirkungsfragen kommunaler Verwaltungsmodernisierung in Deutschland befasst und wurden diese Aktivitäten durch eine entsprechende Forschungsförderung, allen voran durch die Schwerpunktprogramme der Hans-Böckler-Stiftung, finanziell unterstützt. Im Ergebnis ist heute die Zahl der Evaluationen, die sich mit den institutionellen Effekten der lokalen Verwaltungsmodernisierung in Deutschland befassen, kaum noch zu überschauen (siehe im Überblick Kuhlmann 2004). Zum einen liegen Umfragen zum Implementationsstand des NSM vor, die von den Kommunalverbänden selbst initiiert, konzipiert, durchgeführt und ausgewertet wurden. Sie stellen eine Form der internen Evaluation dar, weisen jedoch das Defizit auf, dass die Bereitschaft eines kommunalen Akteurs, sich „gegenüber seinem kommunalen Spitzenverband gegebenenfalls als modernisierungs'lahm' zu ‚outen‘“ (Wollmann 2000: 219), gering ist und die Ergebnisse somit eher das Spiegelbild kommunaler Selbstdarstellungsabsichten sind. Hinsichtlich der vorliegenden externen Evaluationen kommunaler Verwal-

⁵ DFG-Projekt „Wandel europäischer Lokalsysteme: Wirkungen institutioneller Dezentralisierungspolitik in Deutschland, Frankreich und Großbritannien“ der Ruhr-Universität Bochum und der Humboldt-Universität zu Berlin unter Leitung der Verfasserin und Prof. Dr. Jörg Bogumil (siehe www.uni-potsdam.de/db/lis_regierungssystem_brd/index.php?article_id=132&clang=0). [19.3.2008]

tungsmodernisierung gibt es ebenfalls zwei markante Defizite. Zum einen haben die überwiegend qualitative Begleitforschung und die Fallanalysen die methodische Beschränkung, dass sich aus ihnen nur wenig Aussagen über den Verbreitungsgrad der beobachteten Befunde (also etwa Umsetzung von Instrumenten) ableiten lassen. Zum anderen bestand eine Schwachstelle der bisherigen Wirkungsuntersuchungen zur lokalen Verwaltungsmodernisierung, ähnlich wie auf den anderen föderalen Ebenen, darin, dass keinerlei systematische Aussagen über Performanzeffekte der Reformmaßnahmen (Performanzevaluation) möglich waren, und lediglich die Wirkungsebene der „institutionellen Veränderungen“ betrachtet worden ist.

Inzwischen ist allerdings im Bereich der externen Performanzevaluationen kommunaler Verwaltungsmodernisierung empirisch „nachgelegt“ worden. Vor allem ist hier auf die im Jahre 2007 publizierte NSM-Evaluation in deutschen Kommunen zu verweisen, mit der erstmals eine umfassende und flächendeckende Wirkungsanalyse der Verwaltungsmodernisierung in deutschen Kommunen vorgelegt worden ist. Auf der Grundlage einer repräsentativen Umfrage und intensiver Fallstudien wurden sowohl die institutionellen als auch die Performanzeffekte der Reform untersucht, so dass die Studie eine Kombination aus Institutionen- und Performanzevaluation darstellt⁶. In der Gesamtbilanz wird einerseits konstatiert, dass zahlreiche positive Reformeffekte nicht von der Hand zu weisen sind, die sich etwa in positiven Einschätzungen der lokalen Akteure im Rahmen der Umfrage niederschlagen und auch an einschlägigen „Leistungsparametern“ abzulesen sind (Verbesserungen im Serviceniveau, in der Kundenorientierung, Bearbeitungszeit, Prozessgestaltung, Leistungsqualität etc.). Andererseits ist nach mehr als zehn Jahren Verwaltungsmodernisierung noch immer eine NSM-Implementationslücke festzustellen und kann von flächendeckender und allumfassender NSM-Reform bei Weitem keine Rede sein. Hinzu kommt, dass einige Kommunen aufgrund negativer Folgewirkungen oder ausbleibender positiver Effekte wesentliche Reformelemente (etwa „über-dezentralisierte“ Strukturen) wieder zurückbauen und es teils sogar zur Re-Bürokratisierung kommt. Denn bei näherem Hinsehen fällt auf, dass die besonders wirksamen Reformschritte nicht dem Konzeptrepertoire des neuen Managementmodells entstammen, sondern auf „altbekannten“ traditionellen Reformansätzen basieren, die durch die NSM-Diskussion „wiederbelebt“ wurden. Daher kann weder von einem unmittelbaren und direkten Zusammenhang zwischen NSM und Leistungsverbesserungen noch von NSM-bedingten Einsparungen die Rede sein. Allerdings gibt es eher mittelbare Effekte dergestalt, dass der NSM-Diskurs ein günstiges „Modernisierungsklima“ geschaffen und quasi ein Gelegenheitsfenster (*window of opportunity*) geöffnet hat. Damit wurde der Weg geebnet, um viele der schon früher diskutierten Reformansätze wieder aufzugreifen und umzusetzen. Was dagegen die eigentlichen betriebswirtschaftlich-manageriellen Kernelemente des NSM angeht, fällt die Einschätzung deutlich kritischer aus.

⁶ Die Evaluation wurde von der Hans-Böckler-Stiftung finanziert und von 2004-2006 von Jörg Bogumil (Ruhr-Universität Bochum – Projektleitung), Stephan Grohs, Anna Ohm (Universität Konstanz) sowie der Verfasserin durchgeführt. Beteiligt waren darüber hinaus Werner Jann, Christoph Reichard (Universität Potsdam); Leo Kißler (Universität Marburg) und Hellmut Wollmann (Humboldt-Universität Berlin). Zu Konzeption und Ergebnissen der Studie siehe ausführlich Bogumil et al. 2006; 2007).

4 Evaluation institutioneller Politik im föderalen System: Erklärungsversuche zur Ebenen-Varianz

Insgesamt bleibt festzuhalten, dass der Stand institutionenorientierter externer Evaluation in Deutschland auf der kommunalen Ebene am weitesten fortgeschritten ist. Dies hat damit zu tun, dass – institutionell und funktional bedingt – die „Lücke“ zwischen Reformrhetorik (talk) und Reformpraxis (action) auf der lokalen Ebene am geringsten ausgeprägt ist, so dass sich auch substanzielle Wirkungen von Verwaltungsmodernisierung hier eher feststellen lassen. Der Typus der „heuchelnden Organisation“ (Brunsson 1989) scheint sich in den Kommunen dadurch, dass sie unmittelbar problemlösend und bürgernah tätig sind, weniger auszuprägen als auf der Landes-, vor allem aber auf der Bundesebene. Der vergleichsweise weit vorangeschrittene Evaluationsstand im Bereich der lokalen Verwaltungsmodernisierung ist funktional damit zu begründen, dass das Gros der Verwaltungstätigkeit und des Gesetzesvollzuges in Deutschland bei den Kommunen liegt (ca. 85 % aller Bundes- und Landesgesetze und der größte Teil des EU-Rechts; vgl. Schmidt-Eichstaedt 1999: 330). Der Reformdruck lastet somit auf der kommunalen Verwaltung am schwersten und kann von dieser kaum ignoriert oder durch bloße „Rhetorik“ und Programmatik überdeckt werden, was auf die anderen Ebenen bislang weitaus weniger zutraf. Das „empirische Material“ für verwaltungspolitisch interessierte Evaluationsforschung ist vor diesem Hintergrund in den Kommunen sehr viel reichhaltiger, umfassender und substanzieller, was sich in den zitierten Studien zur kommunalen Ebene bei gleichzeitigem Fehlen äquivalenter Untersuchungen auf den anderen Ebenen auch deutlich widerspiegelt. Last but not least, dürfte ein Grund für den unterschiedlichen Stand von Verwaltungsreformevaluationen auf den verschiedenen Ebenen des deutschen Regierungssystems darin zu suchen sein, dass die politische Brisanz solcher Untersuchungen auf der Bundesebene als höher empfunden wird und man sich ihnen deshalb massiver entgegenstellt, als auf der Landes- und Kommunalebene. Bedenkt man, dass unter konkurrenzdemokratischen Bedingungen schlechte Reformbilanzen in den Händen des politischen Gegners diesem eine willkommene Munition im politischen Wettbewerb bieten, so erklärt sich auch aus dieser Sicht die Zurückhaltung der Bundesexekutive in Fragen objektiver (externer) Wirkungsmessung ihrer eigenen Verwaltungspolitik. Ähnliches trifft auf die Länder zu, wenngleich einige Landesrechnungshöfe hier inzwischen einen bemerkenswerten „Kurswechsel“ vollzogen haben. Zwar erweist sich auch auf kommunaler Ebene die konkurrenzdemokratische Logik in weiten Teilen als politisch-institutionelle „Bremse“ der Wirkungsforschung (vor allem in Lokalsystemen mit ehemals „Norddeutscher Ratsverfassung“). Jedoch dringen hier infolge der Kommunalverfassungsreformen zunehmend Elemente lokaler Konkordanzdemokratie vor, durch die der Parteienwettbewerb teilweise abgeschwächt wird. Damit bewirkt auch diese Entwicklung, dass die Institutionalisierung einer internen Evaluationsfunktion wie auch die mögliche Ausbreitung externer Wirkungsanalysen auf kommunaler Ebene begünstigt werden. Dies verhält sich zu der Feststellung konsistent, dass in stark parlamentarisierten, konkurrenzdemokratischen Systemen der Institutionalisierung von Evaluation eher Grenzen gesetzt sind als in präsidialen oder stärker konkordanzdemokratischen Verfassungskontexten (vgl. Derlien 1997: 9-11). Inwieweit die sich abzeichnende weitere „Präsidialisierung“ des deutschen Lokalsystems zu

einem Vordringen evaluativer Verfahren führen wird und möglicherweise gar einen politisch-parlamentarischen Ruf nach „mehr Wirkungstransparenz“ nach sich zieht, bleibt jedoch vorerst abzuwarten.

5 Fazit und Ausblick

Wie die vorstehende Bilanz institutionenorientierter Wirkungsforschung im deutschen Kontext zeigt, lässt sich einerseits ein Zusammenhang zwischen Reformaktivität und Evaluationsintensität und -breite ausmachen. Je weiter man im föderalen System „absteigt“, desto besser stehen die Chancen, „substanzielle“ Ergebnisse und Wirkungen von Verwaltungsmodernisierung festzustellen und desto fortgeschrittener ist auch der Evaluationsstand institutioneller Politik, vor allem im hier interessierenden Bereich der Binnenmodernisierung. Andererseits lässt sich über die verschiedenen Ebenen des föderalen Systems hinweg nach wie vor ein Defizit im Bereich der Performanzevaluation ausmachen. Über die „performance“ der durch Verwaltungsmodernisierung neu geschaffenen Institutionen herrscht in weiten Teilen fortwährend Unkenntnis, wobei das „Wissensdefizit“ wiederum auf der Bundesebene am größten, bei den Ländern und Kommunen inzwischen etwas geringer ist. Gleiches trifft auf die Frage einer verlässlichen Kosten-Nutzen-Bilanz institutioneller Politik zu, die vor dem Problem steht, dass es kaum systematische Angaben zu den tatsächlichen Kosten der Verwaltungsmodernisierung gibt (siehe hierzu Kuhlmann/Wollmann 2006). Vor diesem Hintergrund muss die Ausweitung der empirischen Forschung zu diesen weitergehenden Wirkungsfragen institutioneller Politik als eine entscheidende Herausforderung und Aufgabe der institutionell interessierten Evaluationsforschung sowie der Politik- und Verwaltungswissenschaft insgesamt angesehen werden.

Evaluation institutioneller Politik in Österreich

Seit in den späten 1980er Jahren in den westlichen Ländern die sog. „dritte“ Welle der von den Doktrinen des New Public Management (NPM) beeinflussten Verwaltungsreformen einsetzte, ist die Evaluation institutioneller Veränderungen ein „taken-for-granted-phenomen“ der Verwaltungsreformbemühungen (Dahler-Larsen 2005: 618). Die NPM-Ideologie fordert die Transparenz über die Abläufe, Kosten und besonders die Wirkungen des Verwaltungshandelns (wie des gesamten öffentlichen Sektors), was *unter anderem* mit Hilfe von Evaluationen eingelöst werden soll (Dahler-Larsen 2005: 619). Ungeachtet ihrer nominellen Prominenz und hohen funktionellen Bedeutung ist die Evaluation nicht einfach zu definieren (Hansen 2005: 325). Breit definiert ist eine Evaluation eine methodisch geplante Aktivität mit dem Ziel, die politischen Entscheider und die breite Öffentlichkeit regelmäßig mit „guten“ Daten über die Ergebnisse und Wirkungen des Handelns von öffentlichen Verwaltungen (unter Einschluss ihrer Strukturen) zu informieren und sie anzuregen, sich stärker für reformorientierte Veränderungen zu engagieren (Skinner 2004). „Accountability“, Lernen und Reformen sind heute typische Anliegen der Evaluation institutioneller Veränderungen im politischen System, wovon man sich die Verbesserung der Effizienz und Effektivität der staatlichen Leistungsproduktion erwartet (Hansen 2005: 325). Weiterhin hat die Evaluation eine Legitimationsfunktion für die politischen und administrativen Akteure. Sie ist ein Symbol für das evidenzbasierte Verwaltungshandeln und verbreitet die für die NPM-inspirierte Verwaltungsmodernisierung insgesamt typische „accounting aura“ (Greve 2003: 276), was nicht automatisch ihre „echte“ Anwendung in institutionellen Reformprozessen bedeuten muß (Hammerschmid 2006: 83, 88; Naschold 1995).

1 Politisch-institutionelle Einflussfaktoren

In Österreich ist die Evaluation institutioneller Veränderungen fast ausschließlich im Kontext der NPM-Verwaltungsmodernisierung verankert,¹ wobei ihr praktischer Stellenwert maßgeblich von politisch-institutionellen Systemmerkmalen beeinflusst wird:

- Kompetenzverteilung: Der österreichische Föderalismus (z. B. Beyme 2007) teilt die Aufgaben der Gesetzgebung zwischen Bund und neun Ländern und die Aufgaben der Vollziehung zwischen Bund, neun Ländern und (rund 2350) Gemeinden. Die Länder- und Gemeindeverwaltungen vollziehen eigene und „fremde“ Verwaltungsaufgaben.

¹ Ein interessantes Beispiel wäre der Österreichische Staatsschuldenausschuß, der die Budgetpolitik der Gebietskörperschaften sachverständig evaluiert.

Während nur Bund und Länder die Entscheidungen über Verwaltungskompetenzen treffen können, können Maßnahmen der organisatorischen Verwaltungsmodernisierung von den drei staatlichen Ebenen relativ autonom gestaltet werden („Verwaltungspluralismus“).

- Finanzprobleme: Fiskalpolitischer Natur ist der Umstand, dass die Ausgaben und Schulden der öffentlichen Haushalte seit Jahrzehnten relativ unabhängig vom Konjunkturverlauf wachsen (Fleischmann 2005), was ihre Eindämmung zum Dauerthema macht. Angesichts ihrer Größe kommt dabei den sog. Verwaltungsausgaben der Gebietskörperschaften eine besondere Bedeutung zu.²
- Handlungsstil: Die Mischung zwischen hochgradiger Verrechtlichung und Bürokratisierung einerseits bei gleichzeitiger Informalisierung sowohl der Gesetzgebung und der Vollziehungsaktivitäten ist typisch für das politisch-administrative System Österreichs. Die kalkulierte Nutzung informaler Mechanismen und Instrumente erleichtert sach- und parteipolitisch ausgewogene konsensuale Lösungen (vgl. Lehmbruch 1967), was auf Kosten der Effizienz gehen kann. Demgegenüber setzt Evaluation idealerweise eine Systemrationalität voraus, die der Transparenz ein großes Gewicht gibt, um durch die Problemeinsicht „bessere“ Lösungen zu begünstigen
- Regierungspolitik: Jede Regierungserklärung nach 1945 sprach sich für die Verwaltungsreform aus. Seit den 1960er Jahren des vergangenen Jahrhunderts folgen Leitsymbolen der Verwaltungsreform aufeinander: Verwaltungsvereinfachung (1960er), Demokratisierung der Verwaltung (1970er und 1980er) oder das *wirkungsorientierte* Verwaltungsmanagement in den 1990er Jahren, das an Schweizer Vorbilder³ angelehnt ist. Hierzulande spricht man von „leistungsorientierter Verwaltungsführung“ (Dearing 2003: 83-84)⁴. Um die Mitte der 1990er Jahre tritt das NPM-Leitbild seinen Siegeszug an (Dearing 2003; Pleschberger 2004; Hammerschmid/Meyer 2005a). Seit rund zehn Jahren operieren die Programme, Strategien oder Maßnahmen der Verwaltungsmodernisierung von Bund, Ländern und Gemeinden im NPM-Reformkorridor. Inhaltlich nicht neu, sondern mit „neoliberalen“ rhetorischen Akzenten angereichert, propagierten die von der ÖVP-FPÖ-Koalitionsregierung 2000 und 2003 vorgelegten Regierungsprogramme das Ziel, den „Verwaltungsstaat“ in den „Leistungsstaat“ zu transformieren, mit folgenden zentralen Adjektiven: „effizienter, schlanker, sparsamer und innovativer“ (2000) bzw. „noch bürger- und wirtschaftsfreundlicher und effizienter“ (2003). Die „Leistungsstaat“-Rhetorik der Jahre seit 2000 ist von der neuen „großen“ Koalitionsre-

² Der Beitritt Österreichs zur EU am 1.1.1995 rückte die Ausrichtung der nationalen Haushalte von Bund, Ländern und Gemeinden auf plafondierte europäische Fiskalziele in den Vordergrund, wobei allerdings bereits Art. 13 Abs. 2 B-GV normiert, dass die öffentlichen Haushalte an der „Sicherstellung des gesamtwirtschaftlichen Gleichgewichts“ auszurichten sind.

³ Dabei werden Unterschiede erkennbar, weil die Schweizer Variante einen Schwerpunkt auf die parlamentarische Steuerung der Leistungsproduktion legt, während in Österreich ein deutlicher Fokus auf der Leistungsberichterstattung liegt.

⁴ Die Palette der Leitbild-Konjunkturen wird national und international um die „Public Governance“ (Bauer et al. 2005) bereichert. Wie weit es hier um echte Innovationen geht (bzw. gehen soll), oder nur um das Re-Labeling von „alten“ Phänomenen (Bürgerbeteiligung, Zukunftswerkstätten etc.) mag hier dahingestellt bleiben.

gierung zwischen SPÖ und ÖVP (2007) mittlerweile rhetorisch abgemildert⁵ und der Schwerpunkt in Richtung Qualitätssicherung verlagert worden.

Die genannten Systemmerkmale können den Stellenwert von Evaluationen entweder eher herabsetzen (Informalität, Konsens) oder eher stärken (Fiskaldruck, NPM-Dominanz).

2 Evaluation und Verwaltungsmodernisierung in Bund, Ländern und Gemeinden

Die folgende empirische Analyse untersucht die österreichische Verwaltungsmodernisierung entlang von zwei einfachen evaluationsspezifischen Fragestellungen (vgl. Hansen 2005: 323-324).

- Sind Evaluationen im Rahmen der Verwaltungsmodernisierung institutionalisiert („evaluation in“), wobei zwischen Evaluationen im engeren Sinne (z. B. die Hereinnahme des Instrumentes der Evaluation in die Verwaltungsmodernisierung selbst) und im weiteren Sinne (typischerweise beträfe dies die Instrumente Controlling, Leistungsvereinbarung, Berichterstattung etc.) zu unterscheiden ist.
- Welche institutionellen Reformen wurden von „wem“ und „wie“ hinsichtlich ihrer Wirkungen gemessen und bewertet („evaluation of“).

Bei der Beantwortung der Fragen muß berücksichtigt werden, dass zur neueren österreichischen Verwaltungsmodernisierung nur einige deskriptive Überblicke (Duve 2006; Hammerschmid 2006; Promberger et al. 2004; Dearing 2003; Herbeck 2005) und empirisch-analytische Detailuntersuchungen (Hammerschmid/Meyer 2005a; 2005b; 2006) vorliegen. Die folgende empirische Skizze zum Stellenwert der Evaluation in der österreichischen Verwaltungsmodernisierung berücksichtigt – soweit vorhanden – Veröffentlichungen, diverse Verwaltungsstudien und Daten aus einer eigens für die Analyse durchgeführten *explorativen Umfrage*.⁶

⁵ Das Regierungsprogramm kündigt für Ende 2007 eine „Verwaltungs- und Staatsreform“ an, als deren Zielkriterien Effizienz, Kundenorientierung und Rechtssicherheit genannt werden.

⁶ In den Monaten Januar und Februar 2007 führte der Verfasser gleichlaufende E-Mail-Kontakte mit den für Verwaltungsreformmaßnahmen verantwortlichen Verwaltungsmitarbeitern in österreichischen Bundesländern (mit Ausnahme Wien) und den Landeshauptstädten (mit Ausnahme Wien), in denen die Angeschriebenen ersucht wurden den Stand der Verwaltungsmodernisierung darzustellen bzw. Quellenmaterial zu übersenden. Weitergehend wurde nach konkreten Evaluationen gefragt. Von den angeschriebenen Ländern antworteten sechs Länder (Burgenland, Kärnten, Salzburg, Steiermark, Tirol, Vorarlberg und von den angeschriebenen Städten lediglich zwei (Bregenz, Linz). Der Informationsgehalt der Auskünfte ist sehr unterschiedlich. Die Informationen fließen in anonymisierter Weise in die Analyse ein.

2.1 Bundesebene

Einen „symptomatischen“ Blick auf die österreichische Evaluationssituation erlauben Ausführungen einer federführenden Akteurin der Verwaltungsreform auf der Bundesebene. Dearing betont den möglichen Reformnutzen einer Evaluation für die Qualitätssicherung großer Verwaltungsreformvorhaben (Dearing 2003: 100). Bereits das *Verwaltungsmanagementprojekt* von 1988-1992 (Obermair 1999) sah eine „Projektevaluation“ vor (Dearing 1992: 304). Beim folgenden *VerwaltungsInnovationsProgramm* (Obermair 1999) zu Ende der 1990er Jahre hätte es „Ergebnisberichte mit versuchten Evaluationen“ (Dearing 2003: 100) gegeben⁷, was auf die einfache Darstellung des zeitlichen und finanziellen Umsetzungsstandes der beschlossenen Reformmaßnahmen hinauslief (vgl. Rossmann 2002: 209)⁸. Mit dem Aufschwung der Verwaltungsmodernisierung zum zentralen regierungspolitischen Reformprojekt mit Beginn des Jahres 2000 nehmen die Bekenntnisse zur Evaluation zu: sei es der „gesamten Reformvorhaben“ (Regierungsprogramm 2000), der „Flexibilisierungsklausel“ (Dearing 2003: 100), der seit Anfang der 1990er Jahre durchgeführten „Ausgliederungen“⁹, der „Aufgabenreform“ von 2000¹⁰ (Riess-Passer 2002) oder der *VerwaltungsInnovationsProgramme* der Jahre seit 2000 („VIP I“, „VIP II“ und VIP III¹¹) des Bundes (Bundeskanzleramt 2006: 7).

Etwas pointiert ausgedrückt kennt die österreichische Verwaltungsmodernisierung die „vorgesehene“, „geplante“ oder „angekündigte“ Evaluation, die das Handeln der österreichischen Bundesverwaltung mit der erwähnten „accounting aura“ (Greve 2003: 276) umgibt. Im Folgenden wird in exemplarischer Weise auf praktische Reformprojekte mit evaluativen Ansätzen eingegangen, um die Umsetzung der „Evaluationsrhetorik“ zu beschreiben.

2.1.1 Leistungskennzahlen in der Bundesverwaltung

Die Modernisierung des Rechnungswesens der Bundesverwaltung war Anfang des laufenden Jahrzehnts in der Sicht eines prominenten Verwaltungsreformpolitikers das zentrale „Stück Verwaltungsreform“ (Finz 2002: 10) schlechthin, nach Beratersicht das „Kernstück

⁷ Die Autorin verweist auf einen Internet-Link, der nicht (mehr) aufrufbar ist. Interessant ist auch, dass die Internetpräsentation des Bundeskanzleramtes zur Verwaltungsreform das Stichwort „Evaluation“ nicht anführt.

⁸ Die zitierte Quelle zeigt mit Stand April 2002 eine Übersicht auf Basis eines Ministerratsvortrages vom 22. April 2002 und führt aus, dass damals 47 Prozent der Maßnahmen mit etwa 26 Prozent der geplanten Einsparungen „umgesetzt“ seien.

⁹ Laut Regierungsprogramm 2003: 35, Punkt 20 „Verwaltungsreform“; Dearing (2003: 99) schreibt, dass das österreichische Finanzministerium die „wichtigsten Ausgliederungen“ evaluiert „hat“ (!), was durch intensive Recherchen im österreichischen Finanzministerium (Februar 2007) nicht bestätigt werden konnte.

¹⁰ An die damalige Kommission wurde die Erwartung gestellt, einen „entsprechenden Evaluierungsprozess“ der beabsichtigten „Aufgabenreform“ durchzuführen, was schon mangels Aufgabenreform im Sinne der Kommission unterbleiben musste.

¹¹ Im Rahmen des Projektmanagements der VerwaltungsInnovationsProgramme (z. B. „VIP III“) ist von „Projektevaluation“ die Rede, was im Wesentlichen in der Praxis nicht mehr bedeutet, als dass über eine Datenbank die Entwicklung der von den Ressorts eingegebenen Reformprojekte laufend aktualisiert (und von den Beteiligten) eingesehen werden kann (vgl. Vortrag an den österreichischen Ministerrat vom 8.7.2003).

der Verwaltungsinnovation“,¹² das die finanziellen Ströme in Verbindung mit den „echten“ Leistungen¹³ der Verwaltung abbilden sollte. Nach einer fast zehnjährigen Reformgeschichte¹⁴ konnte der „Produktivbetrieb“ einer einheitlichen Kosten- und Leistungsrechnung in der Bundesverwaltung 2005/2006 aufgenommen werden. Es werden drei Leistungsarten unterschieden: interne Leistungen¹⁵, externe Standardleistungen¹⁶ sowie andere externe Leistungen¹⁷, die mit Leistungskennzahlen operationalisiert werden. Den Leistungen werden anteilmäßig geschätzte Personalkosten und „Overheadkosten“ zugeschlagen. Die Leistungskennzahlen der Ressorts sind in einer Datenbank „hinterlegt“, die verwaltungsintern mit Hilfe einer Recherchefunktion online nach neun Leistungskriterien verglichen werden können, was die Voraussetzung für eine Art „Selbstevaluation“ seitens interessierter Verwaltungsmitarbeiter schafft. Wirkungsseitig erwartet man sich u. a. drei Effekte: Führen und Entscheiden unter Berücksichtigung „wahrer“ Kosten, ein „benchmarking zwischen den Bundesministerien“ (Finz 2002: 9), Steigerung des Kostenbewußtseins über die finanziellen Folgen der Verwaltungsmaßnahmen und der eigenen Verwaltungskosten.

Nach vorliegenden Informationen wurde weder eine summative oder formative Evaluation des „Produktivbetriebes“ durchgeführt. Eine zurückliegende Prüfung des österreichischen Rechnungshofes (Rechnungshof 2005) erlaubt einen selektiven, ressortbezogenen Einblick in „historische“ Umsetzungsaspekte des Reformprojektes. Dem Bericht zufolge ermittelte das für die Verwaltungsreform federführende Bundeskanzleramt selbst erst seit 2002 und damals noch in wenigen Aufgabenfeldern einfache „Leistungskennzahlen“ (wie z. B. die Nutzungshäufigkeiten des von ihm verwalteten *Rechtsinformationssystem*s), qualitative oder wirksamkeitsbezogene Leistungskennzahlen fehlten. In Reaktion auf die kritischen Hinweise des Prüforgans äußerte das Ressort damals die Absicht, „auf Basis der bundesweiten Kosten- und Leistungsrechnung Effizienzzahlen sowie für jede Abteilung repräsentative Leistungskennzahlen“ erstellen zu wollen und darauf aufbauend mit jeder Abteilung „Leistungsvereinbarungen“ (2005) zu schließen und ein Leistungscontrolling (2006) einzuführen, das die Steuerung der Verwaltungsleistungen auf Grundlage der Kostenrechnung erlauben würde (Rechnungshof 2005: 13). Nach den aktuellen Informationen¹⁸ sei die Kosten- und Leistungsrechnung „in allen Ressorts umgesetzt“. Tatsächlich hat jedes Ressort ein „Handbuch“ erarbeitet (teilweise mit Hilfe von Beratungsunternehmen), in dem es „seine“ Informationen und instrumentellen Ansätze zusammenstellt. Die ganz unterschiedlichen Zugänge und Leistungsstandards müssen aktuell noch standardisiert werden, was unter entschei-

¹² So der Projektbericht Public Management der Beratungsfirma ImPlus, die das hier angesprochene Projekt betreute (2001).

¹³ Die Ressorts erheben seit langer Zeit viele und detaillierte Kennzahlen, die „fast ausschließlich“ der Inputmessung dienen (Personalkosten pro Ressort, EDV-Arbeitsplatz-Kosten etc.), leistungsorientierte Kennzahlen waren nicht verbreitet (Dearing 2003).

¹⁴ Die Stationen vom ersten „Absichts“-Beschuß des Ministerrates (Juli 1996), weiteren „Absichts-Beschlüssen“ im Oktober und November 2000 und der Verordnung des Bundesministers für Finanzen vom Dezember 2004 (BGBl. II Nr. 526/2004) sind zusammengefasst in einem „Wahrnehmungsbericht“ des österreichischen Rechnungshofes (Rechnungshof 2005) nachzulesen.

¹⁵ Wie z. B. Budgetplanung, Dienstreisemanagement.

¹⁶ Hier sind z. B. Bürgerservice, Öffentlichkeitsarbeit gemeint.

¹⁷ Sie können zahlreicher Natur sein, die für Bürger, Unternehmen etc. erbracht werden.

¹⁸ E-Mail von Frau Mag. Angerer, Staatssekretariat für Regionalpolitik und Verwaltungsreform, vom 31.8.2007 an den Verfasser.

dender Federführung des Finanzministeriums durchgeführt wird, das ein einheitliches Bundescontrolling anstrebt. Die Kosten- und Leistungsrechnung ist bis heute nicht als Managementinstrument einsetzbar¹⁹ und das Budget ist (noch) das zentrale Steuerungsinstrument des Verwaltungshandelns. Die „echte“ Anwendung der Kosten- und Leistungsrechnung als Führungsinstrument soll mit dem Inkrafttreten der neuen Haushaltsordnung des Bundes im Jahre 2009 möglich werden,²⁰ unter der Voraussetzung der entsprechenden politischen und rechtlichen Festlegungen im Detail.

2.1.2 Flexibilisierungsklausel

Die sog. Flexibilisierungsklausel („Flexi-Klausel“) ist ein „Pilotprojekt“ zur finanziellen Flexibilisierung der Verwaltungsführung (Promberger et al. 2005; Duve 2006: 420-421; Dearing 2003: 93-94). Die „Flexi-Einheiten“ wenden zahlreiche NPM-Instrumente an. Als „NPM-Verwaltung“ können sie Mittel umschichten, erwirtschaftete Überschüsse teilweise einbehalten und für Investitionen oder zum Ausgleich budgetärer Fehlbeträge im Folgejahr verwenden. Im Jahre 2006 nahmen 14 nachgeordnete Einheiten von sieben Ministerien freiwillig am Pilotprojekt teil.

Aufgaben der internen Evaluation erfüllt ein „Controlling-Beirat“, der als „verlängerter“ Arm des „Heimministeriums“ die laufende Einhaltung der mit den „Flexi-Einheiten“ vereinbarten „harten“ Fiskal- und Personalziele beobachtet und „evaluiert“.²¹

Eine seit kurzem vorliegende und publizierte externe Evaluation des projektbegleitenden Beraters beschreibt das Projekt – „die Speerspitze“ (Promberger et al. 2005: 14) der österreichischen Verwaltungsreform – erwartungsgemäß pauschalierend positiv, wie auch Verwaltungsakteure (Dearing 2003: 94; Herbeck 2005: 251) und an ihnen anschließend externe Autoren (Duve 2006: 421-422). Das Reformprojekt habe eine „Aufbruchstimmung“ in den beteiligten Verwaltungseinheiten erzeugt (Promberger et al. 2005: 14; 32), während der Rechnungshof die Schwierigkeiten der Operationalisierung und Messung der strategischen und fachlichen Managementziele anführt (vgl. Rechnungshof 2005: 4).

2.1.3 Leistungsberichte

Das Mitte der 1990er Jahre gestartete Reformprojekt „Leistungsbericht der österreichischen Bundesverwaltung“ fügt sich in den internationalen Trend des „performance information“ (Pollitt 2006) ein und ist ein Beispiel für eine interne, jedoch öffentlich zugängliche Selbstevaluation. Seit Mitte der 1990er Jahre erscheinen Leistungsberichte in unregelmäßiger Ab-

¹⁹ Auskunft von Herrn Ministerialrat Ihle (Finanzministerium) in einem Telefonat mit dem Verfasser am 7.9.2007.

²⁰ Mit diesem Zeitpunkt sollen Vierteljahresberichte über die Kostenentwicklung, Vergleiche der internen und externen Leistungen möglich werden, was evaluative *Potenziale* eröffnet.

²¹ Er besteht aus einem Vertreter des Finanzministeriums, des „Hausministeriums“ sowie aus einem externen Experten der Betriebswirtschaft.

folge, und der letzte Leistungsbericht datiert mit dem Jahr 2004²², ohne dass dieser Sachverhalt in die öffentliche Kritik geraten wäre.

In der Sicht der das Projekt begleitenden „Proponentin“ soll der Leistungsbericht „wirksame Ergebnisse“ der Verwaltungstätigkeiten darstellen und einen „Impuls“ in Richtung leistungsorientierter Steuerung der Verwaltung geben (Dearing 2003: 90; auch Promberger et al. 2004: 36-38). Die Rede ist von „steuerungsrelevanten Daten“ in Form von „allgemeinen Leistungskennzahlen“, die für die „BürgerInnen relevant sind“ (Dearing 2003: 89-90) und der Öffentlichkeit einen „guten Überblick über die Vielfalt und Größenordnung öffentlicher Leistungen geben“²³ sollen. Die Publizierung sogenannter „interner Leistungskennzahlen“ (z. B. für Personal, EDV-Betreuungskosten etc.) ist nicht vorgesehen.

Eine externe Evaluation (Rechnungshof 2005) stellt am Beispiel des Berichtes des Jahres 2003 zahlreiche instrumentelle Mängel heraus. Jedes Ressort berichtet demnach über drei bis fünf „Kernaufgaben“, die aber hinsichtlich ihres Umfangs und ihrer funktionellen oder finanziellen Bedeutung enorm variieren würden. Man berichtet „Unvergleichbares“, was unter anderem darauf zurückzuführen ist, dass die Ressorts ihre Leistungsbereiche selbst auswählen konnten (wie sie auch die Kennziffern und das Datenmonitoring „autonom“ besorgten). Die berichteten finanziellen Inputdaten stünden ohne Bezug zu den berichteten Kennzahlen. Hinzuzufügen ist, dass Sollzielwerte für die Leistungserfüllung nur punktuell angegeben werden, was die Gesamtbewertung der Performanz nicht erlaubt. Das Beispiel macht deutlich, dass die Autonomie von Ressorts bei der Berichterstattung über ihre Leistungen eher kontraproduktiv sein dürfte, andererseits gäbe es ohne Autonomie der Ressorts vermutlich keinen Leistungsbericht.

Erwähnenswert ist, dass die im Leistungsbericht berichteten Leistungskennzahlen sich nicht mit den Kennzahlen im Rahmen der „Kosten- und Leistungsrechnung“ (siehe oben) decken.

Die Leistungsberichte wurden in der Vergangenheit der Regierung (Ministerrat) vorgelegt, die sie routinemäßig „zur Kenntnis“ nahm, sie nicht näher „evaluierte“ und auch nicht für ihre Öffentlichkeitsarbeit verwendete. Eine Zuleitung an das Parlament unterblieb.

2.2 Länder

Die Verwaltungsmodernisierung in den österreichischen Bundesländern ist ebenfalls NPM-geleitet, wobei der erklärte Schwerpunkt vieler Projekte auf der Wirkungsorientierung liegt. Während zu Anfang des Jahrzehnts Absichtserklärungen, konzeptionelle Strategiedokumente und erste (unabgeschlossene) Umsetzungsschritte mit sogenannter „Reformrelevanz“ (Rechnungshof 2002: 8) überwogen, ist seitdem der praktische Reifegrad der Projekte gestiegen.

²² Das Fehlen weiterer Leistungsberichte seit 2004 wird u. a. mit der Absicht begründet, die Haushaltsplanung auf dreijährige Perioden auszudehnen und die Haushaltsplanung stärker an Leistungskennzahlen auszurichten.

²³ Behelf Verwaltungsreform zum Budgetvoranschlag 2006.

Strategische Reformdokumente erwähnen „die“ Evaluierung explizit selten²⁴. Wenn angeführt, dann sollen die Reformmaßnahmen evaluiert werden, freilich „intern“. Festzuhalten ist weiter eine tendenzielle Evaluationsskepsis, derzufolge eine Evaluation unter Umständen bei beträchtlichen Kosten „ohne produktives Ergebnis“ bleiben könnte, was den eingeleiteten Reformprozess zum Stillstand bringen würde.

Bei den evaluationspraktischen Entwicklungen ist keine durchgängige Linie in den Ländern erkennbar. Zu erwähnen ist die an wenigen Beispielen getestete Gesetzesfolgenabschätzung für gesetzliche Maßnahmen der Länder (Schäffer 2005b).

Ende der 1990er Jahre wurde das Pilotprojekt „Budget-Center“, das die Erfahrungen mit der dezentralen Budgetbewirtschaftung der nach Umfang geringen Sachausgaben in den regionalen staatlichen Behörden „testen“ sollte, „positiv“ evaluiert (Platzer 2004: 33), was zu seiner Einführung als Regelpraxis bei diesen Behörden führte. Anfang 2000 wurden im Bundesland Kärnten landesintern 39 Produktgruppen der regionalen staatlichen Bezirksverwaltungsbehörden miteinander verglichen und hohe Produktivitätsunterschiede festgestellt (Platzer 2004: 37).

Im Jahre 2003 begannen drei österreichische Bezirkshauptmannschaften mit externer Beraterunterstützung das Projekt „Voneinander – miteinander lernen“, das ihre Standard-Leistungen, „Produktionsprozesse“ und Maßnahmen vergleichend untersuchte und aufbauend Vorschläge zur nachhaltigen Verbesserung ihres kundenbezogenen Verwaltungshandelns entwickelte (Biwald 2005: 45). Zu letzterem zählen unter anderem die Einführung einer neuen Kennzahl „informelles Verwaltungshandeln“ oder die Vereinbarung, daß Beschwerden von Verwaltungskunden innerhalb von zwei Wochen beantwortet werden müssen.²⁵ Entgegen der verbreiteten Praxis der „Verschlusshaltung“ der Vergleichsergebnisse seien die Ergebnisse dem projektbegleitenden Berater zufolge „einer breiten Öffentlichkeit“ zugänglich gemacht worden, um „Transparenz über maßgebliche Qualitätsumstände“ darzustellen (Biwald 2005: 45). Das Projekt wurde 2005 mit dem österreichischen Verwaltungspreis 2005 gewürdigt.²⁶

Das weithin interessanteste Beispiel ist die Evaluation des „New Public Management Pilotprojekt“ in Vorarlberg.²⁷ Es integriert auf Grundlage eines Beschlusses der dortigen Landesregierung (2002) die in der Verwaltung isoliert praktizierten NPM-Kerninstrumente Leistungsvereinbarungen, Berichtswesen und Wirkungsanalyse-Programm („WaP“²⁸) in einem „Dachprogramm“. Ebenfalls mit politischem Beschluß wurde die Evaluation des „WaP“ „durch einen unabhängigen, nicht am Pilotversuch beteiligten Experten“ nach dem Ende seiner Testphase (2005-2006) empfohlen (vgl. Promberger/Zimmermann 2007: 9). Die externe Evaluation wurde bereits während der Testphase von einem universitätsnahen Beratungsinstitut durchgeführt und Anfang 2007 in einem publizierten Bericht dokumen-

²⁴ Zu erwähnen sind hier TIVES – Tiroler-Verwaltungs-Entwicklungs-Strategie (2005) und das Dokument Entwicklungsstrategien in der Vorarlberger Landesverwaltung (2003).

²⁵ Näheres in der Aussendung der Salzburger Landeskorespondenz vom 14.12.2005.

²⁶ Nach Salzburger Landeskorespondenz vom 14.12.2005.

²⁷ Das Projekt wurde unter Beizug bekannter externer NPM-Experten der Universität St. Gallen entwickelt.

²⁸ Im Zentrum steht das Instrument der Balanced-Scorecard, das eingesetzt wird, um interne und externe Wirkungsdimensionen des Verwaltungshandelns vorzugeben, die von den Mitarbeitern selbst auf einer Skala von 1 bis 5 bewertet wurden. Es wurden auch Nachhaltigkeitsdimensionen einbezogen (das sog. „WaP-Plus“) (Promberger/Zimmermann 2007).

tiert (ebd.). Der Evaluation zufolge ist das „WaP“ ein sinnvolles praktikables Instrument, um binnenorganisatorisch die Wirkungen geplanter und laufender „Verwaltungsleistungen und -projekte“ (ebd.) abzuschätzen. Doch vermeidet der Bericht Aussagen über die materiellen Wirkungen des WaP-gesteuerten Verwaltungshandelns (Promberger/Zimmermann 2007: 117). Unter anderem wurden 13 mehrheitlich leitende Verwaltungsmitarbeiter in sechs Einheiten der hoheitlichen und dienstleistenden Landesverwaltung, die freiwillig mitwirken, mit einem Gesprächsleitfaden interviewt,²⁹ um deren Erfahrungen und Meinungen zum Instrument zu ermitteln, was Informationen mit Evaluationsrelevanz sind. Die Ergebnisse der Befragung sind im Projektbericht allerdings nicht dokumentiert. Statt genauer „Empirie“ finden sich zahlreiche konzeptionelle und praxeologische Überlegungen, die das funktionelle Reformpotential des Instrumentes präskriptiv und positiv beschreiben.

Im Rahmen einer *internen* formativen *Evaluation* wurde die Kohärenz des positiv evaluierten WaP-Instrumentes mit den NPM-Instrumenten Leistungsvereinbarung und Berichtswesen geprüft.³⁰

2.3 Kommunen

Der österreichische Kommunalsektor ist seit Mitte der 90iger Jahre im ersten Blick in breiter Front auf den NPM-inspirierten Pfad der Verwaltungsmodernisierung eingeschwenkt (Pleschberger 2004; Promberger et al. 2004: 100-116). Manche strukturelle NPM-Maßnahmen (z. B. das Outsourcing kommunaler Aufgabenwahrnehmung) sind im Kommunalsektor seit langer Zeit Praxis. Seit etwa zehn Jahren sind Maßnahmen zur Verbesserung der Kundenorientierung der Kommunalverwaltung verbreitet. Daneben findet sich ein breites experimentelles Maßnahmenspektrum,³¹ unter anderem mit „Leistungsstandards“ (bereits KDZ 2001), Balanced Scorecards, Kontraktmanagement oder Leistungsvergleichen (Beispiele in Bauer et al. 2003).

Maßnahmen der NPM-konformen Reform des Haushalts- und Rechnungswesens, welche die Voraussetzung für die Anwendung weiterer Steuerungsinstrumente sind, sind in den vielen kleinen österreichischen Gemeinden (Matschek 2004: 177) und selbst in den größeren österreichischen Städten bisher nicht flächig verbreitet.

Insgesamt ist es erstaunlich, dass in den (insgesamt wenigen) strategischen Reformdokumenten auf die Evaluation so gut wie nicht Bezug genommen wird. Praktische Evaluationen gibt es nur bei Anlegung eines sehr weiten Verständnisses, das die Instrumente Controlling und Bürgerbefragungen als Evaluationen betrachtet.

Ein prominenter Ansatz einer Evaluation im engeren Sinne sind die für die Kommunen typischen interkommunalen Leistungsvergleiche (Pleschberger 2004). Seit 1995 haben (optimistisch geschätzt) allenfalls zwei bis drei Prozent aller österreichischen Kommunen Erfahrungen mit dem Instrument gesammelt, die bisher wissenschaftlich erst annäherungs-

²⁹ Das „WaP“ wurde 2006 in acht Organisationseinheiten für 22 Leistungen durchgeführt, ab 2007 sollen weitere Anwender dazu kommen (Promberger/Zimmermann 2007: 120).

³⁰ Eine Beschreibung der Ergebnisse soll in einer Broschüre im September 2007 erscheinen.

³¹ Einen guten Einblick in diese Richtung bietet die Tagung des Österreichischen Städtebundes am 28.9.2004 in Wien (unter kdz.or.at).

weise (Pleschberger 2004: 307-317; rein deskriptiv Promberger et al. 2004: 105-107) beschrieben wurden und gelegentlich in der Beraterperspektive beziehungsweise in der Sicht der Verantwortlichen (Bindeus et al. 2005) berichtet worden sind. Würde man an die vorliegenden Projekte übliche methodische Standards (z. B. hinsichtlich empirischer Dokumentierung, „Prozessgenauigkeit“, externe Begleitung) anlegen, müsste wahrscheinlich die Mehrheit der Beispiele nicht als Leistungsvergleiche qualifiziert werden.

Die diversen Leistungsvergleiche sind unter Datengesichtspunkten intransparent. Die Promotoren eines aktuellen interkommunalen Vergleichsprojektes betrachten es sogar als einen „Erfolgsfaktor“ derartiger Projekte, dass die Kennzahlen und die Ergebnisse des Vergleichs im Kreise der Beteiligten verbleiben, um die interne Vertrauensbasis nicht zu gefährden (Bindeus et al. 2005: 37).

3 Ansätze für ein „government“ mit Evaluation

Der deutsche Verwaltungsforscher Christoph Reichard kommt in einem Dreiländervergleich der Verwaltungsreformen in Österreich, Deutschland und der Schweiz (abgeschwächt) zum Schluß, dass es „weitgehend an neutralen und seriösen Evaluierungen des tatsächlich erreichten Reformstandes und der damit erzielten Ergebnisse und Wirkungen“ fehle (Reichard 2006a: 48). Das vergleichend konstatierte „Evaluationsdefizit“ (Reichard 2006a: 50; auch Proeller 2006: 35, 36) trifft für Österreich *cum grano salis* zu (auch Promberger et al. 2006: 178; Hammerschmid 2006). Es zeigt sich eine große Schere zwischen der – vor allem – rhetorischen und konzeptionellen NPM-Relevanz der Evaluation und ihrer praktischen Anwendung (wobei die Wirkungsdimension ohnehin separat zu betrachten wäre). Das vielleicht skeptisch klingende österreichische „Evaluationsbild“ kann näher „ausdifferenziert“ werden.

Verwaltungsinsidern zufolge sei unter den österreichischen Politikern die „Bereitschaft für eine echte Evaluation von Reformen noch nicht stark ausgeprägt, einerseits wegen der kurzen Legislaturperioden und andererseits wegen fehlender Einsicht in den Nutzen solcher Methoden in der Verwaltung“ (Dearing 2003: 100). Man setzt zumindest im Bereich der institutionellen Modernisierung lieber qualitative und quantitative Ziele als die Zielerreichung selbst zu evaluieren (Naschold 1995: 17). Andererseits verzichtet man nicht auf das Reformsymbol der Evaluation. Sie kommt in den strategischen Reformdokumenten und weiteren relevanten Darstellungen – zumindest gelegentlich – vor, meistens in Verbindung mit dem Controlling und Bürgerbefragungen. Während das Controlling ebenübergreifend angewendet wird, sind Bürgerbefragungen vor allem ein kommunales Anliegen.

Wegen ihrer Einbettung in das Controlling sind die Evaluationen steuerungspolitisch „präformiert“ und als Instrumente der Verwaltungsführung konzipiert. Die sehr interessante externe *formative Evaluation* („WaP“) gibt in beispielhafter Weise Aufschlüsse über die instrumentelle Funktionalität eines NPM-Instruments, will jedoch über seine materiellen Wirkungen in der Umwelt keine Aussagen treffen. Eine vergleichbare Logik zeigen die als *summative externe Evaluationen* anzusehenden „Wahrnehmungsberichte“ der Rechnungshöfe von Bund und Ländern über die Umsetzung von Reformmaßnahmen. Die österreichischen Rechnungshöfe sind ein unabhängiges Organ der Gesetzgebung, das aus eigenem Antrieb

oder auf Ersuchen einer gesetzgebenden Körperschaft unter Heranziehung festgelegter Prüfmaßstäbe³² organisatorische Abläufe in der öffentlichen Verwaltung (und ihre finanziellen Effekte) prüft. Seit rund einem Jahrzehnt richten die österreichischen Rechnungshöfe ihre Prüftätigkeit (einschließlich der Empfehlungen) immer stärker am NPM-Paradigma aus³³, was „Märkte“ für das New Public Management in Österreich vorbereitet (Platzer 2004: 39-40).

Die von den Rechnungshöfen untersuchten Wirkungsdimensionen sind in deutlicher Weise (weil gesetzlich gefordert) auf die organisatorisch-technische Binnenperspektive der Verwaltung verengt und lassen „externe“ materielle Wirkungsdimensionen außer Acht.

Zusammengefaßt sind die (wenigen) externen Evaluationen – und „selbstverständlich“ die internen – auf die Steuerungsinteressen der Verwaltungsführung zugeschnitten. Sie ergänzen und verstärken die prinzipielle Steuerungsperspektive der NPM-Instrumente (z. B. Controlling, Leistungsvereinbarungen, Kundenbefragungen). Von Interesse ist auch, daß die methodische Seite, die Ergebnisse und die Verwendung von Evaluationen teilweise für „Dritte“ nicht zugänglich sind (mit der Ausnahme der Berichte des Österreichischen Rechnungshofes), wofür kommunale Benchmarking-Projekte gute Belege abgeben.

Übergreifend betrachtet ist die geringe methodische Qualität der Evaluationen zu erwähnen, was auch damit zusammenhängt, daß private kommerzielle Berater die Dienstleistung anbieten und die Rechnungshöfe ihrerseits auf methodische Standards einen geringen Wert legen.

³² Anzulegen sind die ziffernmäßige Richtigkeit, die Rechtmäßigkeit, die Sparsamkeit, die Wirtschaftlichkeit und die Zweckmäßigkeit (insgesamt fünftes Hauptstück B-VG).

³³ Die ausgesprochenen Empfehlungen legen den öffentlichen Verwaltungen die Einführung des NPM-Vollmodells nahe (z. B. Rechnungshof 2002).

Evaluation institutioneller Politik in der Schweiz, Deutschland und Österreich im Vergleich

1 Inhalt und Evaluation institutioneller Politik im Dreiländervergleich

Die drei Länderstudien befassen sich mit institutioneller Politik. Diese versucht über die Veränderung des organisatorischen Institutionengefüges einer Körperschaft die Machtverteilung zwischen Akteuren und Prozessen zu ändern (vgl. Beitrag von Sager in diesem Band).

Institutionelle Politik befasst sich einerseits mit **politischen Institutionen** (wie Föderalismus, halbdirekte Demokratie oder Konkordanzregierung). Dieser Aspekt wird in den Länderberichten allerdings nur im Bericht über die Schweiz behandelt. Sager (in diesem Band) zeigt auf, dass entsprechende Studien für ihre Vergleiche vor allem das „föderalistische Labor“ mit 26 sehr unterschiedlichen Kantonen und mit mehr als 2700 ebenfalls in Bevölkerungsgrösse und Territorium äusserst unterschiedlichen Gemeinden nutzen. Solche Studien beinhalten häufig eine Reflexion darüber, ob und inwieweit die besonderen schweizerischen politischen Institutionen (noch) zeitgemäss sind und inwieweit sie reformiert werden sollten. Allerdings haben die politischen Institutionen auch ein grundlegendes Beharrungsvermögen, womit solche Evaluationen eher den Charakter von Grundlagenforschung haben.

Andererseits umfasst institutionelle Politik die **Institutionenpolitik**, das heisst, sie ist stärker auf der Meso-Ebene von Verwaltungsstrukturen und Organisationsmerkmalen von Körperschaften angesiedelt. Als einzige Studie behandelt diejenige über die Schweiz Reformen der Regierung, des Parlaments, des Territoriums (Gemeindefusionen), des Finanzausgleichs und der Justiz (siehe auch Widmer 2002: 106; Widmer/Rieder 2003: 204) und deren Evaluation. Die beiden anderen Länderberichte sind fokussiert auf **Verwaltungsreformen** und deren Evaluierung, weshalb wir uns in der Folge auf diese Form der Institutionenpolitik **beschränken** werden.

Die drei Länderberichte (Sager, Sabine Kuhlmann sowie Pleschberger in diesem Band) zeigen, dass zahlreiche Verwaltungsmodernisierungen durch das New Public Management (NPM) inspiriert sind. Im Vergleich der drei Staaten und ihrer drei Staatsebenen sind folgende Besonderheiten zu verzeichnen:

Tabelle 1: Verwaltungsreformen und deren Evaluation

| Reformen | Schweiz | Deutschland | Österreich |
|---------------------------------|---|---|--|
| Bundesebene | | | |
| Inhalt | Reformen vor allem in den sog. „FLAG-Ämtern“ ¹ | zahlreich Reformprojekte, hohe Ansprüche, geringe faktische Reformneigung | verschiedene, stark auf Leistungskennzahlen und -berichte ausgerichtete Reformen |
| Evaluation | externe Evaluation der FLAG-Reform | Evaluation des Umbaus der Bundesagentur für Arbeit, viele interne Berichte, Aktivitäten des Bundesrechnungshofs | Evaluationen wurden lediglich angekündigt, kritische Bilanzen des Österreichischen Rechnungshofs |
| Ebene der Länder/Kantone | | | |
| Inhalt | Reformen in rund der Hälfte der Kantone | weit verbreitete Verwaltungsreformen (Umfang nicht quantifiziert) | |
| Evaluation | zahlreiche externe und interne Evaluationen | interne Evaluationen sowie einzelne länderübergreifende Evaluationen, Berichte einzelner Landesrechnungshöfe | Evaluationsskepsis, bloss vereinzelte echte Evaluationen (Pilotprojekte) |
| Ebene der Kommunen | | | |
| Inhalt | NPM-orientierte Verwaltungsreformen in rund 5 % der Gemeinden | breite Verwaltungsmodernisierung, verschiedene Leistungsvergleiche | |
| Evaluation | kaum Evaluationen | viele interne und externe Evaluationen, neuerdings umfassende Performanzevaluation | kaum Evaluationen |

Aus dem Vergleich schälen sich folgende Gemeinsamkeiten und Unterschiede von Verwaltungsmodernisierungen und ihrer Evaluation heraus:

- Tendenziell sind auf den unteren staatlichen Ebenen mehr Reformaktivitäten zu verzeichnen als auf der obersten Ebene. Auffallend ist indessen das weitgehende Fehlen substanzieller Verwaltungsreformen auf kommunaler Ebene in der Schweiz.
- Österreich zeichnet sich insgesamt durch ein tieferes Niveau an Evaluationsaktivitäten aus als Deutschland und insbesondere die Schweiz.
- In allen drei Ländern dominieren formative sowie auf interne Prozesse fokussierte Evaluationen. Mit Ausnahme einer vergleichenden deutschen Länder- sowie einer Kommunalstudie wurde in den drei Studien über keine Evaluationen der externen Wirkungen von Verwaltungsreformen berichtet.

¹ FLAG = Führen mit Leistungsauftrag und Globalbudget.

2 Der Einfluss unterschiedlicher Föderalismusmodelle auf Verwaltungsreformen und deren Evaluation

2.1 Dezentraler und unitarischer Föderalismus

Im Vergleich der drei Länder fallen vor allem die Unterschiede bei den Reformaktivitäten auf kommunaler Ebene zwischen der Schweiz (wenig, vereinzelt) und Deutschland/Österreich (flächendeckend) auf. Wie lassen sich diese Unterschiede erklären?

Zwar sind alle drei Länder föderalistisch aufgebaut, doch folgen Deutschland und Österreich einem unitarischen und die Schweiz einem dezentralen Muster. In Deutschland haben die Länder vergleichsweise mehr Macht bei der Politikgestaltung auf Bundesebene als in der Schweiz die Kantone (siehe Braun 2003). Letztere haben jedoch, auch dank des Steuerföderalismus, vor allem bei der Politikumsetzung mehr Spielräume als die deutschen (und österreichischen) Länder, welche diesbezüglich vor allem als administrative Einheiten wahrgenommen werden (Braun 2003: 73, 85). In Deutschland und Österreich sind die Bundesländer und (für Österreich nur bedingt) die Kommunen² historisch sowie auf Grund von Territorialreformen homogener als die Gebietseinheiten in der Schweiz. Letztere weisen stärkere Disparitäten (in Bevölkerungsgrösse, Gebietsgrösse sowie politischer und administrativer Organisation) auf (vgl. Watts 1999: 64). Dies wirkt sich auch auf die Steuerung und Verbreitung von Verwaltungsmodernisierung und deren Evaluation aus.

2.2 Steuerung der Verwaltungsmodernisierung und Evaluation auf kommunaler Ebene

In **Deutschland** sind Verwaltungsreformen für die Kommunen von einer gemeinsamen Institution entwickelt worden. In Deutschland tonangebend war und ist die von Städten, Gemeinden und Kreisen gemeinsam getragene Kommunale Gemeinschaftsstelle für Verwaltungsmanagement (KGSt). Das von ihr entwickelte sogenannte Neue Steuerungsmodell (NSM) hat flächendeckend eine Anziehungskraft als Reformleitbild entfaltet und ist von den Kommunen ganz oder teilweise übernommen worden. Die KGSt sowie der Städtetag und Landkreistag haben auch Erhebungen zum Umsetzungsstand in Auftrag gegeben. Durch die Förderungsaktivitäten der Hans-Böckler-Stiftung sind zahlreiche Evaluationen der Verwaltungsmodernisierung der deutschen Kommunen durchgeführt worden, zuletzt auch der eigentlichen Wirkungen (vgl. Beitrag von Sabine Kuhlmann in diesem Band, Abschnitt 3.3.).

In **Österreich** und der **Schweiz** wird hingegen die Verwaltungsmodernisierung auf kommunaler Ebene dezentral gesteuert. Es gibt nicht einmal ein „best practice“-Modell. Dies wäre angesichts der Heterogenität der Kommunen auch gar nicht realitätsgerecht. Der Rückstand der Schweizer Kommunen bei Verwaltungsreformen und ihrer Evaluierung gegenüber Deutschland dürfte mit der sehr starken Fragmentierung und Kleinheit der Gemeinden, mit der Schwäche der beiden Kommunalverbände (Gemeinde- und Städtever-

² In Deutschland finden sich allerdings grosse Unterschiede zwischen dem Süden mit weniger Gemeindefusionen und Nordrhein-Westfalen mit umfangreichen Gebietsreformen entsprechend dem nordeuropäischen Gebietsmodell.

band) und mit dem Fehlen einer gemeinsamen Modernisierungsvision (wie Neues Steuerungsmodell) zu tun haben. Schwierig zu erklären ist allerdings der Rückstand zu Österreich, welches ebenfalls relativ kleine und wenig homogene Kommunen aufweist. Es ist nicht ausgeschlossen, dass ein Teil der Unterschiede auf länderspezifische Unterschiede im Verständnis von „Verwaltungsreformen“ zurückzuführen sind. Zu beachten ist zudem, dass sich Schweizer Gemeinden im Vergleich zu ihren österreichischen Partnern möglicherweise stärker als politische Entitäten mit entsprechenden Reformprioritäten (z. B. politische Rechte, Verkleinerungen von Gemeindeparlament und -exekutive, regionale Zusammenarbeit) denn als Verwaltungseinheiten mit entsprechendem (NPM-)Reformschwerpunkt verstehen.

2.3 Steuerung der Verwaltungsmodernisierung und Evaluation auf Ebene der Länder bzw. Kantone

Auf der mittleren Ebene sind die **Verwaltungsreformen** in allen drei Ländern weitgehend selbständig entwickelt worden. Dennoch folgen sie einer ähnlichen, nämlich einer NPM-Logik. Diese lose Form der Koordination erklärt sich durch den kleinen Kreis von Wissenschaftlern, welche betriebs- und verwaltungswissenschaftliche Ansätze der Verwaltungsmodernisierung propagieren, durch den engen Kreis von Beratern sowie durch die Imitation von Reformen der Vorreiter durch andere Gebietskörperschaften.

Auch die **Evaluation** der Verwaltungsreformen wird weitgehend dezentral gesteuert. Dabei waren die Länder in **Deutschland** recht aktiv. Gestützt auf die in den einzelnen Ländern erstellten Studien ist inzwischen auch eine zusammenfassende und vergleichende Übersicht über den Stand der Verwaltungsmodernisierung in den deutschen Bundesländern erstellt worden (vgl. Beitrag von Sabine Kuhlmann in diesem Band, Abschnitt 3.2.). In der **Schweiz** haben verschiedene Kantone nach dem Jahr 2000 externe Evaluationen ihrer Verwaltungsmodernisierung durchführen lassen. Anstösse kamen von einem Netzwerk (Arbeitsgruppe Verwaltungsmodernisierung der Schweizerischen Evaluationsgesellschaft). Die Auftragnehmer mehrerer externer Evaluationen waren identisch. In **Österreich** sind auf Länderebene wenige Evaluationen zu verzeichnen; ja es wird eine verbreitete Evaluationskepsis konstatiert (vgl. Beitrag von Pleschberger in diesem Band, Abschnitt 2.2.).

2.4 Wettbewerb innerhalb des Systems oder Wettbewerb der Systeme?

Aus den Länderberichten schälen sich implizit zwei unterschiedliche Formen des Wettbewerbs heraus. Bei Verwaltungsreformen werden in Deutschland (Kuhlmann 2005: 3-8; 2007: 3-6) und Österreich (Beitrag von Pleschberger in diesem Band, Abschnitte 2.1.3. und 2.3.) im Gegensatz zur Schweiz bei interkommunalen Vergleichen recht häufig **Leistungsindikatoren** oder -kennzahlen eingesetzt. Solche Leistungsvergleiche setzen voraus, dass die Kommunen ihre administrative Gliederung und ihr Rechnungswesen einander stark angeglichen haben. Sind diese Voraussetzungen gegeben, ist ein Wettbewerb oder zumindest ein Erfahrungsaustausch „innerhalb des Systems“ möglich. Die entsprechenden Voraussetzungen

sind vor allem in Deutschland und teilweise in Österreich etwas besser gegeben als in der Schweiz, welche föderalistisch erheblich stärker dezentralisiert ist (Watts 1999: 79, 46). Auch in Deutschland und Österreich fallen indessen die unterschiedlichen Strukturen, Funktionen und Gliederungen erschwerend ins Gewicht und haben dazu geführt, dass Vergleichsringe aufgelöst wurden. In der Schweiz sind interkommunale Leistungsvergleiche wegen der überaus grossen Heterogenität der Kantone und Gemeinden mit zusätzlichen Schwierigkeiten verbunden. Die schweizerischen Kantone und Kommunen setzen deshalb eher auf einen umfassenden „**Wettbewerb der Systeme**“, welcher die Steuern und die Leistungen einschliesst. Einzelne Gemeinden bieten unterdurchschnittliche Leistungen an, ja gebärden sich als Trittbrettfahrer, können sich aber dank tiefer Steuern als Standort attraktiv platzieren. Es ist nicht erstaunlich, dass in der Schweiz Evaluationen institutioneller Politik stark durch umfassende vergleichende Studien geprägt sind, die nicht auf ein einzelnes, vorgegebenes Ziel ausgerichtet, sondern breiter und mehrdimensional angelegt und in ihren Schlussfolgerungen weniger normativ und handlungsorientiert, sondern eher deskriptiv und interpretierend gehalten sind.

3 Unterschiede im Stand der Evaluation institutioneller Politik: Notwendigkeit oder Zufall?

3.1 Erklärungen für Unterschiede in der Institutionalisierung von Evaluation

Der Dreiländervergleich institutioneller Politik und ihrer Evaluation wirft Fragen auf: Warum sind die Schweiz und Deutschland im Bereich der Evaluation weiter fortgeschritten als Österreich? Warum sind die unteren Ebenen zum Teil eher reformfreudig als die oberste Ebene? Warum führen einzelne Gliedstaaten eher Evaluationen durch und andere nicht? Damit stellt sich die Frage, ob die Unterschiede sich systematisch erklären lassen.

Die Erklärung von Unterschieden im Stand der Evaluation von Ländern ist generell ein schwieriges Feld. Eine Gruppe von Wissenschaftler hatte anfangs der 1990er Jahre die „Zwei-Wellen-Theorie“ aufgestellt. Entsprechend ihren Befunden war die erste Welle der Evaluation in den 1960er und 1970er Jahren sozialreformerisch geprägt, die zweite ab den 1980er Jahren durch Budgetrestriktionen (Derlien 1990b). Später und breiter angelegte ländervergleichende Studien (21 Staaten) einer ähnlichen Gruppe von Wissenschaftler zeigen indessen, dass sich die Herausbildung der Evaluation nur durch den Rückgriff auf zahlreiche weitere Faktoren einigermassen plausibel erklären lässt (Derlien/Rist 2002: 442-446).

Auch in den Länderstudien Deutschland und Österreich werden Hypothesen zu förderlichen oder hinderlichen Rahmenbedingungen für die Evaluation institutioneller Politik geäussert. Wir werden sie kurz darstellen und miteinander vergleichen sowie in Bezug setzen zur vorher erwähnten breiter abgestützten Studie.

3.2 Konkordanzdemokratie als Chance oder Hindernis der Evaluation

Der Beitrag von Sabine Kuhlmann für Deutschland (in diesem Band, Abschnitt 4.) stellt einen fortgeschritteneren Evaluationsstand im Bereich der lokalen Verwaltungsmodernisierung im Vergleich zu den entsprechenden Ansätzen auf Ebene Länder und Bund fest. Dieser Fortschritt wird erstens damit begründet, dass das Gros der Verwaltungstätigkeit auf dieser Ebene anfällt. Gegenbeispiel wäre allerdings die Schweiz, in welcher zwar auf kommunaler Ebene ebenfalls der Hauptharst der Verwaltungstätigkeit anfällt, jedoch der Evaluationsstand (im Vergleich zu den anderen Ebenen) aus den weiter oben diskutierten Gründen (siehe oben, Abschnitt 2.2. dieses Beitrags) gering ist.

Der Fortschritt wird zweitens damit begründet, dass die auf der Bundes- und Länderebene dominierenden konkurrenzdemokratischen Bedingungen geringe Anreize für Evaluationen bieten, weil kritische Befunde Wasser auf die Mühlen der politischen Opposition wären. Die in verschiedenen Kommunen bestehende Konkordanzdemokratie könne hingegen die Institutionalisierung der Evaluation begünstigen. Diese These ist kongruent mit anderen Beobachtungen, welche in die gleiche Richtung deuten, sei es, dass sie den Informationsbedarf von Konkordanzsystemen generell als hoch ansehen (Widmer 2007: 89), sei es, dass Konkordanzsysteme im Vergleich zu parlamentarischen Konkurrenzsystemen auf das Parlament weniger disziplinierend wirken und dadurch eine grössere Rezeptivität für konkurrierende Informationen und gegebenenfalls für negative Evaluationsbefunde ermöglichen (Bussmann 2007: 3).

Allerdings ist auch Gegenthesen die Plausibilität nicht völlig abzusprechen. Im Beitrag von Pleschberger über Österreich (in diesem Band, Abschnitt 1.) wird nämlich davon ausgegangen, dass konsensuale Lösungen auf Kosten der Systemrationalität gingen, wodurch der Stellenwert der Evaluation gesenkt würde. In gleiche Richtung geht die Argumentation von Spinatsch (2002: 390), wonach Evaluationen Transparenz über die Wirkungen einer Politik schaffen, welche in durch Konkordanz geprägten Verhandlungssystemen nicht erwünscht ist. Damit wird aber von einer (unseres Erachtens zu) pessimistischen Sicht von Konkordanzsystemen ausgegangen, nämlich dass darin Kompromisse um jeden Preis erzielt werden müssen. Eine etwas optimistischere Sicht geht demgegenüber davon aus, dass die den Konkordanzsystemen inhärenten intensiven Verhandlungsprozesse zu einer höheren Interessenberücksichtigung und Rationalität führen und letztlich mit dem Grundanliegen der Evaluation durchaus vereinbar sein können.

3.3 Erklärungen zum „Fall Österreich“: Ambivalenz von NPM-orientierten Reformen

Die geringe Verwendung von Evaluationen in Österreich im Vergleich zu den Nachbarländern ruft nach einer Erklärung. Im Beitrag von Pleschberger (in diesem Band, Abschnitt 1.) werden fördernde (Fiskaldruck, NPM-Reformen) und hemmende (Informalität, Konsens) Faktoren für Evaluationen identifiziert. Inwieweit diese Faktoren sich auf Österreich – anders als auf andere Länder – auswirken, wird aber nicht plausibel dargelegt. Vergessen wird zudem die Tatsache, dass die Wirkungen institutioneller Faktoren selten vollständig determiniert, sondern kontextabhängig sind. Wir wollen dies am Beispiel NPM-orientierter Re-

formen zeigen. Nach Ansicht von Pleschberger (in diesem Band) erhöhen diese tendenziell den Stellenwert von Evaluationen. Beobachtbar sind jedoch auch gegenteilige Auswirkungen.

NPM-Reformen können einerseits – entsprechend der Argumentation von Pleschberger – zum Anlass genommen werden, um **Evaluationen in Reformvorhaben einzubauen** und um zu erfassen, ob diese Wirkungen erbringen. Beispiele dafür finden sich bei der deutschen Kommunalreform und in der Schweiz bei kantonalen Reformen sowie auf Bundesebene bei der FLAG-Reform. Solche NPM-Reformen sind einer selbst-reflexiven Modernität verpflichtet, die sich ihres Erfolges jeweils kritisch vergewissert. Die Selbstreflexivität bezieht sich dabei nicht bloss auf die betriebliche Ebene, sondern auch auf die Policy-Ebene und damit auf die eigentlichen Politikwirkungen einer (institutionellen) Politik.

Dieser positive Zusammenhang zwischen NPM-Reformen und Evaluation ist in der Wirklichkeit allerdings nicht durchwegs vorzufinden. NPM-orientierte Verwaltungsreformen können andererseits nämlich sehr wohl auch **ohne Evaluation** auskommen. Die Verwaltungsreformen auf der österreichischen Bundesebene sind dafür ein Beispiel. Wie im Beitrag von Pleschberger (in diesem Band, Abschnitte 2.1. und 3.) gezeigt wird, bedienen sich solche Reformen zwar häufig der Evaluationsrhetorik. Sie beschränken sich aber faktisch auf betriebliche Aspekte. Was die Schweiz angeht, so zeigt sich, dass bisher keine empirische Evidenz gefunden werden konnte, dass NPM-orientierte Reformen in Kantonen von einer Zunahme an Evaluationen begleitet sind. Im Kanton Bern ist mit Verweis auf die NPM-Reformen gar ein breit angelegtes Projekt zur geleiteten Selbstevaluation von Staatsbeiträgen (ergänzt durch punktuelle externe Fremdevaluationen) auf Eis gelegt worden³. Zwar hat auch diese Variante der NPM-Reformen eine selbstreflexive Komponente eingebaut und möchte sich ihres Erfolges vergewissern. Sie ist aber auf die managerialen Aspekte fokussiert und vernachlässigt die eigentlichen Politikwirkungen, das heisst die Auswirkungen der Reformen auf die Wirtschaft, Gesellschaft und auf die Umwelt, was ja die eigentliche *raison d'être* staatlichen Handelns bildet. Instrumentell wird stärker auf Leistungsindikatoren und weniger auf Evaluationen gesetzt, auch wenn für solche Überprüfungsmechanismen häufig der Begriff „Evaluation“ verwendet wird (vgl. auch Derlien/Rist 2002: 454).

3.4 Fazit

Insgesamt zeigt sich, dass die für den unterschiedlichen Evaluationsstand zwischen den Ländern sowie zwischen den Staatsebenen vorgelegten Erklärungen zum Teil untereinander und zum Teil mit Befunden aus anderen Länderstudien divergieren. Damit lassen sich die vorgefundenen Unterschiede kaum plausibel erklären. Eine Möglichkeit wäre, nach aussagekräftigeren und empirisch breiter abgestützten Faktoren zu suchen. Wir erachten allerdings diese Variante als wenig aussichtsreich. Wir gehen vielmehr davon aus, dass institutionelle und andere Gegebenheiten (z. B. personelle Konstellationen, historische Ereignisse) in sehr komplexer Weise auf den Stand der Evaluation einwirken. Können wir die Entwick-

³ Projekt „Erfolgskontrolle von Staatsbeiträgen“ (ERKOS) 1998-2004 (vgl. www.fin.be.ch/site/print/kaio-amt-informatik-organisation/kaio-projekte/kaio-erkos-einfuehrung.htm [05.10.2007])

lung der Evaluation in einem Jahrzehnt beobachten, wird sich möglicherweise zeigen, dass Evaluationen in modernen Gesellschaften zwar „unverzichtbar“ (Bussmann 2004: 15) sind (was sich in einer hohen Korrelation zwischen BSP/Kopf und der Institutionalisierung bzw. Praxis der Evaluation ausdrückt), ihr genauer Realisierungszeitpunkt und ihre konkrete Verwirklichung aber kaum umfassend erklärbar und damit weitgehend „zufällig“ sind.

Raumentwicklungspolitik

Evaluierung von Raumentwicklungspolitik in Deutschland

1 Begriffsverständnis¹

Raumentwicklungspolitik hat ihren Ausgangspunkt darin, dass wirtschaftliche, soziale und ökologische Entwicklung auch von der Organisation des Raumes abhängt (OECD 2001: 134-145). Während eine Reihe von Politiken raumbezogene Effekte haben, gibt es auch politische Ansätze, bei denen die Beeinflussung der räumlichen Entwicklung im Vordergrund steht. Insbesondere lassen sich Raumplanungs- und Regionalpolitik unterscheiden:

Raumplanungspolitik umfasst jene politischen Maßnahmen, Strukturen und Prozesse, bei denen es darum geht, „Leitbilder eines anzustrebenden, idealen Zustandes des Raumes zu entwickeln und die Voraussetzungen für ihre Verwirklichung zu schaffen“ (Turowski 2005: 894). Primärer Gegenstand politischer Entscheidungen im Bereich der Raumplanungspolitik ist somit die Nutzung des Raumes, der Fläche. Das klassische Instrument hierzu sind Planungsverfahren.

Regionalpolitik umfasst jene Maßnahmen, Strukturen und Prozesse, „die darauf abzielen, die Struktur einer Volkswirtschaft anders zu gestalten, als sie sich aufgrund der marktwirtschaftlichen Prozesse ergeben hätte“ (Eckey 2005: 933), um so die relative Bedeutung der Regionen in der Volkswirtschaft zu verändern. Der primäre Gegenstand der Regionalpolitik ist die Beeinflussung wirtschaftlicher Aktivitäten in ihrer räumlichen Verteilung. Klassische Instrumente sind die finanzielle Förderung und die Bereitstellung wirtschaftsrelevanter Infrastrukturen.

¹ Das in diesem Abschnitt kurz vorgestellte Begriffsverständnis ist gemeinsam abgestimmt und liegt diesem sowie den beiden anderen Länderbeiträgen zum Themenfeld Raumentwicklungspolitik (Thierstein sowie Hummelbrunner/Meier) in diesem Band zu Grunde.

Tabelle 1: Raumentwicklungspolitik: Raumplanungspolitik und Regionalpolitik – Überblick

| | Raumplanungspolitik | Regionalpolitik |
|------------------------|--|--|
| Gegenstand | Entwicklung von Leitbildern eines anzustrebenden Zustandes des Raumes und Schaffung der Voraussetzung ihrer Verwirklichung | Beeinflussung der relativen Bedeutung von Regionen in einer Volkswirtschaft |
| Primärer Ansatz | Planung der Raumnutzung | Anreize zur Veränderung der räumlichen Verteilung wirtschaftlicher Aktivitäten |

2 Nationale Besonderheiten – zur Struktur des Politikfeldes Raumentwicklungspolitik in Deutschland

Beide Teilbereiche des Politikfeldes Raumentwicklungspolitik haben in Deutschland eine lange Tradition. Sie sind sowohl wechselseitig aufeinander bezogen, als auch von eigenständigen Rationalitäten geprägt.

Ziel der *Raumplanungspolitik* ist es, den „Gesamtraum der Bundesrepublik Deutschland und seine Teilräume [...] durch zusammenfassende, übergeordnete Raumordnungspläne und durch Abstimmung raumbedeutsamer Planungen und Maßnahmen zu entwickeln, zu ordnen und zu sichern“ (Raumordnungsgesetz – ROG, § 1). Das ROG bindet die öffentlichen Stellen bei der Durchführung von „raumwirksamen Politiken“ an die Ziele der Raumordnung. Die Regionalpolitik fällt zusammen mit einer Reihe anderer Politiken (Finanzausgleich, Arbeitsmarktpolitik, etc.) unter die „raumwirksamen Politiken“, die durch Raumordnungspläne koordiniert werden sollen.

Die *Regionalpolitik* ist ihrem Selbstverständnis nach „Teil der allgemeinen Wirtschaftspolitik. Primäre Zielsetzung der Regionalpolitik im Rahmen der Gemeinschaftsaufgabe ist es, dass strukturschwache Regionen durch Ausgleich ihrer Standortnachteile Anschluss an die allgemeine Wirtschaftsentwicklung halten können und regionale Entwicklungsunterschiede abgebaut werden.“ (Deutscher Bundestag 2006b: 8). Zwar muss die Regionalpolitik auch mit den „Zielen und Erfordernissen der Raumordnung“ (Gesetz über die Gemeinschaftsaufgabe „Verbesserung der regionalen Wirtschaftsstruktur“ – WiStruktG § 2) übereinstimmen, das Hauptziel ist aber die „Verbesserung der regionalen Wirtschaftsstruktur“ (WiStruktG § 1).

In diesen Definitionen wird das für Deutschland typische Spannungsfeld zwischen beiden Politiken sichtbar: Einerseits erhebt die Raumordnungspolitik einen Koordinierungsanspruch gegenüber den raumwirksamen Politiken einschließlich der Regionalpolitik. Andererseits entzieht sich die Regionalpolitik diesem Koordinierungssinn wenigstens teilweise, indem sie als eigenständiger Teil der Wirtschaftspolitik aufgefasst wird.

Für die historische Entwicklung lassen sich – nach einigen Jahren der Kompetenzklärung zwischen Kommunen, Ländern und Bund in der Nachkriegszeit – für Raumplanungspolitik und Regionalpolitik ähnliche Phasen unterscheiden. Mit dem Raumordnungsgesetz (1965) wurde die Grundlage für die Raumplanungspolitik, mit dem Wirtschaftsstrukturge-

setz (1969) für die Regionalpolitik gelegt. Ab etwa Mitte der 1970er Jahre führte der Rückgang des wirtschaftlichen Wachstums, aber auch eine zunehmende Mobilisierung gesellschaftlicher Protestbewegungen dazu, dass die Leistungsfähigkeit der beiden Politiken zunehmend in Frage gestellt wurde. Gleichzeitig wurden für die Regionalpolitik Strukturen diagnostiziert, die eine Reform erheblich erschweren („Politikverflechtungsfalle“, siehe Scharpf 1985; Hoppe/Voelzkow 1999). Bis zur Wiedervereinigung 1989 standen Raumplanungs- und Regionalpolitik unter Begründungsdruck. Teilweise wurde dieser Druck durch das Aufkommen der europäischen Regionalpolitik, die andere, modernere Förderansätze vertrat, verstärkt. Staatliche und kommunale Planungen wurden zurückgefahren, die Gemeinschaftsaufgabe zunehmend in Frage gestellt.

Ab 1989 wurden mit der Wiedervereinigung beide Politikansätze – die Raumplanungs- wie die Regionalpolitik – stabilisiert, da sie im Zuge des Vereinigungsprozesses politisch alternativlos waren. Gleichzeitig setzten aber auch Reformtrends ein: Im Zusammenhang mit den Debatten um Verwaltungsreformen werden auch in der Raumplanungspolitik Aspekte der Ergebnisorientierung und des New Public Management verstärkt diskutiert. Für die Gemeinschaftsaufgabe rückt neben der „Export-Basis-Theorie“ zunehmend der Gedanke endogener Potenziale in den Fokus. Während dies in der Raumplanungspolitik zu einer Entflechtung und Vereinfachung der Planungsverfahren einerseits sowie zu einer Zunahme informeller Planungen andererseits führte, gewinnen in der Regionalpolitik Ansätze der Netzwerk- und Clusterförderung sowie eigenständige Maßnahmen der Länder an Gewicht. Eine grundlegende Reform beider Politiken fand – anders als etwa in der Schweiz – nicht statt.

Sowohl in der Raumplanungspolitik wie auch in der Regionalpolitik ist die für Deutschland typische Aufteilung und teilweise Verflechtung von Kompetenzen der Bundes- und der Länderebene anzutreffen. Dabei ist die für die Raumordnungspolitik in Deutschland wichtigste Ebene die der Landes- und Regionalplanung. Auf nationaler Ebene werden – nach einem einmaligen Versuch in den 1980er Jahren – keine Entwicklungspläne mehr aufgestellt. Die für die großräumige Rahmenplanung entscheidende Ebene ist die Landesplanung. Im Bereich der Regionalpolitik kommt hingegen der Gemeinschaftsaufgabe „Verbesserung der regionalen Wirtschaftsstruktur“ (GRW) eine wichtige Koordinierungsfunktion zu, da sie einheitliche Rahmenbedingungen für die Förderpolitiken von Bund, Ländern und Gemeinden schafft, während die primäre Zuständigkeit für die Regionalpolitik bei Bund und Ländern liegt. Die Koordinierungsfunktion erstreckt sich auf die Festlegung der Fördergebiete und Förderhöchstsätze, Fördertatbestände und -regelungen. Die GRW bildet den Orientierungspunkt aller regionalpolitischen Instrumente in Deutschland.

Die Regionalpolitik ist sowohl auf Ebene des Bundes als auch der Länder verschränkt mit der europäischen Regionalpolitik, insbesondere aus dem Europäischen Fonds für Regionale Entwicklung (EFRE). Durch die Anforderung, die europäischen Fördermittel aus nationalen Quellen kofinanzieren, entwickelt sich ein System, in dem bestehende Förderinstrumente aus europäischen und nationalen Quellen gespeist werden. Gleichzeitig wirkt die EU-Förderung als mögliche Innovationsquelle (vgl. etwa Voelzkow/Hoppe 1996).

3 Evaluierung von „Raumentwicklungspolitik“ – Rahmenbedingungen

Evaluationen sind weder in der Raumordnung (ROG) noch in der Regionalpolitik (WiStruktG) gesetzlich vorgeschrieben. Das ROG definiert eine Reihe von Grundsätzen der Raumordnung, mit denen die Leitvorstellung einer nachhaltigen Entwicklung beschrieben wird. Es konzentriert sich im Weiteren jedoch darauf, Verfahren und Zuständigkeiten bei der Erstellung und Umsetzung von Planungen zu regeln. Weder werden überprüfbare Ziele definiert, noch ein System der Evaluierung geschaffen. Bemerkenswert ist, dass der im ROG festgeschriebene regelmäßig vorzulegende Raumordnungsbericht über die „Auswirkungen der Politik der Europäischen Gemeinschaft auf die räumliche Entwicklung des Bundesgebietes“ informieren soll, im Bereich der nationalen Aktivitäten aber lediglich über die durchgeführten Maßnahmen und ihre räumliche Verteilung. Auch die Regionalpolitik des Bundes sieht kein umfassendes Evaluierungssystem vor. In den Haushaltsordnungen des Bundes und der Länder wird grundsätzlich eine Wirtschaftlichkeitsprüfung der jeweiligen Maßnahmen gefordert. Eine inhaltliche Begleitung und Bewertung, die im Sinne von Evaluation auch in die fachlichen Entscheidungsprozesse rückgekoppelt würde, liegt damit nicht vor: „Doch ist in der Bundesrepublik Deutschland regelmäßig in den Gesetzen und Programmen keine Erfolgskontrolle vorgeschrieben.“ (Mäding 1995: 226).

Eine Evaluationspflicht wird lediglich durch den Einsatz der europäischen Strukturfondsmittel begründet. Aufgrund der Festlegungen in den einschlägigen Verordnungen sind für EU-Mittel Evaluierungen vorzunehmen. Da diese Verpflichtungen unmittelbar gelten, ist auch der Mitteleinsatz in Deutschland der Evaluierung unterworfen. Damit wird aber auch der als nationale Kofinanzierung verausgabte Teil der Strukturfondsprogramme ‚evaluierungspflichtig‘. Da die GRW in einer Reihe von Bundesländern mit Strukturfondsmitteln kofinanziert wird, entsteht somit ein indirekter Evaluierungszwang für weite Teile der Regionalpolitik, der aber in der Praxis teils nur „symbolisch“ implementiert wird (Lang 2003).

Im Zusammenhang mit den Diskussionen um Subventionsabbau wird verstärkt die Prüfung des Erfolgs von Fördermitteln gefordert (Deutscher Bundestag 2006a). Auf Ebene der Länder wird besonders dort, wo Planungs- oder Regionalpolitiken eine tiefgreifende Reform durchliefen, auch die Evaluation gestärkt. Ein Beispiel ist die jüngste Reform der Regionalpolitik in Brandenburg, die intensiv wissenschaftlich begleitet und bewertet wird. Allerdings wird der Bedarf an Erfolgskontrolle meist vorrangig als eine Erweiterung der Wirtschaftlichkeitsprüfung nach Haushaltsordnung verstanden. Einzelne Landesrechnungshöfe drängen verstärkt auch auf Wirkungsanalysen der Politiken. Es ist noch nicht absehbar, inwieweit tatsächlich eine, auf die inhaltlichen Wirkungen der Förderung abzielende systematische Evaluation angestrebt wird. Generell ist die skizzierte Entwicklung eher im Bereich der Regionalpolitik als der Raumordnungspolitik erkennbar.

4 Evaluierung in der Raumplanungspolitik

Eine Bilanz der Raumordnungspolitik hinsichtlich ihrer Evaluierung fällt ernüchternd aus:

„Eine systematische Evaluierung der Raumplanung auf den verschiedenen Entscheidungsebenen in Deutschland (Bundesebene, 16 Bundesländer, ca. 75 Regionen, z.T. Landkreise) fehlt bisher sowohl für diese klassischen raumplanerischen Aufgaben als auch für die neuen Instrumente“ (Hübler 2002: 12).

Mit Hilfe der beim Bundesamt für Bauwesen und Raumordnung geführten „laufenden Raumb Beobachtung“ ist eine umfangreiche und detaillierte Beobachtung räumlicher Entwicklungsprozesse in Deutschland möglich. Diese wird jedoch für Evaluationszwecke bislang kaum genutzt. Die laufende Raumb Beobachtung ist ein Monitoring-, kein Evaluierungsinstrument. Auch die Raumordnungsberichte, die ab 1997 als Ressortberichte des Bundesministeriums für Verkehr, Bau und Stadtentwicklung vorgelegt wurden, erfüllen diese Funktion nicht. Zwar konnten gegenüber den vorangegangenen Berichten im Bericht 2005 (Deutscher Bundestag 2005) in einzelnen Bereichen Fortschritte erzielt werden (etwa im Bereich der Nachhaltigkeit räumlicher Entwicklungen), insgesamt bleibt aber die Einschätzung von Hübler (2002: 18) gültig, dass der Raumordnungsbericht zur Evaluierung der Wirksamkeit der Bundesraumordnung nur „begrenzt taugt“, weil auf die „Ermittlung von Wirkungen und Nebenwirkungen weitgehend verzichtet wird“.

Einzelne Ansätze zu Evaluierungen werden von Hübler (2002: 19) ebenfalls als ungenügend eingestuft. Dies gilt für die Evaluierungen von Zentrale-Orte-Konzepten, wie auch für die Evaluierung von Nachhaltigkeitskonzepten von Regionalplänen. Eine intensivere wissenschaftliche Begleitung erfahren in den letzten Jahren vergleichsweise neue Instrumente der Planung (Regionalentwicklungskonzepte, Runde Tische, Agenda21-Prozesse, Regionalmanagements, etc. – vgl. die Beiträge in Sedlacek 2004b). Wenn jüngst eine steigende Relevanz der Evaluierung ausgemacht wird (etwa Sedlacek 2004a; Strubelt 2004) liegt dies aber wenigstens teilweise außerhalb der Raumplanungspolitik: Die Evaluierungsrelevanz wird besonders im Zusammenhang mit modellhaften Förderprogrammen des Städtebaus und der Stadtentwicklung gesehen (Soziale Stadt, Stadtumbau Ost und West, experimenteller Wohnungs- und Städtebau) und nicht auf der Ebene der Planungen.

Was die Akteursstrukturen angeht, so sind für die Evaluierung der Raumplanungspolitik vor allem jene Akteure zu nennen, die für die wissenschaftliche Beschäftigung mit der Raumentwicklung stehen: Als großes wissenschaftliches Netzwerk ist dies allen voran die Akademie für Raumforschung und Landesplanung, sowie die darin größtenteils vertretenen raumwissenschaftlichen Forschungsinstitute und universitären Einrichtungen. Auftraggeber von evaluationsbezogenen Arbeiten sind das Bundesamt für Bauwesen und Raumordnung, sowie auf Länderebene vor allem die für die Landesplanung zuständigen Ressorts. Daneben spielen auch auf regionaler Ebene formierte Akteure aus Planungs- und Entwicklungsprozessen eine Rolle.

Hinsichtlich der Methoden der Evaluierung ist für die Raumplanung kaum eine allgemeine Aussage möglich. Zugenommen haben in jüngster Zeit die Evaluierungen neuer Planungs- und Entwicklungsinstrumente (Regionale Entwicklungskonzepte, Agenda21

etc.), wobei auch partizipative Methoden eingesetzt werden. Auch Sozialstruktur- und Netzwerkanalysen bieten methodische Ansatzpunkte.

Generell wird aber die Nutzung von Evaluationen im Bereich der Raumordnungspolitik skeptisch eingeschätzt. „Speziell die räumliche Planung nutzt das Instrument und die Verfahren der Evaluation in der Praxis noch viel zu wenig und vor allem nicht systematisch genug, um zu einer aussagekräftigen Wirkungsanalyse und -kontrolle ihrer Pläne zu kommen“ (Wollmann 2005a: 279).

5 Evaluierung in der Regionalpolitik

Für die Darstellung des Standes der Evaluierung in der Regionalpolitik wird zwischen der Gemeinschaftsaufgabe als dem dominierenden Instrument, und innovativen Ansätzen unterschieden. Eine besondere Rolle spielt außerdem der Bereich, in den EU-Mittel fließen.

5.1 Gemeinschaftsaufgabe „Verbesserung der Regionalen Wirtschaftsstruktur“

Die GRW wird einem dreistufigen System aus Vollzugs-, Zielerreichungs- und Wirkungskontrolle unterworfen. Die Vollzugskontrolle wird beim Bundesamt für Wirtschaft und Ausfuhrkontrolle geleistet, die Zielerreichungs- und Wirkungskontrolle werden im Rahmen von gesondert vergebenen Forschungsprojekten unternommen (siehe etwa Schalk/Untiedt 1996; 2000; Bellmann/Schwengler 2000; Koller et al. 2001; KfW 2001; Koller 2005). Nur im Einzelfall wird die Umsetzung der Gemeinschaftsaufgabe auf Länderebene nochmals gesondert untersucht (ifo 1999; GEFRA et al. 2004).

Für die Zielerreichungs- und Wirkungskontrolle ist das Bundeswirtschaftsministerium der entscheidende Akteur. Nur im Ausnahmefall treten Länderministerien als Auftraggeber auf. Die Analysen werden vorwiegend durch die großen wirtschaftswissenschaftlichen Institute, das Institut für Arbeitsmarkt- und Berufsforschung sowie einzelne spezialisierte private Akteure erbracht.

Auf der Methodenseite bestehen vor allem zwei Zugänge. Als einzelwirtschaftlicher Ansatz wird – entweder aus den Förderdaten oder durch Befragung – die Datengrundlage auf der Ebene der einzelnen Unternehmen genutzt. Diese Ansätze werden zunehmend mit Matching- und Kontrollgruppenverfahren kombiniert. Ein zweiter Ansatz wählt statt der einzelwirtschaftlichen die gesamtwirtschaftliche Betrachtungsweise – meist mit Hilfe ökonomischer Modelle.

In einer Bestandsaufnahme zur Erfolgskontrolle der deutschen Regionalpolitik kommen Lammers und Niebuhr (2002) zu einer wenig zufrieden stellenden Einschätzung. So könne für die Investitionsförderung zwar davon ausgegangen werden, dass zusätzliche Investitionen und „in geringerem Umfang“ zusätzliche Arbeitsplätze geschaffen werden, über den entscheidenden Aspekt der Mitnahmeeffekte bestünden aber nur ungenügende Kenntnisse. Das Autorenduo kritisiert, dass für den gesamten Bereich der Infrastrukturförderung wie auch für die nicht-investive Förderung im Rahmen der Gemeinschaftsaufgabe keine befriedigende Bewertung vorliegt. Effizienzfragen würden bislang kaum berücksich-

tigt. Der Stand hinsichtlich der Erfolgskontrollen ist somit „generell sicherlich nicht zufriedenstellend“ (Lammers und Niebuhr 2002: 59).

5.2 Innovative regionalbezogene Maßnahmen

Im Bereich der Regionalpolitik haben in den letzten Jahren Programme mit regionalem Fokus und starker Vernetzungskomponente nicht unerhebliches Gewicht gehabt (etwa InnoRegio, Lernende Regionen, etc.). Diese Programme verfolgen meist innovative Politikansätze und haben häufig auch das Design eines Modellprogramms: Sie werden zunächst befristet aufgelegt und mit einer intensiven Begleitforschung auf ihre Ergebnisse und Wirkungen hin untersucht. Dementsprechend finden im Rahmen dieser großen und umfangreichen Forschungsvorhaben auch Evaluierungsfragen Berücksichtigung (vgl. etwa Deutsches Institut für Erwachsenenbildung et al. 2004; BMBF 2005). Der in diesem Zusammenhang häufig verwendete Begriff „Begleitforschung“ ist von „Evaluierung“ meist nicht sauber abgegrenzt. Begleitforschung impliziert ein eher klassisch wissenschaftliches Verständnis. Stärker als bei der Evaluierung steht das Erkenntnisinteresse – weniger der Nutzen – im Vordergrund. Folglich ist „Begleitforschung“ in aller Regel begrifflich auch auf klassische Sozialforschung ausgerichtet, während Evaluierung vor allem auch partizipativere Methoden mit einschließen kann.

Auftraggeber sind hier vorrangig die zuständigen Bundesministerien (v. a. Wirtschaft und Bildung), teils auch Landesministerien. Auftragnehmer sind in der Regel große Forschungsinstitute oder Beratungsfirmen, die häufig in Kooperation mit anderen Bietern auftreten, da sehr komplexe Evaluierungsdesigns ein breit aufgestelltes Team erfordern. Gegenüber der GRW-bezogenen Evaluierung werden hier verstärkt die Netzwerkstruktur und Kooperationsbedingungen in den jeweiligen regionalen Strukturen erfasst.

5.3 EU-Strukturfonds und Länderpolitiken

Aufgrund der von den Strukturfondsverordnungen der EU abgeleiteten Evaluierungspflicht dürfte es in diesem Bereich mit die intensivste Evaluierungsaktivität geben. Toepel und Schwab (2005) beziffern die Anzahl der Evaluierungsstudien für die Strukturfonds in Deutschland insgesamt in der Förderperiode 2000 bis 2006 auf mindestens 210. Selbst wenn man davon nur die im engeren Sinne regionalpolitischen Programme (EFRE, Interreg, Urban) betrachtet, dürften doch deutlich über hundert Evaluierungsstudien durchgeführt worden sein.

Der Schwerpunkt der strukturfondsbezogenen Evaluierungen liegt auf Länderebene. Sie sind im Gegensatz zu den Evaluierungen der GRW in der Regel Multi-Instrumenten-Evaluierungen, die nicht nur Instrumente der Investitionsförderung, sondern auch der FuE- und Infrastrukturförderung abdecken. (vgl. illustrativ IfS et al. 2003; 2005). Diese Evaluierungen brechen die im Bereich der GRW-Evaluierung kritisierte Engführung auf die Investitionsförderung wenigstens teilweise auf. Gleichwohl geht diese Verbreiterung des Evaluationsfokus häufig auf Kosten der Tiefe der Untersuchung.

Auftraggeber sind in erster Linie die zuständigen Ressorts auf Länderebene. Die Evaluierungen werden entweder durch freie, meist private oder als An-Institute organisierte Forschungseinrichtungen oder durch eher beratungsorientierte, häufig international agierende Unternehmen erbracht. Anders als in den 1990er Jahren spielen mittlerweile die großen Wirtschaftsforschungsinstitute kaum noch eine Rolle.

Was die Methoden angeht, so sind die aus der Umsetzung der Förderung gewonnenen Daten zumeist die zentrale Grundlage. Für ausgewählte Bereiche der zu untersuchenden Programme werden ergänzende Befragungen durchgeführt, die teilweise auch Vergleichsgruppendesigns umfassen. Daneben kommen unterschiedliche Methoden zur Evaluierung auf Programmebene statt auf Instrumentenebene (also auf einer höheren Aggregationsstufe) zum Einsatz (makroökonomische Modelle, Scoringmodelle, Entwicklungspfadanalysen etc.). In geringerem Umfang spielen auch partizipative Verfahren eine Rolle.

Eine Gesamteinschätzung ergibt eine im internationalen Vergleich geringe Entwicklungsdynamik: „In the German context, this practice has already seen uneven development, and overall has flourished less than in some other contexts.“ (Taylor et al. 2001: 356.)

6 Evaluationssetting und Evaluationscommunities

Vom Setting her überwiegen in allen Bereichen externe Evaluationen. Bei modellhaften Programmen (Städtebau, Regionenprogramme) werden diese meist als laufende „Begleitforschung“ umgesetzt. Im Bereich der europäischen Programme hängt das Setting stark von den EU-Vorgaben ab. In der Förderperiode 2000 bis 2006 wurde vorwiegend eine Abfolge von Ex-Ante-Evaluierung, Halbzeitbewertung, Aktualisierung der Halbzeitbewertung sowie gegebenenfalls Ex-post-Evaluierung beauftragt. Für die Förderperiode 2007 bis 2013 zeichnet sich eine deutlich stärkere Bedeutung kontinuierlich begleitender Evaluationen ab.

Internationale Kooperationen und Beteiligungen in den Evaluationsteams sind die Ausnahme. Sie kommen soweit ersichtlich vor allem auf zwei Wegen zustande. Zum einen wenn die Kooperation hinsichtlich spezifischer Methoden erforderlich ist. Dies ist etwa bei makroökonomischen Modellierungen gelegentlich der Fall. Zum anderen treten auch Akteure aus Österreich, Dänemark oder der Schweiz alleine oder in Kooperation mit deutschen Akteuren als Anbieter von Evaluationen auf.

Was die Akteursstrukturen im Bereich von Begleitforschung und Evaluierung angeht, lassen sich drei eigenständige Communities identifizieren:

Eine erste Gruppe im Bereich der Raumordnungspolitik besteht aus der Akademie für Raumforschung und Landesplanung, dem Bundesamt für Bauwesen und Raumordnung, den großen raumwissenschaftlichen Forschungsinstituten, fachlich einschlägigen Universitätsinstituten sowie einer Anzahl von unterschiedlich großen privaten Instituten und weiteren Akteuren.

Eine zweite Gruppe, die besonders aus wirtschaftswissenschaftlicher Perspektive Wirkungen regionalpolitischer Instrumente diskutiert, gruppiert sich um die großen Wirtschaftsforschungsinstitute, das Institut für Arbeitsmarkt- und Berufsforschung, einige Universitätsinstitute sowie eine geringe Anzahl privater Institutionen und Akteure.

Eine dritte Gruppe von Akteuren agiert im Bereich der Strukturfondsförderung. Hier sind kleine und mittlere private gegenüber öffentlich finanzierten Akteuren – Universitäten oder Institute – stärker vertreten.

Zwischen den Akteuren, die Evaluierung von Raumordnungspolitiken betreiben und den Communities der Regionalpolitik und Strukturfonds besteht wenig Kontakt. Andererseits überschneiden sich die Communities mit Bezug auf Regionalpolitik und Strukturfonds erheblich.

7 Fazit: Stand der Evaluationskultur

Trotz einer teilweise erheblichen Anzahl von Evaluierungsstudien – etwa im Bereich der Strukturfondsförderung – sowie intensiver wissenschaftlicher Wirkungsforschung – etwa im Bereich der Regionalpolitik – ist der Stand der Evaluationskultur im Politikfeld Raumentwicklungspolitik insgesamt ungenügend. Eine systematische Rückkopplung von Politikergebnissen in die Entscheidungsprozesse findet weder in der Raumordnungs- noch in der Regionalpolitik statt.

Evaluiert wird vorrangig dort, wo es neue und modellhafte Instrumente gibt. In der Regionalplanung sind dies etwa Regionalentwicklungskonzepte, im Bereich der Regionalpolitik Programme wie InnoRegio oder Lernende Regionen. Hier ist meist eine Begleitforschung vorgesehen, die in der Regel starke Evaluationselemente aufweist. Dagegen ist im Bereich der Regional- und Landesplanung ebenso wie im Bereich der Standardinstrumente der Regionalpolitik Evaluation wenig verbreitet. Als Gründe hierfür werden für die Raumplanung „mangelhafte Umsetzungsorientierung“ und „Langfristigkeit der Zielhorizonte“ (Wollmann 2005a: 279) genannt. Auf einen zentralen Punkt weisen Lammers und Niebuhr (2002: 60) bei der Diskussion des Standes der Erfolgskontrolle in der Regionalpolitik hin: Evaluierbarkeit erfordert die genaue Benennung von Zielen. Sowohl in der Raumordnungs- wie auch in der Regionalpolitik bleiben Zielformulierungen meist so unspezifisch, dass eine belastbare Bewertung der Zielerreichung kaum möglich wird. Insgesamt steht im betrachteten Politikfeld noch eine Steuerung über Verfahren im Vordergrund. Dies äußert sich in der Raumplanungspolitik in einem starken Fokus auf die Planerstellung und in der von Wollmann beklagten fehlenden Umsetzungsorientierung, in der Regionalpolitik in einer Dominanz der Steuerung über fiskalische und budgetäre Kenngrößen anstelle von Informationen über Ergebnisse und Wirkungen.

Die gesamte Raumentwicklungspolitik ist bislang nicht auf Evaluierbarkeit ausgelegt. In der Raumplanungspolitik wird realiter die Durchführung des Planungsprozesses, und nicht die Umsetzung seiner Ergebnisse zum Erfolgskriterium. In der Regionalpolitik wird Erfolg häufig als Verausgabung der Mittel und nicht als deren Wirkung verstanden. Solange dies der Fall ist, sind die Möglichkeiten der Evaluierung wie auch ihre Bedeutung im Politik- und Implementationsprozess gering.

Vor diesem Hintergrund ist es nicht erstaunlich, dass die Regierungen und Verwaltungen auf Bundes- und Länderebene wenig bis gar keine internen Ressourcen im Bereich der Evaluation vorhalten. Wenn doch, sind diese häufig nicht im Kernbereich der Verwaltungen angesiedelt sondern als temporäre, kurzzeitig wieder abbaubare Unterstützungsleistungen.

Damit ist auch die Dauerhaftigkeit der bisher erzielten Lerneffekte im Zusammenhang mit Evaluierungen bisher nicht gesichert – ein ‚Rückfall‘ ist jederzeit möglich.

Fortschritte gibt es – soweit ersichtlich – nur in den Bereichen, in denen in nennenswertem Umfang evaluiert wurde. Für die EU-Strukturpolitiken gibt es spürbare Lerneffekte: Evaluation wird weniger als Kontrollinstrument, sondern zunehmend als hilfreiche Unterstützung angesehen. Ein begünstigender Faktor für die Entwicklung der Evaluationskultur ist neben dem vorwiegend von der EU ausgeübten Druck die jüngst zumindest auf Bundesebene, aber auch in einzelnen Ländern stärker geforderte Erfolgskontrolle.

Was die Methoden angeht, werden Vergleichsgruppendesigns vor allem im Bereich der unternehmensbezogenen Investitionsförderung eingesetzt. Ansonsten wird die Raumentwicklungspolitik kaum mit experimentellen Evaluierungsdesigns untersucht. Teilweise erlauben makroökonomische Modelle eine Abschätzung der Nettowirkungen. In belastbaren Methoden für die Erfassung von Nettowirkungen, aber auch in Verfahren zur Aggregation unterschiedlicher Ergebnisse im Falle von Multi-Instrumenten-Evaluierungen liegen die großen methodischen Herausforderungen. Teilweise sind – etwa im Bereich der Raumplanungspolitik – auch mögliche Wirkungsbeziehungen noch zu klären.

Es fällt auf, dass es sehr wenig Forschung zu Evaluation im Politikfeld Raumentwicklungspolitik gibt. Aus der raum- und regionalwissenschaftlichen Forschung widmen sich – unabhängig von der fachlichen Provenienz – nur einzelne diesem Thema. Während insbesondere für die Regionalpolitik die Erfassung der Programmwirkungen innerhalb der wirtschaftswissenschaftlichen Diskurse eine Rolle spielt, ist das zweite Charakteristikum der Evaluierung, die Rückkopplung von Wirkungswissen in den Politikprozess, kaum Gegenstand wissenschaftlicher Auseinandersetzung.

Qualitätssicherung und Methodenentwicklung verlaufen weitgehend unsystematisch. Im Falle der wirtschaftswissenschaftlichen Evaluierungen der Regionalpolitik über die fachwissenschaftliche Community, im Falle der strukturfondsbezogenen Evaluierungen über die dort relevanten Foren auf nationaler und europäischer Ebene. Foren zum Austausch werden vor allem organisiert in zwei Arbeitskreisen der DeGEval (Stadt- und Regionalentwicklung sowie Strukturpolitik), sowie in weiteren fachwissenschaftlichen Veranstaltungen. Evaluationsstandards spielen bislang praktisch kaum eine Rolle – wobei ihre Wahrnehmung und Relevanz steigt, wie jüngst wiederholte Verweise in Ausschreibungsunterlagen illustrieren.

Generell ist somit Evaluierung im Politikfeld Raumentwicklungspolitik noch nicht in die politische Kultur integriert. Evaluiert wird vorrangig bei innovativen Instrumenten und dort, wo der Einsatz von EU-Mitteln dazu zwingt. Politikformulierung und –steuerung sind bislang aber nicht systematisch auf die Nutzung von Evaluationswissen angelegt. Doch gibt es neue Ansätze – betrieben vor allem von der Bundesregierung sowie von Landesrechnungshöfen – die auf eine bessere Erfolgskontrolle zielen. Zur Verstärkung dieses Trends könnten eine bessere Nutzung der vorhandenen Evaluierungserfahrungen und besonders eine stärker ausgebauten Forschung über Evaluierung beitragen.

Evaluation von Raumentwicklungspolitik in Österreich

1 Struktur des Politikfeldes „Raumentwicklungspolitik“ in Österreich¹

Die Bundesverfassung konstituiert Österreich als einen föderativen und spezifisch dezentralisierten Staat. Der Begriff Raumordnung selbst findet sich nicht in der Bundesverfassung, sie ist dort nicht als Staatsaufgabe verankert. Auf Grund des Verfassungsgrundsatzes einer Generalzuständigkeit der Länder für nicht ausdrücklich dem Bund übertragene Hoheitsaufgaben (Bundesverfassungsgesetz B-VG) fällt hoheitliche Raumordnung in die Zuständigkeit der Länder, mit Ausnahme jener Planungsmaßnahmen, die in die ausdrückliche Zuständigkeit des Bundes fallen (z. B. Eisenbahnwesen, überregionales Straßennetz, Bergwesen, Forstwesen, Wasserrecht; ÖROK 1998: 23).

Raumordnung und Raumplanung wird in Österreich also von Bund, Ländern und Gemeinden wahrgenommen. Sie ist damit kompetenzrechtlich eine komplexe Materie, bei welcher der Bund aufgrund der sektoralen Zuständigkeiten, die Länder aufgrund ihrer umfassenden Planungsbefugnis tätig werden. Allerdings gibt es in Österreich keine „Rahmenkompetenz“ des Bundes. Landesgesetze bilden die gesetzliche Grundlage für die überörtliche und örtliche Raumplanung. Die überörtliche Raumplanung stellt die verbindliche Rahmenplanung für die örtliche Raumplanung der Gemeinden dar. Die Vollziehung der örtlichen Raumplanung fällt nach dem B-VG in den eigenen Wirkungsbereich der Gemeinden (ÖROK 2002: 4).

Bei derart vernetzten Materien kommt der gegenseitigen Abstimmung und weitgehenden Kooperation große Bedeutung zu. Um den Notwendigkeiten der Zusammenarbeit zwischen den drei Ebenen von Gebietskörperschaften zu entsprechen wurde 1971 die Österreichische Raumordnungskonferenz (ÖROK) geschaffen. Sie ist eine von Bund und Ländern beim Bundeskanzleramt (BKA) eingerichtete Organisation zur Koordination der Raumordnung auf gesamtstaatlicher Ebene. Das Beschlussorgan umfasst unter dem Vorsitz des Bundeskanzlers alle Bundesminister und Landeshauptmänner sowie die Präsidenten des Städte- und des Gemeindebundes und mit beratender Stimme auch der Interessensvertretungen. Unterhalb der politischen Konferenz sind Gremien bzw. Arbeitsausschüsse eingerichtet, die von der Beamtenebene beschickt werden und vorbereitende bzw. spezifische Aufgaben erfüllen (vgl. ÖROK 1998: 74).

Ein Spezifikum dieser Aufgabenteilung zwischen Bund, Ländern und Gemeinden ist es, dass alle Gebietskörperschaften im Rahmen ihrer privatwirtschaftlichen Tätigkeit raumwirksame Maßnahmen setzen können, für deren Ausgestaltung die Verfassung keine Nor-

¹ Zum Begriffsverständnis von „Raumentwicklungspolitik“ vgl. den zwischen den Autoren abgestimmten ersten Abschnitt im Beitrag von Schwab in diesem Band.

men setzt (vgl. ÖROK 1992: 7). Das ist besonders für den Bereich der Regionalpolitik von Bedeutung. Bund und Länder (teilweise auch Gemeinden) betreiben mit Hilfe verschiedener Programme die Förderung der regionalen Entwicklung (vgl. ÖROK 1998: 27).

Während diese Fördertätigkeiten von Bund und Ländern anfangs überwiegend nicht abgestimmt vorgenommen wurde, erfolgte dies seit den 1980er Jahren teilweise koordiniert. Einschneidend war der Beitritt Österreichs zur Europäischen Union im Jahr 1995, seit dem einerseits die EU-beihilfenrechtlichen Vorschriften für die österreichischen Regionalförderungen gelten und sich andererseits die österreichischen Gebietskörperschaften an der EU-Strukturfondspolitik beteiligen.

Die Vorgaben von europäischer Ebene, die EU-Regionalpolitik im Rahmen mehrjähriger, integrierter Programme durchzuführen, führte zu einer systematischen und zwischen Bund und Ländern koordinierten Vorgangsweise, was eine neue Qualität darstellte. Im Kontext der österreichischen EU-Regionalpolitik nimmt die ÖROK eine wichtige Funktion zwischen innerstaatlicher und europäischer Ebene ein.

Bei den regionalen Zielprogrammen, auf die hier primär fokussiert werden soll, wurde die Aufgabe der Verwaltungsbehörden – und damit auch die Hauptzuständigkeit für die Durchführung von Evaluationen – auf die Landesebene übertragen.

Bedingt durch das komplexe österreichische Fördersystem, das für einzelne Projekte finanzielle Beiträge von verschiedenen Förderstellen auf Bundes-, Landes- oder kommunaler Ebene ermöglicht, sind in den Programmen für jede „Maßnahme“ sogenannte „Maßnahmenverantwortliche Förderstellen“ (MF) festgelegt, die für die Koordinierung der an einer Maßnahme beteiligten Förderstellen verantwortlich sind. Die Abwicklung der EU-Förderung auf (Einzel-)Projektebene (Antragstellung, Genehmigung, Abrechnung) erfolgt somit durch die jeweils zuständige Bundes- oder Landesförderstelle.

Um die Verantwortlichkeiten bei der Implementierung der regionalen Strukturfondsprogramme für die Periode 2000-2006 auch innerhalb Österreichs klarer zu regeln, wurden diese – aufbauend auf die Erfahrungen mit den Strukturfondsprogrammen der Periode 1994/95-1999 und unter bestmöglicher Nutzung der bestehenden Verwaltungsstrukturen und -kapazitäten – in einer Vereinbarung gemäß § 15a BV-G zwischen Bund und Ländern² festgelegt.

2 Evaluation von „Raumentwicklungspolitik“

2.1 Rahmenbedingungen, rechtliche Aspekte und Zuständigkeiten

In Österreich gibt es keine Tradition zur Durchführung von Evaluationen im Bereich der Raumentwicklungspolitik. So gibt es weder im Bereich der Raumplanungspolitik noch der Regionalpolitik diesbezügliche nationale gesetzliche Vorschriften.

² BGBl 147/21.12.2002 – Vereinbarung zwischen dem Bund und den Ländern gemäß Artikel 15a B-VG über Regelungen zur partnerschaftlichen Durchführung der Regionalprogramme im Rahmen der EU-Strukturfonds in der Periode 2000-2006.

Im Rahmen der Haushaltsordnungen sowie der Aufgaben des Rechnungshofes sind Überprüfungen der Gebarung der öffentlichen Hand (unter anderem) hinsichtlich ihrer Sparsamkeit, Wirtschaftlichkeit und Zweckmäßigkeit vorgesehen (Rechnungshofgesetz, BGBl Nr. 144/1948 in der Fassung BGBl I Nr. 39/1999), was aber weniger Evaluationen im Sinne des gegenständlichen Artikels entspricht (vgl. Pichler in diesem Band).

Die Verpflichtung zur Evaluation im Bereich der Regionalpolitik in Österreich erfolgte erstmals 1995 zum Zeitpunkt des Beitritts zur Europäischen Union mit der Beteiligung an den EU-Strukturfondspolitik der Periode 1994/5-1999 (vgl. Artikel 6 der Rahmenverordnung 2081/93). Damit kann die Strukturfondspolitik als treibende Kraft für dieses Betätigungsfeld in Österreich bezeichnet werden (vgl. ÖROK 2003: 44).

2.2 Raumplanungspolitik

Im Bereich der Raumplanungspolitik kann in Österreich nicht von einer systematischen Anwendung des Instruments der Evaluation gesprochen werden, sondern es kommt eher punktuell zu aktuellen Fragestellungen zum Einsatz.

Beispielhaft kann hier die 2004 im Auftrag des Amtes der Vorarlberger Landesregierung durchgeführte Evaluierung der raumordnungsrechtlichen Regelungen für Einkaufszentren in Vorarlberg angeführt werden. In dieser Studie wurden die Vorschriften Vorarlbergs mit jenen der anderen Bundesländer verglichen und deren Steuerungswirkung untersucht (TU Wien 2004).

Ebenso zur Frage der Größen von Handelsbetrieben führte 2006 das Büro DI Tischler im Auftrag des Amtes der Steiermärkischen Landesregierung eine Evaluierungsstudie durch. Ziel war es, basierend auf dem Vergleich der unterschiedlichen gesetzlichen Regelungen ausgewählter Bundesländer Empfehlungen hinsichtlich eventueller Abänderungen des Raumordnungsgesetzes des Landes zu erhalten (Fleck 2006).

Ein breiteres Untersuchungsspektrum hat die 1997 durchgeführte Evaluation des „Österreichischen Raumordnungskonzeptes 1991“, welche in die Empfehlungen mündete, das Raumordnungskonzept neu zu gestalten sowie auf die Erweiterung des Aufgabenfeldes, die Erhöhung der Aufgabenkomplexität, die Verschärfung mancher Probleme, die Erweiterung der Grundsätze und Ziele sowie auf die Grenzen der Steuerung räumlicher Prozesse Bedacht zu nehmen (Kurzfassung veröffentlicht in ÖROK 1999: 9-63).

Weiter sind an dieser Stelle die Instrumente der „Strategischen Umweltprüfung“ (SUP) sowie der „Umweltverträglichkeitsprüfung“ (UVP) zu nennen, die beide auf gesetzlichen Grundlagen³ beruhen und auf die Feststellung von Auswirkungen von Plänen und Programmen bzw. von Projekten im Umweltbereich abstellen.

³ Bundesgesetz über die Prüfung der Umweltverträglichkeit (Umweltverträglichkeitsprüfungsgesetz 2000 (UVP-G 2000) in der gegenwärtigen Fassung) sowie Richtlinie 2001/42/EG des Europäischen Parlaments und des Rates über die Prüfung der Umweltauswirkungen bestimmter Pläne und Programme vom 27. Juni 2001.

2.3 Regionalpolitik

Nationale Förderprogramme und -instrumente

In den letzten Jahren wurden zwei zentrale Säulen des Konzepts der endogenen Regionalentwicklung in Österreich evaluiert:

- „Förderungsaktion Eigenständige Regionalentwicklung – FER“: Diese seit 1979 bestehende Förderungsaktion des BKA diene der finanziellen Förderung von Projekten in strukturschwachen Regionen sowie der Unterstützung regionaler Betreuungsstrukturen. Ziel der Evaluierung war zu analysieren, wie im Zeitraum 1979-1999 eigenständige Regionalentwicklung in Österreich mit dem Instrument FER umgesetzt wurde und welche Entwicklungsdynamiken es darin gab. Dazu wurde das Instrument analytisch positioniert und anhand von einzelnen Komponenten sowie der gesamten Entwicklungsdynamik im Evaluierungszeitraum bewertet (Joanneum Research/InTeReg 2000).
- Regionalmanagement: Mit dem Beitritt Österreichs zur Europäischen Union 1995 wurden vom BKA gemeinsam mit den Bundesländern Regionalmanagementstrukturen zur Unterstützung der Regionen für die Nutzung der EU-Strukturfondsmittel gegründet. Ziel der Evaluierung war eine Bilanz der Pionierphase dieser Strukturen von 1995 bis 2002, mit dem Fokus auf Systembeziehungen, Wirkungen, Berufsbildentwicklung und Veränderungsoptionen (Draxl et al. 2004).

Insgesamt lag der Schwerpunkt der Evaluierungstätigkeit bei Förderinstrumenten, die an der Schnittstelle von Regional- und Innovationspolitik angesiedelt sind und teilweise einen koordinierten Einsatz von Bundes- und Landesmitteln vorsehen. Den höchsten Raumbezug weist dabei die Regionale Impulsförderung auf (RIF 2000): Dieses Programm bündelt die regionalpolitisch relevanten Förderaktivitäten des Bundesministeriums für Verkehr, Innovation und Technologie (BMVIT) für Regionen, Unternehmen und Impulszentren (Technologie-, Gründer- und Innovationszentren). Es wurde 2003 einer Zwischenevaluierung unterzogen (im wesentlichen Vollzugs- und Wirkungsanalyse, gekoppelt mit internationalem Benchmarking), derzeit läuft die Endevaluierung (Technopolis 2005).

Darüber hinaus wurden einige Evaluierungen durch Länder und Regionen durchgeführt, wobei der Schwerpunkt bei der Bewertung von Entwicklungsstrukturen lag. So wurden in einigen Bundesländern (Kärnten, Steiermark, Tirol) spezifische Evaluierungen von Regionalmanagements durchgeführt (Draxl et al. 2004; Draxl/Schuh 2005; ÖSB Consulting 2004; ÖAR Regionalberatung 2005). In Niederösterreich erfolgte eine Evaluierung der Regionalen Innovationszentren (RIZ) in Bezug auf ihre Inkubatorfunktion und die Rolle als Impulsgeber für regionale Entwicklung und Innovation (ARC systems research/ÖAR Regionalberatung 2004). Es gibt unseres Wissens keine – über das regionale Umfeld hinaus bekannten – Evaluierungen von Umsetzung und Wirkung regionaler Entwicklungsprogramme.

EU-Regionalpolitik

Im gesamten Themenfeld „Raumentwicklung, Regional- und Strukturpolitik“ kann der Bereich der EU-Regionalpolitik für sich in Anspruch nehmen, in Österreich jener mit der meist ausgeprägten Evaluationstätigkeit von Programmen zu sein. Dabei ist die in den Strukturfondsverordnungen enthaltene diesbezügliche Verpflichtung sicherlich als eine der treibenden Kräfte zu betrachten. Zu Beginn noch eher als Pflichtübung verstanden, wird dieses Instrument von den Verantwortlichen zunehmend auch über das vorgeschriebene Maß hinaus eingesetzt.

Österreich war in der Periode 2000-2006 an 25 Strukturfondsprogrammen beteiligt und für jedes der Programme wurden jeweils eine Ex-ante-Evaluierung sowie eine Halbzeitbewertung durchgeführt. Ende 2005 wurden die Halbzeitbewertungen wie in den Verordnungen vorgesehen aktualisiert. Auf den dabei gewählten koordinierten und begleitenden Ansatz wird in Punkt 1.5 näher eingegangen. Neben diesen Hauptprogrammen wurden in einigen Bundesländern (z. B. Salzburg, Oberösterreich, Kärnten) auch die EFRE Programme für innovative Maßnahmen einer Evaluierung unterzogen, mit denen u. a. Erkenntnisse für die Gestaltung der neuen Programme für das Ziel „Regionale Wettbewerbsfähigkeit“ gewonnen wurden (z. B. Convelop/Technopolis 2007).

3 Beteiligte, Evaluierungsdesigns und -methoden

Im Bereich der Regionalpolitik sind die Auftraggeber für Evaluierungen nationaler Förderprogramme in erster Linie das Bundeskanzleramt (Abteilung Raumordnung und Regionalpolitik) sowie die einschlägigen Fachministerien (BMVIT und BMWA), in einigen Fällen auch die zuständigen Abteilungen der Landesregierung. Bei den Strukturfonds-Programmen sind die Auftraggeber die (mit Ausnahme von INTERREG) auf Ebene der Landesregierungen angesiedelten Verwaltungsbehörden.

Die Auftragnehmer sind zumeist private Beratungsunternehmen oder außeruniversitäre Forschungseinrichtungen. Der Kreis der in Frage kommenden Anbieter ist relativ überschaubar, kennt sich gut und kooperiert häufig. In Österreich besteht in diesem Bereich eine kleine und spezialisierte „evaluation community“, die dabei disziplinenübergreifend tätig ist (z. B. im Bereich der Innovationspolitik). International agierende Beratungsunternehmen, Universitäten oder Wirtschaftsforschungseinrichtungen sind als Auftragnehmer kaum vertreten.

Internationale Kooperationen und Beteiligungen sind eher die Ausnahme. Sie erfolgen in erster Linie, um Spezialwissen einzubringen oder als externe Sichtweise beziehungsweise internationale Expertise, zum Beispiel mit Blick auf die europäische Perspektive und die Entwicklungen in anderen Mitgliedstaaten. Eine Sonderform stellt die Kooperation mit deutschen oder Schweizer Partnern bei den Programmen grenzüberschreitender Zusammenarbeit dar.

Die Settings sind ausschließlich externe Evaluationen, die bei den Strukturfonds-Programmen vorwiegend begleitend zur Programmumsetzung angelegt sind. Dadurch werden externe Bewertungen durch interne Reflexionsprozesse der beteiligten Programmakteu-

re ergänzt, wobei bewusst unterschiedliche Perspektiven und Sichtweisen eingeholt werden.

Hinsichtlich Methoden handelt es sich im Bereich Strukturfonds zumeist um die Auswertung und Interpretation von Monitoringdaten, ergänzt durch Befragungen von Projekten oder Akteuren. Bei den begleitenden Evaluierungen dominieren ad-hoc-Analysen für spezifische Fragestellungen sowie partizipative Designs mit einem fließenden Übergang zur Prozessmoderation oder zum Coaching der Programmakteure.

Evaluationsstandards haben keinen großen Stellenwert. Gemeinsame Standards werden noch am ehesten durch die Koordination und Abstimmung der Evaluierungsarbeiten im Rahmen der ÖROK erreicht, die bereits bei den Vorbereitungen zu den Ausschreibungen ansetzen und sich bis hin zur gemeinsamen Diskussion der Ergebnisse erstreckt.

Bei den nationalen Förderprogrammen handelt es sich vorwiegend um aggregierte Betrachtungen der einzelnen Förderfälle. Die Ergebnisanalysen werden auf Basis von Primärdokumenten durchgeführt (Förderanträge und Projektberichte), ergänzt durch Befragungen von Förderempfängern, ExpertInnen, intermediären Stellen oder Programmverantwortlichen. Vergleichs- oder Kontrollgruppenansätze werden kaum angewendet.

Bei den beiden Evaluationen der Instrumente endogener Regionalentwicklung wurden innovativere Methoden angewendet: Bei der FER-Evaluation wurde explizit auf die Erschließung von „verdecktem“ Wissen geachtet und im Sinne eines konstruktivistischen Wirklichkeitsverständnisses die Konzeption und Durchführung aus unterschiedlichen Blickwinkeln heraus betrachtet, um zu nützlichen Informationen über das Förderinstrument und seine Auswirkungen zu gelangen. Bei der Evaluierung der Regionalmanagements wurde ein systemischer Ansatz verfolgt, in dem Hypothesenbildung, Erhebungen und Feedbackschleifen mit Regionalmanagements und relevanten Umwelten spiralförmig verknüpft wurden. Neben der Steuerungsgruppe gab es spezifische Reflexionsformate (Workshops, Großgruppen, Feedback durch ausländische ExpertInnen). Der Fokus lag dabei sowohl auf der individuellen Ebene der Regionalmanager als auch auf jener der Regionalmanagement – Organisationen.

4 Good Practice und Einbindung in internationale Diskurse

Als es in der Periode 2000 bis 2006 im Jahr 2002 um die Frage der bestmöglichen Gestaltung der Durchführung der von der Verordnung verpflichtend⁴ vorgeschriebenen Halbzeitbewertungen ging, spielte von Anfang an die Frage der Nützlichkeit sowie das Lernen aller Beteiligter eine entscheidende Rolle. Ausgehend von den in der Periode 1995 bis 1999 gewonnenen Erfahrungen sollte Evaluation in der neuen Programmperiode verstärkt als Lern- und Reflexionsinstrument verwendet werden.

Dafür wurde die Durchführung von *begleitenden Evaluationen* im Rahmen eines koordinierten Ansatzes gewählt. Dieser Schritt, der über das von der Europäischen Kommission vorgeschriebene Mindestmaß hinausging, bedeutete einen Versuch, das verpflichtend vor-

⁴ Artikel 42 der Verordnung (EG) Nr. 1260/1999 des Rates vom 21. Juni 1999 mit allgemeinen Bestimmungen über die Strukturfonds.

geschriebene Instrument Evaluation mit dem größten Nutzen für die Bedürfnisse des Programms anzuwenden. Die Erkenntnisse, die aus der Betrachtung zur Halbzeit und somit auf einen Zeitpunkt bezogen gewonnen worden sind, sollten im Rahmen von begleitenden Evaluierungen für programmspezifische Fragestellungen genutzt werden.

Bei diesen Überlegungen wurde auch das in der EU dominante „Logical Framework Model“ kritisch hinterfragt. Dieses Modell zielt auf die Herstellung einer Gesamtrationalität ab bzw. geht davon aus, dass eine Gesamtrationalität sowie eine Ziel-Mittel-Wirkungslogik existiert, die mit Evaluation überprüft und gefördert werden kann (vgl. Strohmeier/Holzinger 2004: 7).

Aus der Sicht von systemischen Ansätzen sind solche als lineare Abfolgen verstandene Programmabläufe überholt. Regionalentwicklungspolitik ist vielmehr als Prozess zu verstehen, der von den unterschiedlichen Handlungslogiken einer Vielzahl von Beteiligten bestimmt wird und diese zu berücksichtigen hat (vgl. Strohmeier/Holzinger 2004: 7).

Als Steuerungselement wurde eine *Koordinations- und Arbeitsplattform Evaluierung* („KAP-EVA“) eingerichtet, die den Informations-, Wissens- und Meinungsaustausch zwischen Auftraggebern und Auftragnehmern, den Programmen und unterschiedlichen involvierten Ebenen ermöglicht hat. Die Koordinations- und Arbeitsplattform bestand im Kern aus den für die Programme verantwortlichen Verwaltungsbehörden und deren jeweiligen EvaluatorInnenteams und wurde im Bedarfsfall um weitere Akteure erweitert (wie z. B. Förderstellen, Experten).

Der Kommunikationsprozess war dabei auf die gesamte Programmlaufzeit ausgelegt. Während der Zeitraum 2002 bis 2003 primär den Halbzeitbewertungen gewidmet war, standen anschließend die begleitenden Evaluierungen im Mittelpunkt. Dieses Modell ermöglichte die im Rahmen der Halbzeitbewertung gewonnenen Erfahrungen im weiteren Verlauf für programmspezifische Fragestellungen primär der Verwaltungsbehörden zu nutzen und damit den Zeitraum bis zu den verpflichtend vorgeschriebenen Aktualisierungen der Halbzeitbewertungen Ende 2005 zu füllen.

Die Lernprozesse der KAP-EVA fanden auf mehreren Ebenen statt. Einerseits wurden technische Fragen der Programmabwicklung, des Monitorings, etc. behandelt. Überwiegend fand das Lernen auf einer methodisch-theoretischen Ebene statt. Von den horizontalen Themen wie Umwelt und Chancengleichheit über methodische Themen wie das Wirkungsmonitoring bis hin zu thematischen Studien zu Regionalmanagements, Innovationswirkungen und anderen vertiefenden Untersuchungen war dafür ein breites Themenspektrum gegeben (vgl. Strohmeier/Holzinger 2006: 16-17).

Als ein Ergebnis dieser Lernprozesse resultieren neue Überlegungen für die Begleitung und Bewertung komplexer Programme. Beim Ansatz des *Wirkungsmonitoring* werden statt einzelner quantitativer Indikatoren jene Prozesse beobachtet, die für das Erreichen von Zielen und Ergebnissen ausschlaggebend sind. Für diese Prozesse werden zunächst Annahmen in Form von Wirkungsdiagrammen getroffen, deren Eintreffen durch die laufende Auswertung von Informationen (z.B. durch Projektanträge oder Berichte) validiert wird. Damit werden neben den quantitativen auch qualitative (Prozess-)Indikatoren erfasst, und die Programmverantwortlichen verfügen über ein Steuerungsinstrument, das ihnen zeitnah Informationen über das wahrscheinliche Erreichen der Programmziele liefert (Hummelbrunner 2004).

Für diesen Ansatz wurden zunächst die theoretischen Grundlagen erarbeitet und die praktische Umsetzung anhand einiger Pilotanwendungen getestet. Derzeit erfolgt eine breitere Anwendung im Rahmen der Vorbereitung bzw. ex-ante Evaluierung einiger Strukturfondsprogramme. Für die Periode 2007 bis 2013 wurde neben der Ausrichtung der Strukturfondspolitik auf die Lissabon-Strategie von der Europäischen Ebene auch ein verstärkt strategischer Ansatz eingeführt. Als neues Element wurde das strategische Kohäsionskonzept geschaffen, das aus den strategischen Kohäsionsleitlinien auf europäischer Ebene sowie den Nationalen Strategischen Rahmenplänen der einzelnen Mitgliedstaaten (in Österreich: „STRAT.AT“) besteht und den Rahmen für die operationellen Programme bildet.

Auf Grund der guten Erfahrungen wurde auch die Erstellung dieses *Nationalen Strategischen Rahmenplans* 2007 bis 2013 (STRAT.AT) durch eine *Ex-ante Evaluierung* begleitet, obwohl zum Zeitpunkt der Entscheidung dafür noch unklar war, ob es eine diesbezügliche Verpflichtung dazu gibt. Das Ziel der Ex-ante Evaluierung war, in verschiedenen Phasen Feedback auf die Ausarbeitungen des STRAT.AT zu geben, um damit einen Beitrag zur Erhöhung der Stringenz der Planung zu leisten. Die Vorgangsweise sah eine enge Verzahnung mit dem Strategiefindungs- und Ausarbeitungsprozess vor (Joanneum Research/InTeReg 2005).

Die Parallelführung der Ex-ante Evaluierung zur Erstellung des STRAT.AT erlaubte, Ergebnisse und Empfehlungen der Ex-ante Evaluierung direkt in den Strategiefindungsprozess einfließen zu lassen, wobei das Klima von einem konstruktiven Dialog geprägt war. Im Verständnis der Ex-ante EvaluatorInnen bestand die Hauptaufgabe darin, eine unabhängige Beurteilung des STRAT.AT in einem engen Rückkopplungsprozess sicherzustellen, wobei eine zusätzliche externe Außensicht in Form einer internationalen Expertise des European Policy Research Centers (EPRC) eingeholt wurde.

Die Gestaltung als partnerschaftlich-reflexiver Prozess in einem vernetzten Setting mehrerer Beteiligter hat Lernprozesse bei den beteiligten Stakeholdern unterstützt und steigerte damit deren Fähigkeit, die Einbettung und Kohärenz sowie potenzielle Implikationen zu erkennen. Dabei verschaffte die laufende Dokumentation der Empfehlungen die notwendige Transparenz im Zusammenspiel von Strategiefindung und Evaluierung.

Durch die Beteiligung an den einschlägigen Gremien und Veranstaltungen der EU wird eine enge Einbindung in den europäischen Diskurs gewährleistet. Dabei nimmt der „österreichische Weg“ der Evaluation in einigen Bereichen auch eine Pionierrolle ein. So wurde der Ansatz einer begleitenden Evaluierung (nach anfänglicher Skepsis) von der EU Kommission mittlerweile als verpflichtender Standard für die neue Programmperiode festgelegt. Österreich ist einer der wenigen Mitgliedsstaaten, wo die Evaluierung auch bei der Erstellung der neuen strategischen Grundlagen angewendet wird.

Darüber hinaus nehmen österreichische EvaluatorInnen auch an den einschlägigen Tagungen und Arbeitskreisen der DeGEval und damit am fachlichen Diskurs in Deutschland teil (in dieser Hinsicht wirkt sich das Fehlen einer eigenen Evaluierungsgesellschaft in Österreich als Vorteil aus).

5 Stand der Evaluationskultur und Ausblick

Innerhalb weniger Jahre ist es in Österreich gelungen, ohne nennenswerte Evaluierungstradition im Bereich Regionalpolitik eine neue und wirksame Evaluationskultur zu entwickeln. Die zentrale Rolle spielte dabei die Koordinations- und Arbeitsplattform Evaluierung („KAP-EVA“), besonders der damit institutionalisierte Erfahrungs- und Meinungsaustausch zwischen Auftraggebern und EvaluatorInnen. Durch die programmübergreifende Diskussion von Evaluierungskonzepten und -ergebnissen konnte ein gemeinsames Verständnis von Evaluationsprozessen gefördert werden, was wiederum zu einer Qualitätssteigerung führte. Dieses in einer flexiblen Struktur gestaltete Netzwerk wurde auch international positiv bewertet und kann beispielhaft für eine offene Form des Wissenstransfers zwischen den beteiligten Ebenen angeführt werden.

Die verbesserte Evaluierungskultur manifestiert sich auch in effektiven Lernprozessen auf Ebene der evaluierten Programme, sodass Empfehlungen in sehr hohem Ausmaß und relativ rasch umgesetzt wurden. Die durchgeführten Evaluierungen und das damit verbundene Finden von neuen und besseren Methoden gaben wichtige Anstöße für Verwaltungslernen (vgl. Strohmeier/Holzinger 2006: 16).

Es ist noch zu früh um feststellen zu können, in wieweit die positiven Erfahrungen mit lernorientierten Evaluationen im Bereich der Strukturpolitik und die pilothaft erprobten innovativen Methoden auch breiter diffundieren und zu einer Veränderung der Evaluierungskultur im gesamten Bereich der Raum- und Regionalpolitik führen. Durch die überschaubare und stark vernetzte „evaluation community“ bestehen dafür zumindest gute Voraussetzungen.

Die nächste Herausforderung stellt die Gestaltung der begleitenden Evaluierungen der Programme der Periode 2007 bis 2013 dar. Hier gilt es, die gewonnenen Erfahrungen mit der verstärkt strategischen Ausrichtung zu verknüpfen und dabei die flexibleren Vorgaben von europäischer Ebene zum Anlass zu nehmen, einen möglichst großen inhaltlichen Mehrwert für die beteiligten Programmpartner sowohl der nationalen als auch regionalen Ebene zu erzielen.

Evaluation von Raumentwicklung, Regional- & Strukturpolitik in der Schweiz

1 Zur Struktur des Politikfeldes „Raumentwicklungspolitik“ in der Schweiz²

Die räumliche Entwicklung in der Schweiz befindet sich in einem profunden Umbruch. Dieser schlägt sich sowohl in den Politikfeldern nieder, die sich mehr auf die Entwicklungsperspektive der Räume konzentrieren, als auch auf die eigentliche raumplanerische hoheitliche Politik. Raumentwicklungspolitik in der Schweiz nannte sich bis vor kurzem „Raumordnungspolitik“. Obwohl die gleichen Politikfelder gemeint waren, hat in der Übergangszeit der Jahrtausendwende eine Neuorientierung eingesetzt (Thierstein 2002). Auch in diesem Feld ist erkannt worden, dass der traditionelle hoheitliche Durchgriff auf die gesellschaftlichen und räumlichen Strukturen schwächer wurde. Lange Zeit eingesetzte Instrumente sind stumpf geworden, unter anderem weil sich ihr Anwendungsobjekt verändert hat.

1.1 Die „Frühzeit“

Die „Geistesgeschichte“ der Schweizer Regionalpolitik zerfällt in zwei Teile. Zur ersten Generation der Instrumente gehören das aus einer Branchenkrise entstandene Hotelkreditgesetz (1966), das Investitionshilfegesetz (IHG 1974), das Bundesgesetz für die Gewährung von Bürgschaften im Berggebiet (1976) und der Bundesbeschluss zur Förderung wirtschaftlich bedrohter Gebiete (1978). Dies waren Antworten auf den technologischen Wandel und die durch die erste Ölpreiskrise ausgelöste Wirtschaftsrezession Anfang der 1970er Jahre. Das 1970 aufgelegte Bundesgesetz zur Verbesserung der Wohnverhältnisse im Berggebiet ist eng mit dem IHG verbunden. Verwaltungsexterne Evaluationen liegen aus dieser Zeit kaum vor.

Nach 1978 war es lange ruhig, die Instrumente etablierten sich und wurden im Detail nachgebessert. Erst der Beginn der Wirtschaftsrezession in den 1990er Jahren brachte neuen Schwung. Gleichzeitig entwickelte die Europäische Union neue regionalpolitische Instrumente auf der Basis aktueller Erfahrungen und theoretischer Neuerungen zur räumlichen

¹ Ich danke meinem Mitarbeiter Dipl. Geograf Stefan Lüthi für die kompetente Zusammenstellung der neueren Evaluationen zum Thema.

² Zum Begriffsverständnis von „Raumentwicklungspolitik“ vgl. den zwischen den Autoren abgestimmten ersten Abschnitt im Beitrag von Schwab in diesem Band.

Entwicklung – INTERREG, LEADER, URBAN und andere – es entstand Druck von innen und von aussen. Mit der Beteiligung an der EU-Gemeinschaftsinitiative INTERREG II entstand 1995 eine zweite Generation an Instrumenten, bestehend neben INTERREG aus dem an das LEADER-Programm der EU angelehnte RegioPlus und dem sektoral-orientierten Tourismusförderungsinstrument InnoTour (beide 1997). Parallel wurden auch der Bundesbeschluss zur Förderung wirtschaftlich bedrohter Gebiete sowie das IHG vor dem Hintergrund der veränderten Förderlogik revidiert. Zu dieser Logik gehören die zeitliche Begrenzung der Instrumente und die Verpflichtung zur Evaluation.

1.2 Der „Perspektivenwechsel“

Die Raumordnungspolitik verstand sich lange Zeit als Engpass- oder Schwachstellen-Politik, gerichtet auf strukturschwache, eher periphere Räume sowie den ländlichen Raum und das Berggebiet. Eine grundlegende Neuorientierung der Regionalpolitik setzte in den 1990er Jahren ein. Ursachen waren unter anderem eine Ernüchterung über die bescheidenen Erfolge der bisher stark engpass- und ausgleichsorientierten Politik sowie die im Zuge der Liberalisierung erneuerte Grundsatzfrage nach der Legitimität staatlicher Eingriffe in die Wirtschaftsentwicklung. Diese Feststellung gilt sowohl für die Regionalpolitik der OECD und der EU als auch für die Schweiz (OECD 2002; Thierstein/Behrendt 2001).

1.3 Ein Paradigma gewinnt Fahrt

Regionalpolitik ist neu ausgerichtet auf alle Regionen, nicht nur die benachteiligten. Es geht darum, alle Regionen in die Lage zu versetzen, dass sie ihre Entwicklungspotentiale maximieren können. Auf 'künstlich' induzierte Entwicklung soll verzichtet werden. Investitionen sollten auf Grund von positiven Ertragserwartungen in einer Region getätigt werden und nicht um von grosszügigen staatlichen Anreizen zu profitieren. Dies bedeutet aber keinen abrupten Verzicht von Unterstützungsleistungen. Die schwächsten Regionen und Kantone profitieren weiter vom bundesstaatlichen Finanzausgleich und bestimmten finanziellen Anpassungen, um bestehenden Verpflichtungen nachzukommen. Es geht allerdings nicht mehr darum, in allen Regionen das gleiche Mass an Infrastruktur aufrechtzuerhalten. Vielmehr kommt es darauf an, vorteilhafte Rahmenbedingungen für die Unternehmensentwicklung zu sichern. Da eine solche Entwicklung von den Charakteristika der betreffenden Regionen abhängt, sollte eine entsprechende Infrastrukturpolitik auf Regionstypen aufbauen.

Das Bewusstsein ist gewachsen, dass die bisher kleinmassstäblich verstandene regionale Strukturpolitik mit dem relativ geringen Mitteleinsatz wenig ausrichten kann. Andere Sektoralpolitiken, wie die Verkehrs-, die Energie-, Landwirtschafts- oder neuerdings die Bildungspolitik bewegen weit mehr staatliche Mittel mit raumrelevanten Auswirkungen. Die Liberalisierung der Märkte sowie des Welthandels, die wachsende Integration von exportorientierten Unternehmen in globale Standortnetze sowie die Liberalisierung von öffentlichen Dienstleistungen beförderten einen Gesinnungswandel (Thierstein et al. 2006).

1.4 Das Bundesgesetz über die Regionalpolitik

Eine Expertengruppe hatte es unter der Federführung des Staatssekretariats für Wirtschaft (seco) unternommen, die Regionalpolitik neu zu konzipieren (Expertenkommission 2003). Das Gremium war mehrheitlich der Ansicht, dass trotz des vorliegenden Bundesbeschlusses zur Neugestaltung des Finanzausgleichs und der Aufgaben zwischen Bund und Kantonen (NFA) eine explizite Regionalpolitik auch in Zukunft notwendig sei. Die Neue Regionalpolitik (NRP) nimmt das oben skizzierte Paradigma teilweise auf und zielt auf die Stärkung der Wettbewerbsfähigkeit und der regionalen Wertschöpfung. Dieses Ziel soll über drei strategische Ausrichtungen erreicht werden (dabei wird die erste durch die beiden weiteren flankiert):

- Förderung von Initiativen, Programmen und Projekten zur Stärkung der Innovationskraft, Wertschöpfung und Wettbewerbsfähigkeit der Regionen,
- Stärkung der Kooperation und die Nutzung von Synergien zwischen Regionalpolitik und raumwirksamen Sektoralpolitiken des Bundes,
- Aufbau und Betrieb eines integralen Wissenssystems über die Regionalentwicklung, die Aus- und Weiterbildung für regionale Akteure und die systematische Evaluation der Wirksamkeit.

Zentrale Reform im Sinne von Governance ist ein mehrjähriges Umsetzungsprogramm für den Zeitraum von 2007 bis 2013, das an Leistungsvereinbarungen mit den Kantonen gekoppelt ist. Diese Ausführungen lassen erkennen, welche Anforderungen an eine problemorientierte Evaluationsstrategie der verantwortlichen Politikbereiche gestellt sind. Es fehlen zu den meisten oben aufgeführten Entwicklungen Ex-post oder Ex-ante Evaluationen.

2 Evaluationsbezogener Kontext – Rechtlicher Rahmen und Zuständigkeiten

Die Bedeutung des Begriffs „Raumentwicklungspolitik“ definiert sich entlang der Einteilung der politischen Ressorts in der Schweiz. Die Raumentwicklungspolitik im engeren Sinne ist in der Schweiz der Oberbegriff für die vom Bundesamt für Raumentwicklung verwaltete Raumplanung und die Regionalpolitik, die beim seco als Teilgebiet der Standortförderung angesiedelt ist. Die Raumentwicklung im engeren Sinne macht einen sehr geringen Teil der Transferzahlungen des gesamten Bundeshaushaltes aus. Aufwendungen für Umwelt und räumliche Entwicklung wurden für das Jahr 2003 mit rund 2,8 Prozent der gesamten Transfersumme budgetiert. Davon entfallen 0,6 Prozent auf die Aufgaben der räumlichen Entwicklung, inklusive dem Schutz der ländlichen Attraktivitätswerte und der Aufwendungen für Investitionshilfe für Berggebiete (OECD 2002: 132).

Die rechtliche Verankerung der Regional- und Raumplanungspolitik hat in der Schweiz keine lange Tradition. Die Raumplanung wurde erstmals 1969 in der Schweizer Bundesverfassung verankert. 1980 folgte das heute noch gültige Bundesgesetz über die

Raumplanung (RPG). Die Regionalpolitik erhielt 1974 ihre erste explizite Rechtsgrundlage mit dem Bundesgesetz über die Investitionshilfe für das Berggebiet (IHG). Erst das auf den 1. Januar 2008 in Kraft getretene neue Bundesgesetz über die Regionalpolitik vereint die regionalpolitischen Fördermassnahmen in einem konsistenten Gesetzeserlass.

Das Raumplanungsgesetz beruht in seiner inhaltlichen Ausrichtung auf dem Verfassungsgedanken von 1969 – Orientierung an der landwirtschaftlichen Bodennutzung und der kriegswirtschaftlich geprägten Vorsorge. Mit Ausnahme kleiner Revisionen steht erst jetzt eine grundlegende Überarbeitung des Gesetzes an. Raumplanung in der Schweiz ist gekennzeichnet durch eine ausgesprochen starke kommunale Planungshoheit.

Die Schweizer Evaluationspraxis wird neuerdings geprägt durch die Verpflichtung zur Evaluation von Massnahmen des Bundes (vgl. Mader in diesem Band). Die Schweizer Beteiligung an INTERREG hat das Land mit der in der Strukturfonds-Verordnung der EU geforderten Zwischen- und Schlussevaluationen vertraut gemacht.

3 Überblick über vorliegende Evaluationen der Raumentwicklung, Regional- und Strukturpolitik

Evaluationen in der Regional- und Raumplanungspolitik waren lange Zeit keine gängige Praxis. Eine der Ausnahmen bildet die Evaluation aus dem Jahre 1992 des ordnungspolitisch sehr umstrittenen Bundesbeschlusses zur Förderung wirtschaftlich bedrohter Gebiete (Boulianne/Rey 1992). Erst die Schweizer Beteiligung an der EU-Gemeinschaftsinitiative INTERREG II ab Mitte der 1990er Jahre sowie die jahrelange Diskussion um eine neue Regionalpolitik haben zu einer Systematisierung und Intensivierung von Wirksamkeitsuntersuchungen geführt. In diesem Beitrag sind Evaluationen wichtiger raumrelevanter Politikfelder berücksichtigt, die ab der Jahrtausendwende durchgeführt wurden. Bewusst zählen dazu einige Arbeiten zur Verkehrs-, Siedlungs- und Agglomerationspolitik. Die räumlichen Auswirkungen dieser Fachpolitiken sind in der Regel erheblich tief greifender als es die eigentliche Regional- und Raumplanungspolitik je erreichen könnte. Die meisten vorliegenden Arbeiten konzentrieren sich auf die Regionalpolitik; sie sind externe Evaluationen und als Ex-Post Evaluationen auf Bundesebene angelegt.

3.1 Die Regionalpolitik

Mit der EU-Gemeinschaftsinitiative INTERREG setzt für die Schweiz die Notwendigkeit zur systematischen Evaluierung ein. Die gesamte Schweizer Programmbeteiligung wird in Zwischen- und Schlussevaluationen untersucht. In gleicher Weise wird mit den einzelnen Programmen der regionalen Teilräume von INTERREG vorgegangen (Behrendt et al. 2003; Devouassoux/Pfister Giauque 1998; Schleicher-Tappeser et al. 1999; Transversal/SEREC 2000). Im Hinblick auf die jeweils kommende Programmperiode fordert die EU-Kommission eine Ex-ante Evaluation des neu aufzustellenden Operationellen Programms der einzelnen Förderregionen (Scherer et al. 2001).

Die erste Schlussevaluation der Schweizer Beteiligung an INTERREG II erfolgte 1999 (Thierstein et al. 1999). Sie stellt fest, dass die finanzielle und operationelle Unterstützung den Ausbau der grenzübergreifenden Zusammenarbeit massgeblich vorangetrieben hat. Dies zeigt sich besonders in den Regionen mit einer weniger langen Tradition in der Zusammenarbeit. Im Hinblick auf die Integrationspolitik deckt sich das Programm INTERREG II mit der Absicht des Bundes, die Akteure an der schweizerischen Grenze bei der Zusammenarbeit mit ihren europäischen Nachbarn zu unterstützen. Diese Aussage wird in der Schlussevaluation von INTERREG III gestützt, die neu auch die räumlich ausgreifende transnationale Zusammenarbeit untersucht hat (Schnell/Pfister Giauque 2006; von Stokar et al. 2003). INTERREG II und III leisteten einen wichtigen Beitrag zur grenzüberschreitenden und transnationalen Zusammenarbeit und haben die Grundidee bei den öffentlichen Partnern gut etabliert. In INTERREG III hat sich auch die Beteiligung der Binnenkantone verstärkt. Es konnte ein Informations-, Abstimmungs- und Lernprozess unter den teilnehmenden Grenzregionen in Gang gesetzt werden (Schnell/Pfister Giauque 2006; von Stokar et al. 2003).

Bezüglich seiner Wirkungen wird INTERREG III als Querschnittsprogramm mit Schwerpunkt (Mikro-) Integration bewertet. Der primäre Effekt von INTERREG ist ein Kennenlernen und das Schaffen von Vertrauen. Wo dieses vorhanden ist, kann INTERREG auch ein regionalwirtschaftliches oder raumordnungspolitisches Instrument sein. Die integrationspolitische Komponente von INTERREG nimmt eine grosse Bedeutung für die territoriale Kooperation allgemein ein und ist somit eine Grundlage für die Erreichung regionalpolitischer oder raumordnungspolitischer Ziele (Schnell/Pfister Giauque 2006). Dem Förderinstrument INTERREG wird eine gute vertikale Abstimmung unter den Programmverantwortlichen und Projektträgern zugebilligt. Damit hat sich INTERREG als ein Experimentierfeld für grösserräumige territoriale Kooperation von Schweizer Kantonen erwiesen.

Übereinstimmend räumen die Evaluatoren die Schwierigkeit ein, die sehr abstrakten und kaum operationalisierten Ziele der Integrationspolitik und Regionalentwicklung genauer zu „messen“. Häufig ist der Kausalitätszusammenhang eines individuellen Projektes mit dem übergeordneten Ziel schwer herzustellen. Zum zweiten sind im Grunde die nicht direkt als Förderziele erzeugten Wirkungen wie ‚Vertrauen zwischen grenzübergreifenden Institutionen‘ oder ‚Kennenlernen anderer Verwaltungskulturen‘ als die hauptsächlichen und längerfristig wirksamsten Effekte zu werten.

In den Jahren 2002 bis 2004 wurden zu einzelnen Förderinstrumenten begleitende oder Ex-post Evaluationen vorgelegt, deren Hauptaussagen in einem Gutachten für die bundesrätliche Expertengruppe zur Neugestaltung der Schweizer Regionalpolitik vorweggenommen wurden (Thierstein/Behrendt 2001). Das harte Ringen der bisherigen Vorteilsnehmer in den Berggebieten und ländlichen Räumen hatte es dem federführenden seco aber geraten erscheinen lassen, nachträglich eine ausführliche Ex-post Evaluation vorzulegen.

Das dienstälteste Schweizer Förderinstrument der Regionalpolitik – das IHG – wurde letztmals im Jahre 2004 evaluiert (Bieger et al. 2004). Die Evaluation hält fest, dass die Regionalpolitik im Berggebiet als Teil der staatlichen Wirtschaftspolitik zu begreifen ist und die Investitionshilfe auf wirtschaftliche Wachstumsimpulse im Sinne von Entwicklungsinfrastrukturprojekte und auf unternehmensnahe Basisinfrastruktur zu orientieren sei. Die kleinräumige Regionalpolitik soll weitergeführt und die regionalen Organisationsstrukturen im Sinne der neuen Regionalpolitik erhalten werden. Die bisher sehr kleinteilig geschnittenen

Förderregionen sollten nach Funktionsräumen und zentralen Orten neu abgegrenzt werden. Das Konzept der ‚Lernenden Region‘ (Florida 1995; Thierstein et al. 2000) sollte im Berggebiet kombiniert werden mit einem wissensorientierten Ansatz des Regionalmanagements, das die teilweise unbefriedigende Leistung der Regionalsekretariate verbessern helfen soll (Bieger et al. 2004).

Für das lernprozessorientierte Förderinstrumente RegioPlus liegt die Zwischenevaluation vor (Bättig et al. 2002). Danach sind in RegioPlus kaum überprüfbare Wirkungsziele zu finden. Das neue Bundesgesetz zur Regionalpolitik enthält ebenfalls recht breit gefasste Ziele. Diese machen es gemäss einer Studie im Auftrag des seco nicht „glaubwürdig nachvollziehbar“, regionalwirtschaftliche Wirkungen regionalpolitischer Programme mit einem Indikatorenset abbilden zu können. Diese erhofften Wirkungen lägen im nicht-messbaren Bereich. Eine Abbildung der intangiblen Wirkungen im Sinne eines Indikatorensets mit den bestehenden Anforderungen der Datenverfügbarkeit und -vergleichbarkeit sei nicht möglich (EBP 2007: 40).

Methodisch zu einem ähnlichen Fazit gelangt die Evaluation des Bundesbeschlusses zugunsten wirtschaftlicher Erneuerungsgebiete eines zweiten zentralen politischen Förderinstrumentes (von Stokar et al. 2004). Dieses war aufgrund seiner einzelbetrieblichen Förderkomponente stets ein Dorn im Auge ordnungspolitisch argumentierender Kritiker. Die Evaluatoren haben trotz der oben vorgebrachten Kausalitäts- und Messbarkeitsprobleme versucht, mit einer Kombination von sekundärstatistischer Auswertung, multivariater statistischer Analyse und Interviews, ein differenziertes Urteil über die Wirkungen zu erarbeiten. Es fällt ernüchternd aus. Der Bundesbeschluss zeichnet sich durch hohe Mitnahmeeffekte sowie hohe Opportunitätskosten bei den Steuererleichterungen des Bundes aus. Insgesamt ist ein mässiges Kosten-Wirksamkeits-Verhältnis in Bezug auf die tatsächlich gehaltenen oder geschaffenen Stellen zu erkennen (von Stokar et al. 2004).

Die Ergebnisse der neueren Evaluationen zur Schweizer Regionalpolitik ergeben ein ähnliches Fazit, das wie oben erwähnt bereits 2001 in einer Zielerreichungsüberprüfung dieses Politikfeldes vorgelegen hat (Thierstein/Behrendt 2001). Die Evaluatoren hatten damals vorliegende Evaluationen ausgewertet (neben bereits zitierten Arbeiten unter anderem Deschenaux/Erlanger 1998; Smeral/Keller 2000; Thierstein et al. 1999). Die späteren, ausführlicheren Evaluationen zur bisherigen Regionalpolitik stiessen den wichtigen Lerneffekt an, dass das seco Pilotprojekte im Hinblick auf die NRP initiierte. Eine verwaltungsinterne Zwischenevaluation kommt zum Schluss, dass diese Pilotprojekte in einigen Kantonen eine Dynamik und Auseinandersetzung mit der Neuen Regionalpolitik ausgelöst haben (Egli 2006).

3.2 Die Raumplanungspolitik

In der Raumplanungspolitik gab es bis vor kurzem keine einzige Evaluation, weder als Vollzugs-, Zielerreichungs- noch als Wirkungsuntersuchung. Erst Anfang 2007 wurde im Hinblick auf die Überarbeitung des Raumplanungsgesetzes (RPG) die Evaluation einer internationalen Expertengruppe mit „Beobachtungen und Anregungen“ zur Raumplanung und Raumentwicklung vorgelegt (Scholl et al. 2006). Angesichts der weit reichenden Conse-

quenzen des RPG muss die Evaluation methodisch und inhaltlich als problematisch bezeichnet werden. Gespräche mit Schlüsselpersonen alleine können keine quantitative und zeitreihenbezogene Analyse ersetzen; ebenso fehlt jegliche Governance-Analyse oder Vollzugsuntersuchung. Dies wiegt umso schwerer, als räumliche Nutzungsplanung in der Schweiz eine Verbundpolitik darstellt, über deren Wirksamkeit noch immer kaum etwas Quantitatives und Strategisches bekannt ist.

3.3 Die Agglomerationspolitik

Städte und Agglomerationen als Politikgegenstand wurden in der Schweiz vom Bund erst Ende der 1990er Jahre „entdeckt“ (Bundesrat 1999). Die Wirkungsabschätzung von neun Agglomerations-Modellvorhaben und deren Auswirkungen auf die wirtschaftliche Entwicklung stellt die erste Studie über einen Teilbereich der Schweizer Agglomerationspolitik dar (Thierstein et al. 2004). Kurz darauf folgte ein verwaltungsinterner Zwischenbericht über die Schweizer Agglomerationspolitik. Darin wird festgestellt, dass die Lancierung der Agglomerationspolitik in vielen Agglomerationen und ihren zugehörigen Kantonen eine positive Dynamik ausgelöst habe. Wo diese Dynamik schon vorher bestand, konnte sie der Bund positiv unterstützen und fördern. Die Agglomerationspolitik des Bundes brauche deshalb zurzeit keine grundsätzliche Richtungskorrektur (ARE/seco 2006).

3.4 Weitere raumrelevante Sektoralpolitiken

Zahlreiche Sektoralpolitiken wirken wie vorne angesprochen massgeblich auf den physischen Raum ein. Hier werden exemplarisch einige ausgewählte Beispiele vorgestellt. Die Wechselbeziehungen zwischen Siedlungsentwicklung und Verkehrsinfrastruktur prägt die Raumentwicklung stark. Das Bundesamt für Raumentwicklung (ARE) liess dazu ein Wirkungsmodell entwickeln und beauftragte externe Evaluatoren, vier Ex-post Fallstudien zu erarbeiten. Ausgewählt wurden die S-Bahn Zürich als urbanes System des öffentlichen Verkehrs (Güller et al. 2004), der Autobahntunnel Vue-des-Alpes (von Fischer et al. 2006), der Eisenbahntunnel Vereina (Aliesch et al. 2006) sowie die Strassenverkehrsinfrastruktur in der Magadinoebene (Giacomazzi et al. 2004). Die vier Evaluationen weisen allesamt unterschiedliche Wirkungen von Verkehrsinfrastruktur auf Siedlung und Raum nach. Gemeinsam ist ihnen die Feststellung, dass die komplexe Kausalität mit der eingesetzten Methodik eines konzeptionellen Wirkungsmodells nicht abschliessend beantwortet werden könne. Bislang ist dies der einzige Versuch, komplexe Phänomene von Siedlungs-, Verkehrsinfrastruktur- und Wirtschaftsentwicklung sowie öffentlichem Finanzhaushalt systematischer auszuleuchten.

3.5 Die Auftraggeber und die Auftragnehmer

Die Auftraggeber lassen sich an einer Hand abzählen: das Staatssekretariat für Wirtschaft (seco), das Bundesamt für Raumentwicklung (ARE) sowie programmseitig Verantwortliche von INTERREG. In einzelnen Fällen liegen Evaluationen vor, die durch grössere Kantone – etwa Zürich zum Flughafen (Peter et al. 2005) – oder durch Fachvereinigungen der Regionsinteressen beauftragt wurden (Vodoz et al. 2006). Als Auftragnehmer tritt ebenfalls eine kleinere Gruppe von einerseits sozialwissenschaftlich ausgerichteten, privaten Beratungsbüros auf – mehrheitlich in der Deutschschweiz angesiedelt – und andererseits anwendungsorientierten Hochschulinstitutionen – zum Beispiel an der Universität St. Gallen (IDT-HSG) oder an der Eidgenössischen Technischen Hochschule Lausanne (C.E.A.T.). Auffallend, dass bislang keine im Ausland angesiedelten Evaluations- und Beratungsfirmen am Markt aufgetreten sind. Der enge persönliche Kontakt zwischen Schweizer Evaluatoren und Auftraggebern sowie die meist geringen Honorarvolumen – die keine Ausschreibung nach WTO-Richtlinien nötig machen – erzeugen wohl eine marktabstottende Wirkung.

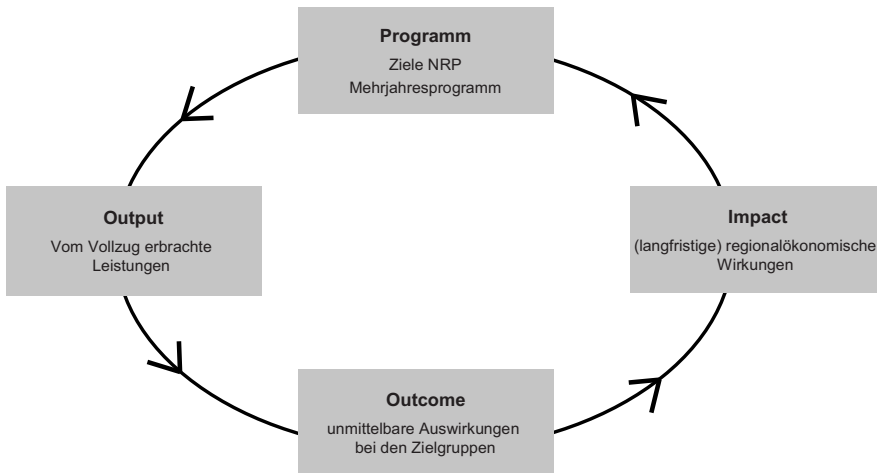
4 Good practice und Einbindung in internationale Diskurse

Im Bezug auf Evaluationssettings und Methoden sehen wir in der Raumplanung mangels systematischer Arbeiten keine Trends. Die Evaluationen aus der Regionalpolitik oder Standortentwicklungspolitik widerspiegeln den vorne geschilderten inhaltlichen Paradigmenwechsel. Die regionalpolitischen Instrumente der „ersten Generation“ konzentrierten sich auf den Ausbau der Basis-Infrastruktur und das Senken von Kapitalkosten.

Die Erfahrungen mit der ersten Generation regionalpolitischer Instrumente, die Erkenntnisse zur Vielschichtigkeit regionaler Entwicklung und die Einsicht, dass die Wirkungen der einzelbetrieblichen und Infrastrukturförderung nur sehr bescheiden waren, haben die Grenzen der finanziellen Förderung einzelner Regionen und Unternehmen gezeigt. Sowohl in den anderen OECD-Ländern als auch in der Schweiz wurde mit einer Veränderung der Instrumente auf die zunehmende Komplexität und Geschwindigkeit der Wirtschaftsentwicklung reagiert. Neben einer Veränderung der Ziele und Mittel sind alle regionalpolitischen Instrumente der zweiten Generation zeitlich begrenzt, branchenübergreifend, lern- und partizipationsorientiert ausgerichtet. Entsprechend schwieriger sind solche eher „weichen“ Instrumente auf ihre Wirkungen hin zu überprüfen.

Zahlreiche aktuellere Evaluationen im Kontext der neuen Regionalpolitik arbeiten explizit mit Ziel-, Leistungs- und Wirkungsindikatoren sowie einem konzeptionellen Wirkungsmodell. Diese abstrakte Vorstellung hilft einen stringenten Zusammenhang aufzuzeigen (Abbildung 1) zwischen den eingesetzten Fördermitteln (Input), den Projektergebnissen (Output), die von den Zielgruppen genutzt werden (Outcome) und die dadurch längerfristige Wirkungen erzielen, die über das Projekt hinaus reichen (Impact).

Abbildung 1: Konzeptionelles Wirkungsmodell für die Regionalpolitik der Schweiz



Wirksamkeits- und Wirkungsuntersuchungen spielen aufgrund des Paradigmenwechsels in der neuen Regionalpolitik eine wachsende Rolle. Die NRP lässt den Kantonen wesentlich mehr Spielraum, bedeutet aber für den Bund eine grosse Herausforderung bezüglich seiner Förderpolitik. Das neue Bundesgesetz sieht vor, dass der zielgerichtete und koordinierte Mitteleinsatz in Zukunft verstärkt über Controlling-, Monitoring- und Evaluationsmassnahmen sichergestellt werden muss. Für diese Wirksamkeitsüberprüfung erarbeitete das Staatssekretariat für Wirtschaft (seco) bis Mitte des Jahres 2007 ein Set von Indikatoren zur Begleitung der verschiedenen Umsetzungsschritte (EBP 2007). Dieser kombinierte Instrumenteneinsatz deutet darauf hin, dass sich auch in der Schweiz eine internationale Evaluationspraxis festsetzt, die wesentlich auf die Erfordernisse der EU-Strukturfondspolitik zurückgeht (Bachtler/Wren 2006).

5 Stand der Evaluation

Die Kleinheit der Schweiz bringt es mit sich, dass sich die meisten in der Raumentwicklungspolitik tätigen Evaluatoreninnen und Evaluatoren wiederkehrend bei unterschiedlichen Gelegenheiten treffen, ohne dass eine fest gefügte „Evaluation Community“ vorhanden wäre. In der Bundesverwaltung existiert seit einigen Jahren ein locker gestricktes, themenübergreifendes „Netzwerk Evaluation“. Zahlreiche Evaluatoren sind Mitglied in der Schweizerischen Evaluationsgesellschaft SEVAL (www.seval.ch), die in ihrer Evaluatoren-Datenbank auch ein Tätigkeitsfeld „Raumordnung“ führt. Raumbezogene Spezialisten treffen sich in der Schweizerischen Studiengesellschaft für Raumordnung und Regionalpolitik (www.rorep.ch). Doch in keinem mir bekannten Falle sind die Evaluatoren ausschliesslich in der Raumentwicklung tätig. Viele Evaluationsnehmer sind sowohl in der generellen Poli-

tikberatung als auch in fachbezogener Auftragsforschung tätig, was in der Schweiz Garant zu sein scheint für disziplinenübergreifende Vernetzung und befruchtenden Austausch.

Die Evaluationsstandards der SEVAL (Widmer et al. 2000) dürften vielen Evaluatorinnen und Evaluatoren bekannt sein, Anwendungsstandard oder fester Bestandteil von Ausschreibungen sind sie nicht. Die hier verwendeten Evaluationen nehmen in der Regel nicht explizit auf die SEVAL-Standards Bezug.

Ein Wermutstropfen im Urteil über den heutigen Zustand rührt von der Nicht-Mitgliedschaft der Schweiz in der EU her. Der dort übliche, hoch stehende Austausch zwischen hohen Beamten und der angewandten Wissenschaft findet zu oft ohne Schweizer Beteiligung statt. Dies mag sich hemmend auf die Diffusion neuer inhaltlicher und methodischer Entwicklungen in der Evaluationspraxis auswirken.

Eine neue Generation von gut ausgebildeten Hochschulabsolventinnen und -absolventen mit zum Teil internationaler Erfahrung verleiht der öffentlichen Verwaltung sowohl auf Bundes- wie auf Kantonsebene einen willkommenen Qualifizierungsschub. Die Evaluationskultur kann durch das Wechselspiel zwischen anspruchsvoller Nachfrage und anspruchsvollem Angebot nur gewinnen.

Einschränkend sind die oft geringen Budgets für Wirkungsuntersuchungen zu nennen. Aufgrund der Vergaberichtlinien werden häufig auch wichtige Themen unter der Schwelle von Franken 50'000 gehalten, damit Direktvergaben oder eingeschränkte Ausschreibungen möglich sind. Dies bringt vor allem private Büros aus Interesse an der Thematik dazu, Angebote unter den Selbstkosten einzureichen. Auf die Dauer kann es für Auftraggeber wie für Auftragnehmer nicht befriedigend sein, wenn Pflichtenheft und Budget im Missverhältnis stehen.

6 Bilanz

In der schweizerischen Raumentwicklungspolitik sind Evaluationen von Programmen und Instrumenten zum Standard geworden. Die seit den späten 1990er Jahren laufende Diskussion über eine Neuorientierung der Schweizer Regionalpolitik hat das Bewusstsein geschärft – zusammen mit dem Verfassungsauftrag zur Wirksamkeitsüberprüfung von Bundesmassnahmen – dass Evaluationen fester Bestandteil des Politikzyklus sind. Den in der NRP formulierten Zielen mangelt es an inhaltlicher Operationalisierbarkeit. Dies ist zwar nicht neu, macht jedoch die Aufgabe zur Entwicklung eines für Controlling, Monitoring und Evaluation geeigneten Indikatorenkatalogs nicht einfacher (EBP 2007).

Methodisch zeichnet sich eine Anlehnung an gängige europäische Vorgehensweisen in den Struktur- und Regionalpolitikevaluationen ab. Neuland beschreitet man allerdings überall mit der Tatsache, dass Raumentwicklung zunehmend auf institutionelle Lernprozesse und intangible Ziele ausgerichtet ist. Es gilt weiter auszuleuchten, welche Konsequenzen dies für das Design von formativen Evaluationen sowie den Einsatz von Prozess- und qualitativen Indikatoren entfaltet.

In der Raumentwicklungspolitik liegt der Schwerpunkt der bislang vorliegenden Evaluationen auf Ex-Post Untersuchungen zur Regionalpolitik. Raumplanung, die Dynamiken urbaner Räume, die Überlagerung von funktional verflochtenen Handlungsebenen auf un-

terschiedlichen räumlichen Massstäben sowie eine inhaltlich breiter gefasste Standortentwicklungspolitik hat bislang vom hier beschriebenen ‚Evaluationstrend‘ noch wenig gespürt. Der Begriff ‚Raumentwicklung‘ macht deutlich, dass ein Prozess im Gange ist, der eine Vielzahl von Akteuren auf unterschiedlichen räumlichen Ebenen einschliesst. Zentral ist dabei die Erkenntnis, dass Politik hoheitlich-territorial gebunden ist, während die Systeme der überwiegend privaten Leistungserstellung meist funktional orientiert sind. Hoheitliches und privates Handeln müssen daher in ihrem wechselseitig bedingten, funktionalen Wirkungsgefüge gesehen und analysiert werden.

Die Neue Regionalpolitik stellt zwar die regionale Wettbewerbsfähigkeit in den Mittelpunkt des Fördergeschehens. Räumlich bleibt man hingegen bei den bisherigen Vorteilsnehmern. Die räumliche Dynamik spielt heute in hoch verdichteten urbanen Räumen und in Metropolregionen (Thierstein et al. 2006). Für diese Räume gibt es keine nennenswerte Fachpolitik, die landläufig zur Raumentwicklungspolitik mit gezählt würde. Aber genau dieses räumliche Zusammenwirken von Fachhochschul(standort)-, Innovations- und Standortentwicklungspolitik schafft heute Voraussetzungen, damit die kleine, offene Volkswirtschaft im internationalen Standortwettbewerb mithalten kann.

Die Evaluation räumlicher Entwicklung im Vergleich der drei Länder

1 Kontext

In Deutschland, Österreich und der Schweiz sind Bund, Länder/Kantone und Gemeinden seit langem bestrebt, die räumliche Entwicklung zu planen, zu ordnen und wo nötig zu steuern: in Wachstumsgebieten durch lenkende Eingriffe, in wirtschaftlich schwachen Gebieten durch Ausgleichs- und Fördermassnahmen. Globalisierung der Wirtschaft, absehbare demographische Einbrüche und das Thema Nachhaltigkeit haben in allen drei Staaten eine Diskussion über die künftige Ausrichtung der räumlichen Entwicklung und den Einsatz der verfügbaren Ressourcen ausgelöst. In der Schweiz hat dies zur Neuausrichtung der Regionalpolitik und – zumindest in Ansätzen – der Raumentwicklungspolitik geführt. Ähnliche Bestrebungen laufen auch in den anderen Ländern (vgl. z. B. Sinz 2006).

Raumordnung und Raumplanung liegen in der Schweiz und Österreich in der Kompetenz der Länder/Kantone und Gemeinden. Gleichwohl haben beide Länder gesamtstaatliche Leitbilder. In Deutschland hat der Bund dagegen trotz Kompetenz seit langem kein Bundesraumordnungsprogramm mehr ausgearbeitet. In allen drei Ländern beeinflussen die Sektoralpolitiken des Bundes gleichwohl die räumliche Entwicklung erheblich.

Bei der Regionalpolitik sind alle drei Ebenen involviert, die Gemeinden meist über Gemeindeverbände. Viele Programme sind Verbundprogramme, die aus mehreren Töpfen auf den drei Verwaltungsebenen gespeist werden und damit den Eigengesetzlichkeiten der „Politikverflechtung“ (Scharpf et al. 1976; 1980) unterliegen.

In den drei Ländern gibt es Standardprogramme zur regionalen Wirtschaftsförderung, daneben experimentelle und Modell-Programme. In Deutschland und Österreich kommen Projekte im Rahmen des europäischen Strukturfonds hinzu. Über die Interreg-Programme und -projekte ist auch die Schweiz daran beteiligt. Hier unterliegen alle drei Länder der von der EU vorgegebenen Evaluationspflicht und den damit verbundenen Routinen wie Halbzeit-, Schluss- und ex-ante-Evaluationen.

In der Schweiz hat der „Neue Finanzausgleich“ (Schweizerischer Bundesrat 2003; Rey 2000) zur Entflechtung der Geldflüsse und einer Klärung der Zuständigkeiten von Bund und Kantonen (sowie Städten) geführt. In Österreich stellt die Österreichische Raumordnungskonferenz (ÖROK) ein wichtiges Bindeglied zwischen den drei Ebenen dar. In Deutschland führt die gewachsene Aufgabenverteilung zwischen Bund (Gesetzgebung mit Länderbeteiligung) und Ländern (Verwaltung) zu einer gewissen Wirkungsferne der Bundesadministration (Benz 2005).

2 Die Evaluation von Raumplanung und Raumentwicklungspolitiken

Schon die wechselnd verwendeten Begriffe „Raumplanung“, „Raumordnung“, „Raumentwicklung“ zeigen, dass dieser Politikbereich nicht so leicht zu fassen ist. Die wenigen Evaluationen in diesem Politikbereich konzentrieren sich auf die Auswirkungen spezifischer Projekte (z. B. Einkaufszentren, Infrastrukturbauten) und städtebaulicher Modellvorhaben auf die räumliche Entwicklung. Die Instrumente von Raumplanung und Raumordnung und deren Wirksamkeit werden dagegen nicht thematisiert. Allein in Österreich wurde auch Inhalt und Vorgehensweise beim Raumordnungsprogramm evaluiert. Überwiegend beschränkt man sich auf (wertende) Raumb Beobachtung.

Die geringe Zahl und spezifische Ausrichtung von Evaluationen scheint systemisch bedingt. Raumplanungen sind Konditionalprogramme mit unvollständigen Wirkungsketten. Sie koordinieren relevante Aspekte ex ante und schaffen Rahmen- und Genehmigungsbedingungen und allenfalls Anreize für andere Sektoralpolitiken oder Private. Sie erzielen Wirkungen erst über die Investitionsentscheide und die Ressourcenbereitstellung Dritter. Ohne Aktionen Dritter bleiben Raumplanungen Papier.

Wirkungen der Raumplanung können nur im Rahmen konkreter Vorhaben anderer staatlicher oder privater Akteure evaluiert werden: an dem, was realisiert, was aufgrund von Auflagen geändert oder was verhindert worden ist (vgl. auch Holzinger/Landsteiner 1993). Raumordnungspolitische Aspekte sind für die Entscheidungsträger immer nur ein Entscheidungskriterium unter anderen.

In der Praxis verlaufen Wirkungen oft umgekehrt, so dass (grössere) private Investitionsentscheide zu Anpassungen der Raumplanung führen. Im Rahmen von Public-Private-Partnerships oder dem „Zürcher Modell“ (Eisinger/Reuther 2007) werden solch kooperative Planungen kultiviert. Positive Wirkungen werden über Aushandlungsprozesse erzielt. Wie und wie gross die Wirkungen sind, lässt sich nicht vorherbestimmen.

Das Paradigma der nachhaltigen Raumentwicklung steht noch am Anfang seiner konzeptionellen und operationalen Klärung (vgl. z. B. Baccini/Oswald 2003); es ist noch weit entfernt von sicherer Evaluierbarkeit.

3 Die Evaluation der Regionalpolitiken

In allen drei Länderbeiträgen nimmt die Evaluation der Regionalpolitik den grösseren Raum ein. Denn im Unterschied zur Raumplanung hat Regionalpolitik eigene Programme, die unmittelbar etwas bewirken und in den ausgewählten Zielgebieten etwas in Gang setzen sollen.

3.1 Paradigma – Wirkungsketten – Indikatoren

Auf den ersten Blick können Evaluationen regionalpolitischer Programme von klaren Ziel-Mittel-Relationen ausgehen. Doch wie die Länderbeiträge zeigen, ist die Sachlage nicht so einfach.

Auf der obersten Ebene geht es um die leitenden Maximen regionalpolitischer Programme. Bei der „Neuorientierung der Regionalpolitik“ ist in der Schweiz darüber ausführlich diskutiert worden. Auch für ein deutsches Bundesland wie Brandenburg hat es ganz praktische Folgen, wenn die Maxime „Gleichwertigkeit der Lebensräume“ ersetzt wird durch die „dezentralisierter Konzentration“ oder wenn „Entleerungsräume“ nicht mehr per se als Übel angesehen werden. Ähnlich verschob sich die Regionalpolitik für die schweizerischen Berggebiete von der Förderung der Infrastruktur auf die Verbesserung der regionalen Wertschöpfung durch Kooperationen.

Für die Evaluation regionalpolitischer Programme ergeben sich daraus Konsequenzen.

- Abhängig vom wegweisenden Paradigma werden andere Ziele und Wirkungen angestrebt und evtl. andere Zielgruppen angesprochen.
- Je nach Paradigma greifen die eingesetzten Ressourcen jeweils in andere Wirkungszusammenhänge ein, ja andersartige Ressourcen und Vorgehensweisen können zweckmässig werden.
- Je nach dem werden andere Indikatoren relevant, um kritische Knoten im Wirkungsgefüge und die erzielten Haupt- und Nebeneffekte zu erfassen.

In der Realität sind die Mittel für Evaluation häufig so beschränkt, dass Wirkungszusammenhänge nur oberflächlich erfasst werden können und die verfügbaren Indikatoren nur Teilaspekte widerspiegeln: quantifizierbare Wirkungen eher als qualitative.

3.2 Vorgehensweisen und Adressaten

Bei der Evaluation von regionalpolitischen Projekten wird in den drei Ländern ähnlich vorgegangen. Man stützt sich zum einen auf „harte“ Daten und zum anderen auf Interviews mit Betroffenen und Beteiligten. Adressaten sind die zuständige Zentralverwaltung auf Bundes- und Länder-/Kantonsebene sowie die Bevölkerung und Wirtschaft in der betroffenen Region.

Welche Daten geeignet und verfügbar sind, ist situationsabhängig und muss vorher geklärt werden, was nicht immer geschieht. Nachträgliche Erhebungen und Rekonstruktionen sind meist zu aufwendig. Abstrakte Vorgaben von Indikatorenkatalogen wie zum Beispiel bei den EU-Evaluationen mögen aus Gründen der Vergleich- und Aggregierbarkeit notwendig sein; sich darauf zu beschränken ist fragwürdig, doch aus Budgetgründen oft unumgänglich.

Die Interviews dienen dazu, qualitative Wirkungen und Wertungen zu erfassen. Anhand der gewonnenen Informationen lassen sich Ziele und Prioritäten der Beteiligten herausfinden, Wirkungsverkettungen zusammenfügen sowie kritische Erfolgsfaktoren und Schwachpunkte erkennen.

Solche Informationen sind für den Projektfortgang nützlich, um ergriffene Massnahmen anzupassen und allenfalls andere Schwerpunkte zu setzen. In vielen Fällen, vor allem in Österreich, ist man daher zu begleitenden Evaluationen übergegangen, die ins Projekt-Coa-

ching übergehen. Auf diesen Erfahrungen baut auch die EU-Empfehlung auf, vermehrt begleitende Evaluationen einzusetzen.

Nach den österreichischen Erfahrungen sind Schlussevaluationen weiterhin wichtig. Ihnen obliegt es, Bilanz über die (positiven und negativen) Wirkungen und die Wirksamkeit und Effizienz der eingesetzten Ressourcen zu ziehen. Im günstigen Fall gehen begleitende, ex post und ex ante-Evaluation des Folgeprojektes – wie in Österreich praktiziert – ineinander über und befruchten sich gegenseitig. Für die übergeordnete Ministerebene werden Resultate einzelner Projektevaluationen aggregiert und Schlüsse für Programme gezogen.

Vor allem in Deutschland werden auch ökonometrische Modelle eingesetzt, um die Wirksamkeit von Programmen zu bemessen. Modellhaft verknüpfte Indikatoren für Inputs, Outputs und Rahmenbedingungen sollen aufzeigen, in welchem Masse die angestrebten Wirkungen auf gesamtwirtschaftlicher Ebene tatsächlich eintreten. Die Grenzen solcher Modellaussagen sind bekannt: Sie sind abstrakt; sie sind nur so gut wie die zugrundeliegenden Indikatoren; es sind black box-Modelle, die nichts über die internen Wirkungszusammenhänge aussagen; sie lassen kaum kleinräumige Aussagen zu. In Österreich werden sie daher für regionalpolitische „Evaluationen“ gar nicht, in der Schweiz nur punktuell eingesetzt.

3.3 Auftragnehmer und Auftraggeber

Auftragnehmer regionalpolitischer Projektevaluationen sind in allen drei Ländern vor allem kleine bis mittelgrosse, private und universitätsnahe Institute. Ihr Tätigkeitsfeld konzentriert sich auf das eigene Land, bedingt durch die bestehenden Beziehungsnetze.

Das Angebot an qualifizierten Evaluatoren ist zwischenzeitlich in allen drei Ländern ausreichend. Allerdings führt der starre Evaluationsrhythmus der EU-Programme immer wieder zu Kapazitätsengpässen. Der Übergang zu begleitenden Evaluationen verspricht eine gleichmässige Auslastung.

Der Engpass für die Weiterentwicklung und Verbreitung regionalpolitischer Evaluationen liegt vor allem auf Auftraggeberseite. In den Verwaltungen der drei Länder sind Fachleute knapp, die Projekte zielorientiert ausschreiben und im Folgenden den Evaluations- und Rückkopplungsprozess organisieren und vorantreiben können.

Allein in der zuständigen Abteilung der österreichischen Bundeskanzlei scheint sich zwischenzeitlich ein produktiver Erfahrungsschatz mit regionalpolitischen Evaluationen angesammelt zu haben. In Deutschland bestehen in der Verwaltung weiterhin Abneigungen gegen die „verwaltungsfremde“ Wirkungsorientierung. Daher gibt es auf Bundes- wie auf Landesebene kaum Evaluationsfachleute in Schlüsselpositionen, sondern allenfalls in peripheren, temporären Stellen. In der Schweiz beschränken sich regionalpolitische Evaluationen bislang auf Einzelfälle.

3.4 Nutzen der Evaluationspflicht

Wer sich an regionalpolitischen Programmen der EU beteiligt, unterliegt einer Evaluationspflicht. Das gilt mittelbar auch für die involvierten Stellen in der Schweiz. Hier enthalten zudem die revidierte Bundesverfassung sowie einige Gesetze eine Evaluationspflicht.

Von den drei Ländern unterliegt Deutschland dieser Evaluationspflicht im Rahmen der europäischen Strukturfondsprogramme am längsten. Dies hat bislang wenig ausgelöst. Evaluationsberichte werden zwar pflichtgemäss eingeholt, Evaluation bleibt jedoch oft ein symbolischer Akt. Externe Wirkungen interessieren die Verwaltung wenig, die Politiker nur mit positiven Vorzeichen.

Die Schweiz ist seit den Interreg-Programmen dieser Evaluationspflicht ausgesetzt. Es war hier ein Novum, dass sämtliche Projekte routinemässig evaluiert werden mussten. Man konnte sich mit den Ansprüchen, Routinen und Restriktionen einer derartigen Evaluationspflicht vertraut machen. Inhaltlich haben diese Evaluationen bislang nicht allzu viel gebracht. Schon das grenzüberschreitende Kennenlernen wurde als Erfolg gewertet. Gemeinsame wirkungsorientierte Projekte waren dagegen spärlich.

Auch Österreich unterliegt seit dem EU-Beitritt dieser Evaluationspflicht. In der ersten Runde unterzogen sich die zuständigen Stellen dieser Pflicht ohne grosse Begeisterung: viel Aufwand, wenig Ertrag. Entscheidend war die Schlussfolgerung daraus: Wenn schon evaluiert werden muss, dann so, dass es uns selber was nutzt, selbst wenn der Aufwand etwas grösser wird. In der Folge haben sich in Österreich regionalpolitische Evaluationen etabliert und weiterentwickelt: mit dem Übergang zu begleitenden Evaluationen, dem Einbezug unterschiedlicher Stakeholder und dem Projekt-Coaching, mit der Verknüpfung von ex post- und ex ante-Evaluationen als Grundlage für die nächste Generation der Programme. Doch: neue Programmrichtlinien der EU entwerteten Erkenntnisse aus der vorausgegangenen Programmphase und niemand interessierte sich mehr für diese Evaluationen.

Was kann man daraus lernen:

- Die auferlegte Evaluationspflicht kann Anstösse geben, aber sie ist kein Selbstläufer.
- Erst wenn Evaluationen zum wohlverstandenen Eigennutz eingesetzt werden, entwickeln sie auch positive Eigendynamik.
- Entscheidend ist die Wirkungs- und Effizienzorientierung der Beteiligten.
- Diskontinuitäten unterminieren Evaluationsanstrengungen.

3.5 Rückkopplungen und Lernen der Verwaltung

Die Chancen, dass eine Rückkopplung von Evaluationsresultaten in den Policy-Prozess gelingt, sind nach den Ausführungen der Referenten recht unterschiedlich. Wo – wie in Deutschland – Evaluationen mehr oder minder pro forma durchgeführt werden, sind sie von vornherein gering.

In der Schweiz sind zunächst Rückkopplungen auf Ebene der Chefbeamten durch die schlanke Verwaltung vorgegeben. Denn die Programmverantwortlichen begleiten meist gleichzeitig die entsprechenden Projekt- und Programmevaluationen. So wird amtsintern

zur Kenntnis genommen, was gut und nicht so gut gelaufen ist, und welche Verbesserungen empfohlen werden. Gleiches geschieht meistens auch in der betreffenden Region. Nur endet dann meist der Reflektionsprozess – oder wird von der Politik überholt. Bei der systematischen Auswertung und Synthetisierung der Resultate mehrerer Evaluationen steht man in der Schweiz – aus meiner Sicht – erst am Anfang.

In Österreich sind diese Rückkopplungsprozesse am weitesten gediehen und am besten organisiert. Mehrere Faktoren scheinen das zu begünstigen (vgl. auch Benz/Fürst 2002):

- Regionalpolitik ist in der Bundeskanzlei angesiedelt; die Stellen sind seit den Anfängen unter Kreisky in den 1960er Jahren sehr qualifiziert besetzt.
- Die Österreichische Raumordnungskonferenz (ÖROK) leistet Abstimmung über die föderalen Ebenen hinweg.
- In der Koordinations- und Arbeitsplattform Evaluierung (KAP-EVA) tauschen sich ebenenübergreifend Auftraggeber und Auftragnehmer sowie weitere Beteiligte aus.
- Für jedes Projekt gibt es eine „massnahmenverantwortliche Förderstelle“ auf Bundes- oder Landesebene. Damit sind die Zuständigkeiten und Verantwortlichkeiten geklärt.
- Durch die begleitende Evaluation findet eine laufende Rückkopplung von Evaluationsresultaten in den Projektablauf statt.
- In Grossgruppenveranstaltungen werden die Erkenntnisse aus verschiedenen Projekten ausgetauscht und diskutiert; die Diskussion wird gezielt auf übergreifende Fragestellungen gelenkt.
- Ex post- und ex ante-Evaluationen werden verknüpft und liefern gemeinsam Grundlagen für die Strategieentwicklung der folgenden Phase.

Die erforderliche Zeit für solche Reflektionsprozesse leistet man sich auf EU-Ebene noch nicht. Wie anders wäre es möglich, dass neue regionalpolitische Programme bereits ausgeschrieben werden, bevor die Evaluationen der vorausgegangenen vorliegen?

4 Schluss

Der sinnvolle Einsatz regionalpolitischer Evaluationen will gelernt sein. Das haben die drei Länderstudien gezeigt. Die Voraussetzungen sind unterschiedlich und je nachdem mehr oder minder günstig.

Sachverstand, Wirkungsorientierung und organisatorische Logistik sind drei wichtige Erfolgsfaktoren. Nicht der Evaluator und die Evaluation stehen im Mittelpunkt, sondern die zielgerichtete, möglichst wirksame und effiziente Durchführung von Massnahmen, die aus Sicht unterschiedlicher stakeholder konsensfähig und zukunftsfruchtig sind. In diesem Rahmen erhalten regionalpolitische Evaluationen ihren Wert – als Dienstleistung innerhalb des vielschichtigen, föderalistischen Policy-Prozesses.

Soziale Arbeit

Evaluation in der Sozialen Arbeit in Deutschland. Entwicklungslinien und Besonderheiten der Evaluationsdebatte am Beispiel der Kinder-, Jugend- und Familienhilfe

1 Einleitung

Evaluationen in der Sozialen Arbeit lassen sich bis in die Anfänge sozialpädagogischer Forschung zurückverfolgen; ob man nun bis zu den Studien zur Armutsforschung von Thomas von Aquin zurückgehen will (vgl. Müller 1978: 15) oder lediglich um mehrere Jahrzehnte zurückblickt etwa auf die Arbeiten von Alice Salomon und Siddy Wronski (vgl. Heiner 2001: 484). In einem weit gefassten Verständnis von Evaluation ist das Nachdenken und Forschen über Bedingungen, Umsetzungsformen, Zielgruppen, Ergebnisse und Wirkungen mit dem Ziel der Verbesserung fachlichen Handelns ein inhärenter Bestandteil Sozialer Arbeit. Bei einem enger gefassten Begriff von Evaluation – wie er diesem Beitrag zugrunde liegt – als systematischer und transparenter Untersuchung des Nutzens oder Wertes eines Gegenstandes auf der Grundlage empirisch erhobener Daten (vgl. DeGEval 2002: 13) handelt es sich jedoch um ein jüngerer Phänomen, das vor allem im letzten Jahrzehnt beachtlich an Aufmerksamkeit gewonnen hat.

Die Entwicklungslinien der Evaluationsdebatte werden in diesem Beitrag am Beispiel der Kinder-, Jugend- und Familienhilfe seit der Reformphase in den 1960er und frühen 1970er Jahren mit Schwerpunkt auf den letzten zehn bis 15 Jahren skizziert. Anzumerken ist, dass die Kinder- und Jugendhilfe als eines der großen Praxisfelder Sozialer Arbeit in Deutschland auch allgemein fördernde, bildende und präventive Angebote (z. B. Freizeitangebote, Familienbildungsstätten) umfasst. Der Terminus Soziale Arbeit wird weder eindeutig noch einheitlich verwendet (vgl. Hering/Münchmeier 2000: 11) und bislang hat sich keine bestimmte Systematik der Praxisfelder in Deutschland etabliert (vgl. Homfeld/Schulze-Krüdener 2003: 2). Um eine Verortung der Kinder- und Jugendhilfe innerhalb des Feldes der Sozialen Arbeit, die bis heute nicht geklärt ist (vgl. Bock 2002: 299), dennoch in systematisierender Absicht zu ermöglichen, kann Werner Tholes Unterscheidung vier großer sozialpädagogischer Praxisfelder herangezogen werden: Neben dem hier interessierenden Handlungsfeld unterscheidet er die Altenhilfe, sozialpädagogische Angebote im Gesundheitssystem (bspw. Sozialpsychiatrische Dienste, Werkstätten für behinderte Menschen, Kurhäuser) und das „sehr diffuse Praxisfeld der erwachsenenbezogenen Sozialen Hilfen“ (bspw. Schuldnerberatung, Resozialisierungsmaßnahmen, Unterkünfte für Obdachlose; vgl. Thole 2002: 20-22).

2 Strukturen und Besonderheiten der Kinder-, Jugend- und Familienhilfe in Deutschland

Die Strukturen, Angebote und Leistungen sowie Zuständigkeiten und Grundsätze der Kinder- und Jugendhilfe sind im achten Sozialgesetzbuch (SGB VIII; vgl. Wiesner 2006) verankert. Sie richtet sich mit direkt helfenden Angeboten an junge Menschen sowie ihre Familien, wenn Benachteiligungen auftreten oder akute individuelle Schwierigkeiten erkennbar werden. Sie umfasst zugleich allgemeinerzieherische, fördernde und unterstützende Aufgaben. Darüber hinaus kommt einer anwaltschaftlich verstandenen Kinder- und Jugendhilfe die Funktion zu, im politischen Raum für positive Lebensbedingungen und eine kinder- und jugendfreundliche Umwelt einzutreten (vgl. Jordan 2005: 12-14). Ihre Leistungs- und Aufgabenbereiche umfassen gemäß der gesetzlichen Untergliederung:

- Jugendarbeit (Jugendverbandsarbeit, offene Jugendarbeit, kulturelle Jugendbildung), Jugendsozialarbeit (Jugendberufshilfe, Schulsozialarbeit), erzieherischen Kinder- und Jugendschutz;
- Förderung der Erziehung in der Familie (allgemeine Förderung, Beratung, gemeinsame Wohnformen für Mütter/Väter und Kinder u. a.);
- Förderung von Kindern in Tageseinrichtungen und in Kindertagespflege;
- Hilfe zur Erziehung (ambulant, teilstationär, stationär), Eingliederungshilfe für seelisch behinderte Kinder und Jugendliche, Hilfe für junge Volljährige;
- andere, teils hoheitliche Aufgaben (u. a. Mitwirkung in gerichtlichen Verfahren).

Ein Schlüssel für das Verständnis der Besonderheiten der Evaluationsdebatte in der Kinder- und Jugendhilfe – gerade auch im Vergleich zu angrenzenden Feldern wie Bildung, Gesundheit oder Arbeitsmarkt – liegt in ihren spezifischen Strukturen sowie ihren grundlegenden fachlichen Prinzipien und Grundsätzen. Dabei ist hervorzuheben, dass die Kinder- und Jugendhilfe föderal strukturiert ist; die jeweiligen Zuständigkeiten bedingen Gegensätze und Ziele von Evaluationen. Die Aufgaben des Bundes sind begrenzt auf Kompetenzen der Gesetzgebung (vgl. Wiesner 2006: 11-14) sowie Aufgaben der überregionalen Anregung und Förderung der Jugendhilfe (§ 83 SGB VIII). Evaluationen sind daher vorrangig im Bereich der Förderung von Modellprogrammen und der Gewinnung übertragbarer Erkenntnisse angesiedelt (vgl. Haubrich et al. 2007; Dietzel/von Troschke 1988), teils auch bezogen auf Infrastrukturförderung (bundeszentrale Träger, Dach- und Fachorganisationen) und Gesetzgebung. Die Länder fördern gemäß § 82 SGB VIII die Tätigkeit der öffentlichen und freien Träger, mit dem Ziel der Weiterentwicklung der Jugendhilfe sowie eines gleichmäßigen Ausbaus der Angebote, und sie unterstützen die Jugendämter und Landesjugendämter in ihren Aufgaben; zudem bestimmen sie die Ausführungsgesetze zum SGB VIII sowie die Gesetze zur Tagesbetreuung von Kindern.

Die Umsetzung, also die Planung und Förderung der Aufgaben und Leistungen der Jugendhilfe obliegt der kommunalen Selbstverwaltung. Organisatorisch sind diese Aufgaben im Jugendamt gebündelt. Eine Besonderheit ist die Zweigliedrigkeit des Jugendamtes, bestehend aus der Verwaltung und dem Jugendhilfeausschuss. Im Jugendhilfeausschuss

sind neben Mitgliedern der Vertretungskörperschaft auch Vertreter/-innen der anerkannten Träger der freien Jugendhilfe stimmberechtigt. Diese Besonderheit trägt der starken Stellung der freien Träger in Deutschland Rechnung. Gemäß dem SGB VIII ist die öffentliche Jugendhilfe verpflichtet, partnerschaftlich mit einer Vielfalt von freien Trägern (wie Wohlfahrtsverbänden, Jugendverbänden, Kirchen u.a.) zusammenzuarbeiten (§ 3, § 4), sie zu fördern, an Entscheidungen zu beteiligen (vgl. Jordan 2005: 251) und ihre Selbstständigkeit hinsichtlich Zielsetzung und Durchführung ihrer Aufgaben anzuerkennen. Neben dem Grundsatz der Partnerschaft und der Trägerpluralität gilt in der Kinder- und Jugendhilfe das Subsidiaritätsprinzip, das den Vorrang der freien Jugendhilfe betont. Demzufolge sollen öffentliche Träger von eigenen Maßnahmen absehen, wenn die freie Jugendhilfe Aufgaben übernimmt. Gleichwohl obliegt dem öffentlichen Träger die Gewährleistungspflicht und die Gesamtverantwortung (§ 79 SGB VIII).

Diese weitgehende Autonomie der freien Träger und die partnerschaftlichen Strukturen der Aufgabenwahrnehmung prägen die spezifische Rahmenbedingungen für Evaluationen, die sich von anderen Feldern grundlegend unterscheiden und in der Planung und Umsetzung von Evaluationen auf der kommunalen, der Landes- und der Bundesebene zu berücksichtigen sind.

Ein weiteres wichtiges Prinzip der Kinder- und Jugendhilfe ist die Partizipation der Adressaten/-innen. So haben Kinder und Jugendliche das Recht, an allen sie betreffenden Entscheidungen der öffentlichen Jugendhilfe beteiligt zu werden (§ 8 SGB VIII). Sofern dies auch die Qualität der Angebote und deren Bewertung betrifft, ergibt sich für Evaluationen die Frage, wie den Interessen junger Menschen und ihrem Partizipationsrecht in geeigneter Form Rechnung getragen werden kann.

3 Die Anfänge der Evaluationsdebatte in der Reformära

Deutschland gehört zu den Ländern, in denen sich bereits früh Evaluationsaktivitäten etabliert haben (Derlien 1997: 4-5). Eine Vorreiterrolle kommt dem Bildungsbereich zu, was nicht ohne Auswirkungen für das benachbarte Feld der Kinder- und Jugendhilfe blieb. Getragen von einem parteiübergreifenden Reformkonsens, dem Vertrauen in die Möglichkeit wissenschaftlich basierter politischer Steuerung und einem Verständnis von Bildung als unentbehrliches Investitionsgut wurden in diesem Politikfeld schon in den 1960er Jahren Sozialexperimente mit institutionellen Alternativen als Mittel einer rationalen politischen Entscheidungsfindung eingesetzt (vgl. Hellstern/Wollmann 1984a: 39-42).

Für den außerschulischen Bereich der Jugendarbeit und Jugendhilfe sind diese Entwicklungen bedeutsam, weil die Kindertagesbetreuung als Handlungsfeld der Jugendhilfe im Verständnis des Bildungsrats als Elementarbereich des Bildungswesens in die Reformbemühungen einbezogen war. Vor allem aber hatte die Debatte zu Ansätzen und Methoden pädagogischer Begleitforschung erkennbaren Einfluss auf die Sozialpädagogik. Gleichwohl ließen sich die dort gewonnenen Erfahrungen nicht einfach übertragen. Jugendhilfe und Jugendarbeit standen weniger im Mittelpunkt öffentlicher Aufmerksamkeit, auf staatliche Planung wurde zunächst weitgehend verzichtet. Die Reformbemühungen setzten erst mit einiger zeitlicher Verzögerung ein und so konnten bis zum Ende der Reformära nur wenige

Großforschungsprojekte realisiert werden (vgl. Grohmann 1997: 91-98). Angesichts der föderalen, subsidiären und partnerschaftlichen Strukturen der Jugendhilfe ließ sich auch die Idee des Experimentierens mit institutionellen Alternativen nicht in gleicher Weise umsetzen. Während in der Bildungsforschung wissenschaftliche Begleitforschung in erster Linie als (quasi-)experimentelle Wirkungsforschung verstanden wurde, wurden diese Ansätze im Hinblick auf ihre Anwendbarkeit auf Soziale Arbeit von Anfang an in Frage gestellt. Vor dem Hintergrund des sich in den Sozialwissenschaften in den 1960er Jahren vollziehenden „Positivismusstreits“ (Kritischer Rationalismus vs. Kritische Theorie) und einer aus der mangelnden Anwendbarkeit und Praxisrelevanz von Forschungsergebnissen entstandenen Kritik an der so genannten „positivistischen Kontrollforschung“ (bspw. Hoschka et al. 1978: 85) wurden in den 1970er Jahren Ansätze der Handlungsforschung (Moser 1975; Altrichter/Gstetter 1993) oder Innovationsforschung entwickelt, die auf eine kontinuierliche Verbesserung der Praxis abzielten.

Evaluation fand – soweit sie der Fachöffentlichkeit zugänglich gemacht wurde – bis Anfang der 1990er Jahre überwiegend im Kontext von wissenschaftlichen Begleitstudien zu Modellvorhaben statt (vgl. Heiner 2001: 486). Nach dem Abklingen der Reformeuphorie, einer zunehmenden Skepsis gegenüber der Verwendbarkeit von Forschungsergebnissen für Reformprozesse und vor dem Hintergrund der durch die Erdölkrise 1973 ausgelösten Rezession unterlagen Modellvorhaben ab Ende der 1970er Jahre veränderten Erwartungen. Der Staat zog sich weitgehend aus der Gestaltung positiver Lebensbedingungen mittels einer „aktiven Politik“ (Mayntz/Scharpf 1973a) zurück und konzentrierte Fördermaßnahmen auf die Erprobung spezifischer Handlungsansätze zur Bewältigung aktueller Probleme in einzelnen Praxisfeldern (bspw. Jugendarbeitslosigkeit, Integration von Migrant*innen) (vgl. Heiner 2001: 485). Fragen der Verwendung staatlicher Budgets und der Entscheidung über Prioritäten angesichts knapper Ressourcen rückten in den Vordergrund.

Da Modellvorhaben bis heute vorrangig an jeweils aktuellen und sozialpolitisch brisanten Themen orientiert sind und weniger an grundlegenden fachlichen Fragestellungen oder gar der Entwicklung von Evaluationskonzepten, konnte sich allerdings kaum eine eigenständige Evaluationstradition etablieren (vgl. Heiner 2001: 486). Wissenschaftliche Begleitforschung war lange Zeit der gängigere Begriff und wurde erst in jüngerer Zeit durch den Begriff „Evaluation“ abgelöst. Beide Begriffe sind wenig trennscharf, ihre Gemeinsamkeiten und Unterschiede werden kaum diskutiert und doch meinen sie nicht das Gleiche. Wissenschaftliche Begleitforschung¹ versteht sich in erster Linie als anwendungsorientierte Sozialforschung, Evaluation hingegen – folgt man den Standards für Evaluation (DeGEval 2002; Widmer et al. 2000) sowie Begründungen einer Evaluationstheorie (Shadish et al. 1991) – erfordert eine konzeptionelle und theoretische Auseinandersetzung mit Fragen der Bewertung und der Nutzung (vgl. Haubrich 2009).

¹ Für wissenschaftliche Begleitforschung gibt es im Übrigen weder konzeptionell noch sprachlich ein englischsprachiges Pendant.

4 Neue Steuerung, Qualitätsentwicklung und Selbstevaluation – die 1990er Jahre

In den 1990er Jahren setzte sich die Ökonomisierung des Wohlfahrtsstaates weiter durch, wenn auch anfangs – angesichts der Vereinigung der beiden deutschen Staaten und des 1991 in Kraft getretenen neuen Kinder- und Jugendhilfegesetzes (SGB VIII) – mit wenig Aufmerksamkeit bedacht. Einen entscheidenden Einschnitt stellte die Einführung neuer Steuerungsmodelle (KGSt 1993) dar. Im Jahr 1994 verdeutlichte die kommunale Gemeinschaftsstelle für Verwaltungsvereinfachung (KGSt) Elemente outputorientierter Steuerung exemplarisch für den Bereich der Kinder- und Jugendhilfe (KGSt 1994). Leistungen sollten in standardisierter Form als Produkte beschrieben und die Kosten zurechenbar werden, damit bei der Vergabe von Fördergeldern auf dem Markt die günstigsten Leistungen eingekauft werden könnten (vgl. auch Sabine Kuhlmann in diesem Band).

Diese Entwicklungen waren folgenreich für die Evaluationsdebatte. Nicht zuletzt hierin wurzelt die für die Kinder- und Jugendhilfe charakteristische enge Verzahnung von Evaluation und Qualitätsentwicklung. Schließlich erfordert die Beschreibung von Produkten ihre inhaltliche Konkretisierung und damit einhergehend eine Bestimmung und Bewertung ihrer Qualität. Ursprünglich aus der Betriebswirtschaft stammende Konzepte des Qualitätsmanagements, der Qualitätssicherung und der Qualitätsentwicklung hielten Einzug in die Kinder- und Jugendhilfe. Allerdings lag von Anfang an der Schwerpunkt auf Qualitätsentwicklung (in Abgrenzung zu Kontrolle und Management) als kontinuierlicher Prozess der ständigen Weiterentwicklung und Verbesserung von Qualität.²

In diesem Kontext gewann Selbstevaluation als Sonderweg in der Sozialen Arbeit an Bedeutung. Bereits Ende der 1980er Jahre wurde sie in einem kleinen Kreis interessierter Sozialpädagogen/-innen als Instrument der Förderung fachlicher Professionalität (Heiner 1988; vgl. Spiegel 1997: 32) diskutiert. Im Rahmen der QS-Initiative des Bundesministeriums für Familie, Senioren, Frauen und Jugend (BMFSFJ) wurde sie einer breiten Fachöffentlichkeit mit praktischen Hilfestellungen zur Verfügung gestellt und wurden begleitend Fortbildungsmodule für Fachkräfte entwickelt. Auch im Kontext neuer Steuerungsmodelle fand Selbstevaluation Eingang als ein Instrument der Qualitätsentwicklung. Für die Kinder- und Jugendhilfe lag hierin eine Möglichkeit, Qualitätskriterien selbst zu bestimmen und nicht von außen vorgeben zu lassen.

Ausgehend von der Sozialen Arbeit haben sich Konzepte der Selbstevaluation mittlerweile auch in anderen Feldern etabliert; dieser Verbreitung tragen u.a. die „Empfehlungen zur Anwendung der Standards für Evaluation im Handlungsfeld der Selbstevaluation“ Rechnung (DeGEval 2004b; Müller-Kohlenberg/Beywl 2003). Der Selbstevaluation kommt im deutschsprachigen Raum eine vergleichsweise größere Bedeutung zu als in der internationalen Debatte. Der Begriff (self-evaluation oder auto-évaluation) ist zwar auch dort anzutreffen, es ist jedoch eher unüblich, Fremd- und Selbstevaluation einander konzeptionell kontrastierend gegenüberzustellen.³ Häufiger trifft man auf die Unterscheidung externe und

² Vgl. insbesondere die QS-Reihe des BMFSFJ (<http://www.qs-kompodium.de/>).

³ Für Hinweise auf englisch- und französischsprachige Quellen und zur Konkretisierung des Argumentes an dieser Stelle möchte ich Thomas Widmer herzlich danken.

interne Evaluation. Selbstevaluation in der Sozialen Arbeit in Deutschland meint aber eine sehr spezifische Form interner Evaluation. Sie ist definiert als die Evaluation der eigenen Arbeit durch Fachkräfte⁴ (vgl. Heiner 1988: 7).

5 Aktueller Stand und Trends der letzten Jahre

Qualitätsentwicklung – und so mittelbar auch Evaluation – wurde im Jahr 1999 mit der Verpflichtung der öffentlichen und freien Träger, im Bereich der Hilfe zur Erziehung Leistungs-, Entgelt- und Qualitätsentwicklungsvereinbarungen abzuschließen, zu einem gesetzlich verankerten Bestandteil der Kinder- und Jugendhilfe (§ 78a-g SGB VIII). Zugleich verschiebt sich der Fokus der Beurteilung von Jugendhilfeleistungen von den Produkten zu den Wirkungen. Die mit Modellen neuer Steuerung verbundene Intention, Leistungsbeurteilung mit Steuerung zu verknüpfen, setzt sich in Bezug auf Wirkungsmessung oder Wirkungsbeobachtung fort. Begriffe wie „wirkungsorientierte Steuerung“ (z. B. Schröder/Kettiger 2001), „wirkungsorientierte Jugendhilfe“⁵, „wirkungsorientierte Evaluation“ (vgl. Beywl et al. 2004) und neuerdings auch „evidence-based“ haben Konjunktur; doch dies sollte nicht als lediglich neue „Mode“ verkannt werden. Dahinter verbirgt sich ein zunehmendes Interesse der Auftraggebenden an durch Evaluation kodifiziertem, verlässlichem und wenn möglich steuerungsrelevantem Wissen (Lüders/Haubrich 2006: 14).

Der Elfte Kinder- und Jugendbericht des Bundes (BMFSFJ 2002a) markiert ein Trendwende: Er plädiert für „konsequente Evaluation“, zumal „(...) die Kinder- und Jugendhilfe zu wenig begründetes Wissen über die Auswirkungen der eigenen Praxis besitzt“ (ebd.: 255). Er hebt hervor, dass Anwendungsfelder und Geltungsbereich der Selbstevaluation begrenzt sind und eine Ergänzung durch externe Evaluationen erforderlich ist. Allerdings fehlten „(...) im deutschsprachigen Raum die dafür notwendigen Instrumente, Verfahren und Standards (...)“ (ebd.: 256). Deutlicher formuliert es der Koalitionsvertrag zwischen CDU/CSU und SPD: „Jugendhilfe sollte sich auch unter Effizienzgesichtspunkten weiterqualifizieren; dringend muss die Lücke im Bereich der Jugendhilfe-Wirkungsforschung geschlossen werden (...)“ (o. A.: 2006)

Entwicklungslinien und Trends werden umso unschärfer, je aktueller sie sind; schwer ist zu erkennen, was dauerhaft bleiben wird oder nur kurzlebig ist. Dies gilt umso mehr, als die Evaluationsdebatte bis heute vergleichsweise unübersichtlich ist. Nicht nur der Begriff Evaluation wird mittlerweile inflationär verwendet, auch die Qualität von Evaluationsaktivitäten differiert erheblich und häufig erscheinen Berichte lediglich als graue Materialien oder werden der Fachdebatte überhaupt nicht zugänglich gemacht. Vor diesem Hintergrund gibt es am Deutschen Jugendinstitut Bemühungen, einen Überblick über das Feld zu erlangen. Handlungsfeldübergreifend sind das Projekt „Jugendhilfe und sozialer Wandel“ sowie die Recherchen der Konzeptstelle „Strategien und Konzepte externer Evaluation in der Kinder- und Jugendhilfe“ (Projekt eXe; www.dji.de/evaluation) über externe Evaluati-

⁴ Diese Definition der Selbstevaluation unterscheidet sich von anderen deutschsprachigen Definitionen (insbesondere Widmer 2006b: 87-88; vgl. zur Diskussion unterschiedlicher Definitionen Haubrich/Lüders 2004: 327).

⁵ Vgl. die Schriftenreihe des ISA zum gleichnamigen Modellprogramm (www.wirkungsorientierte-jugendhilfe.de).

onsstudien hervorzuheben. Letztere zeigen nicht nur, dass die üblichen Recherchewege über Zeitschriften und Datenbanken das Feld nicht hinreichend erschließen und einer Ergänzung bedürfen. Sie lassen vor allem erkennen, dass Evaluationsberichte vorrangig auf die Darstellung der Erkenntnisse zielen und einen mehr oder weniger ausführlichen Methodenteil beinhalten, wohingegen nicht immer deutlich wird, welcher Evaluationsansatz zur Anwendung kam, welche Evaluationszwecke (formativ, summativ, wissensgenerierend) verfolgt wurden, wie sie zu Bewertungen gelangen und welche Beteiligtegruppen in welcher Weise in die Steuerungsarchitektur der Evaluation eingebunden waren. Diese Ergebnisse verweisen auf einen Entwicklungsbedarf, zumal Evaluationsdesigns und -berichte nicht nur Auskunft über die zu Grunde liegenden Theorien der Wissensgenerierung geben sollten, sondern auch über die Verfahren der Bewertung des Gegenstandes (Programme, Politik, Organisationen etc.) und gegebenenfalls über die angestrebte Nutzung oder Verwendung der Evaluationsergebnisse.

Während sich in der deutschen Fachdebatte ein gewisser Nachholbedarf gegenüber der internationalen Diskussion abzeichnet, weitet sich zugleich das Feld der Evaluation in der Praxis schnell aus, wie die Ergebnisse einer bundesweiten Querschnitterhebung bei Jugendämtern und Einrichtungen stationärer Hilfe zur Erziehung durch das Projekt „Jugendhilfe und sozialer Wandel“ zeigen (Pluto et al. 2007). So gaben 79 Prozent der befragten Jugendämter an, Kenntnisse über die Durchführung von Evaluationsstudien in ihrem Jugendamtsbezirk zu haben – im Osten mehr als im Westen und in den Städten häufiger als auf dem Land. Dabei kannten sie Evaluationen der Hilfe zur Erziehung zu 61 Prozent bei öffentlichen und zu 75 Prozent bei freien Trägern, gefolgt von Jugendarbeit (56 % und 52 %) und Kindertagesbetreuung (28 % und 12 %). Überwiegend werden Formen der Selbstevaluation genannt (Jugendamtsbefragung: 75 % Selbstevaluationen, 34 % externe Evaluationen, 12 % Peer-Evaluationen; Einrichtungsbefragung: 88 % Selbstevaluationen, 38 % auch Fremdevaluationen). Während Evaluation als fachliches Konzept – was auch immer im Einzelnen darunter verstanden wird – offensichtlich in der Praxis angekommen ist, zeigt sich der Studie zufolge in der Umsetzung des § 78a-g SGB VIII gerade im Hinblick auf Qualitätsentwicklungsvereinbarungen eine Diskrepanz zwischen den gesetzlichen Vorgaben und ihrer Umsetzung; Defizite werden in diesem Kontext auch bezüglich der Partizipation von Jugendlichen erkennbar.

Hinter diesen allgemeinen Tendenzen verbergen sich disparate Entwicklungen in den einzelnen Handlungsfeldern, wobei hier nur schlaglichtartig wichtige Aspekte in den drei eben genannten Feldern mit hoher Evaluationsaktivität beleuchtet werden können.

Im Feld der Hilfe zur Erziehung, neben der Kindertagesbetreuung der kostenintensivste Bereich, sind zwei für die Evaluationsdebatte folgenreiche Studien hervorzuheben – die JES-Studie (BMFSFJ 2002b) und die JULE-Studie (BMFSFJ 1998) – sowie aktuell die Evaluation des Modellprogramms „Wirkungsorientierte Jugendhilfe“, die auf der Grundlage eines quasi-experimentellen Designs überprüfen will, ob wirkungsorientierte Leistungs-, Entgelt- und Qualitätsentwicklungsvereinbarungen nach § 78a-g SGB VIII wirkungsvolle Hilfe ermöglichen (vgl. Polutta 2006). In vielen Jugendämtern und Einrichtungen werden Qualitätssicherungs-, Controlling- oder Monitoringsysteme etabliert (z. B. EVAS, WIMES), wobei die Erfahrungen zeigen, dass die Suche nach geeigneten Instrumenten noch nicht als abge-

schlossen gelten kann und v.a. die Prozesse ihrer Implementation im Zusammenwirken von öffentlicher und freier Jugendhilfe entscheidend sind.

Im Bereich Kindertagesbetreuung liegt der Schwerpunkt auf Qualitätsentwicklung. Evaluation wird – abgesehen von wenigen namhaften Evaluationsstudien (z. B. Zimmer et al. 1997; Conrad et al. 1997) – überwiegend in diesem Kontext thematisiert. Einen zentralen Stellenwert haben dabei die 1999 gestartete „Nationale Qualitätsoffensive im System der Tageseinrichtungen für Kinder – NQI“ des Bundes sowie Qualitätsoffensiven der Träger von Kindertageseinrichtungen, in deren Rahmen unterschiedliche Zugänge der Qualitätssicherung entwickelt wurden (vgl. Dittrich 2007). Als eine Besonderheit in diesem Feld ist hervorzuheben, dass eine regelmäßige Überprüfung des Standes der Umsetzung des Gesetzes zum stufenweisen Ausbau der Kindertagesbetreuung im SGB VIII verankert wurde (§ 24a; vgl. Deutscher Bundestag 2006c). Aktuell ist Bildung im Elementarbereich ein zentrales Thema, wobei jedoch Forschungsfragen gegenüber Evaluationsthemen dominieren.

Obwohl Jugendarbeit im Vergleich weitaus weniger kostenintensiv ist, finden laut Jugendamtsbefragung gerade hier viele Evaluationsaktivitäten statt. Ausschlaggebend dafür dürfte nicht zuletzt der Wirksamkeitsdialog Nordrhein-Westfalen sein: Das Bundesland verpflichtete 1999 alle Träger, die aus Landesmitteln gefördert oder kofinanziert werden, an dem landesweiten Wirksamkeitsdialog als Voraussetzung einer weiteren Förderung partnerschaftlich mitzuwirken. Diese Initiative hat viel in Bewegung gebracht und letztlich den Bereich der Jugendarbeit gestärkt, da transparent und öffentlich verhandelbar wurde, was in diesem Handlungsfeld geleistet wird (vgl. Beywl et al. 2001).

6 Anmerkungen zu Perspektiven und Herausforderungen

Für eine Zwischenbilanz, um Herausforderungen für die Zukunft zu benennen, können an dieser Stelle nur wenige ausgewählte Aspekte hervorgehoben werden. Qualität, Evaluation und Wirkungsfragen sind gleichsam zu einem selbstverständlichen Moment professionellen Handelns geworden, auch wenn dabei noch einige inhaltlich-fachliche Fragen ungelöst sind.

Nach einer anfänglich defensiven Haltung gegenüber der Qualitätsdebatte wurde der Begriff schon bald inhaltlich gefüllt und als anschlussfähig an frühere Diskussionen zu Professionalität und Handlungskompetenz erachtet. Allerdings drängt sich zugleich der Eindruck auf, dass dessen ungeachtet in der Praxis Qualitätssicherungssysteme nach bekannten Modellen (wie ISO 9000 oder EFQM) etabliert oder bestehende Dokumentationssysteme übernommen wurden (vgl. Flösser/Oechler 2004). Eine zentrale Herausforderung liegt hier in der widersprüchlichen Anforderung, einerseits Ressourcen sparen zu wollen und eine Vereinheitlichung von Vergleichsmöglichkeiten zu erreichen, indem man anderenorts entwickelte Modelle übernimmt, und andererseits alle Beteiligten und ihre fachlichen Perspektiven dauerhaft einzubinden, um Akzeptanz und Nutzen zu erhöhen und Widerstände abzubauen. Möglicherweise ist diese aktuell erkennbare Lücke u.a. auch einer fachlichen Engführung der Evaluationsdebatte in der Kinder- und Jugendhilfe geschuldet, die vor allem auf Programmevaluation fokussiert. Dabei wird – bspw. auch im Unterschied zur Evaluation von Hochschulen (vgl. Uwe Schmidt in diesem Band) – organisatorischen As-

pekten häufig wenig Beachtung geschenkt, wenngleich sie als Rahmenbedingungen und Ressourcen für die Umsetzung pädagogischer Programme von zentraler Bedeutung sind.

In den vergangenen Jahren hat vor allem die Frage nach den Wirkungen der Kinder- und Jugendhilfe an Bedeutung gewonnen. Die Beantwortung dieser Frage erfordert Evaluation; ihr kann nicht allein mit Qualitätsentwicklung begegnet werden. Wenngleich Wirkungen in der Sozialen Arbeit notwendiger Bestandteil pädagogischen Denkens sind, birgt die zunehmende Anbindung von Steuerungsprozessen an Wirkungsfragen die Gefahr einer wenig hilfreichen Komplexitätsreduktion. Schließlich sind professionelles Handeln wie auch die Strukturen und Verfahren in der Sozialen Arbeit ja gerade dadurch charakterisiert, dass Entscheidungen unter Unsicherheit getroffen werden, Wissen fallspezifisch angewendet wird und mit der Eigenaktivität der Klienten/-innen gerechnet werden muss (vgl. Dewe/Otto 2001). In der Suche nach Antworten auf die Wirkungsfrage steht die Kinder- und Jugendhilfe noch am Beginn der Debatte.

Inwieweit es gelingen kann, angemessene Antworten auf solche Herausforderungen zu finden, hängt sicherlich auch davon ab, ob und inwiefern Evaluation institutionell stärker verankert wird und der Professionalisierungsprozess voranschreitet. Auch sind die Folgen der Föderalismusreform für die Qualitätsdebatte von einiger Bedeutung, wenngleich sie bislang noch kaum absehbar sind.

Evaluation in der Sozialarbeit in Österreich – auf dem Weg zu einem neuen Verständnis und aktiver Auseinandersetzung

Dieser Beitrag zeichnet die laufenden Diskussionen und Themen in der Sozialarbeit in Österreich nach und identifiziert wesentliche gegenwärtige Geschehnisse und Grundströmungen und darin die Relevanz von Evaluationsaktivitäten. Ein verifiziertes Abbild von Evaluationen der Sozialarbeit in Österreich zu liefern ist jedoch noch nicht möglich – denn derzeit markiert ein Umbruch die Sozialarbeit und im Gefolge deren Evaluationsagenden.

1 Kontextuelle Dimensionen für Evaluationen in der Sozialarbeit in Österreich

In Österreich sind Sozialwesen, Sozialarbeit und Sozialwirtschaft in der letzten Dekade herausgefordert, Finanzierung und Kontinuität der Dienstleistungen zu gewährleisten. Kontextuelle Ereignisse hierin sind die Veränderungen in den Finanzierungs- und Ressourcenstrukturen und den rechtlichen Rahmenbedingungen in der Leistungserbringung der Sozialarbeit zum einen; und der Umbau der Sozialarbeitsausbildung seit 2002 zum anderen. Professionsverständnis, Fachpraxis als auch Evaluationsverständnis und -praxis werden reformuliert.

Finanzierungs- und Ressourcenstrukturen, die sich durch Einnahmen von der öffentlichen Hand, durch selbst erwirtschaftete Mittel als auch durch Spenden und Sponsoring kennzeichnen, verlagern sich. In Österreich wurden im Jahr 2002 Sozialausgaben in der Höhe von 62,5 Mrd. Euro geleistet (Statistik Austria 2006). Bereits Mitte der 1990er Jahre waren in den Organisationen des Sozialwesens im Durchschnitt die selbst erwirtschafteten Anteile – inklusive jener über Leistungsverträge – beinahe so hoch wie Förderungen durch die öffentliche Hand. Spenden und Sponsoring spielen eine geringere Rolle (Loidl-Keil 2002). Ökonomisierungsprozesse spiegeln sich in den veränderten Finanzierungsstrukturen, deren Auswirkungen auf die organisationalen Orientierungen und Verfahrensweisen – auch zur Leistungsbeurteilung und Evaluation – massiver wahrnehmbar werden. Legitimationsgrundlagen ändern sich, betriebswirtschaftliche Kalküle und Fragen nach Effektivität und Effizienz gelangen in den Diskurs.

Organisationsstrukturen ändern sich. Heute haben über drei Viertel der soziale Leistungen erbringenden Organisationen weniger als zehn MitarbeiterInnen, über 95 Prozent weniger als fünfzig MitarbeiterInnen (Statistik Austria 2006). Wir beobachten eine „Verkleinerung“ der Organisationen, in denen soziale Dienstleistungen erbracht werden. Zudem

weiten sich anteilig privat-rechtlich organisierte Einrichtungen gegenüber öffentlich-rechtlichen aus.

Hinsichtlich der rechtlichen Rahmenbedingungen sind zuerst spezifische Rechtsquellen (z. B. Sozialhilfegesetz, Kinder- und Jugendhilfegesetz, Bewährungshilfegesetz, Pflegegesetz) je nach Handlungsfeld zu nennen. Evaluation spielt darin eine kaum genannte Rolle; Qualitätssicherung hingegen wird zunehmend ein rechtlich formulierter Anspruch (Loidl-Keil/Viechtbaur 2003). Da in weiten Bereichen der Sozialarbeit eine Trennung von öffentlich und privat organisierten Leistungen besteht (z. B. in Jugendfürsorge und Jugendarbeit), ist die Vertragsgestaltung wesentlich. Als einschlägige Rechtsquelle hierfür gilt das Bundesvergabegesetz zur Ausgestaltung der Rahmenbedingungen von Sozialarbeit und deren Evaluation (Kiendl-Wendner/Schwar 2007). Ausformulierungen bezüglich Evaluationen bleiben allerdings unkonkret bzw. offen. Das Vergaberecht ist nur in einem eingeschränkten Umfang anwendbar, was für die Akteure im Bereich sozialer Dienstleistungen Erleichterungen und mehr Flexibilität im Verfahren im Einzelfall bringen soll. Aus der Sicht der Organisationen im Sozialbereich (Tauber 2006) brachte aber bereits das Bundesvergabegesetz 2002 nachteilige Effekte für die Praxis mit sich. Die Qualität des Dienstleistungsangebotes wird durch die Dominanz von Auftraggebern durchwegs mehr und mehr von diesen selbst bestimmt.

Fachpraxis und Profession der Sozialarbeit in Österreich befinden sich gegenwärtig durch den Umbau der Sozialarbeitsausbildung von vormals Akademien zu nunmehr Fachhochschulen in einem Akademisierungsprozess, welcher für Sozialarbeitspraxis und Evaluationen beziehungsweise Forschungen in diesem Fachgebiet von nachhaltiger Bedeutung sein dürfte. Dieser Wandel führt innerhalb der Profession der SozialarbeiterInnen einerseits zu enormen Spannungen – inhaltlichen Positionierungen, Dequalifizierungsaspekten, neu zu positionierendem Verhältnis von Theorie und Praxis –, andererseits zu einer „Verwissenschaftlichung“ des Berufes in heute ein modularisiertes System von Bakkalaureat und Master. Die Sozialarbeit wird zusehends als akademische Ausbildung bzw. akademischer Beruf wahrgenommen (Laskowski et al. 2006). Zwar wird eine Akademisierung nicht einer Professionalisierung der Evaluationen in der Sozialarbeit gleichzusetzen sein, doch dürften Chancen zu einer Ausprägung von Evaluationsdiskurs und -aktivitäten bestehen.

2 Evaluationsaktivitäten in Fachpraxis und Sozialforschung

Evaluationsaktivitäten finden auf unterschiedlichen Ebenen statt, ausgehend von der Fachpraxis in der KlientInnenarbeit selbst, von betrieblichen Erfordernissen und Interessen, von administrierenden oder finanzierenden Behörden auf Bundes- als auch Landesebene, vereinzelt von Dachorganisationen oder forschungs- und wissenschaftsfördernden Fonds beziehungsweise Abteilungen von Bund oder Ländern (z. B. Jugendwohlfahrt, Arbeitsmarktservice).

Zur Aufklärung der Frage nach Evaluationsaktivitäten würde der Blick auf die sozialarbeiterische Praxis einerseits und auf die sozialarbeiterische Forschung andererseits zu richten sein. Zur Erfassung der Evaluationsagenda in der Sozialarbeit in Österreich existiert

meinen Recherchen nach aktuell weder ein groß angelegtes Forschungsprojekt, noch eine zentrale Stelle, die die Evaluationsaktivitäten in der Sozialarbeit ausweisen würde.

Betrachten wir die sozialarbeiterische Praxis vor Ort. Evaluationen gehören darin inzwischen zum täglichen Arbeitsalltag, sie dienen vorrangig der Dokumentation des eigenen beruflichen Handelns. So kommen in der Fachpraxis der Sozialarbeit Erhebungsbögen, Beobachtungsinstrumente oder computergestützte Monitoring-Systeme zum Einsatz. Zusehends greifen Dokumentations- und Monitoring-Systeme (auch) wissenschaftlich fundierte Verfahren und Forschungsdesigns auf. Zum integrierten Repertoire zählen professionelle Standards zur (Selbst-)Evaluation oder systematisch eingesetzte Evaluationsdesigns jedoch keineswegs, weder in der Fachpraxis – in der zumeist selbstgestrickte Verfahren eingesetzt werden – noch in der Beforschung der Sozialarbeit.

Die sozialarbeiterische Praxis macht in ihrem Bemühen um einheitliche Standards zur Qualitätssicherung und Evaluation wohl Fortschritte, aber nicht nur. Einerseits werden Datenbanksysteme entwickelt und installiert, die dann Österreichweit für Monitoring und Verbleibsanalysen genutzt werden können (z. B. Data-Ware-House des Arbeitsmarktservice). Andererseits gibt es Beispiele (z. B. Pflegekinderwesen) von zuvor Bundesländer übergreifenden Dokumentationen – welche ein Material für Evaluationen und weiterführende Evaluationsforschung gegeben haben –, die dann aufgelöst werden und nunmehr in den Bundesländern unterschiedlich oder nicht mehr betrieben werden (Loidl-Keil et al. 2004).

Zur Evaluationsdurchführung werden zunehmend Kooperationen mit Universitätsinstituten und Fachhochschulstudiengängen eingegangen. Häufig erfolgen „Evaluationen“ dann unter anderem in Form von Seminararbeiten, Projektarbeiten oder Diplomarbeiten. Bei solchen Evaluationen handelt es sich zum Teil um beauftragte – oder zumindest „gutgeheißene“ – und weitaus unterfinanzierte Evaluationsvorhaben. Inwieweit dabei tatsächlich professionellen Evaluationsstandards genügende Evaluationen geleistet werden und diese theoretisch wie methodisch fundiert sind, sei kritisch in Frage gestellt.

Werden Evaluationsvorhaben im Bereich der Sozialarbeit als Studien finanziert und beauftragt, dann bieten hierfür universitäre Institute, nunmehr besonders die Fachhochschulstudiengänge für Sozialarbeit, sowie auch außeruniversitäre und privatwirtschaftliche Forschungs-, Evaluations- und Beratungsunternehmen ihre Dienste an. Als die größten Institute im außeruniversitären Sozialforschungsbereich, welche in einer weiteren Auslegung auch Agenden der Sozialarbeit evaluieren, können das Institut für Höhere Studien (IHS), das Wirtschaftsforschungsinstitut (WIFO) als auch das Forschungszentrum Seibersdorf (ARC) genannt werden. Sozialarbeitsspezifisch jedoch sind diese keineswegs ausgerichtet. Des Weiteren gibt es etwa 25 bis dreißig überdauernd aktive, kleinste und kleine private Sozialforschungseinrichtungen, die Evaluationen von Sozialarbeit in ihrem Portfolio haben. Auch die Szene der Unternehmensberatung greift die Thematik bereitwillig mit ihren Angeboten zu Qualitätssicherung und Zertifizierung auf. Spezifische Einrichtungen für Evaluationen und Beforschungen der Sozialarbeit stehen – insbesondere mit den Ressourcen der Fachhochschulstudiengänge oder Internet-Plattformen (wie www.sozialarbeitsforschung.at) – derzeit erst in ihren Anfängen.

Regelmäßig kann in Evaluationen bzw. Evaluationsforschungen¹ in der Sozialarbeit weiter beobachtet werden, dass eine Interessensgruppe nicht präsent ist. Die Stimme der KlientInnen fehlt. ExpertInnen werden befragt, SozialarbeiterInnen, GeschäftsführerInnen, TherapeutInnen und andere Professionisten interviewt – aber nur allzu selten werden die Kinder, Jugendlichen, Älteren, Behinderten, Prostituierten, Haftentlassenen gehört. In den implizierten Verzerrungen sehe ich großen Bedarf an methodologisch-kritischen Debatten.

Eine für die Fachpraxis und -ausbildung und für die zu erwartende Professionalisierung der Evaluationen in der Sozialarbeit ermutigende Entwicklung zeichnet sich im fixen Platz von Evaluationen im Fachhochschulwesen ab. Studiengänge unterziehen sich alle Jahre einer Evaluation seitens des Fachhochschulrates und alle fünf Jahre einer Bedarfs- und Akzeptanzanalyse. Mit der kontinuierlichen Präsenz von Evaluationen in Lehre und Studium der Sozialarbeit werden Ansprüche und Anforderungen an Evaluationen für Studierende und später PraktikerInnen, für Lehrende und Forschende prägnanter.

3 Stand der Evaluation von Sozialarbeit in Diskurs und Projekten

Pointiert gefasst befinden sich Evaluationen in der Sozialarbeit in Österreich zwischen selbstgestrickter Selbstbeurteilung und administrativem Auftragsvollzug; auch wenn die Notwendigkeit zur Evaluation und zu deren Professionalisierung erkannt ist. Häufig mutieren vormals schlicht KlientInnen- und MitarbeiterInnenbefragungen oder Arbeitsaufzeichnungen – mit Helmut Kromrey „Akzeptanzforschung“ (2006a: 106) – genannte Aktivitäten nun zu „Evaluationen“, ohne dass aber konzeptuell, methodologisch oder methodisch etwas geändert würde. Solcherart Evaluationen werden durchgeführt, doch mangelt es noch an spezifischer Evaluationsforschung und ausgewiesener wissenschaftlicher Literatur zur Sozialarbeitsevaluation. Evaluationsstudien beziehungsweise -ergebnisse werden als „Graue Literatur“ nur langsam oder – etwa in Bewahrung von Auftraggeberinteressen – meist gar nicht publik. Es bedarf einer neuen Evaluationskultur.

Im Diskurs der Profession Sozialarbeit war die Evaluationsthematik marginal präsent. Aus einer Durchsicht von 24 Ausgaben des Organs des Österreichischen Bundesdachverbands der Sozialarbeiterinnen und Sozialarbeiter, der Zeitschrift SiÖ, der Jahre 2001 bis 2006 – mit über 200 Beiträgen – wird dies evident. Der Diskussionsstand darin zur Evaluationsthematik ist ernüchternd: es war nicht ein Beitrag auffindbar, in dem diese explizit aufgegriffen worden wäre. Natürlich sind die Beurteilung der Sozialarbeit, Qualitätssicherung, Leistungsmessung, die Erbringung von Wirkungsnachweisen, Fragen der Planung und Steuerung und hierfür geeigneter Instrumente thematisiert. Evaluationen, Beurteilungen, Einschätzverfahren sind virulente Themen, impliziert in der Professionsdebatte oder rudi-

¹ In dieser Unterscheidung folge ich den Auslegungen von Suchman (1967, zit. nach Wottawa/Thierau 2003: 13) oder auch Scriven (1991: 139), demnach ich „Evaluation“ als eine unsystematisch auf alltagstheoretische bzw. berufspraktische Erfahrung begründete Beurteilung und „Evaluationsforschung“ als eine ausdrücklich in wissenschaftlicher Forschungslogik begründete systematisierte Beurteilung verstehe; eine andere begriffliche Auslegung findet sich in den Standards der DeGEval (2002: 36), wonach „Evaluationsforschung“ eine übergeordnete Profession zur Sammlung und Entwicklung von „Evaluationen“ und im Sinne einer Beforschung von Evaluationen zu verstehen wäre.

mentär ausgedrückt in Belangen der Methodenkompetenzen der SozialarbeiterInnen (Popp 2003: 15), oder in Beiträgen zu Controlling und Finanzierungspolitik (Soratori 2004: 28; 2005: 33). Auf einer Bundestagung im Jahr 2004 erarbeitete ein Arbeitskreis Fragen der Schlüsselkompetenzen in der Profession „Soziale Arbeit“ beziehungsweise zu den sozialarbeiterischen Grundqualitäten; in der Zusammenfassung (Tötschel 2004: 23) kommt Evaluation nicht vor. In einem den Entwicklungen und Perspektiven der Sozialen Arbeit in Österreich gewidmeten umfangreichen Band (Knapp 2004) widmet sich keiner der 34 Beiträge explizit dem Thema Evaluation. Die Absenz des Themas Evaluation – hier im berufsvertretenden Verband und in neueren Publikationen – stimmt nachdenklich; die Sozialarbeit als Profession enthält sich in Evaluationsdiskurs und -aktivität einer Positionierung und nimmt ihre Agenden erst zögerlich profiliert wahr.

Aktuell initiieren einzelne Entwicklungs- und Forschungsprojekte im Nahbereich sozialarbeiterischer Leistungen Anstöße zur Intensivierung und Ausdehnung von Evaluationen. So bilden insbesondere EQUAL-Partnerschaften² einen potentiellen Rahmen, um Agenden der Sozialarbeit und deren Evaluation zu befördern. Für die Entwicklung von Qualitätskriterien in der Sozialarbeit in Österreich könnte die Entwicklungspartnerschaft „Donau-Quality in Inclusion“ bedeutsam werden (www.donau-quality.at). Sie zielt auf die Erarbeitung von Qualitätskriterien und Richtlinien für deren Anwendung im Sozialarbeitsbereich, um Grundlagen für Ausschreibungen nach dem Prinzip „BestbieterInnen“ zu erstellen – aufbauend auf der aktuellen Vergabepaxis, in der Entscheidungen immer mehr zu Gunsten von „BilligbieterInnen“ getroffen werden. Die Projektagenda liegt in der Stärkung von sozialen Organisationen und der Einführung von Qualitätskriterien zur Bestimmung von Ausschreibungen. Dabei wird die Situation aufgegriffen, dass Organisationen im Sozialbereich überwiegend auf Aufträge der öffentlichen Hand angewiesen sind. Finanzierungen solcher Dienstleistungen werden zunehmend nach dem Vergaberecht durchgeführt. Es ist zu erwarten, dass pauschale Förderungen (Subventionen) von Organisationen der Sozialarbeit in Zukunft immer häufiger von einer direkten Bezahlung der sozialarbeiterischen Leistung abgelöst werden. Diese erfolgt über Leistungsverträge, Honorare oder Tagsätze für ambulante und stationäre Betreuungen. Darstellung und Bewertung von Preis und Qualität einer Leistung werden immer maßgeblicher für Ausschreibung, Angebotserstellung und Beauftragung. Um Ausschreibungen nach dem Prinzip „BestbieterInnen“ zu ermöglichen, braucht es die Formulierung von qualitätsorientierten Kriterien für sozialarbeiterische beziehungsweise sozialwirtschaftliche Dienstleistungen, die in einem öffentlichen Diskurs vermittelt werden. Die Sozialarbeit ist herausgefordert, ihre Leistungen zu kommunizieren und objektivierbare Qualitätskriterien zu entwickeln.

Zwar hat sich die Evaluationsthematik in den letzten Jahren verfestigen können, doch der Institutionalisierungsgrad von Evaluationsforschung und -profession in der Sozialarbeit in Österreich ist immer noch als gering bzw. separiert einzustufen. Evaluationsforschende

² www.equal-esf.at (21. Juli 2007): „EQUAL ist eine europäische Gemeinschaftsinitiative mit dem Ziel der Bekämpfung von Diskriminierung und Ungleichheiten im Zusammenhang mit dem Arbeitsmarkt. Die Gemeinschaftsinitiative wird in Österreich vom Bundesministerium für Wirtschaft und Arbeit in Kooperation mit dem Bundesministerium für soziale Sicherheit, Generationen und Konsumentenschutz umgesetzt. Mit der Unterstützung der Abwicklung des Programms – im Auftrag der Bundesministerien – sowie als Ansprechpartner für alle EQUAL Interessierten wurde das EQUAL BÜRO ÖSTERREICH beauftragt. (...)“

beginnen sich erst zu organisieren, unter anderem als Mitglieder in der DeGEval. Da sie überwiegend aus anderen Disziplinen – wie der Soziologie, Psychologie, Sozialpädagogik oder auch Betriebswirtschaftslehre – stammen, sind sie, wenn, disziplinär in ihren Fachgesellschaften organisiert. Fachliche Evaluationsstandards sind weder bei Beauftragenden noch Durchführenden von Evaluationen in der Sozialarbeit ausreichend bekannt. Dem Stellenwert evaluationsspezifischer Theorie und Methodik ist noch deutlich mehr Durchdringung der Evaluationspraxis zu wünschen.

4 Verbindungen mit anderen Evaluationsfeldern und internationalen Entwicklungen

Das Angebot von sozialen Diensten in Österreich, in denen Sozialarbeit zum Einsatz kommt, umfasst eine breite Palette: Familienhilfe und Dorfhelferinnen, berufliche Rehabilitation, arbeitsmarktnahe Angebote, Sucht und Drogenberatung, Bewährungshilfe, Angehörigenberatung, Besuchsdienste, Frühförderungen, Heimhilfe, mobile therapeutische und psychosoziale Dienste, Notruftelefon, organisierte Nachbarschaftshilfe, Sozialservice und viele andere Leistungen. Das Zielgruppenspektrum sind (nach ihren Anteilen) Kinder und Jugendliche, Menschen mit Behinderungen, ältere Menschen, Angehörige, psychisch Kranke, Arbeitslose, Familien, Pflegebedürftige, Kranke, Frauen, Sucht-Kranke, Verschuldete, Gewaltopfer, Ausländerinnen, Schwangere, Haftentlassene, Obdachlose und andere (Loidl-Keil/Posch 2006). Da der Beruf Sozialarbeit an Sozialakademien gelehrt wurde und bis 2005 keinen Hochschul- beziehungsweise Universitätsabschluss voraussetzte, waren und sind höher qualifizierte Positionen in den genannten Tätigkeitsfeldern der Sozialarbeit vorrangig von anderen Quellberufen (v. a. PsychologInnen, SoziologInnen, PädagogInnen, JuristInnen) besetzt – was eine disziplinär eigenständige Entwicklung von Evaluationen in der Sozialarbeit erschwert haben mag.

Die Evaluationsfelder, mit denen Evaluationen in der Sozialarbeit am stärksten assoziieren, sind zuallererst in arbeitsmarktnahen Bereichen und im Gesundheitswesen, in psychosozialen Diensten und beruflicher Rehabilitation zu sehen; in der Arbeit mit Frauen, MigrantInnen, Kindern und Jugendlichen und im Bildungswesen. Deutlicher als Evaluationen bzw. Evaluationsforschungen verbreiteten sich Bemühungen um Qualitätssysteme und Leistungsbestimmungen – im Gefolge von Leistungsvergaben und -vereinbarungen.

Hinsichtlich der Internationalisierung von Sozialarbeit spielen deren Akademisierung, transnationale Projekte – überwiegend EU-(ko)finanziert – internationaler Studierenden- und Lehrendenaustausch, Kooperationen mit ausländischen Fachhochschulen und Universitäten, als auch Initiativen zum fachlichen Austausch in Berufsverbänden eine maßgebliche Rolle. Zudem ist eine höhere Beteiligung im internationalen Fachdiskurs zu erwarten, etwa in bestehenden internationalen als auch durch die Einführung neuer Fachzeitschriften in Österreich (www.soziales-kapital.at). Mit diesen Entwicklungen wird auch der internationale Austausch zu Evaluationen in der Sozialarbeit befördert werden; als wesentlich erachte ich zudem die Mitgliedschaften von in Österreich tätigen Evaluationsbetreibenden in Fachgesellschaften wie der DeGEval.

5 Wird die Sozialarbeit zugänglicher und aktiver für Evaluationen?

Gegenwärtig redefinieren sich Sozialarbeit und deren Evaluationsagenda in Österreich, wir haben es mit neuen Ordnungen zu tun. Sozialarbeiterische Leistungen unterliegen aktuell zusehends Vergaberichtlinien, die zwar qualitative Verbesserungen intendieren, doch in der Praxis häufig zu kontrastierenden Entwicklungen führen. Hödl (2007) argumentiert, wie grundsätzlich schwierig bzw. unmöglich eine Objektivierung sozialarbeiterischer Leistungen ist. Ausschreibung, Anbotslegung und Vergabe werden zu einem Darstellungsritual. Leistungsdarstellung – oder besser „Darstellungskunst“ – riskiert zum entscheidenden Qualitätskriterium zu werden, da sie zum auswahlentscheidenden Kriterium vor der eigentlichen Leistungserstellung wird. Ein neues Tätigkeitsfeld ist dies nicht so sehr für EvaluatorenInnen, sondern für Public Relations- und Marketing-SpezialistInnen.

Regelungen, mit denen Qualitäten erhöht und Wirksamkeiten verbessert werden sollen, benachteiligen strukturell kleinere Organisationsformen und sprechen deutlich eine Marktsprache. Geht es dann um die Beurteilung sozialarbeiterischer Leistungsqualitäten und Leistungswirkungen, werden verschiedenste Professionen und Berufe aktiv. Wir beobachten zudem isomorphe Prozesse – mit noch abzuwartenden Effekten. Vor allem aus Industrie und Wirtschaft werden Konzepte und Techniken zur Leistungsfeststellung und -beurteilung voreilig übernommen, die Praxis droht ob der herbeigeredeten selbstverständlichen Eignung industrieller bzw. wirtschaftlicher Instrumente auch für sozialarbeiterische Leistungen in vielen Fällen gegenüber den Konsequenzen von Qualitätshandbüchern und Zertifizierungen für Denkweisen und Handlungstechniken abzustumpfen. An die Stelle des klientInnen-orientierten Engagements tritt der brave Einhaltungswille gesetzter Ordnungen, an die Stelle des Glaubens an die situative Kraft der Hilfsentfaltung der Glaube an die korrekte Arbeitsschrittausführung im Interventionsplan. Es wird etwas gemacht, weil es so im Handbuch steht, und nicht aufgrund einer Einschätzung in der Begegnung mit einem Menschen in Not. Bezugspunkte und Kriterien zur Beurteilung, respektive Evaluation, der Sozialarbeit ändern sich substantiell – in der Sprache von Max Weber treten bürokratische an die Stelle moralischer Prinzipien.

Die Sozialarbeit als Profession in Österreich blieb in Evaluationsdiskurs und Evaluationsforschung bislang weitgehend abstinert. Vermeintlich, denn der Evaluationsdiskurs ist zwar mitnichten gering ausgeprägt, kaum disziplinär oder institutionalisiert und unspezifisch. Er vollzieht sich aber mittelbar im Auseinandersetzen mit und Finden theoretisch fundierter Positionen. Denn Zugangsweisen und Verständnisse von Evaluationen sind von Theorien und Auslegungen abhängig. Evaluationen und evidenzbasierte Beurteilungen sozialarbeiterischer Leistungen können nicht losgelöst von den zugrunde gelegten Interessen, Ausrichtungen oder bewusst gewählten Parteilichkeiten betrachtet werden. Einig ist sich die Profession, dass Evaluationen wissenschaftlich betrieben werden müssten. Wie Prinzipien der Wissenschaftlichkeit ausgelegt und erreicht würden, welche Evidenz herangezogen oder wie Verlässlichkeit methodisch hergestellt wird, dazu finden wir allerdings recht unterschiedliche Positionierungen vor.

Evaluation heißt Urteilsfindung. Darin fließen Vorstellungen über Gerechtigkeit, beobachtete und wünschbare Lebensumstände, Reflexionen über Bedürftigkeiten und erforderliche Interventionen, Diskrepanzen der eigenen Erfahrung mit denjenigen Lebenswirklich-

keiten von Dunkelheit und Desillusion, mit denen Sozialarbeit zu tun hat, unter anderem mit ein, ebenso wie individuell variierende Wissenschafts- und Evaluationskompetenzen. Ein Urteil ohne Standpunkt und Wertreferenzen ist nicht möglich. Evaluation von Sozialarbeit kann kein Selbstzweck sein, sondern soll Urteilsfindung und Verständnis von Beschaffenheit und Zusammenwirken relevanter Dimensionen in unserer Handlungspraxis fördern. Gerade wenn Evaluation aufklärend für die Fachpraxis sein soll, wird sie theoretisch gut fundiert sein und die Komplexität des sozialarbeiterischen Leistungsgeschehens erhalten müssen (Loidl-Keil/Posch 2006). Zur Evaluation von Sozialer Arbeit in Österreich stehen noch etliche Anstrengungen aus; und es bleibt zu wünschen, dass die Soziale Arbeit in Österreich für Evaluationen zugänglicher und aktiver wird.

Evaluationen in der Sozialen Arbeit in der Schweiz: Eine Einschätzung der Situation aus professionsorientierter Sicht

1 Einführende Bemerkungen und Evaluationskontexte

Der vor allem in Deutschland vorangetriebenen Debatte um Qualität, Qualitätsstandards, Rahmenrichtlinien oder Zertifizierung ist es zu verdanken, dass sich die Evaluationstätigkeit in der Sozialen Arbeit von einem reinen Untersuchungsauftrag zu einem mehrdimensionalen Ansatz mit dem Anspruch, *Evaluationsforschung* zu sein, verbreitert hat. In die Schweiz ist dieser Anspruch und auch die damit verbundene Fachdiskussion erst in Teilen diffundiert. Von einer systematischen Evaluationstätigkeit sollte deshalb noch nicht gesprochen werden.

Dies betrifft auch Programme und Projekte der Kantone und Gemeinden in den Handlungsfeldern der Sozialen Arbeit. Obwohl mit Artikel 170 der Bundesverfassung verlangt wird, dass die Massnahmen des Bundes auf ihre Wirkungen hin zu überprüfen sind, profitieren die Handlungsfelder der Sozialen Arbeit erst punktuell: So formuliert das Bundesamt für Gesundheit (BAG) die Absicht, alle in der Schweiz existierenden niederschweligen Hilfsangebote im Drogenbereich zu evaluieren und das Bundesamt für Justiz lässt die geförderten Modellversuche im Straf- und Massnahmenvollzug evaluieren. Hinzu kommt, dass eine Verpflichtung zur Evaluation auf der Ebene der Kantone und Gemeinden mit dem Bundesverfassungsartikel nicht verbunden ist, was dazu beiträgt, „dass die Kantone in Bezug auf Evaluationen ... eher passiv sind und die Tendenz haben, die Aktivitäten anderer Kantone zu ignorieren.“ (Kaufmann 2006: 3). Somit sollte im Moment von einer *sich entwickelnden* schweizerischen Evaluationsforschung in den Handlungsfeldern der Sozialen Arbeit gesprochen werden.

Dies, aber auch der Umstand, dass keine spezifische Datenbank vorliegt, welche die Evaluationen in den Handlungsfeldern der Sozialen Arbeit bündelt, sowie das Fehlen einer Fachdiskussion mit entsprechender Publikationstätigkeit (auf die sich der vorliegende Beitrag hätte beziehen können), macht eine Fokussierung des vorliegenden Beitrages notwendig. Betrachtet werden hier Evaluationen, die einen wissenschaftlichen Anspruch deklarieren, unter anderem indem sie einer Publikationsaufforderung nachkommen (insbesondere durch Eintrag in eine Forschungsdatenbank). Evaluationen, die unveröffentlicht sind beziehungsweise durch eine gezielte Nachfrage bei privaten Büros, Einzelpersonen oder Auftraggebern hätten eruiert werden müssen, sind hier nicht berücksichtigt. Dagegen werden alle Evaluationen einbezogen, die sich auf die Handlungsfelder der Sozialen Arbeit (Handlungsfeldbezug) beziehen, also nicht nur solche Evaluationen, die aus dem Kreis der

Sozialen Arbeit selbst (Disziplinbezug) entstanden. Damit wird also auch die Evaluationstätigkeit anderer Disziplinen (Psychologie, Jurisprudenz etc.) berücksichtigt, soweit sie sich auf Handlungsfelder der Sozialen Arbeit beziehen (z. B. Evaluation der öffentlichen Drogenabgabestelle durch das Institut Sozial- und Präventivmedizin).

Insofern handelt es sich bei den folgenden Aussagen in erster Linie um eine Standortbestimmung der wissenschaftlichen Evaluationstätigkeit in den Handlungsfeldern der Sozialen Arbeit.

2 Darstellung der Evaluationsaktivitäten

2.1 Methodisches Vorgehen

Zur Skizzierung der Evaluationslandschaft in den Handlungsfeldern der Sozialen Arbeit wurde ein kombiniertes Vorgehen aus Forschungsdatenbankrecherche, Dokumentenanalyse sowie Schlüsselpersonenbefragung realisiert:

Forschungsdatenbankrecherchen beziehen sich auf die Projektdatenbank SIDOS (Schweizerisches Informations- und Datenarchiv) sowie die Datenbank des Schweizerischen Nationalfonds SNF. Der Eintrag in SIDOS erfolgt durch die Evaluierenden selbst (Selbstdекlaration), impliziert also generell keine Einschränkung auf Evaluationen von Universitäten oder Hochschulen. Der Eintrag in die Datenbank des SNF dagegen wird durch die Institution selbst vorgenommen und bedingt, dass das eingetragene Projekt durch eine der Fördermassnahmen des Bundes unterstützt wird. Berücksichtigt wurden alle Einträge in den Datenbanken bis zum Stichtag 30. November 2006. Gesucht wurde jeweils separat sowie kombiniert mit thematischen Schlüsselbegriffen („Evaluation“, „Wirkung“, „Wirksamkeit“) sowie entlang den Handlungsfeldern („Soziale Arbeit“, „Sozialarbeit“, „Sozialpädagogik“, „Soziale Animation“, „Sozialpolitik“).

Die Dokumentenanalyse (Stichtag 30. November 2006) bezog sich auf das Bulletin der Schweizerischen Evaluationsgesellschaft (SEVAL) und sollte Hinweise auf Evaluationsaktivitäten ausserhalb der in den Eintragungen der Forschungsdatenbanken aufgeführten Evaluationen geben. Auf diese Weise konnten 75 weitere Evaluationen erfasst werden, welche die Arbeitsfelder der Sozialen Arbeit betreffen.

Weil vermutet wurde, dass sich Evaluationen nicht ausschliesslich mit den oben genannten engen Suchbegriffen finden lassen, wurden im Rahmen eines Versands im Dezember 2006 Schlüsselpersonen der schweizerischen Hochschulen für Soziale Arbeit angeschrieben und aufgefordert, a) die Evaluationstätigkeit der Organisation darzustellen und entsprechende Studien zur Verfügung zu stellen und b) die Anfrage weiterzuleiten (Schneeballsystem). Auf diesem Weg konnten weitere 37 Studien aus Institutionen der Sozialen Arbeit in die Analyse eingehen. Diese Studien stellen allerdings Evaluationen der Sozialen Arbeit dar; hier dominiert also der Disziplinbezug den Handlungsfeldbezug.

Insgesamt standen so für eine folgende Systematisierung 112 Evaluationsstudien zur Verfügung. Aufgrund des Vorgehens bilden dabei umfangreichere Studien aus dem Hochschulbereich einen Schwerpunkt; Evaluationsstudien aus der Praxis sind dagegen kaum vertreten.

2.2 Evaluationsaktivitäten: eine Systematisierung

Nach *thematischen Schwerpunkten* ist der Gesundheitsbereich Schwerpunkt der Evaluationsaktivitäten (siehe Tabelle 2 am Schluss des Beitrags). 49 Evaluationen untersuchen weitgehend gleichberechtigt Programme, Organisationen respektive Organisationseinheiten sowie Kampagnen. Eher weniger stehen Evaluationen zu Generalthemen, Fortbildungen oder biographischen Verläufen zur Verfügung. Evaluationen von Beschäftigungs- und Integrationsprogrammen sowie arbeitsmarktpolitischer Massnahmen weisen darauf hin, dass mit insgesamt 15 Evaluationsstudien auch das Handlungsfeld Berufseinmündung und Arbeitsmarkt von Wichtigkeit ist. Eine ähnliche Evaluationsdichte haben die Themen Stadt/Wohnen (11 Evaluationsstudien), Strafvollzug/Devianz (10), physische und psychische Beeinträchtigung/Traumatisierung (9) sowie das Thema Familie und Heimwesen (7). Eher weniger Evaluationsstudien beschäftigen sich mit den sozialen Sicherungssystemen wie Sozialhilfe; hier sind vier Evaluationsstudien nachgewiesen.

Nach *Bereichen* systematisiert sind es vor allem regional verortbare Programme, Projekte und Institutionen, die evaluiert werden. In nahezu allen thematischen Schwerpunkten finden sich hier Konzentrationen der Evaluationstätigkeit. Auffallend wenige Evaluationen legen ihren Schwerpunkt auf Methoden der Sozialen Arbeit.

Nach *Ausführenden* sind es insbesondere die Institute der Universitäten (und hier vor allem Institute für Sozial- und Präventivmedizin sowie für Suchtforschung) und der Hochschulen für Soziale Arbeit, die Evaluationen durchführen. Diese Institutionen decken 72.3 Prozent aller erfassten Evaluationen ab. In einem zweiten Cluster (23.2 %) finden sich private Büros, Stellen der Verwaltung sowie die Pädagogischen Hochschulen. Wenige Evaluationsaktivitäten (4.5 %) werden von anderen Instituten der Fachhochschulen sowie Nichtregierungsorganisationen durchgeführt. Dieser Befund sollte mit Bedacht interpretiert werden, weil davon ausgegangen werden kann, dass es vor allem die Hochschulen sind, die Einträge in die Forschungsdatenbanken vornehmen. Dass sich die universitäre Evaluations-tätigkeit und die der Fachhochschulen für Soziale Arbeit thematisch unterscheiden, wird in der Kreuztabellierung deutlich. Während die Universitäten ihren Schwerpunkt auf die Themen Gesundheit/Sucht/HIV und Aids legen (27 von insgesamt 43 Evaluationen), ist die Evaluationstätigkeit der Fachhochschulen für Soziale Arbeit diversifiziert über nahezu alle erfassten Themen.

Finanziert werden Evaluationsstudien vor allem von den kantonalen und städtischen Verwaltungen sowie dem Bundesamt für Gesundheit (51 %). In einem zweiten Cluster finden sich die Träger sowie die beteiligten Institutionen selbst und das Programm DoRe des Schweizerischen Nationalfonds (20.5 %). Organisationseinheiten der Bundesämter sowie thematische Programme des Schweizerischen Nationalfonds und Stiftungen finanzieren rund 18 Prozent der Evaluationsstudien.

3 Stand der Evaluationen: Eine Einschätzung aus professionsorientierter Sicht

Die Analyse und Systematisierung von Evaluationen sollte auch im Hinblick auf die Frage vorgenommen werden, inwieweit die Soziale Arbeit als wissenschaftliche Disziplin von der Evaluationstätigkeit profitiert. Gemeint ist damit, ob es sich um reine Auftragsforschung handelt, also die Fragestellung und Zielsetzung der Auftrag gebenden Institution im Vordergrund steht, oder ob und in welcher Form Aspekte einer „wissensgenerierenden“ Evaluation, die eine über den direkten Nutzen für den Auftraggeber hinausreichende Entwicklung des Fachs anstrebt, sichtbar werden (siehe Haubrich et al. 2005: 4).

Dies zu betrachten ist aus professionsorientierter Perspektive interessant, da es sich als illusorisch erwiesen hat, von einem direkten Transfer wissenschaftlichen Wissens in berufliches Handeln auszugehen (siehe dazu Gredig et al. 2003). Diese Annahme, die eine expertokratische Vorstellung der Wissensverwendung vertritt, ist einem Modell gewichen, das Unterschieden zwischen wissenschaftlichem Wissen, Erfahrungswissen und professionellem Wissen Rechnung trägt. Wissenschaftliches Wissen wird hier mit einem Fallverständnis relationiert, oder anders ausgedrückt: Das Verhältnis von sozialwissenschaftlicher Theorie und professioneller Handlungspraxis erfährt durch eine auf Wissensgenerierung abzielende Evaluationstätigkeit eine Überbrückung.

Wie sehr kann dieses Ziel einer Wissensgenerierung jenseits des konkreten Evaluationsauftrages in bisherigen Studien berücksichtigt werden? Um diese Fragen zu beantworten, soll auf eine einfache Systematik mit zwei Kriterien Rückgriff genommen werden:

1. Einbringen einer eigenen Fragestellung, d.h. im Fragekatalog der Evaluation wird explizit eine disziplinspezifische Frage formuliert, die sich vom eigentlichen Evaluationsgegenstand abhebt;
2. Einbettung der Ergebnisse in den nationalen und internationalen Diskurs mit dem Ziel, Ergebnisse durch eine Auseinandersetzung mit theoretischen Diskursen von ihrer historischen Einmaligkeit zu entkern: entweder durch das Aufzeigen des Standes der Forschung im Hinblick auf das Evaluationsthema und das Wiederaufnehmen im Fazit des Evaluationsberichtes; oder durch Verallgemeinerungen wie das Entwickeln eines Modells oder die Frage nach der Übertragbarkeit der unmittelbaren Evaluationsergebnisse.

In die Analyse einbezogen wurden 26 Evaluationsstudien aus den Hochschulen für Soziale Arbeit, die durch die Mailingumfrage vorgelegt wurden (siehe Tabelle 3 am Schluss des Beitrags).

Tabelle 1: Typen von Evaluationsstudien aus professionsorientierter Sicht

| Typ Evaluationsstudie | Finanzierungsquelle | Eigene, professionsspezifische Fragestellung; explizite Formulierung | Theoretische Einbettung | | Verallgemeinerung | |
|--|--|--|--|--|---|----------------------------|
| | | | Stand der Forschung | Im Fazit aufgegriffen | Modellentwicklungen | Frage nach Übertragbarkeit |
| Auftragsorientiert wissensverwendend (N=16) | Träger | Keine | „Ausgangslage“ | Nein | Nein | Nein |
| Auftragsorientiert wissensvermittelnd (N=7) | Träger, Stiftungen, Schweizerischer Nationalfonds (DoRe) | Angedeutet | „Stand der Forschung“ „Zum theoretischen Hintergrund“ | Nicht konsequent | Nur für konkreten Evaluationsgegenstand | Angedeutet |
| Professionsorientiert wissensgenerierend (N=3) | Schweizerischer Nationalfonds (NFP, DoRe) | Nahezu gleichberechtigt mit dem Evaluationsgegenstand | „Stand der Forschung“ | Ja, und daraus neue Fragestellungen abgeleitet | Mit Bezug auf Stand der Forschung | Oft |

Aus professionsorientierter Sicht charakterisieren demnach drei Typen von Evaluationen die aktuelle schweizerische Evaluationslandschaft der Sozialen Arbeit:

Typ „Auftragsorientiert wissensverwendend“

Dieser Typ von Evaluationen wird in der Regel von den Trägern in Auftrag gegeben, die selbst für die Finanzierung aufkommen oder an andere Stellen (vor allem die Verwaltung) Gesuche auf finanzielle Unterstützung stellen. Die Fragestellungen der Evaluationen richten sich an den konkreten Bedürfnissen der Auftraggeberin aus, in vielen der analysierten Fälle handelte es sich um Evaluationen, die die Gesamtwirkung eines Projektes (Impact) oder deren Prozess- und Strukturqualität messen sollen. Fragestellungen wie „Werden die Ziele des Projektes erreicht?“, „Welche sichtbaren Wirkungen haben die Interventionen?“, „Wird die Präsenz der Teams von der Bevölkerung wahrgenommen?“ „Wie verändert sich das Verhalten der Zielgruppe?“ et cetera (Willi et al. 2005: 4-5) oder „Was hat das Kinderbüro erreicht und was kann es bewirken?“ (Fabian 2004: 7) zeigen die Themen- und Ortsgebundenheit der Evaluationen.

Eine eigene, aus der Perspektive der Sozialen Arbeit als wissenschaftliche Disziplin adaptierte Fragestellung wird in diesem Evaluationstyp nicht berücksichtigt, es dominiert die Kundenorientierung: „Die vorliegende Evaluation hat die Funktion, nach dem Pilotversuch Bilanz zu ziehen und Grundlagen für die Entscheidung ‚Weiterführen oder Abbruch des SIP-Projektes‘ zu erarbeiten“ (Willi et al. 2005: 6); „... wobei es einerseits darum ging, die Coachingpraxis detailliert zu beschreiben und deren Stärken und Schwächen aufzude-

cken, andererseits herauszufinden, welche Wirkungen (beabsichtigte und unbeabsichtigte) durch das Coachingprogramm bei den verschiedenen Beteiligten ausgelöst werden“ (Kobi 2004: 2); „Wie soll sich das kantonale basel-städtische Gesamtkonzept für schulintegrierte und schulnahe familienergänzende Tagesbetreuung gestalten und mit welcher Strategie ist es unter den gegebenen Voraussetzungen bestmöglich realisierbar?“ (Schaffner 2003: 8).

Das methodische Vorgehen „ummantelt“ diese Fragestellungen. Auch wenn die Methoden der Datenerhebung eine grosse Bandbreite aufweisen und von quantitativen Methoden (standardisierte Befragungen z. B. bei Steiner et al. 2006) bis zu qualitativen Methoden (z. B. Beobachtung bei Willi et al. 2005; halbstandardisierte Leitfadeninterviews bei Riedli 2003, Leitfadengespräch bei Mattes/Fabian 2007), von Primärerhebungen bis zur Auswertung von Sekundärdaten reichen, liegt das Schwergewicht auf den qualitativen Methoden, vor allem dem Leitfadengespräch und der Dokumentenanalyse.

Die Datenauswertung ist entlang den Fragestellungen der Evaluationsauftraggebenden aufgebaut und schliesst mit einem Fazit, das die Fragen nochmals übersichtsartig verdichtet. Weil es keine Berichtsteile gibt, die den Stand der Forschung aufarbeiten, werden die Ergebnisse der Evaluation nicht in einen verallgemeinerbaren Zusammenhang gestellt resp. wird die Frage nach einer Übertragbarkeit der Ergebnisse nicht diskutiert.

Weil sich die theoretischen Kenntnisse der Auftragnehmenden nicht in explizierter Form in den Berichten wiederfinden, wird dieser Evaluationstyp als *wissensverwertend* bezeichnet, d.h. das vorhandene Wissen wird durch das Evaluationsteam selbst infolge einer Reflexion des Erhobenen direkt und unmittelbar verwertet. Die Auftraggebenden partizipieren an diesem Wissen nicht dadurch, dass es im Bericht diskutiert wird, sondern durch die Adaption auf das konkrete Beispiel durch „wissende“ Evaluatorinnen und Evaluatoren.

Typ „Auftragsorientiert wissensvermittelnd“

Ebenfalls von Seiten der Träger in Auftrag gegeben, unterscheiden sich diese Evaluationen vom obigen Typ namentlich bezüglich der Einbettung des Evaluationsgegenstandes in einen theoretischen Rahmen. „Zum Stand der Forschung“ oder „Theoretischer Hintergrund“ lauten die Überschriften der Kapitel, in denen die Auftragnehmenden versuchen, die Fragestellungen in einen weiteren, oftmals sozialpolitischen Diskurs zu stellen: „Aufgrund des breiten Spektrums der Mobilen Jugendarbeit drängt sich die Frage nach einer Definition des Ansatzes auf. Eine Definition ermöglicht ein gemeinsames Begriffsverständnis und regelt den gemeinsamen Begriffsgebrauch“ leitet Fabian (2002: 3) seine Evaluation der Mobilen Jugendarbeit Basel ein, um sich anschliessend auf der Basis der vorhandenen Literatur mit Grundtypen, Grundprinzipien und Methoden Mobiler Jugendarbeit auseinanderzusetzen und sie gegenüber anderen Formen von Jugendarbeit (Schulsozialarbeit und Jugendarbeit im offenen Freizeitbereich) abzugrenzen. Drilling et al. (2006) stellen ihrer Evaluation der Schulsozialarbeit im Fürstentum Liechtenstein drei Begründungen von Schulsozialarbeit (sozialpolitisch, sozialpsychologisch, bildungstheoretisch) voran und diskutieren diese unter Bezugnahme auf die aktuelle Theoriedebatte in der schulbezogenen Jugendhilfe.

Vor diesem Hintergrund des im Evaluationsbericht *explizit* diskutierten Standes der Forschung werden in einem anschliessenden Teil die konkreten Fragen der Auftraggebenden bearbeitet. Auf eine aus der Forschungsperspektive der Evaluierenden generierte eigene Fragestellung wird auch in diesem Typ verzichtet, vielmehr geht es darum, dem Auf-

traggebenden das Spektrum der Thematik aufzuzeigen, damit dieser die spätere Einordnung seines Projektes in einen Gesamtzusammenhang nachvollziehen kann. Weil es dabei vor allem um die Vermittlung von Wissen (an die Auftraggebenden) geht, wird dieser Typ Evaluation als *auftragsorientiert wissensvermittelnd* bezeichnet. „Das Hauptziel der vorliegenden Evaluation besteht darin, ein *vertieftes Verständnis* der Wirkungsweise dieses Systems der offenen Jugendarbeit zu erreichen“, so Sommerfeld et al. (2006: 7; Hervorhebung im Original). Entsprechend wird aus den qualitativen Ergebnissen ein Wirkungsmodell entwickelt, das die Systemdynamik der „Midnight Basketball“-Projekte und ihre Wirkungsweise im Hinblick auf die von „Midnight Basketball“ ausgewiesenen Zieldimensionen „Chancengleichheit“, „Gesundheitsförderung“, „Gewaltprävention“ und „soziale Integration“ aufzeigen soll.

Methodisch zeichnen sich die Evaluationsstudien durch ihre Triangulation aus. Vor allem treten zusätzlich zu den qualitativen Methoden (siehe Typ 1) quantitative Methoden hinzu, insbesondere die standardisierte Befragung, zumeist in Form des Fragebogens (z. B. bei Fabian/Gutmann 2004, Fabian et al. 2006).

Auch wenn in diesem Typ Evaluationen der Darstellung des Forschungsstandes breiter Raum gegeben wird, werden die Ergebnisse im Teil „Fazit“ nur ansatzweise wieder auf diesen Forschungsstand zurückgeführt. Wo Verallgemeinerungen realisiert werden (z. B. bei Sommerfeld et al. 2006 oder Keller et al. 2003) beziehen sich diese auf Modellentwicklungen für den konkreten Evaluationsgegenstand. Die Frage nach der Übertragbarkeit der Ergebnisse auf andere Handlungsfelder der Sozialen Arbeit thematisiert dieser Evaluationstyp kaum. Die Konsequenzen für professionelles Handeln beschränken sich auf das evaluierte Projekt, gehen also in der Regel nicht über den konkreten Evaluationsgegenstand hinaus.

Typ „Professionsorientiert wissensgenerierend“

Drei der 26 in die Analyse einbezogenen Studien gehören dem Typ „professionsorientierter wissensgenerierender Evaluationen“ an (Sommerfeld/Castelli 2006; Eser Davolio/Drilling 2007; Kundert/Sommerfeld 2001).

Markantester Unterschied zu den bisher behandelten Typen ist die Anreicherung des Auftrages durch eigene Fragestellungen. „Bei der Evaluation der Quartierkommission Länggasse wurde die übergeordnete Frage gestellt, welche Bedingungen nötig sind, damit die Mitwirkung von Betroffenen in der Stadtplanung und somit die Gestaltung des Stadtteils gelingt“, schreiben Kundert und Sommerfeld (2001: 14), und später: „Welche Anhaltspunkte werden bei der Evaluation der Quartierkommission Länggasse gefunden, wie Verwaltungen und Betroffene die Forderung nach Mitwirkung in anderen Berner Stadtquartieren und weiteren Städten erfolgreich umsetzen können (Frage nach der Übertragbarkeit des Beispiels)?“ (ebd.: 14). Eser Davolio und Drilling (2007) fragen „Was kann Soziale Arbeit in Bezug auf Rechtsextremismus ausrichten? Unter welchen Bedingungen? Mit welchen Ansätzen?“, und versuchen, aus den Evaluationsergebnissen „Guidelines“ zu entwickeln, die Städte und Gemeinden im Umgang mit rechtsextremistischen Vorfällen anleiten.

Wie sehr neben den von den Auftraggebenden formulierten Fragestellungen der Evaluation die auf die Profession Soziale Arbeit abzielende Fragestellung (die aus dem Kreis der Evaluierenden stammt) Einfluss auf die Evaluationstätigkeit hat, formulieren Sommer-

feld und Castelli (2006: 25), deren Arbeit quasi als idealtypisch für diesen Evaluationstyp gelten kann, in ihrer Evaluation zweier Wohnsiedlungen in Zürich: „Die erste Zielsetzung besteht also darin, die in der Zwischenzeit stattgefundenene Entwicklung der in den Siedlungen vorgefundenen Sozial- und Wohnformen zu beschreiben und vor allem im Hinblick darauf zu analysieren, ob sich dort Hinweise finden lassen, ob und wenn ja welche Faktoren dazu beigetragen haben, die Tendenz zur ‚sozialen Verarmung‘ des sozialen Raums nachhaltig zu durchbrechen... Diese Zielsetzung soll zur Weiterentwicklung der theoretischen Grundlagen und Konzeptionen und Verfahren der sozialraumorientierten Sozialen Arbeit beitragen und adressiert neben der Sozialen Arbeit die Stadtsoziologie, die Stadtplanung und die Sozialpolitik.“

Als Verallgemeinerungen werden Modelle vom sozialen Raum (Sommerfeld/Castelli 2006: 15), von tragfähiger Alltagskultur (ebd.: 14) entwickelt bzw. anhand der Evaluationsergebnisse präzisiert, im Fazit dargelegt und diskutiert. Im Gegensatz zu den Evaluationen der anderen Typen findet sich bei Sommerfeld und Castelli im Schlusskapitel ein Rückbezug auf den in den einleitenden Kapiteln dargestellten Stand der Forschung sowie die Formulierung offener Fragen bzw. eines Ausblicks: „Fazit: Es lässt sich festhalten, dass aufgrund unserer Ergebnisse und der Interpretation dieser Ergebnisse die sozialpolitischen Hoffnungen in den kleinräumigen Sozialraum als Quelle sozialen Kapitals zur Bearbeitung von sozialen Problemen als überzogen bezeichnet werden müssen.“ (ebd.: 96) Und später: „Wir haben mit der Untersuchung der Reichweite der sozialen Kapitalbildung in privilegierten Verhältnissen einen Beitrag geleistet, der unseres Erachtens die Grenze der aktuell hoch gehandelten Sozialraumorientierung aufzeigt.“ (ebd.: 97) Stattdessen präferieren die Autoren das Konzept der „tragfähigen Alltagskultur“ – „eine notwendige Zielvorstellung, die professionelle Orientierung in den sozialpolitischen Wechselspielen verleiht“ (ebd.: 97).

Insofern wird von *professionensorientiert wissensgenerierenden Evaluationen* gesprochen: Es ist wie eine zweite Forschungsebene, die zum eigentlichen Auftrag hinzukommt, wobei die Fragestellung, die generiert wird, eine Passung zwischen Evaluationsauftrag und Forschungsinteresse der Sozialen Arbeit darstellt (in diesem Zusammenhang fällt auf, dass alle Evaluationsstudien dieses Typs mit einer Forschungsförderung durch eine am engeren Projekt unbeteiligte Institution, hier des Schweizerischer Nationalfonds, einhergehen). Dieses so erworbene Wissen wird „offen“ generiert, d.h. im Bericht unter Rückgriff auf die Literatur diskursiv dargelegt. Die Evaluation von Sommerfeld und Castelli (2006) enthält so drei Zieldimensionen:

- Die Genossenschaftsverwaltungen der Wohnsiedlungen erhalten Antworten über die Soziabilität der Bewohnerinnen und Bewohner (Evaluation im engeren Sinn),
- die Genossenschaftsverwaltungen sowie die am Fachdiskurs interessierten Fachpersonen erhalten anhand eines konkreten Beispiels Einblicke in eine aktuelle fachwissenschaftliche Debatte (Wissenstransfer),
- im Arbeitsfeld „Sozialraum“ tätige Fachpersonen erhalten ein theoriebasiertes Argumentarium, das auch die internationale Debatte berücksichtigt, sowie Überlegungen für weitere Aktivitäten offeriert (Professionalisierung).

4 Fazit

Die Evaluationslandschaft in der Schweiz im Feld der Sozialen Arbeit darzustellen, stösst auf unterschiedliche Schwierigkeiten: Datenverfügbarkeit und fehlender Diskurs über Evaluationen im Bereich Soziale Arbeit sind die beiden gewichtigen Argumente für eine Fokussierung der Ausführungen, wie sie im vorliegenden Beitrag vorgenommen wurde. Folglich gründet die Aussagekraft des Beitrags auf einer Recherche in Datenbanken sowie einer Dokumentenanalyse. Damit einher ging die Beschränkung auf zwar selbstdeklarierte, aber generell doch wissenschaftliche Evaluationen. Dargestellt wurden in einem ersten Schritt Evaluationen, welche die Handlungsfelder der Sozialen Arbeit betreffen, wobei sich zeigt, dass der Gesundheitsbereich vor allem für die Nachbardisziplinen der Sozialen Arbeit ein wichtiger Evaluationsbereich ist. Evaluationen, die aus den Institutionen der Sozialen Arbeit selbst stammen, sind themenunspezifisch. Für letztere Forschung hat insbesondere die Schaffung der speziellen Fördereinrichtung DoRe des Schweizerischen Nationalfonds Perspektiven eröffnet.

In einem zweiten Schritt wurden die von Seiten der Hochschulen für Soziale Arbeit angefertigten Evaluationen einer inhaltlichen Analyse unterzogen. Dabei interessierte speziell, ob und in welcher Form die Soziale Arbeit generell von den Evaluationen profitieren kann. Die vorgeschlagene Drei-Typen-Lösung könnte sich aus professionsorientierter Sicht eignen, Evaluationen zukünftig mit Blick auf den Prozess der Wissensgenerierung zu planen.

Tabelle 2: Systematisierung der Evaluationsstudien nach thematischen Schwerpunkten

| Schwerpunkte und ihre Themen | Auftraggeber/Finanzierung (dominierend) |
|---|---|
| Schule (7), davon: Schulsozialarbeit (4) Mediationsprojekt (1) Mittagstisch (1) Angebote Tagesstruktur (1) | Verwaltungen, DoRe Bundesamt für Gesundheit Erziehungsdepartement BS |
| Berufseinmündung und Arbeitsmarkt (15), davon: Arbeitsmarktpolitik (3) Institutionen (RAV) (1) Beschäftigungsprogramme (4) 10. Schuljahr, Integrationsprogramme (5) Betriebliche Sozialarbeit (1) Methoden (Coaching in der Berufsbildung) (1) | Eigenprojekte Staatssekretariat für Wirtschaft Seco Jugendamt. Seco, Programmträger, KTI Kantonsverwaltungen DoRe DoRe |
| Familie und Heimwesen (7), davon: Fremdplatzierung, Heimwesen (4) Kinderbetreuung (1) Mütter-/Väterberatung (2) | DoRe, Verwaltung Bundesamt für Sozialversicherung DoRe, Städte |
| Armut/soziale Sicherheit (5), davon: Sozialhilfe (4) Obdachlosigkeit (1) | Glückskette, Fürsorgekommission, Eigenfinanzierung, NFP 51 Bundesamt für Gesundheit |

| Schwerpunkte und ihre Themen | Auftraggeber/Finanzierung (dominierend) |
|---|---|
| Gesundheit/Sucht/HIV-AIDS (49), davon: Generalthemen – ambulanter Suchtsektor, organisatorischer Wandel (1) Generalthemen – HIV/AIDS (3) Netzwerke (4) Programm, z. B. Methadon, Alkohol, HIV (10) Kampagnen (6) Stellen/Einrichtungen (10) Präventionsprojekte Schule und Heime (12) Fortbildungen Drogenfachpersonen (1) Biographische Verläufe (2) | BAG u.a. BAG (2), NFP EU, BAG (2), Eigenfinanzierung BAG (7) BAG (2), Schweizerische Gesundheitsstiftung, Eigenfinanzierung BAG (3), Träger BAG (4), DoRe, Stiftungen für Gesundheitsförderung (2), kantonale Verwaltungen (3) Träger BAG (2) |
| Physische und psychische Beeinträchtigung/Traumatisierung (9), davon: Institutionen (6) Projekt (Assistenzdienst, Begleitung) (2) Reintegration (1) | NF, DoRe, BSV, Eigenfinanzierung Eigenfinanzierung NFP |
| Strafvollzug/Devianz (10), davon: Institutionen (3) Projekte: Gemeinnützige Arbeit (3) Projekte: Prävention, Intervention (4) | DoRe, Bundesamt für Justiz, Eigenfinanzierung Bundesamt für Justiz (3) NFP 40+, DoRe, Verwaltung, Träger |
| Stadt/Wohnen (11), davon: Stadtteilentwicklung (4) Wohnungsbau (3) Öffentlicher Raum (4) | KTI, Eigenprojekt, DoRe DoRe (2) Verwaltung, Träger |

Quelle: Eigene Zusammenstellung anhand der Recherchen; die Anzahl der Studien ist jeweils in Klammern angegeben.

Tabelle 3: Die Drei-Typen-Lösung: Einordnung der im Rahmen des Rundmails vorgelegten Evaluationsstudien

| | |
|----------------------------|---|
| Studien des Typs 1: | Fabian (2004) Fabian/Kaegi/Stübi (2004) Fabian/Steiner (2004) Grüter/Schaffner (2006) Gunten (2006) Gunten/Berger-Künzli (2006) Kobi (2004) Mattes/Fabian (2007) Riedli (2003) Schaffner (2002) Schaffner (2003) Schaffner (2004) Schaffner/Fabian/Drilling (2002) Steiner/Fabian/Schnorr (2006) Sommerfeld/Baumgartner (2001) Willi/Neuhäusler/Gabriel-Schärer (2005) |
| Studien des Typs 2: | Baumgartner (2003) Drilling/Müller/Fabian (2006) Fabian (2002) Fabian/Gutmann (2004) Fabian/Steiner/Guhl (2006) Keller/Fabian/Kaegi/Hornung (2003) Sommerfeld et al. (2006) |
| Studien des Typs 3: | Eser Davolio/Drilling (2007) Sommerfeld/Castelli (2006) Kundert/ Sommerfeld (2001) |

Anmerkung: Für ausführliche Angaben siehe Literaturverzeichnis.

Evaluation in der Sozialen Arbeit im Ländervergleich

Die Evaluationsdebatte in der Sozialen Arbeit in Deutschland, in Österreich und in der Schweiz weist Gemeinsamkeiten wie auch Unterschiede auf. In einer vergleichenden Betrachtung rücken grundlegende Themen in den Vordergrund, die gleichermaßen Entwicklungen in allen drei Ländern betreffen, jedoch unterschiedliche Kontexte der Evaluationsdebatte sowie länderspezifische Besonderheiten andeuten. Dieser Beitrag greift zentrale Beobachtungen auf und formuliert Thesen zur Institutionalisierung und Professionalisierung von Evaluation sowie zum Verhältnis föderaler Strukturen des Feldes und regionsübergreifender Evaluationsaktivitäten in der Sozialen Arbeit.¹

1 Institutionalisierung und Professionalisierung der Evaluation

Wiewohl die Geschichte der Sozialen Arbeit weiter zurückreicht, die Sozialarbeitspraxis immer schon beurteilende Verfahrensweisen beinhaltete und auch sozialpädagogische Forschung von Anfang an evaluative Fragen thematisierte, prägte sich eine Debatte um Evaluation erst frühestens ab den 1960er Jahren aus; wobei dies in den drei Ländern unterschiedlich verlief. Während in Deutschland praktische Erfahrungen mit dem Instrument der Evaluation und die Anfänge ihrer Institutionalisierung in den 1960er Jahren datieren, setzten Bemühungen um die Institutionalisierung von Evaluation in der Schweiz erst in den 1970er Jahren ein (vgl. Derlien 1997: 4-5). Allerdings wurde zumindest in Deutschland und der Schweiz damals noch von wissenschaftlicher Begleitforschung, Implementationsforschung oder Programmforschung gesprochen (vgl. bspw. Müller 1978; Mayntz/Scharpf 1973b). In Österreich dagegen befand sich die Soziale Arbeit in diesen Dekaden in einer beruflichen Identitätsumbildung. Weder Sozialarbeit noch Sozialpädagogik waren damals als universitäre Disziplinen ausgeformt – dies ist inzwischen begonnen –, was mit erklären mag, weshalb auch keine nennenswerten Großforschungsprojekte existierten, die der Sozialen Arbeit zuzurechnen wären. Wurden sozialarbeiterische bzw. sozialpädagogische Aktionsfelder erforscht, so geschah dies aus anderen Traditionen und Disziplinen heraus. Evaluation als eigenständiges Thema in der Sozialen Arbeit – und zwar im Rahmen der Thematiken Qualitätssicherung und Legitimation – wurde in Österreich erst in den 1990er Jahren sichtbar (vgl. Meusburger/Paulischin 2007: 169-184). Etwa zur gleichen Zeit fand in Deutschland in der Sozialen Arbeit der begriffliche und konzeptionelle Übergang von der Begleitforschung

¹ Anzumerken ist jedoch, dass bezogen auf die Schweiz die weitgehende Inexistenz von Diskursen in der Fachliteratur zur Sozialen Arbeit eine Validierung der hier getroffenen Aussagen notwendig machte (für Ausnahmen siehe etwa Baumgartner 2000; Landert 1996). Für die Bereitschaft dazu danken wir Edgar Baumgartner, Carlo Fabian und Thomas Widmer.

zur Evaluation statt – ebenfalls im Kontext von Qualitätsentwicklung und „Neues Steuerungsmodell“. In der Schweiz lässt sich ein Übergang von der Projektbegleitung mit evaluativem Anspruch zur Evaluation nicht eindeutig nachzeichnen. Seit den 1970er Jahren finden vor allem in den Bundesämtern für Gesundheit und Energie sowie bei der Direktion für Entwicklungszusammenarbeit Evaluationstätigkeiten statt (Widmer et al. 2001). Charles Landert (2006: 100) spricht für den Bereich Soziale Arbeit von einem „relativ frühen Zeitpunkt“ der Herausbildung von Evaluationswissen und verweist auf das 1996 abgeschlossene Programm des Schweizerischen Nationalfonds „Wirksamkeit staatlicher Massnahmen“.

Die Institutionalisierung von Evaluation blieb allerdings sowohl in Deutschland als auch in Österreich und der Schweiz bis heute relativ fragmentiert in den einzelnen Arbeitsfeldern; eine arbeitsfeldübergreifende systematische Debatte – abgesehen von wenigen Sammelbänden (wie Hellstern/Wollmann 1984b; Müller-Kohlenberg/Münstermann 2000; Stockmann 2006c) und Lehrbüchern (besonders Wottawa/Thierau 2003) – zur Evaluation findet bislang kaum statt. Selbst innerhalb der Sozialen Arbeit verlaufen Evaluationsdiskurse handlungsfeldspezifisch segregiert und in je typischer Weise beeinflusst durch angrenzende Felder. Während Evaluationen Sozialer Arbeit im Gesundheitswesen und in der Altenhilfe deutlich durch Entwicklungen in der Medizin und Psychologie geprägt sind, findet man im Feld der Jugendberufshilfe und im Bereich der beruflichen Resozialisierung eine Nähe zur Evaluation in der Arbeitsmarktforschung. Die Debatte in der Kinder-, Jugend- und Familienhilfe – besonders im Kontext bildender und allgemeinfördernder Angebote – ist in Deutschland, der Schweiz und Österreich speziell durch den angrenzenden Bildungsbereich beeinflusst.

Nicht zuletzt aufgrund dieser Segmentierung der Evaluationsdebatte zwischen den Arbeitsfeldern, aber auch innerhalb der Sozialen Arbeit, steht die Evaluation in Deutschland – und wohl nicht nur in dem hier diskutierten Feld – noch eher am Anfang eines Prozesses der Professionalisierung (vgl. Haubrich/Lüders 2001: 71); für Österreich gilt dies in noch prägnanterer Weise. Obgleich die Nachfrage nach Evaluationen steigt und somit auch zunehmend Bedarf an ausgebildeten und erfahrenen Evaluatoren/-innen besteht, die Funktion der Evaluation mehr und mehr institutionalisiert wird und in – je nach Arbeitsfeld und Land – unterschiedlicher Weise eine gesetzliche Verankerung findet, gibt es bislang nur vereinzelt Ausbildungsrichtungen oder Nachdiplomkurse an Fachhochschulen oder Hochschulen mit dem Schwerpunkt Evaluation. Von durchgehenden Karrieremöglichkeiten für Evaluatoren/-innen im Feld Sozialer Arbeit – auch an Fachhochschulen und Hochschulen – kann man derzeit nicht ausgehen. Erschwerend kommt hinzu, dass Evaluationen angesichts knapper öffentlicher Kassen oftmals nur gering budgetiert werden, was zwar einen Anstieg evaluativer Diplomarbeiten befördert, jedoch nur bedingt zur Stabilisierung von Beschäftigungsmöglichkeiten beiträgt. Monographien und Sammelbände (bspw. Heil et al. 2001; Müller-Kohlenberg/Münstermann 2000; Heiner 1998) finden sich in der Sozialen Arbeit häufiger zum Thema Qualität als Evaluation. Für Österreich etwa ist bemerkenswert, dass sich eine aktuelle, groß angelegte Entwicklungspartnerschaft – „Donau – Quality in Inclusion“ – ausführlich zum Thema Qualität Sozialer Arbeit äußert (EQUAL 2007), hierin Evaluation jedoch lediglich, wenn überhaupt, als implizite Thematik auftaucht und sie kaum spezifiziert wird. Bislang fehlen auch einschlägige Lehrbücher, die den Besonderheiten des Handlungsfeldes Soziale Arbeit Rechnung tragen würden. Hinsichtlich Konzeptionen und

Modellen der Evaluation nimmt die deutschsprachige Debatte häufig Bezug auf die englischsprachige Debatte, allerdings beginnt sich eine eigenständige theoretische und methodologische Evaluationsdebatte namentlich bezüglich qualitativer Evaluationsforschung zu etablieren (Flick 2006; Kuckartz et al. 2007; Bohnsack/Nentwig-Gesemann 2008) – durchaus mit Beiträgen aus dem Feld der Sozialarbeit und Sozialpädagogik.

Wichtige Schritte in Richtung Professionalisierung stellten insbesondere die Gründung der „Schweizerischen Evaluationsgesellschaft“ (SEVAL) im Jahr 1996 und der „DeGEval – Gesellschaft für Evaluation“ in Deutschland und Österreich im Jahr 1997 dar. Der Arbeitskreis „Soziale Dienstleistungen“ in der DeGEval wurde im November 2002 – damals unter dem Namen „Soziale Arbeit/Gesundheitswesen“ – als Forum für den Austausch fachbezogener Fragen der Evaluation konstituiert. Der Arbeitskreis bietet einen Rahmen für Vernetzung, Austausch und Bündelung der Diskussion; allerdings bedarf es noch intensiver Bemühungen, der Segregation der fachlichen Debatten innerhalb der Sozialen Arbeit entgegenzuwirken, zentrale Akteure einzubinden und Vernetzungen mit bestehenden Strukturen zu etablieren.

Diese Fachgesellschaften tragen zur Professionalisierung besonders durch die Entwicklung und Veröffentlichung von „Standards für Evaluationen“ sowie der Empfehlungen zur Anwendung derselben im Kontext der Selbstevaluation, Empfehlungen für die Aus- und Weiterbildung in der Evaluation und für Auftraggebende (vgl. www.degeval.de und www.seval.ch) bei. Doch bleibt anzumerken, dass die Veröffentlichung der Standards wie auch der weiteren Empfehlungen in Deutschland, Österreich und der Schweiz mit dem Beginn eines Prozesses der Professionalisierung einhergehen (vgl. Beywl 1999). So wird auf die Standards für Evaluation in der Sozialen Arbeit zwar mittlerweile durchaus immer häufiger Bezug genommen – etwa in Evaluationsberichten oder bei der Ausschreibung von Evaluationen. Auch befasste sich die Arbeitsgemeinschaft für Kinder- und Jugendhilfe (AGJ) in Deutschland eingehend mit den Standards. Jedoch entfalten die in den Standards beschriebenen Qualitätsanforderungen an Evaluationen nur schrittweise ihre Praxisrelevanz und so mag es wenig verwundern, dass diesen bislang in der Praxis – zumindest soweit es über Evaluationsberichte nachvollzogen werden kann – häufig nicht Rechnung getragen wird, wie die Recherche und Auswertung von Evaluationsstudien im Bereich Soziale Arbeit in Deutschland (www.dji.de/evaluation) und der Schweiz zeigen (vgl. Haubrich und Drilling in diesem Band). Eine entsprechende Recherche und Systematisierung wurde für Österreich noch nicht durchgeführt, doch ist von einer ähnlichen Situation auszugehen. Diese zurückhaltende Integration von Standards in der Evaluationspraxis hat möglicherweise auch darin ihre Ursache, dass etwa gegenüber der nordamerikanischen Evaluationsdebatte – die den theoretischen Hintergrund bildet für die Standards des Joint Committee on Standards for Educational Evaluation (Joint Committee 2006), an denen sich zuerst die SEVAL-Standards (Widmer et al. 2000) und dann die Standards der DeGEval (2002) orientierten, – noch ein gewisser Nachholbedarf besteht. Insbesondere steht noch die Aufgabe an, den fachlichen Anschluss zwischen den Standards für Evaluation und den Debatten innerhalb der Sozialen Arbeit durch eine intensivere Auseinandersetzung in Forschung und Lehre zu erlangen. Dafür sind durchaus Ansatzpunkte gegeben, wenn man bedenkt, dass die amerikanischen Standards in pädagogischen Feldern entwickelt wurden und praktische Anschauungsbeispiele aus Bildung, Erziehung und Sozialer Arbeit umfassen. Für die Weiter-

entwicklung eines länderübergreifenden deutschsprachigen Evaluationsdiskurses in der Sozialen Arbeit könnten angesichts der Heterogenität und Segregation des Feldes gerade die Standards als evaluationsfachliche Standards einen gemeinsamen Bezugspunkt bilden; in ihnen liegt diskursförderndes Potential.

2 Rechtliche Verankerung und Föderalismus

Soziale Arbeit ist eine gesellschaftliche Agenda, die – im Rahmen föderaler und subsidiärer Strukturen – mitunter durch die öffentliche Hand wahrgenommen, finanziert und reguliert wird, weshalb die gesetzlichen und öffentlich-rechtlichen Rahmenbedingungen die Entwicklung von Evaluationsaktivitäten und damit auch einer Evaluationsdebatte in der Sozialen Arbeit maßgeblich beeinflussen. Hierbei ist von Bedeutung, wie die Zuständigkeiten auf der Ebene des Bundes, der Bundesländer oder Kantone und Kommunen strukturiert sind. In allen drei Ländern ist Soziale Arbeit durch föderale und subsidiäre Strukturen gekennzeichnet, die nicht nur zwischen diesen Ländern divergieren, sondern auch nochmals zwischen den Handlungsfeldern Sozialer Arbeit innerhalb der Länder. Qualitäten, Prozesse, Leistungen oder Wirkungen von Maßnahmen sind jedoch grundsätzlicher Art, auch wenn für deren Aufklärung regionale Besonderheiten durchaus mit zu berücksichtigen sind. Ebenso gelten konzeptuelle und methodische Evaluationsaspekte prinzipiell und gegenseitig abhängig, weniger aber regionsspezifisch, womit offenkundig sein dürfte, dass regionsübergreifende Zuständigkeiten und Förderinitiativen bedeutsam für die Entwicklung des Evaluationsfeldes sind. Dies belegen insbesondere die Erfahrungen aus Deutschland, auf Erfahrungen hierzu kann in Österreich und der Schweiz mangels solcher Initiativen noch nicht verwiesen werden.

Wie diese Erfahrungen in Deutschland zeigen, war die – zwar deutlich auf Anregungs- und Gesetzgebungsfunktion begrenzte – Zuständigkeit des Bundes in der Kinder-, Jugend- und Familienhilfe gleichwohl ein wesentlicher Faktor für die Weiterentwicklung der Evaluationsdebatte in Deutschland. Großforschungsprojekte und Reforminitiativen auf Bundesebene waren zugleich Auslöser und Kontext für die in Deutschland vergleichsweise früh beginnende Evaluationsdebatte in der Sozialen Arbeit. Die Bedeutung von Bundesinitiativen für die Entwicklung der Evaluationsdebatte bleibt auch in den Folgejahren erhalten. Die Evaluation und wissenschaftliche Begleitung – mit vergleichsweise guter Ressourcenausstattung – obliegt meist externen Forschungsinstituten. Ihre Bedeutung für den Fachdiskurs ist vor allem auch darauf zurückzuführen, dass die Ergebnisse im Sinne der Anregungsfunktion des Bundes von überregionaler Relevanz sein sollen und gezielt der Fachöffentlichkeit zugänglich gemacht werden. Bislang konzentrierte sich die Bundesförderung vorrangig auf Modellvorhaben, in einzelnen Evaluationsstudien wurde aber auch die Regelpraxis in den Blick genommen (bspw. BMFSFJ 1998; 2002b). Begleitstudien bieten allerdings nur in begrenztem Umfang einen Rahmen, um Evaluationskonzepte oder -modelle weiterzuentwickeln. Daher wurden – ebenfalls auf Ebene des Bundes – Initiativen ins Leben gerufen, um auch konzeptionelle und methodische Anregungen für Evaluationen in der Kinder-, Jugend- und Familienhilfe zu befördern; zu nennen wären die QS-Initiative des Bundesministeriums für Familie, Senioren, Frauen und Jugend (<http://www.qs-kompendium.de/>), die

Nationale Qualitätsoffensive im Bereich der Kindertagesstätten wie auch das „Projekt eXe“ (www.dji.de/evaluation).

In Österreich sah die Situation bezüglich nationaler Förder- bzw. Evaluationsprogramme in der Sozialen Arbeit bislang dürrig aus, es ließe sich also auf den Erfahrungen aus Deutschland durchaus intensiver aufbauen, und in Anlehnung könnten sowohl Forschungsprojekte über Evaluationen als auch Evaluationen von Modellvorhaben konzipiert werden. Außer Frage sei hierbei gestellt, dass Impulse zur Entwicklung von Evaluationen ebenso von kommunaler oder Landesebene ausgehen; eine singuläre Aktivitäten zusammenführende Evaluationsagenda allerdings wird sich kaum ohne nationale Initiativen vorantreiben lassen.

In der Schweiz hat der Bereich der Evaluationen speziell durch die Förderung des Schweizerischen Nationalfonds Impulse erhalten. Förderlich wirkten sich die Programme „Wirksamkeit staatlicher Massnahmen“ (Nationales Forschungsprogramm) sowie ab den 1990er Jahren „Do Research“ (Förderung von Forschung an Fachhochschulen) auf die bereits bestehende Evaluationstätigkeit im Bereich Gesundheit aus. Auch im Bereich Strafvollzug wurden ab den frühen 1990er Jahren alle Modellversuche evaluiert. In den Nationalen Forschungsprogrammen des Schweizerischen Nationalfonds finden sich seit 1996 keine gesonderten Förderungsmaßnahmen mehr für die Evaluationstätigkeit, wohl aber finden sich jeweilige Evaluationscluster innerhalb der Programme: jüngst zum Beispiel im Rahmen der Forschung zu Ursachen von Rechtsextremismus (z. B. Evaluation von Interventionsprogrammen, siehe Eser Davolio/Drilling 2007; Meta-Evaluation von Antirassismusprojekten, siehe Widmer et al. 2007a).

In der bisherigen Argumentation hoben wir die Bedeutung regionsübergreifender und -vergleichender Aktivitäten in ihrer Funktion für die Institutionalisierung von Evaluation hervor. Eine nicht unerhebliche Rolle in der Qualitätsentwicklung wie auch in Evaluationen spielen des Weiteren inhaltlich-fachliche Standards der Sozialen Arbeit. Bemühungen diese zu entwickeln setzen in aller Regel professionsbezogene und deshalb ebenfalls überregionale Initiativen wie auch eine akzeptanzfördernde strukturelle Verankerung voraus. Die förderlichen Aspekte föderaler Strukturen seien hierbei keinesfalls in Abrede gestellt. Letztlich liegt gerade in heterogenen Umsetzungsformen in unterschiedlichen Kontexten auch ein Potential für Vielfalt, Innovation, Vergleichs- und Lernmöglichkeiten (vgl. Landert 2006: 99) – sofern die notwendigen Informationen wiederum vergleichend zusammengetragen und öffentlich zugänglich gemacht werden.

3 Lernen in vergleichender Perspektive

Die Heterogenität der drei Länderbeiträge zum Feld Soziale Arbeit verdeutlicht bereits die Potentiale einer länderübergreifenden vergleichenden Perspektive in einer anschaulichen Form. Letztlich heißt Lernen aus gewonnenen Erfahrungen und Einsichten in den Nachbarländern, von den jeweiligen Besonderheiten und spezifischen Perspektiven zu abstrahieren und Gemeinsamkeiten wie auch Unterschiede als Anlässe zur Reflexion zu nehmen und zu fragen, was dies für die Weiterentwicklung fachlicher Debatten im eigenen Land bedeutet. Einerseits kann man im Vergleich lernen, was sich jeweils als förderlich oder als hinderlich

für die Etablierung von Evaluationsaktivitäten erweist, was man bewahren sollte und wo Veränderungen wünschenswert wären. Andererseits gibt es jenseits dieser heterogenen Rahmenbedingungen fachliche Herausforderungen, die sich in allen drei Ländern in ähnlicher Weise stellen. Dies gilt zum einen für evaluationsfachliche Anforderungen, wie sie in den Standards für Evaluation in allen drei Ländern in weitgehender Übereinstimmung formuliert sind. Es gilt zum anderen aber ebenso für fachliche Fragestellungen und Besonderheiten sozialpädagogischer Untersuchungsgegenstände. Ein Lernen aus Evaluationen kann dabei sowohl für die fachliche – auch länderübergreifende – Weiterentwicklung des Feldes gewinnbringend sein als auch für methodische und konzeptionelle Fragen der Gestaltung und Umsetzung von Evaluationen in der Sozialen Arbeit.

Nutzung der Evaluationsfunktion

Zur Nutzung der Evaluationsfunktion: Partielle Bestandsaufnahme und Pilotstudie in österreichischen und deutschen Bundesministerien

Die „Nutzung der Evaluationsfunktion“ ist nach wie vor, trotz vieler theoretischer und empirischer Arbeiten dazu, keineswegs in zufriedenstellender Weise erreicht. Nicht nur in den 1960er Jahren wurden in den USA für viele Millionen Dollar Evaluationen durchgeführt, deren Ergebnisse nicht umgesetzt und zum Teil von den Auftraggebern sogar ignoriert wurden (Patton 1997); auch heute wird ein Mangel an Nutzung und Rezeption beklagt (Stamm 2003).

Eine wesentliche Voraussetzung für die Nutzung der Evaluationsfunktion ist der Stellenwert, der Evaluation im jeweiligen Setting, Kontext beziehungsweise Land zugeschrieben wird. Dieser Stellenwert bestimmt auch die Anzahl an Evaluationen und ihre institutionelle Verankerung. Im europäischen Raum spielt die Europäische Union zweifellos eine Vorreiterrolle; sie sieht Evaluation als wichtigen Bestandteil der Qualitätssicherung. Williams, de Laet und Stern stellen in ihrem Endbericht „The use of evaluations in the commission services“ (2002) nicht nur den eindrucksvollen Umfang der Evaluationstätigkeit in der Europäischen Kommission dar, sie zeigen auch das Bestreben der Kommission auf, Evaluationen zu nutzen.

Während in der Schweiz die Verpflichtung zur Wirkungsprüfung staatlicher Maßnahmen und damit auch zur Evaluation in der Bundesverfassung verankert ist, ist man in Österreich und Deutschland von einer öffentlichen Thematisierung der Evaluation staatlicher Maßnahmen weit entfernt. Allerdings fehlt es bislang an Studien über die institutionelle Verankerung der Evaluationsfunktion in Bundesministerien. Noch weniger weiß man über Verwendung von Evaluationen und deren Ergebnissen. Aus diesem Grund wurde im Zuge der Vorbereitung des vorliegenden Beitrags eine erste Pilotstudie zur Nutzung der Evaluationsfunktion in deutschen und österreichischen Bundesministerien durchgeführt, über die im zweiten Teil des Beitrags berichtet wird.

Der erste Teil des Beitrags enthält eine Bestandsaufnahme der einschlägigen Literatur zur Nutzung der Evaluationsfunktion. Da diese Bestandsaufnahme nicht vollständig und durch die Perspektive der Autorinnen geprägt ist, haben wir ihr das Attribut „partiell“ vorangestellt. Für einen Literaturüberblick siehe zum Beispiel Shulha und Cousins (1997) sowie Hofstetter und Alkin (2003) beziehungsweise das Standardwerk von Patton (1997).

1 Partielle Bestandsaufnahme zur Nutzung der Evaluationsfunktion

Wir wollen den Nutzen der Evaluationsfunktion aus drei Perspektiven betrachten. Die erste Perspektive betrifft die Evaluation selbst, ihre Durchführung, Berichtlegung und Empfehlungen und damit die Evaluatorinnen und Evaluatoren. Kriterien zur Nützlichkeit von Evaluation sind in den diversen publizierten Standards festgelegt (siehe z. B. Joint Committee 1994; 2006; DeGEval 2002). Die zweite Perspektive betrifft die Nutzung und damit den institutionellen Kontext beziehungsweise die institutionelle Verankerung der Evaluation. Auf die Unterschiede, die diesbezüglich im deutschsprachigen Raum bestehen, wurde bereits einleitend hingewiesen. Die dritte Perspektive betrifft den Nutzen der Evaluation für die jeweiligen Stakeholdergruppen (inklusive der Auftraggebenden), was auch die Verwendung beziehungsweise Umsetzung von Ergebnissen und Empfehlungen inkludiert. Als quasi Oberbegriff werden wir in diesem Beitrag immer von der Nutzung der „Evaluationsfunktion“ sprechen.

Eine analoge Differenzierung wie hier dargestellt nimmt auch Stockbauer (2000a: 16-17) vor (siehe Tabelle 1).

Tabelle 1: Definition von Nützlichkeit, Nutzung und Nutzen (Stockbauer 2000a: 16-17)

| | |
|--------------|---|
| Nützlichkeit | Synonyme: Verwendbarkeit, Verwertbarkeit engl. <i>utility</i> Merkmale von Evaluationen und Evaluationsergebnissen, welche die potenzielle Nutzung beschreiben. |
| Nutzung | Synonyme: Verwendung, Verwertung engl. <i>utilization, (use)</i> Wertfreier Begriff, der den objektiv messbaren Gebrauch der Evaluation und der Evaluationsergebnisse beschreibt. |
| Nutzen | engl. <i>usefulness, (use)</i> Wertmäßige, subjektive Beurteilung der Qualität und Nützlichkeit der Evaluation und Evaluationsergebnisse. |

Zweifellos sind diese drei Perspektiven nicht unabhängig voneinander. So hat zum Beispiel eine Evaluation, die von Anfang an die verschiedenen Stakeholdergruppen einbezieht, größere Chancen, dass ihre Ergebnisse – insbesondere wenn diese zielgruppenspezifisch aufbereitet werden (Owen/Rogers 1999) – auch genutzt werden. Umgekehrt garantiert jedoch auch die institutionelle Verankerung der Evaluationsfunktion nicht, dass die Evaluationsergebnisse zum Beispiel von den Stakeholdern wirklich gemäß der Empfehlungen umgesetzt werden (Stamm 2003); hier spielt eine Vielzahl von Faktoren eine Rolle (Schrader/Helmke 2003).

Während in der anglo-amerikanischen Literatur der Fokus auf „Knowledge Utilization Research“ liegt (siehe Patton 1997) und damit diese Tradition den Fokus auf der ersten Perspektive hat, jedoch auch die weiteren Perspektiven einschließt, liegt im deutschen Sprachraum der Fokus meist eher auf der Rezeptionsforschung, die nach der Bekanntgabe der Ergebnisse der Evaluation ansetzt. Prominent sind hier vor allem Rezeptionsstudien im Anschluss an die PISA-Studie (z. B. Kohler/Schrader 2004).

In der Literatur werden auch zahlreiche Faktoren genannt, die Einfluss auf Nützlichkeit, Nutzung und Nutzen von Evaluation haben. Dazu wird auf Preskill und Caracelli (1997), Balthasar und Rieder (2000) sowie Widmer und Neuenschwander (2004) verwiesen. Einen guten Überblick über (vor allem) die US-amerikanische Forschung geben Hofstetter und Alkin (2003).

1.1 Zur Nützlichkeit – Knowledge Utilization Research

Für die Nützlichkeit von Evaluation werden – wie bereits angeführt – in den verschiedenen Standards Kriterien benannt. Aus Platzgründen wird hier explizit nur auf die DeGEval (DeGEval 2002) und das Joint Committee on Standards for Educational Evaluation (JCSEE; Joint Committee 1994; siehe auch Joint Committee 2006) Bezug genommen. „The utility standards are intended to ensure that an evaluation will serve the information needs of intended users“, so die Ausführung des JCSEE (1994: 23). Gemäß der DeGEval sollen die Nützlichkeitsstandards sicherstellen, „dass die Evaluation sich an den geklärten Evaluationszwecken sowie am Informationsbedarf der vorgesehenen Nutzer und Nutzerinnen ausrichtet“ (DeGEval 2002: 21). Sie benennt dazu acht Nützlichkeitsstandards: Identifizierung der Beteiligten und Betroffenen, Klärung der Evaluationszwecke, Glaubwürdigkeit und Kompetenz des Evaluators/der Evaluatorin, Auswahl und Umfang der Informationen, Transparenz von Werten, Vollständigkeit und Klarheit der Berichterstattung, Rechtzeitigkeit der Evaluation sowie Nutzung und Nutzen der Evaluation (DeGEval 2002: 21-25).

Die anglo-amerikanische Literatur beschäftigt sich mit dieser Thematik unter der Bezeichnung „Knowledge Utilization Research“ (siehe z. B. Owen/Rogers 1999; Patton 1997). Danach wird bereits bei der Planung der Studie sowie im Verlauf der Durchführung darauf geachtet, wie die Ergebnisse für die verschiedenen Stakeholdergruppen aufbereitet und an sie weitergegeben werden können. Als besonders relevant werden die Abfassung des Berichts beziehungsweise der Berichte für unterschiedliche Rezipientengruppen sowie die Formulierung von Empfehlungen hervorgehoben. So wurden zum Beispiel von Hendricks und Papagiannis (1990) Kriterien für die Abfassung von Evaluationsberichten formuliert; Cousins und Leithwood (1993) konzipierten ein Rahmenmodell – „Knowledge Utilization Conceptual Framework“ – das Bedingungen benennt für die effektive Aufnahme und Umsetzung von Evaluationsergebnissen. Diese sind nach Cousins und Leithwood (1993) die „Characteristics of the Source of Information“, das „Improvement Setting“ und der „Interactive Process“, die einander wechselseitig beeinflussen.

Eine zentrale Rolle dafür, dass die Nützlichkeitsstandards auch realisiert werden können, kommt dem Auftraggeber und damit auch dem institutionellen Kontext der Evaluation zu.

1.2 Zur Nutzung – Verwendungstypen und institutioneller Kontext der Evaluation

Nutzung beziehungsweise Verwendung von Evaluationen kann auf unterschiedliche Weise und mit unterschiedlichen Zielen erfolgen. Mittlerweile besteht in der Literatur weitgehend

Einigkeit darüber, welche der in der Vergangenheit zahlreich vorgeschlagenen Typen von Evaluationsverwendungen sich letztlich durchgesetzt haben. Weiss, Murhpy-Graham und Birkeland (2005) beispielsweise nennen folgende drei Typen: „instrumental“, „conceptual“ und „political/symbolic“. Diese Auffassung teilen unter anderem Balthasar (2006), Hofstetter und Alkin (2003) und – schon früher publiziert – Johnson (1998). Balthasar und Johnson ergänzen allerdings auch jenen von Patton (1997) vorgeschlagenen Typ des „process use/process-related use“. Weiss et al. (2005) fügen den drei oben genannten Typen auch „imposed use“ („auferlegte“ Verwendung) als neuartige Entwicklung der Evaluationsnutzung hinzu (siehe Tabelle 2).

Tabelle 2: Typen der Evaluationsverwendung

| Typen der Evaluationsverwendung | Erklärung |
|---------------------------------|--|
| Instrumental use | Indicates „instances where respondents in the study could document the specific way in which the social science knowledge had been used for decision-making or problem-solving purpose.“ (Alkin 2005: 435) |
| Conceptual use | Indicates „instances in which the knowledge about an issue influences a policy maker’s thinking although the information was not put to a specific, documentable use.“ (Alkin 2005: 435) |
| Symbolic use | Indicates instances in which evaluations are used to „justify preexisting preferences and actions“. (Weiss et al. 2005: 13) |
| Process use | „The evaluation process itself could have impact; that is, it could lead to utilization [...] the act of conducting an evaluation and the interactions between the evaluator and various stakeholders can lead to changes in the individuals or in the organisation. Thus it is not the findings of the evaluation that contribute to use but the process itself.“ (Alkin 2005: 435) |
| Imposed use | „It is a type of use that comes about because of pressure from outside.“ (Weiss et al. 2005: 16) |

Einen neuen theoretischen Ansatz liefern Henry und Mark (2003; Mark/Henry 2004). Sie meinen, dass der Begriff „evaluation use“ zu eng gefasst sei und führen den breiter gefassten Begriff „evaluation influence“ als theoretischen Ansatz ein. Anhand dieses neuen Ansatzes soll es gelingen, die Wege von der Durchführung der Evaluation bis zu ihrer Nutzung besser nachzuzeichnen.

Viele Studien, die sich mit der Erforschung der Verwendung von Evaluationen und deren Ergebnissen beschäftigen, erheben in einem ersten Schritt den institutionellen Kontext von Evaluationen. Denn der Verankerung und dem Stellenwert von Evaluation, die sich im institutionellen Kontext niederschlagen, kommt eine wesentliche Rolle für die Nutzung von Evaluation zu. Bei der folgenden Darstellung wird von den institutionellen Kontextfaktoren, die von Balthasar (2006) genannt wurden, ausgegangen und dabei der Bezug zu anderen Studien – Widmer und Neuenschwander (2004) sowie Williams et al. (2002) – hergestellt.

Der erste von insgesamt sechs Faktoren nach Balthasar ist das Vorhandensein einer evaluationsverantwortlichen Stelle. Widmer und Neuenschwander (2004) sprechen in diesem Zusammenhang von „institutional design“ und unterscheiden zwischen „centralized setting“ und „decentralized evaluation setting“. Im centralized setting gibt es neben diver-

sen anderen Abteilungen eine Abteilung für Evaluation. Sie ist beispielsweise für die Initiierung von Evaluationen oder die Verbreitung von Ergebnissen verantwortlich. Im decentralized setting kann jede Abteilung Evaluationen initiieren und sich Hilfe von einer Evaluationsseinheit, die gleichzeitig für alle Abteilungen Unterstützung anbietet, holen.

Der zweite Faktor nach Balthasar (2006) betrifft die Kostenübernahme. Hier geht es darum, ob eine Institution ein eigens ausgewiesenes Budget für Evaluationen besitzt. Die Höhe des Budgets wird auch von Widmer und Neuenschwander (2004) erhoben. Mit dem dritten bis fünften Faktor, der Auslösung einer Evaluation, der Vergabe- und der Publikationspraxis beschäftigten sich neben Balthasar (2006) sowohl Widmer und Neuenschwander (2004) als auch Williams et al. (2002). Der sechste Faktor nach Balthasar betrifft die Frage, ob die Evaluation von externen Spezialisten durchgeführt wurde. Eine zusätzliche wichtige Rolle für die Einschätzung der Evaluationstätigkeit in einer Institution spielt nach Balthasar (1998) auch die Anzahl der durchgeführten Evaluationen innerhalb eines bestimmten Zeitraums.

Auch wenn der institutionelle Kontext kein Garant dafür ist, dass Evaluation einen Nutzen für die jeweiligen Stakeholdergruppen hat, so erhöht eine gute institutionelle Verankerung von Evaluation die Wahrscheinlichkeit dafür.

1.3 Zum Nutzen – Rezeption und Verwendung durch Stakeholder

Zu diesem Thema wurden in Deutschland im Anschluss an die Veröffentlichung der PISA Ergebnisse eine Reihe von Studien unter der Bezeichnung Rezeptionsforschung durchgeführt (Helmke/Schrader 2001; Kohler/Schrader 2004; Schrader/Helmke 2003; 2004). Diese Studien beschäftigen sich unter anderem mit den Voraussetzungen, die auf Seiten der Stakeholder (Lehrpersonen) gegeben sein sollten, damit die Ergebnisse von Evaluation positiv rezipiert werden. Nach Schrader und Helmke (2003) kommt dem Interesse eine zentrale Bedeutung zu; Interesse steigert die Rezeption, führt zu besserem Lernen, zu einer Kompetenzsteigerung und somit zum Aufbau von Fachwissen. Des Weiteren ist es wichtig das Vorwissen, die Einstellung, die Motivation und die Handlungskompetenz der Stakeholder (Lehrpersonen) zu berücksichtigen. Ebenfalls eine wichtige Rolle spielt die generelle Aufgeschlossenheit der Schule gegenüber Evaluation sowie die Qualität der Fortbildung. Hohe Fortbildungsqualität führt bei Lehrpersonen zu mehr Engagement und Interesse für Veränderung (Becker 1983; Eigler/Nenninger 1983; Helmke/Schrader 2001). Eine wichtige Voraussetzung, damit all dies geschehen kann, ist jedoch, dass Ergebnisse von Evaluationsstudien von den Stakeholdern überhaupt bemerkt werden, was nicht automatisch gegeben ist (Kohler/Schrader 2004).

Eine Studie zum Nutzen von Evaluation aus Sicht von (vier) Auftraggebern von Keller-Ebert (2006) ergab, dass diese auch mehrere Jahre nach Abschluss der Evaluation von den gewonnenen Erkenntnissen profitieren und auch Transferwirkungen in andere Projekte stattfinden.

Wenn man die Befunde und Aussagen der Arbeiten zur Nutzung der Evaluationsfunktion durch mündliche Berichte von Evaluatorinnen und Evaluatoren, Auftraggebenden sowie diversen Evaluierten ergänzt, so zeigt sich klar, dass das Potential von Evaluation

hinsichtlich Nützlichkeit, Nutzung und Nutzen bei weitem nicht ausgeschöpft ist. Dies belegt auch die Pilotstudie bei österreichischen und deutschen Bundesministerien, deren Befunde nun dargelegt werden sollen.

2 Pilotstudie zur Nutzung der Evaluationsfunktion an Österreichischen und Deutschen Bundesministerien

Recherchen hatten ergeben, dass in beiden Ländern keine repräsentative Studie vorliegt, die sich mit der Nutzung der Evaluationsfunktion beschäftigt. Aus diesem Grund haben wir den Entschluss gefasst, für diesen Beitrag eine eigene Erhebung auf Bundesebene, das heißt in sämtlichen Bundesministerien und den beiden Bundeskanzlerämtern in Deutschland und Österreich durchzuführen. Aus Zeit- und Ressourcengründen konnte sie nur als Pilotstudie und damit als grobes Screening realisiert werden. Eine weitere Erschwernis ergab sich dadurch, dass in Österreich eine neue Regierung gewählt worden war und die Koalitionsverhandlungen sehr lange Zeit in Anspruch nahmen, wodurch die Ansprechpartner (es wurden z. T. neue Ministerien gebildet) längere Zeit nicht klar waren. Die Erhebung erfolgte schließlich in 14 (allen) deutschen Bundesministerien, in elf österreichischen (zwei von diesen wurden im Zuge der Regierungsneubildung geteilt; die Erhebung wendete sich jedoch an die ehemaligen Bundesministerien) sowie in beiden Bundeskanzlerämtern. Aufgrund der Rahmenbedingungen konnte für die Beantwortung des Fragebogens nur ein Zeitrahmen von vier Wochen gegeben werden. Einige Antworten kamen erst nach der vorgegebenen Frist; in diesem Beitrag werden alle Antworten berücksichtigt.

2.1 Erhebungsinstrument

Aufgrund der genannten Rahmenbedingungen wurde für die Erhebung kein Interview, sondern ein kurzer, neu konzipierter Fragebogen eingesetzt, der sowohl offene als auch geschlossene Fragen enthält. Er wurde in Anlehnung an die Studien zum institutionellen Kontext konzipiert (siehe vor allem Balthasar 2006), jedoch wurden nur einige der angeführten Faktoren erfragt, da das primäre Ziel – eine möglichst hohe Rücklaufquote zu erhalten – durch zu viele und zum Teil nicht beantwortbare Fragen gefährdet erschien.

Folgende Aspekte wurden durch teils offene teils geschlossene Fragen erhoben:

- Durchschnittliche jährliche Anzahl an Evaluationen
- Finanzielles Volumen von Evaluationen
- Maßnahmen zur Weiterleitung von Evaluationsergebnissen an Stakeholder/Nutzer
- Ausmaß der zufriedenstellenden Weiterleitung der Evaluationsergebnisse
- Sicherung der Umsetzung und Anwendung von Evaluationsergebnissen durch die Stakeholder
- Ausmaß der zufrieden stellenden Sicherung des Nutzens von Evaluationen
- Kontakt mit und Interesse an einer Zusammenarbeit mit der DeGEval

Damit lag der Fokus auf den Faktoren Kostenübernahme und Publikationspraxis gemäß Balthasar (2006), wobei der letztere Faktor ausführlicher erfragt wurde. Zusätzlich wurde indirekt auch der erste Faktor nach Balthasar (2006) „Vorhandensein einer evaluationsverantwortlichen Stelle“ durch die Frage nach einem/einer Ansprechpartner/in für Evaluation erhoben.

Der Fragebogen wurde mit einem Begleitbrief verschickt, der von der Erst-Autorin in ihrer Funktion als Vorstandsvorsitzende der DeGEval unterschrieben wurde. Da von den Homepages der Bundesministerien nur in den seltensten Fällen Evaluationsverantwortliche identifiziert werden konnten, wurden die Briefe und der Fragebogen direkt an den/die jeweilige/n Bundesminister/in beziehungsweise an den/die Bundeskanzler/in adressiert. In welcher Form beziehungsweise ob überhaupt eine Weiterleitung des Briefes erfolgte, kann damit nur indirekt aus dem Rücklauf geschlossen werden.

2.2 Ergebnisse der Pilotstudie

Der Rücklauf gibt bereits eine erste Information über den Stellenwert beziehungsweise die Verankerung von Evaluation auf Ebene des Bundes in Österreich und Deutschland. Von den insgesamt 27 ausgeschickten Fragebögen wurden nur sieben ausgefüllt retourniert; in einem Fall wurde ein Teil der erfragten Informationen in einem Brief angegeben (nicht im Fragebogen); in elf Fällen erfolgte keinerlei Reaktion. In weiteren acht Fällen wurde zwar auf die Anfrage reagiert, jedoch konnten die Fragen nicht beantwortet werden. Der Grund war in fast allen Fällen, dass Evaluationen zwar geplant, jedoch bisher noch nicht durchgeführt worden seien.

Tabelle 3 enthält eine detaillierte Übersicht getrennt nach den beiden Ländern. Es zeigen sich deutliche Unterschiede zwischen Deutschland und Österreich hinsichtlich des Rücklaufs. Über Gründe dafür kann nur spekuliert werden. So könnte der deutlich höhere Rücklauf in Österreich möglicherweise daran liegen, dass die Erst-Autorin durch Politikberatung, Berichte über Forschungsergebnisse und Medienanfragen als Expertin in Österreich – nicht jedoch in Deutschland – einen gewissen Bekanntheitsgrad hat.

Tabelle 3: Rückmeldung zum Fragebogen

| Rückmeldung | Österreich | Deutschland |
|---|------------|-------------|
| Vollständig ausgefüllter Fragebogen | 5 | 2 |
| Information (nicht Fragebogen) | 1 | 0 |
| Keine Information vorhanden, Evaluation geplant | 2 | 3 |
| Negative Rückmeldung | 0 | 1 |
| Neutrale Rückmeldung | 1 | 0 |
| Keine Rückmeldung | 3 | 9 |
| Total | 12 | 15 |

Die Auswertung der sieben ausgefüllten Fragebögen ergab (gemäß Angaben der Befragten) eine Schwankungsbreite von drei bis zehn Evaluationen pro Jahr. Das dafür eingesetzte Volumen liegt zwischen 0,2 und 1,4 Millionen Euro. Die Weiterleitung von Evaluationsergebnissen gelinge in 75 bis hundert Prozent der Fälle zufriedenstellend. Die Weiterleitung geschehe vorwiegend in Form von:

- Publikationen (Internet, Intranet, Fachzeitschriften, Presseaußendungen etc.)
- Diskussion unter Einbezug der Beteiligten
- Veranstaltungen (Enquete, Pressekonferenz etc.)

Die Sicherung der Umsetzung der Evaluationsergebnisse gelinge in fünfzig bis 85 Prozent der Fälle zufriedenstellend. Die Sicherung erfolge durch:

- Zielvereinbarungen (Aktions- und Zeitplan)
- Begleitgremien
- Kontrolle durch Dienst- und Fachaufsicht
- Erlässe, Richtlinien und Verträge

Die Frage nach der DeGEval wurde nur in neun Fällen beantwortet. Von den fünf Bundesministerien, die bereits Kontakt mit der DeGEval hatten, haben vier Interesse an mehr Information. Vier Bundesministerien hatten bisher keinen Kontakt, jedoch auch kein Interesse an Informationen über die DeGEval.

3 Resümee und Ausblick

Wie bereits der Literaturüberblick gezeigt hat, ist die Diskrepanz zwischen der Relevanz, die der Nutzung der Evaluationsfunktion zukommt, und ihrer tatsächlichen Realisierung sehr groß. Der vermutlich bedeutsamste Einfluss kommt dabei dem Stellenwert zu, den Evaluation in dem jeweiligen Setting/Land hat. Hier hat die Schweiz mit der Verankerung der Wirksamkeitsprüfung staatlicher Maßnahmen in der Bundesverfassung (Grundgesetz) eindeutig eine Vorreiterrolle inne. Ein hoher Stellenwert von Evaluation erhöht die Wahrscheinlichkeit, dass Evaluationen nicht nur symbolisch oder auf Druck in Auftrag gegeben werden, sondern mit dem Ziel daraus zu lernen und die Ergebnisse respektive Empfehlungen zu nutzen. Ein hoher Stellenwert von Evaluation macht es auch wahrscheinlicher, dass ein entsprechender institutioneller Kontext gegeben ist, der die Standards von Evaluation inklusive deren Nutzung sichert. Somit dient ein hoher Stellenwert sowohl der Nützlichkeit, der Nutzung als auch dem Nutzen von Evaluation.

Leider sind wir in Österreich und Deutschland hinsichtlich der Verankerung von Evaluation und der Etablierung einer Evaluationskultur deutlich von der Schweizer Situation entfernt. Dies zeigen auch die Ergebnisse der durchgeführten Pilotstudie. Optimistisch kann dabei stimmen, dass sie sehr heterogen sind, das heißt in einigen Bundesministerien Evaluation durchaus verankert ist. Gemäß des Rücklaufs und der Befunde kann man die Bundesministerien (inklusive Bundeskanzlerämter) vorsichtig in drei Gruppen einteilen: in (1)

Evaluationsexperten, (2) Evaluationsmotivierte und (3) Evaluationsdistanzierte, wobei jedoch die Experten in der Minderzahl und die Distanzierten in der Mehrzahl sind.

Die Herausforderungen, die sich für eine Evaluationsgesellschaft in so einem Umfeld stellen, sind Förderung von Information und Austausch, Zusammenführung unterschiedlicher Perspektiven und vor allem Professionalisierung von Evaluation. Allen diesen Aufgaben hat sich die DeGEval – Gesellschaft für Evaluation verpflichtet. Neben den Standards für Evaluation (DeGEval 2002) hat sie Empfehlungen zur Aus- und Weiterbildung in der Evaluation (DeGEval 2004a), zur Anwendung der Standards im Feld der Selbstevaluation (DeGEval 2004b) sowie für Auftraggebende von Evaluationen (DeGEval 2007) herausgegeben.

Bei der Verbreitung und Umsetzung der diversen Professionalisierungsmaßnahmen müssen jedoch auch psychologische Faktoren, die Änderungen von Einstellungen und Verhalten als Voraussetzung von Akzeptanz und institutioneller Verankerung von Evaluation erschweren, berücksichtigt werden. Im Besonderen ist hier Änderungsresistenz aus Selbstwertschutz zu nennen. Sie betrifft vor allem Personen mit langer Verweildauer in einer bestimmten Position und hohem Entscheidungsspielraum (z. B. hohe Beamte in Ministerien). Die Änderungsresistenz wird noch verstärkt, wenn ein hoher Unternehmensdruck gegeben ist (Wottawa 2001; siehe auch Spiel 2004).

Unter diesen Bedingungen ergeben sich sowohl praktische als auch wissenschaftliche Aufgaben für Evaluatorinnen und Evaluatoren, die sich den Evaluationsstandards verpflichtet haben. Erstere betreffen unter anderem Maßnahmen zur Etablierung einer Evaluationskultur in Institutionen (Fetterman 1994; Patton 1997; siehe auch Atria et al. 2006). Zu den wissenschaftlichen Aufgaben gehört insbesondere eine systematische Wirkungsforschung. Einen kleinen Beitrag dazu wollen wir selbst leisten. Ausgehend von der hier präsentierten Pilotstudie bereiten wir eine vertiefende Untersuchung vor, in der anhand von systematischen Interviews die Bundesministerien charakterisiert und verglichen werden sollen. Die Erhebung wird vorerst in österreichischen Bundesministerien durchgeführt werden. Wir gehen dabei davon aus, dass derartige Studien selbst eine Intervention darstellen und damit – hoffentlich – einen kleinen Schritt in Richtung Akzeptanz von Evaluation auslösen.

Evaluationen in der Schweiz: Verbreitung und Verwendung

„In the end, the worth of evaluations must be judged by their utility.“ (Rossi et al. 1999: 431)

Evaluationen werden durchgeführt, um zu verwendbaren Ergebnissen zu gelangen. Die Verwendungsorientierung gehört also mit zur Definition von Evaluation (Vedung 1999: 12). Die Frage nach dem Ausmass der Verwendung von Evaluationen durch Verwaltung und Politik ist somit ein Kernthema der Evaluationsforschung. Der vorliegende Beitrag geht am Beispiel der schweizerischen Bundespolitik auf diese Thematik ein; er verfolgt dazu drei Fragen vertieft:

Da die erste Bedingung für die Verwendung von Evaluationen deren Durchführung ist, wird zunächst die Verbreitung der Evaluationstätigkeit in der schweizerischen Bundesverwaltung untersucht. Danach wird die Frage nach der Verwendung von Evaluationen gestellt. Es wird geprüft, ob und in welchem Umfang die realisierten Evaluationen auch genutzt wurden, wobei kurz auf die Faktoren eingegangen wird, die die Verwendung beeinflussen. Schliesslich geht es um die Frage nach den Akteurguppen, von denen die Evaluationen verwendet werden. Dabei interessieren insbesondere Politikerinnen und Politiker – diese signalisierten in der 1999 erneuerten schweizerischen Bundesverfassung ein besonderes Interesse an Evaluationen: „Die Bundesversammlung sorgt dafür, dass die Massnahmen des Bundes auf ihre Wirksamkeit überprüft werden“, wie Artikel 170 der Bundesverfassung festhält.

Angesichts der grossen Bedeutung der Verwendung von Evaluationen erstaunt es nicht, dass sich die evaluationswissenschaftliche Verwendungsforschung in den USA schon seit Jahren als eigenständiger Forschungszweig etabliert hat (vgl. z. B. Cousins/Leithwood 1986; Lee 2000). Auch in Europa sind verschiedene Arbeiten entstanden, die sich mit der Verwendung von Evaluationen auseinandersetzen (z. B. Valovirta 2002; Van der Knaap 1995). In der Schweiz gehen die ersten diesbezüglichen Arbeiten auf die 1980er Jahre zurück (Klöti/Schneider 1989; Zimmermann et al. 1988). In neuerer Zeit haben sich in der Schweiz insbesondere Thomas Widmer und Peter Neuenschwander mit der Institutionalisierung und der Verwendung von Evaluationen befasst (Widmer/Neuenschwander 2004; Widmer et al. 2001).

Grundlage des vorliegenden Beitrags bildet eine breit angelegte Untersuchung über Verbreitung und Verwendung von Evaluationen in der Schweizer Bundesverwaltung (Balthasar 2007). „Evaluation“ wurde dabei definiert als eine wissenschaftliche und empirisch gestützte retrospektive Beurteilung der Konzeption, des Vollzugs und/oder der Wirksamkeit staatlicher Aktivitäten, seien dies Massnahmen, Programme oder Projekte. Evaluationen bewerten staatliche Aktivitäten nach transparenten Kriterien und stellen Kausalzusammen-

hänge zwischen Aktivitäten und Wirkungen fest. Die Untersuchung beinhaltet neben zehn Fallstudien auch eine Erhebung und Befragung zu allen Evaluationen, welche die Bundesverwaltung in den Jahren 1999 bis 2002 abgeschlossen hat. Kontaktiert wurden die Verantwortlichen aller rund 300 in diesen Jahren abgeschlossenen Evaluationen. Bei diesen Personen handelt es sich nicht um *Evaluationsdurchführende*, sondern um *Evaluationsbetreuende*. Anders als in den bisherigen Arbeiten zur Verwendungsforschung wurde nicht eine allgemeine Einschätzung der Nutzung von Evaluationen vorgenommen, sondern es wurde nach deren konkreten Verwendung gefragt. Es wurden zudem nicht nur Evaluationsverantwortliche, sondern auch weitere mögliche Nutzerinnen und Nutzer der Evaluation befragt, wobei sich herausstellte, dass sich die Ergebnisse der beiden Gruppen kaum unterschieden.

1 Verbreitung von Evaluationen

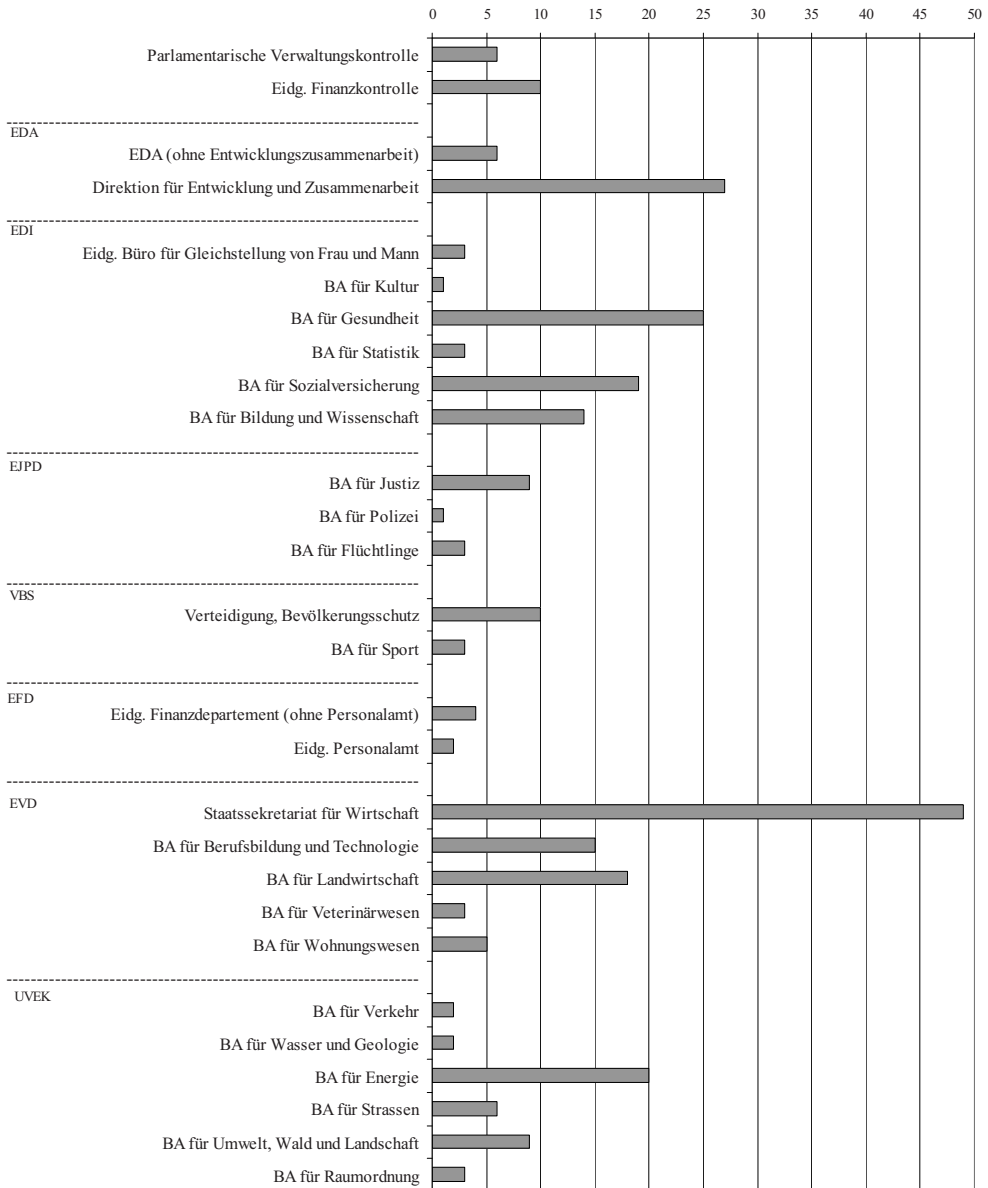
Der Ursprung der Evaluationstätigkeit in der Schweiz liegt zum einen im Nationalen Forschungsprogramm 6 „Entscheidungsvorgänge in der schweizerischen Demokratie“ (Linder 1987) und zum andern in der Arbeitsgruppe Gesetzesevaluation des Eidgenössischen Justiz- und Polizeidepartements (Arbeitsgruppe Gesetzesevaluation 1991). Zimmermann zählte für die Zeit von 1979 bis 1988 42 Evaluationen, die von der Bundesverwaltung in Auftrag gegeben worden waren (Zimmermann 1990: 80). Die diesem Beitrag zugrunde liegende Befragung geht von 308 Evaluationen in den Jahren 1999 bis 2002 aus. Der Vergleich der Zahlen weist auf ein starkes Wachstum der Evaluationstätigkeit in den letzten zwanzig Jahren hin. Tatsächlich hat die Institutionalisierung der Evaluation in der Schweiz auf Bundesebene einen hohen Grad erreicht. Zusammen mit Australien wird der Eidgenossenschaft diesbezüglich sogar ein Spitzenplatz zugeschrieben (Mader 2007). Treibende Kraft dieser Entwicklung waren nicht zuletzt die Bemühungen um eine qualitativ gute Gesetzgebung sowie um eine wirkungsorientierte Verwaltungsführung und -kontrolle.

Die Intensität der Evaluationstätigkeit unterscheidet sich jedoch innerhalb der Bundesverwaltung beträchtlich. In der Befragung wurden 45 Institutionen der Bundesverwaltung berücksichtigt. In rund einem Drittel dieser Institutionen wurden im Untersuchungszeitraum keine Evaluationen durchgeführt. Dazu gehören unter anderen die Bundeskanzlei, das Bundesamt für Polizei und das Bundesamt für Kommunikation. In den übrigen 29 Ämtern und Dienststellen wurden zwischen einer und 49 Evaluationen gezählt. Dies macht die nachfolgende Abbildung 1 deutlich. Sie listet für alle befragten Dienststellen der Bundesverwaltung die Zahl der Evaluationen auf, zu denen der Fragebogen beantwortet wurde. Weil die von uns durchgeführte Befragung einen Rücklauf von über neunzig Prozent erzielte, können wir die Zahl der Evaluationen, die wir in unsere Untersuchung einbeziehen konnten, mit der Zahl der durchgeführten Evaluationen gleichsetzen.

Wie Abbildung 1 weiter zeigt, sind am meisten Evaluationen im Staatssekretariat für Wirtschaft abgewickelt worden. Rund ein Sechstel aller Evaluationen stammt aus dieser Einheit. Sie befassen sich mit wirtschaftspolitischen Themen im Inland und mit der Entwicklungszusammenarbeit. Zahlreiche Evaluationen wurden auch bei der Direktion für Entwicklung und Zusammenarbeit, beim Bundesamt für Gesundheit, beim Bundesamt für Sozialversicherung, beim Bundesamt für Landwirtschaft und beim Bundesamt für Energie durch-

geführt. zwanzig Ämter oder Dienststellen haben zwischen einer und zehn, deren acht zwischen elf und 49 Evaluationen realisiert.

Abbildung 1: Übersicht über die Anzahl der Evaluationen nach Bundesämtern und Dienststellen (1999 bis 2002)



Legende: Grundgesamtheit ist die Anzahl der Antworten in der von Balthasar (2007) durchgeführten Befragung von Evaluationsverantwortlichen.

Einfache Erklärungen für die Unterschiede zwischen den Dienststellen gibt es nicht. Kissling-Näf und Knoepfel gehen davon aus, dass Ämter, die mit dem Management mehrerer Politikbereiche betraut sind, für Lernprozesse offener sind als solche, die nur einen einzigen „verwalten“. Für die beiden Autoren spielen aber auch Flexibilität und Aufnahmefähigkeit, das Alter der Institution und des Politikbereichs sowie der Politiktyp eine Rolle (Kissling-Näf/Knoepfel 1994: 123). Zum Teil lässt sich die unterschiedliche Verbreitung von Evaluationen auf den unterschiedlichen Legitimationsbedarf der Politikbereiche zurückführen. Dabei ist anzunehmen, dass der Legitimationsbedarf neuer Bereiche, wie sie die Umwelt-, die Drogen- oder die Energiepolitik darstellen, grösser ist als bei den traditionellen, etwa der Verteidigung, der Landwirtschaft oder der Alterssicherung. Allerdings sind auch die Kultur- oder die Migrationspolitik Bereiche mit hohem Legitimationsbedarf, obwohl sie nur sehr wenige Evaluationen aufweisen. Weiter werden programmatisch konzipierte Politiken wohl eher evaluiert als der klassische Gesetzesvollzug. Dies könnte die hohe Evaluationsintensität der Direktion für Entwicklung und Zusammenarbeit und des Bundesamts für Gesundheit erklären. Allfällige Einflüsse dieser Art werden aber überlagert von situationsspezifischen Gegebenheiten, wie sie etwa bei Evaluationsprogrammen auftreten können, die in der Schweiz im Bereich der Berufsbildung oder der Krankenversicherung durchgeführt wurden (Bussmann 1997: 32).

Die Verbreitung von Evaluationen spiegelt sich nicht nur in der Zahl der Evaluationen pro Dienststelle wider, sondern auch im Anteil des Budgets, das für Evaluationen aufgewendet wird. Die Untersuchung zeigte, dass 28.4 Prozent der Evaluationen interne und externe Kosten von unter 50'000 Franken auslösten. In 26.6 Prozent der Fälle lagen die Kosten zwischen 50'000 und 99'999 Franken. 21.2 Prozent der Evaluationen kosteten zwischen 100'000 und 250'000 Franken, während die Kosten bei 15.5 Prozent der Evaluationen höher veranschlagt wurden.

Mit bestimmten Annahmen konnte errechnet werden, dass für die in die Untersuchung einbezogenen Evaluationen insgesamt rund 32 Millionen Franken aufgewendet wurden.¹ Für die Zeit von 1999 bis 2002 können die Evaluationsaufwendungen der Bundesverwaltung pro Jahr somit auf rund 8 Millionen Franken geschätzt werden. Dieser Betrag entspricht rund 0.02 Prozent der Ausgaben des Bundes in diesen Jahren. Aus der nachfolgenden Abbildung 2 wird ersichtlich, dass die prozentualen Ausgaben der Bundesämter und Dienststellen für Evaluationen überall unter einem Prozent lagen.²

Die Abbildung 2 zeigt aber, dass sich die Ämter und Dienststellen hinsichtlich der finanziellen Aufwendungen für Evaluationen beträchtlich unterscheiden. So wird etwa im Energiebereich, verglichen mit dem Gesamtbudget, relativ viel evaluiert. Dies liegt daran, dass das im Untersuchungszeitraum laufende Programm „Energie2000“ systematisch beurteilt wurde. Im Bundesamt für Justiz dagegen hatte die Evaluation im Massnahmenvollzug deshalb grosse Bedeutung, weil dort unterschiedliche Modelle erprobt wurden. Das Bun-

¹ Es wurden folgende Annahmen getroffen: Die Durchschnittskosten einer Evaluation von unter 50'000 Franken sind 40'000 Franken, jene einer Evaluation zwischen 50'000 und 99'999 Franken sind 80'000 Franken, jene einer Evaluation zwischen 100'000 und 250'000 Franken sind 120'000 Franken, jene einer Evaluation von über 250'000 Franken sind 350'000 Franken. Für die Evaluationen ohne Kostenangabe wurde angenommen, dass diese 40'000 Franken gekostet haben.

² Die Eidgenössische Finanzkontrolle (EFK) bildet hier als Kontrollorgan eine Ausnahme.

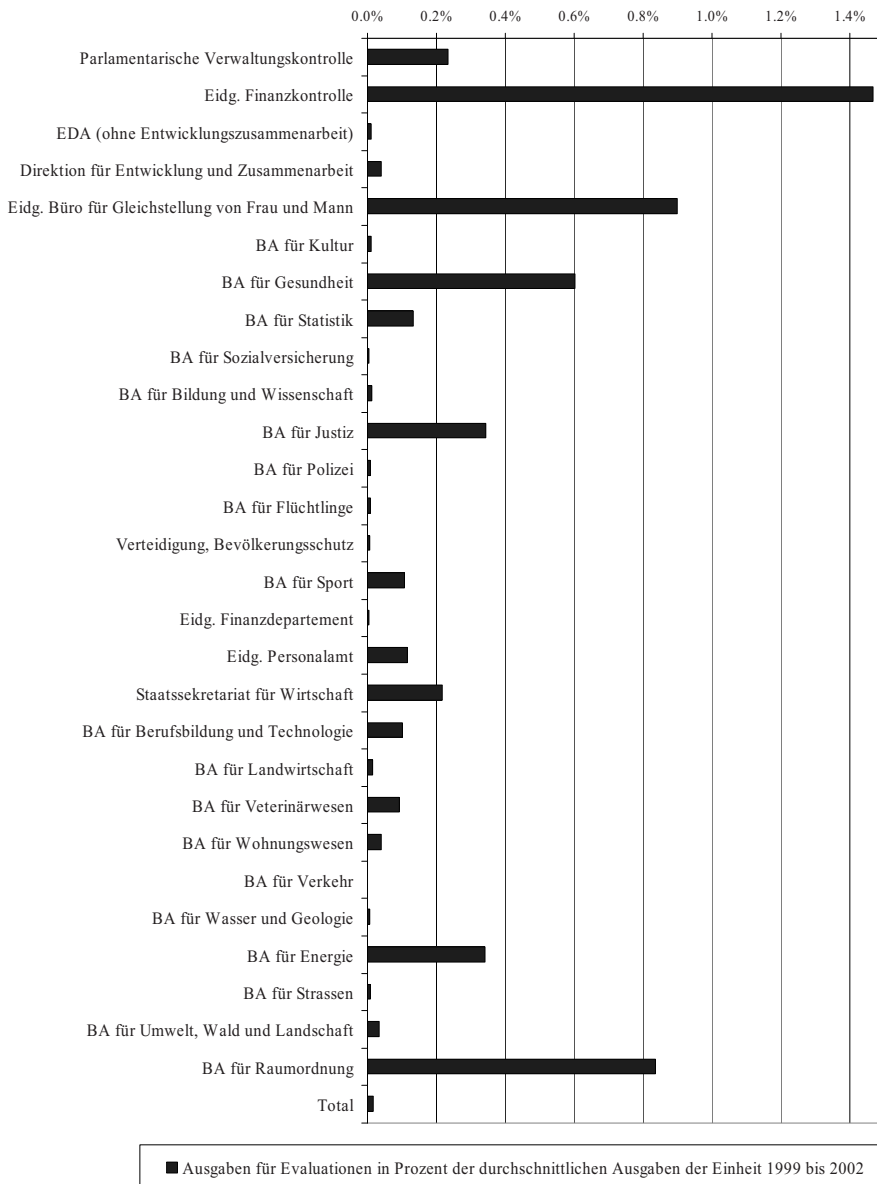
desamt für Gesundheit wiederum legt seit Jahren grossen Wert auf die Evaluation und verfügt über den verwaltungsintern am besten ausgebauten Evaluationsdienst. Das grosse relative Gewicht, das der Evaluation in der Raumordnungs- und der Gleichstellungspolitik zukommt, erklärt sich nicht etwa durch grosse Evaluationsbudgets, sondern gerade durch die bescheidenen Mittel, welche diesen Politikbereichen generell zur Verfügung stehen. Die Tatsache, dass das Eidgenössische Büro für die Gleichstellung von Frau und Mann mit 0.9 Prozent der gesamten Aufwendungen den grössten Anteil der verfügbaren Mittel für Evaluationen verwendet, muss vor diesem Hintergrund also relativiert werden. Die relativ grossen Aufwendungen der Eidgenössischen Finanzkontrolle für die Evaluation erklären sich durch deren Aufgabe als Kontrollorgan des Bundes.

Immer wieder wird auch die Frage aufgeworfen, wie gross die Aufwendungen für Evaluationen etwa sein müssen, um eine ausreichende Beurteilung der Wirkungen zu ermöglichen. Eine der ersten systematischen Studien zu dieser Thematik stammt von Marvin Alkin und Brian Stecher (Alkin/Stecher 1983). Sie untersucht vornehmlich die *Zusammensetzung* der Aufwendungen für Evaluation, nennt aber auch deren prozentualen Anteile am Gesamtbudget von verschiedenen öffentlichen Institutionen. Diese liegen zwischen einem und zehn Prozent des Budgets (Alkin/Stecher 1983: 120). Das schweizerische Bundesamt für Gesundheit sieht zehn bis 15 Prozent der Programmkosten für die Evaluation vor (BAG 1997: 24). Das Aktionsprogramm Energie 2000 wendete etwa 1.3 Prozent des gesamten Budgets für Evaluationen auf (Balthasar 2000: 210).

Gemäss Auskunft des Leiters der zentralen Evaluationsstelle der für den Haushalt der Europäischen Gemeinschaft zuständigen Direktion gab die EU in den letzten Jahren zwischen hundert und 150 Millionen Euro für Evaluationen aus, dies bei einem Budget von 109 Milliarden Euro. Evaluationen machen somit zwischen 0.09 und 0.14 Prozent des EU-Haushalts aus.³ Die ermittelten Ausgaben des Bundes für die 1999 bis 2002 durchgeführten Evaluationen liegen bei durchschnittlich 0.02 Prozent und damit unter sämtlichen genannten Richtgrössen. Im Vergleich zur EU wird in der Schweiz somit relativ wenig Geld für Evaluationen eingesetzt. Ob die Mittel generell ausreichend, zu hoch oder zu tief sind, lässt sich jedoch so nicht beurteilen, weil dies wesentlich von der Fragestellung und von der Zielsetzung der jeweiligen Evaluation abhängt: Pilotprojekte, die im Erfolgsfall flächendeckend eingeführt werden sollen, müssen intensiver analysiert werden als Projekte, die sich bereits seit Jahren bewährt haben; politisch umstrittene Massnahmen sind auf umfangreichere Evaluationen angewiesen als Massnahmen, die über eine breite Akzeptanz verfügen; Programme, die seit langem laufen, aber noch nie evaluiert wurden, müssen eher evaluiert werden als Programme, die vor kurzem bereits beurteilt wurden.

³ Information des Leiters der zentralen Evaluationsstelle der Generaldirektion Haushalt der Europäischen Gemeinschaft anlässlich der Tagung „Evaluation in der Europäischen Gemeinschaft – Aktuelle Praxis und zukünftige Trends“ der Schweizerischen Evaluationsgesellschaft vom 24. November 2004 in Bern.

Abbildung 2: Geschätzte prozentuale Ausgaben für Evaluationen nach Bundesämtern und Dienststellen



Legende: Geschätzte Ausgaben für Evaluationen in Prozent der Ausgaben der Ämter und Dienststellen der Jahre 1999 bis 2002 gemäss Schweizerische Eidgenossenschaft (2003) (vgl. Balthasar 2007)

2 Verwendung von Evaluationen

Im Zentrum der Befragung der Evaluationsverantwortlichen der schweizerischen Bundesverwaltung standen Fragen im Zusammenhang mit der Verwendung der durchgeführten Evaluationen und deren Ergebnissen. Untersucht wurden insbesondere das Ausmass der Verwendung sowie die Faktoren, welche die Verwendung beeinflussen.

2.1 Ausmass der Verwendung

65.8 Prozent der befragten Evaluationsverantwortlichen gingen von einem hohen oder eher hohen Grad der Verwendung der Evaluation aus, die sie begleitet haben. Dieses Resultat ist überraschend hoch, wird aber durch die Ansichten der befragten weiteren Personen, die als mögliche Nutzende der Evaluation in Frage kamen, bestätigt. Von dieser Gruppe sprachen mehr als sechzig Prozent von einer (eher) intensiven Verwendung. Nur zwischen 25 und dreissig Prozent aller Evaluationen waren wenig oder gar nicht verwendet worden. Zwar gibt es in der internationalen Literatur keine breit abgestützten Zahlen, die als Vergleichsgrössen herangezogen werden könnten. Die verfügbaren Hinweise lassen aber den Schluss zu, dass die erhobene Verwendungsintensität auch im internationalen Vergleich als ausgesprochen hoch beurteilt werden kann (Weiss 1998b). Dieses Ergebnis lässt sich einerseits darauf zurückführen, dass es der methodische Ansatz erlaubt hat, sehr gezielt nach den Effekten von einzelnen Evaluationen zu fragen. So wurden nicht bloss generelle Einschätzungen, sondern ganz spezifische Erfahrungen ermittelt. Auf der anderen Seite kann es aber auch sein, dass der gewählte Zugang über die für die Evaluation verantwortlichen Personen zu einer zu positiven Darstellung der Verhältnisse geführt hat, da diese oft selbst in die Durchführung der Evaluation oder der evaluierten Massnahme involviert waren. Allerdings wird die Glaubwürdigkeit der Einschätzungen durch die Tatsache unterstrichen, dass sich die Einschätzungen der ebenfalls befragten weiteren potenziellen Nutzenden nicht wesentlich von denjenigen der Evaluationsverantwortlichen unterschieden.

Die Untersuchung beschränkte sich nicht auf eine generelle Einschätzung der Verwendung der untersuchten Evaluationen, vielmehr wurden auch unterschiedliche *Arten* der Verwendung ermittelt. Dabei wurde unterschieden zwischen instrumenteller, konzeptioneller, prozessbezogener und symbolischer Verwendung. Während es bei der *instrumentellen* Verwendung um die direkte praktische Nutzung wissenschaftlicher Erkenntnisse geht, spricht die *konzeptionelle* Verwendung die indirekten Wirkungen an, und zwar durch „gradual sedimentation of insights, theories, concepts and ways of looking at the world“ (Weiss 1977: 535). Diese Form der Verwendung kommt dadurch zustande, dass die Informationen, die eine Evaluation bereitstellt, von den Entscheidungsverantwortlichen stets im Lichte ihrer eigenen Erfahrungen, ihres Wissens und ihrer Bedürfnisse betrachtet und interpretiert werden (Ledermann 2004: 13). Von *prozessbezogener* Verwendung wird dann gesprochen, wenn sich unter den am Evaluationsprozess Beteiligten ein gemeinsames Verständnis bezüglich relevanter Probleme im Bereich der untersuchten Massnahme entwickelt. Prozessbezogene Verwendung kommt dadurch zustande, dass Akteure allein durch die Teilnahme am Evaluationsprozess – und nicht erst aufgrund der Ergebnisse – Einblicke in das Pro-

gramm gewinnen, die ihr Verhalten beeinflussen (Vedung 1999: 242). Von *symbolischer* Verwendung von Evaluationsergebnissen wird gesprochen, wenn Entscheidungstragende Evaluationen dazu nutzen, ihre Optik zu bestätigen und sich Legitimation zu verschaffen (Henry/Rog 1998: 90). Symbolische Verwendung kann eine positive Konnotation haben, wenn die bestehende positive Einschätzung bestätigt wird. Sie kann aber auch negativ besetzt sein: dann nämlich, wenn die Verantwortlichen aus einer Evaluation gar nichts lernen wollen. In der nachfolgenden Abbildung 3 sind die Resultate der Befragung zu den verschiedenen Arten der Nutzung von Evaluationen dargestellt.

Abbildung 3: Verwendung von Evaluationen

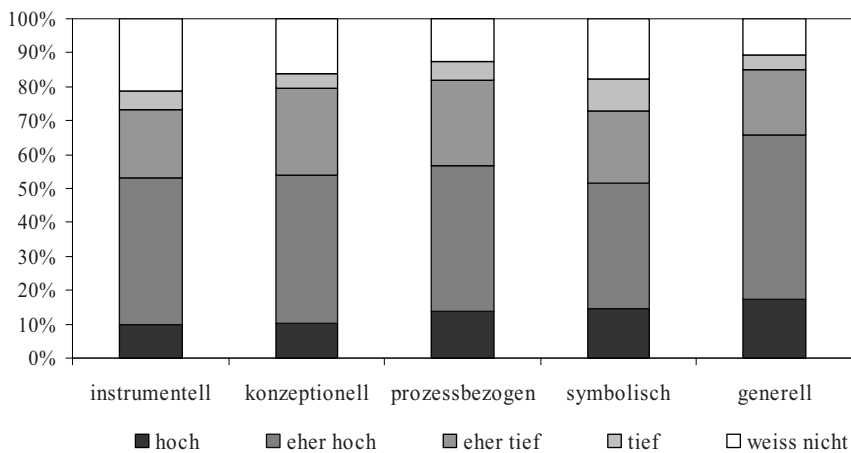


Abbildung 3 zeigt, dass das Ausmass der *prozessbezogenen* Verwendung als überdurchschnittlich hoch eingeschätzt wurde; dieses war gemäss den Angaben der Evaluationsverantwortlichen in 56.8 Prozent der Fälle „hoch“ oder „eher hoch“. Personen, die in die Evaluationsprojekte involviert waren, unterstreichen damit, dass sie im Prozess der Evaluation viel gelernt haben, und dass sie das Gelernte für ihre zukünftige Arbeit brauchen können. Auch die Verwendung der Empfehlungen, das heisst die *instrumentelle* Verwendung, wird als relativ hoch beurteilt. Die Evaluationsverantwortlichen gaben an, dass in 257 von 278 Evaluationen Empfehlungen formuliert worden sind. Bei 136 der 278 Evaluationen wurde von einer (eher) hohen instrumentellen Verwendung ausgegangen, da mehr als die Hälfte der Empfehlungen umgesetzt wurde (48.9 Prozent aller Evaluationen). In 66 Fällen (23.7 Prozent aller Evaluationen) wurde die instrumentelle Verwendung als „(eher) tief“ eingeschätzt, da weniger als die Hälfte der Empfehlungen umgesetzt wurde. Somit trat eine instrumentelle Verwendung sehr häufig auf. Damit bestätigt sich die Beobachtung von Stamm, dass „direkte Rezeption und Nutzung deutlich häufiger festgestellt werden kann als in der einschlägigen theoretischen Literatur behauptet“ (Stamm 2003: 326).

Eine *konzeptionelle* Verwendung trat mindestens ebenso häufig auf wie die instrumentelle. Von den Evaluationsbeauftragten gaben rund 54 Prozent an, dass die konzeptionelle Verwendung „hoch“ oder „eher hoch“ war. Carol Weiss und andere Autoren haben die Ansicht vertreten, dass die Verwendung von Evaluationen in erster Linie über das „enligh-

tenment“ erfolge, also eine konzeptionelle sei (Weiss 1977). Unsere Arbeit zeigt dagegen, dass die konzeptionelle Verwendung nicht häufiger auftrat als die instrumentelle.

Der bedeutendste Unterschied zwischen den Evaluationsverantwortlichen und den weiteren potenziell Nutzenden zeigte sich bei der Einschätzung der *symbolischen* Verwendung. Über fünfzig Prozent der Evaluationsverantwortlichen schätzten die symbolische Verwendung als „(eher) hoch“ ein; bei den weiteren Nutzenden lag dieser Anteil bei fast 65 Prozent. Ob die konstatierte symbolische Verwendung positiv oder negativ konnotiert war, liess sich aus der Befragung nicht eruieren.

2.2 Einflussfaktoren auf die Verwendung

Die Verwendung der Ergebnisse von Evaluationen wird von ganz unterschiedlichen Faktoren bestimmt. Dies machten zahlreiche wissenschaftliche Untersuchungen deutlich (Balthasar 2007). Die diesem Beitrag zugrunde liegende Untersuchung ging von einem Wirkungsmodell aus, das die Verwendung von Evaluationen als Ergebnis des Zusammenspiels von institutionellen Faktoren versteht, nämlich: die Lancierung und Durchführung einer Evaluation, sodann Kontextfaktoren, die das Umfeld der Evaluation betreffen, und schliesslich prozessbezogene Faktoren. Das Modell knüpft an die Arbeiten von Cousins und Leithwood (1986; 1993) an, die kontext- und prozessbezogene Faktoren unterscheiden.

Als von herausragender Bedeutung für die Verwendung einer Evaluation hat sich das Interesse der Entscheidungstragenden an ihr erwiesen. War ein solches vorhanden, verbesserten sich die Chancen auf eine Verwendung massiv. Zudem müssen die Evaluationen so angelegt sein, dass ihnen ein Umsetzungspotenzial zugesprochen wird. Für die Befragten war also wichtig, dass die Evaluation die richtigen Fragen beantwortete, dass sie die Informationen zum richtigen Zeitpunkt lieferte, und dass der Sprachstil, in der die Evaluation abgefasst war, für die Nutzenden verständlich und hilfreich war. Schliesslich hat sich auch eine positive Einschätzung des Kosten-Nutzen-Verhältnisses der Evaluation günstig auf die Verwendung ausgewirkt.

Neben diesen prozessbezogenen Faktoren spielt auch die Institutionalisierungsform eine Rolle. Die Verwendung von Evaluationen scheint jedoch weniger von der Distanz zwischen Evaluierenden und Evaluierten an sich abzuhängen als vielmehr von der Ausprägung einzelner Faktoren, die diese Distanz bestimmen. So haben die Ergebnisse der empirischen Analyse gezeigt, dass die instrumentelle Verwendung, das heisst die Umsetzung der Empfehlungen einer Evaluation gefördert wird, wenn die verantwortlichen Institutionen die Kosten für die Evaluation selber tragen müssen. Demgegenüber ist eine ausgeprägt konzeptionelle Verwendung dann zu erwarten, wenn externe Evaluierende eingesetzt werden. Diese verfügen nämlich in der Regel über breite methodische Kenntnisse sowie Erfahrung in der Durchführung von Evaluationen, was für eine konzeptionelle Verwendung besonders wichtig ist. Die Untersuchung zeigt weiter, dass sich die prozessbezogene Verwendung signifikant verbesserte, wenn die Evaluation durch diejenigen Personen ausgelöst wurde, die sie später auch umsetzten, und/oder wenn sie amtsintern durchgeführt wurde.

Generell stützt unsere Arbeit die These der lernorientierten, interaktionistischen Evaluationstheorien, die besagen, dass eine intensive Verwendung der Resultate dann zu erwar-

ten ist, wenn die Evaluierenden eng mit den Evaluierten zusammenarbeiten. Michael Pattons Forderung, Evaluationen sollten sich ganz an den Bedürfnissen der Nutzenden orientieren, gilt besonders für die prozessbezogene Verwendung (Patton 1997). Nur in Ausnahmefällen ist die Verwendung von Evaluationen stärker, wenn die Distanz zwischen Evaluierenden und Evaluierten zunimmt. Am ehesten ist dies bei Evaluationen der Fall, die konzeptionelle oder symbolische Verwendung finden. Allerdings sollten auch in diesen Fällen die Evaluationen so angelegt sein, dass ihnen von den möglichen Nutzenden ein Umsetzungspotenzial zugesprochen wird. Die in der Tradition von Donald Campbell – zum Beispiel von Davies, Nutley und Smith – vertretene Argumentation, dass die Distanz zwischen Evaluierenden und Evaluierten Unabhängigkeit signalisiere und damit die Akzeptanz der Ergebnisse erhöhe (Campbell/Stamley 1963; Davies et al. 2000), kann aufgrund unserer Untersuchung am ehesten für die symbolische Verwendung bestätigt werden. Entgegen den Vermutungen von Ray Rist zeigt die Untersuchung ausserdem, dass die Glaubwürdigkeit der Information, verglichen mit der Orientierung an den spezifischen Bedürfnissen der Nutzenden, für die Verwendung eine untergeordnete Bedeutung hat (Rist 1994: 200).

3 Nutzung durch die Politik

Eine wichtige Zielgruppe von Evaluationen sind Politikerinnen und Politiker. Bei der Befragung der Evaluationsverantwortlichen haben wir festgestellt, dass bei rund zwanzig Prozent der untersuchten Evaluationen vorgesehen war, dass auch das Parlament deren Ergebnisse verwendet. Ein Beispiel dafür ist der Bericht „Das Kartellrecht: Standortbestimmung“, der von der Parlamentarischen Verwaltungskontrolle zuhanden der Geschäftsprüfungskommission des Nationalrates erarbeitet wurde (Parlamentarische Verwaltungskontrolle 2000). Die Verwendung dieses Berichts wurde in einer Fallstudie vertieft untersucht (Balthasar 2007). Dabei zeigte sich, dass die Geschäftsprüfungskommission mit dem Bericht unter anderem die Absicht verfolgte, „einen gewissen politischen Druck auf den Bundesrat auszuüben. Die Kartellgesetzesrevision war lange geplant, aber die Botschaft brauchte ziemlich viel Zeit, bis sie überhaupt erarbeitet wurde.“ Darüber hinaus wollte sich die Kommission mithilfe des Berichts in die Lage versetzen, die Qualität eines Entwurfs des Bundesrats zur Revision des Kartellgesetzes beurteilen zu können. Als nun aber der Bericht zur Diskussion vorlag, standen zusätzlich weitere Informationsquellen zur Entscheidungsfindung zur Verfügung, namentlich ein Vernehmlassungsentwurf des zuständigen Departements sowie eine vom Bundesrat veranlasste ergänzende Untersuchung. Gegenüber diesen beiden Informationsquellen kam der Bericht der Parlamentarischen Verwaltungskontrolle etwas spät: „Es gab eine Art Konkurrenz. Es gab keine grüne Wiese. [...] Wenn parallel zu einer Evaluation plötzlich etwas in der Verwaltung gemacht wird, wird es schwierig. Das ist von Zeit zu Zeit auch eine Taktik der Verwaltung: Man macht nichts und wenn man sieht, dass sich die Oberaufsicht um ein Thema kümmert, dann beginnt man, Aktivismus zu betreiben. Und wenn das Produkt dann vorliegt, kann man sagen: Wir haben schon alles vorgeschlagen.“

Vor diesem Hintergrund wurde der Bericht nach Meinung der Befragten eher wenig verwendet. Er sei wohl wahrgenommen und in der Botschaft des Bundesrates auch zitiert worden – die Entscheidungsfindung jedoch habe der Bericht kaum beeinflusst. Dessen

Nützlichkeit wurde insbesondere durch den Umstand beeinträchtigt, dass die Verwaltung fast parallel dazu eigene Arbeiten durchführte und ein revidiertes Gesetz zur Stellungnahme unterbreitete. Dadurch kam die Kommission zeitlich unter Druck; sie konnte sich nicht wie sonst üblich vertieft mit dem Bericht auseinandersetzen. Immerhin nahm der Gesetzesvorschlag des Bundesrates in weiten Teilen die georteten Revisionsanliegen auf.

Das Beispiel zeigt, dass es nicht einfach ist, Evaluationen so anzulegen, dass Politikerinnen und Politiker diese auch verwenden. Zum einen stehen parlamentarische Organe in der Regel unter einem hohen Zeitdruck, dem die für Evaluationen notwendigen empirischen Arbeiten kaum entsprechen können. Zum anderen können Politikerinnen und Politiker oft auch auf andere wichtige Informationsinstrumente zur Entscheidungsfindung zurückgreifen. In einer funktionstüchtigen Demokratie liefern auch Parteien und Interessengruppen, administrative und parlamentarische Aufsichtsorgane, Gerichte und Medien permanent Rückmeldungen über den Vollzug und die Wirkungen staatlicher Massnahmen (Janett 2003: 99).

Die Befragung der Evaluationsverantwortlichen hat dennoch ergeben, dass das Parlament die an es gerichteten Evaluationen mehrheitlich auch nutzt. 58.5 Prozent der Evaluationen mit dieser Zielgruppe haben gemäss Einschätzung der Evaluationsverantwortlichen dort auch Verwendung gefunden. Dieselbe Frage haben wir auch den weiteren Nutzenden gestellt. Obwohl in ihrer Einschätzung etwas zurückhaltender, gaben auch sie an, dass das Parlament 45 Prozent der an sie gerichteten Evaluationen genutzt habe. Aus der aktuellen Politik gibt es denn auch etliche Beispiele für die Verwendung von Evaluationen im Parlament. Dazu gehören etwa die Nationalratsdebatte über die Gleichstellung von Frau und Mann, die unter anderem auf der Grundlage eines Berichts des Bundesrates über die Evaluation der Wirksamkeit des Gleichstellungsgesetzes erfolgt ist. Ein weiteres Beispiel ist die Totalrevision des Opferhilfegesetzes, mit der sich der Ständerat befasst hat – sie wurde zuletzt durch eine Evaluation des Vollzugs und der Wirksamkeit dieses Gesetzes ausgelöst (Mader 2007).

Insgesamt darf der Stellenwert von Evaluationen für die Politik aber nicht überschätzt werden. Politikerinnen und Politiker reagieren auf Evaluationen in erster Linie dann, wenn Missstände aufgedeckt werden, wenn sie in ihren eigenen Überzeugungen bestätigt werden, oder wenn sie sich mit Evaluationen profilieren können. Weil diese aber oft zu spät kommen, eignen sich Evaluationen kaum zur Profilierung. So reagieren Politikerinnen und Politiker weniger auf Evaluationen als vielmehr auf Krisen, auf Medienberichterstattungen und auf Interessengruppen, die an sie herantreten.

Obwohl sich Politikerinnen und Politiker also nur selten für die Ergebnisse einzelner Evaluationen interessieren, kann angenommen werden, dass sie den Wert von Vollzugs- und Wirkungsanalysen für die Verwaltung durchaus anerkennen. Politikerinnen und Politiker scheinen Evaluationen grundsätzlich als nützlich zu betrachten. Bestätigt wird dies nicht zuletzt durch die Tatsache, dass der Artikel 170 der Bundesverfassung im Vorfeld der Abstimmung kaum zu Diskussionen Anlass gab.

4 Schluss

Der vorliegende Beitrag setzt sich mit der Verbreitung und Nutzung von Evaluationen in der schweizerischen Bundesverwaltung auseinander. Aufgrund einer umfangreichen Erhebung und Analyse aller zwischen 1999 und 2002 abgeschlossenen Evaluationen konnte festgestellt werden, dass in der Schweiz auf nationaler Ebene eine rege Evaluationstätigkeit herrscht. Von einer flächendeckenden Evaluationskultur kann dennoch nicht gesprochen werden. Vielmehr variiert der Grad der Evaluationstätigkeit innerhalb der Bundesverwaltung sehr stark. Rund ein Drittel der Bundesämter und Dienststellen hat im Untersuchungszeitraum keine Evaluation durchgeführt. Bei einem weiteren Drittel kann höchstens von einer sporadischen Evaluationstätigkeit ausgegangen werden. Ein Drittel der Einheiten der Bundesverwaltung hingegen führt regelmässig Evaluationen durch – diese haben sich so in den letzten Jahren wertvolle Erfahrung in Planung, Begleitung und Umsetzung von Wirkungsuntersuchungen aneignen können. Ein Vergleich der Ausgaben des Bundes für Evaluationen mit jenen der Europäischen Kommission hat allerdings ergeben, dass in der Schweiz relativ wenig Geld für Evaluationen ausgegeben wird.

Wie unsere Untersuchung zeigt, bescheinigen sowohl die befragten Evaluationsverantwortlichen als auch die weiteren möglichen Nutzenden den durchgeführten Evaluationen mehrheitlich eine gute Verwendung. Knapp zwei Drittel der befragten Evaluationsverantwortlichen gingen von einem hohen oder eher hohen Grad der Verwendung aus. Dabei waren es vor allem drei Faktoren, die sich günstig auf die potenzielle Verwendung einer Evaluation ausgewirkt haben: das Interesse der Entscheidungsträger an ihr, eine positive Beurteilung ihres Umsetzungspotenzials sowie eine positive oder ausgeglichene Kosten-Nutzen-Bewertung. Was die institutionellen Rahmenbedingungen betrifft, scheint die Verwendung von Evaluationen weniger von der Distanz zwischen Evaluierenden und Evaluierten an sich abzuhängen, als vielmehr von gewissen Faktoren, die diese Distanz bestimmen. So werden etwa die Empfehlungen einer Evaluation viel eher umgesetzt, wenn die für die Umsetzung verantwortlichen Ämter oder Dienststellen die Kosten der Evaluation selber tragen müssen.

Evaluationen richten sich in erster Linie an die Verwaltung. Von dieser Akteurgruppe werden sie denn auch häufig verwendet. Hingegen scheint es schwierig zu sein, Evaluationen so anzulegen, dass sie auch von Politikerinnen und Politikern verwendet werden. Evaluationen stehen nicht zuoberst auf deren Prioritätenliste, zumal, wenn es sich um retrospektive, das heisst die Vergangenheit betreffende Evaluationen handelt: „Politicians have to muddle through an information jungle – acting more like artists, using intuition and judging all sorts of available information“ (Lemne/Sohlman 2004: 10). Politische Entscheide fallen nach politischen Kriterien, sachliche Kriterien sind relevant, aber nicht dominant.

Nutzung der Evaluationsfunktion im Vergleich

1 Einleitung

Ein zentraler Aspekt professioneller Evaluation, im Evaluationsalltag allerdings oft vernachlässigt, ist derjenige der *Nutzung von Evaluationen*. Was geschieht mit den Evaluationsberichten, mit den Ergebnissen und Erkenntnissen? Werden Schlussfolgerungen und Empfehlungen von den Adressaten und Adressatinnen wahrgenommen und in Planung und Umsetzung von Politiken, Programmen und Projekten berücksichtigt? Werden Prozesse der Optimierung ausgelöst oder unterstützt? Warum werden Evaluationen in Auftrag gegeben und durchgeführt?

Die Gründe, warum der Nutzung von Evaluationen in der Praxis bisweilen wenig Beachtung geschenkt wird – während die Forschung diesem Thema durchaus grössere Aufmerksamkeit schenkt (vgl. Widmer/Neuenschwander 2004: 391) – und was getan werden kann, um die Nutzung zu fördern, wurden an der Tagung ‚Evaluation im deutschsprachigen Raum‘ (2007) diskutiert.¹ Der Beitrag versucht einige der Diskussionsbeiträge und Voten aufzunehmen, sie zu systematisieren und zu verdichten, um sie dann mit Ergebnissen aus der Literatur zu reflektieren und zu kontrastieren.

Ein Blick in die für den deutschsprachigen Raum massgeblichen Evaluationsstandards zeigt, dass dem Nutzenaspekt in den Formulierungen unterschiedliche Wichtigkeiten zugeschrieben werden. In der Übersetzung der Standards des Joint Committee on Standards for Educational Evaluation (2006) wird unter den *Nützlichkeitsstandards* der Standard N7 *Wirkung der Evaluation* aufgeführt: „Evaluationen sollen so geplant, durchgeführt und dargestellt werden, dass die Beteiligten & Betroffenen dazu ermuntert werden, dem Evaluationsprozess zu folgen, damit die Wahrscheinlichkeit steigt, dass die Evaluation genutzt wird“ (ebd.: 87). Die Schweizerische Evaluationsgesellschaft (SEVAL) verwendet eine ähnliche Formulierung (Widmer et al. 2000: 6), während die Deutsche Gesellschaft für Evaluation (DeGEval) in der Bezeichnung des Standards direkt auf den Nutzungs- und Nutzenaspekt eingeht: *N8 Nutzung und Nutzen der Evaluation* (DeGEval 2002: 9). Auch die Beschreibung des Standards ist bei der DeGEval eindringlicher: „Planung, Durchführung und Berichterstattung einer Evaluation sollen die Beteiligten und Betroffenen dazu ermuntern, die Evaluation aufmerksam zur Kenntnis zu nehmen und ihre Ergebnisse zu nutzen“ (DeGEval 2002: 9).

¹ Der vorliegende Beitrag basiert teilweise auf den Kommentaren von Brigitta M. Gadiant, Nationalrätin (Schweiz), Julia Filtzinger, Bundesrechnungshof (Deutschland) und Rupert Pichler, Bundesministerium für Verkehr, Innovation und Technologie (Österreich) sowie den Beiträgen von Spiel/Bergsmann und Balthasar (in diesem Band) und nimmt Aussagen und Voten der Plenumsdiskussion auf.

Aufschluss über die Problematik rund um die Nutzung der Evaluationsfunktion² geben die weiteren Ausführungen zum genannten Standard:

„Selbst mit hoher Professionalität erzeugte Ergebnisse werden häufig nicht genutzt oder ganz anders, als dies in den Vereinbarungen zur Evaluation geplant war. Wichtige Voraussetzungen für eine intendierte Verwendung liegen in einer hohen Durchführbarkeit, Fairness und Genauigkeit von Evaluationen [...]. Besonders die Umsetzung von Schlussfolgerungen oder Empfehlungen hängt stark davon ab, ob die Adressaten und Adressatinnen die Erwartung haben, dass das Evaluationsprojekt ihnen und ihren Vorhaben von Nutzen ist. Dabei ist es nicht selten so, dass unterschiedliche Adressaten auch unterschiedliche Nutzenerwartungen haben, die sich widersprechen oder sich gegenseitig ausschließen können. Eine wichtige Voraussetzung, um eine vereinbarte bzw. intendierte Nutzung und damit auch den Nutzen von Evaluationen zu fördern, ist die angemessene Einbeziehung der verschiedenen Adressaten und Adressatinnen in die Planung und die Durchführung der Evaluation sowie die Aufbereitung von Ergebnissen. Ebenfalls positiv wirkt es sich aus, wenn im Verlauf der Untersuchung kontinuierlich und transparent Rückmeldungen gegeben werden (Stockbauer 2000[b]). Dies fördert, dass abgesehen vom Ergebnisnutzen bereits im Verlauf einer Evaluation ein Prozessnutzen eintritt (Patton 1998).“ (DeGEval 2002: 25).

In dieser Ausführung des Standards werden namentlich prozessbezogene Aspekte verdeutlicht. Die Frage ist, ob auch strukturelle Rahmenbedingungen die Nutzung von Evaluationen fördern können. Andererseits werden in diesem Standard die relevanten Punkte benannt. Deren konsequente Beherzigung in der Praxis würde die Nutzung der Evaluationsfunktion sicher in vielen Fällen stärken.

2 Was ist eine Evaluation? – Die Sicht der Praxis

Bevor man sich der Frage der Nutzung der Evaluationsfunktion widmen kann, ist eine Klärung dessen von Interesse, was eine Evaluation ist beziehungsweise was die Politik und Verwaltung unter Evaluation verstehen oder auch, welche Form von Berichterstattung die Funktion einer Evaluation übernimmt.

Das Schweizer Parlament beispielsweise arbeitet mit verschiedenartigen Berichten, um Sachverhalte und Politiken beurteilen und anstehende Entscheide unterstützen zu können. Der Begriff *Evaluation* taucht dabei explizit selten auf. Entsprechende Dokumente werden meist als Bericht, Prüfung oder Studie bezeichnet. Anzunehmen ist, dass es mehr Evaluationen gibt, als in der parlamentarischen Arbeit als solche bewusst zur Kenntnis genommen werden. Entsprechend ist vermutlich auch die Nutzung der Evaluationsfunktion bedeutender als allgemein angenommen.³ Ähnliches wird aus Österreich bestätigt, wo mit Begriffen

² In Anlehnung an den Beitrag von Spiel/Bergsmann (in diesem Band) wird der Begriff ‚Evaluationsfunktion‘ eher umfassend gemeint. Verschiedene Aspekte spielen hier eine Rolle: Die Evaluation selbst (Durchführung, Berichterlegung, Empfehlungen), die Nutzung der Evaluationsfunktion (institutioneller Kontext und Verankerung) sowie der Nutzen der Evaluation für die Stakeholder (inkl. Verwendung und Umsetzung von Ergebnissen und Empfehlungen).

³ Eine etwas andere Sicht hat Balthasar (in diesem Band), indem er feststellt, dass die Verwaltung häufiger Evaluationsberichte verwendet, die Politik aus verschiedenen Gründen hingegen weniger. Politiker und Politikerinnen würden sich vielmehr auch auf andere, ganz unterschiedliche Informationsquellen abstützen.

wie Bericht oder Information gearbeitet wird. Umgekehrt wird aus Deutschland berichtet, dass beispielsweise das, was als Evaluation deklariert wird, oft nicht den Anforderungen an Evaluationen gemäss Bundeshaushaltsordnung BHO, Paragraph 7 (Wirtschaftlichkeit und Sparsamkeit, Kosten- und Leistungsrechnung) entspricht (Jusline o.J.). Die festgestellten Mängel können vermutlich auf ein nicht einheitliches Verständnis von Evaluation zurückgeführt werden. Was als Evaluation deklariert wird, sind oft Berichterstattungen an die Bundesregierung, die ganz überwiegend auf Selbstauskünfte der politikumsetzenden Administrationen zurück gehen. In der Verwaltung wird eine enorme Zahl von Vorgängen als Evaluation bezeichnet. Die Spannweite reicht dabei von einer normalen Überprüfung des bestimmungsgemässen Einsatzes der Mittel im Rahmen einer Verwendungsnachweisprüfung, bis zu einer umfangreichen gesamtwirtschaftlichen Analyse mit vorgeschalteter Effektivitätskontrolle anhand quantitativer Methoden.

In der Praxis zeigt sich, dass es in gewissen Kontexten viele Evaluationen gibt, ohne dass diese als solche bezeichnet werden. In anderen Kontexten hingegen wird sehr Unterschiedliches als Evaluation bezeichnet, auch wenn es weit von den Anforderungen professioneller Evaluation entfernt ist. Diese Heterogenität an Bezeichnungen für Evaluationen (oder evaluationsähnlicher Berichte) sowie die Spannweite betreffend Ziele, Methodik und Ausführung dieser Berichte zeigt, dass die Einschätzung der Nutzung schwierig ist. Widmer (2008a: 272-273) hält zu recht fest, dass es neben der Evaluation auch verwandte Zugänge gibt und nennt folgende Formen: Audit (Prüfung), Controlling (Führung), Qualitätsmanagement (Führung), Performance Measurement (Leistungsmessung), Monitoring (Beobachtung) und schliesslich die Evaluation (Bewertung). Entsprechend sollte in der Praxis – trotz nicht immer trennscharfen Abgrenzungen der genannten Formen – die jeweils passende eingesetzt werden.

Diese Ausführungen zeigen, dass am gemeinsamen Verständnis der verschiedenen Akteure, was Evaluationen sind, was sie leisten können und was Evaluationen nicht sind, noch gearbeitet werden muss.

3 Die Vielschichtigkeit der Nutzung der Evaluationsfunktion

Welche Funktion oder welchen Einfluss Evaluationen haben können, wird in Politik und Verwaltung unterschiedlich gesehen und ist facettenreich. Eine Verständigung darüber ist indessen sinnvoll, damit die Nutzung optimiert und die Nutzenerwartungen von den Beteiligten besser eingeschätzt werden können. Wie bei Balthasar sowie Spiel/Bergsmann (in diesem Band) schon angesprochen, gibt es zum Thema ‚Nutzung und Nutzen von Evaluationen‘ Grundlagen. Zusammenfassend nennt beispielsweise Brandt (2007: 190-194; vgl. auch Horton et al. 2003: 111-118) folgende Aspekte zur Verwendung von Evaluationen:

- a. *Direkte oder instrumentelle Nutzung:* Entspricht unmittelbarer und praxisbezogener Nutzung von Evaluationsergebnissen durch Auftraggeber/innen und weiteren Stakeholdern zur Entscheidungsfindung oder Gestaltung der Programme.

- b. *Konzeptionelle Nutzung*: Betrifft grundlegende Erkenntnisse über Funktionsweisen eines Programms und bewirkt eher indirekte Veränderungen von Kognitionen, also Überzeugungen, Meinungen und Einstellungen (vgl. auch Weiss 1977: 535).
- c. *Überzeugungs-Nutzung*: Nutzen liegt in der Bekräftigung oder Widerlegung politischer, fest verankerter Positionen.
- d. *Prozess-Nutzung*: Die Durchführung selbst kann einen Nutzen haben (Veränderung von Wahrnehmungen, Förderung von Lernprozessen etc.).

Vedung (1997, zit. nach Widmer/Neuenschwander 2004: 392) ergänzt diese Typologie mit dem Aspekt der *taktischen* Nutzung von Evaluationen. Hier ist gemeint, dass eine Evaluation entweder Zeit gewinnen oder die Übernahme von Verantwortung umgehen will. Es handelt sich hierbei allerdings um eine Nutzung, die von Widmer (2008a: 274) als „stets problematisch“ bewertet wird.

Stamm (2003: 233-279) ihrerseits postuliert vier *empirisch* definierte Typen von Evaluationsnutzung:

- a. *Typ Blockade*: Evaluationen werden aus verschiedenen Gründen mit Widerständen oder konflikthaften Situationen konfrontiert, direkte Auswirkungen bleiben aus.
- b. *Typ Innovation*: Erneuerungsprozesse, Erkenntniszuwachs und Lernprozesse sind möglich.
- c. *Typ Alibi*: Es geht um Evaluationen mit reinem Bestätigungs- und Rechtfertigungscharakter.
- d. *Typ Reaktion*: Reflexhafte, emotionale oder rationale Wirkungen der Evaluation, Entscheidungsprozesse sind schnell, bewusst oder unbewusst.

Mit Blick auf die Verwendung von Evaluationen betont Stamm, dass den vier genannten Typen die Filterung der Befunde gemeinsam ist: „Erstens durch den Filter der subjektiv gefärbten Orientierungen und zweitens durch den Filter der Ergebnisreduktion über eine selektive Auswahl derjenigen Befunde, die am ehesten mit den aktuellen Entwicklungstendenzen kompatibel sind“ (Stamm 2003: 289). Subjektive Selektionsprozesse aber auch zugrunde liegende Einstellungen beeinflussen somit zumindest teilweise, ob und welche Ergebnisse und Erkenntnisse aus Evaluationen wahrgenommen und verwendet werden.

Die zwei skizzierten Typologien zur Nutzung von Evaluationen zeigen, insbesondere wenn man diese noch kombinieren möchte, wie komplex die Bedingungen sind, dass eine Nutzung der Evaluation stattfinden kann.

Aus Sicht der Politik und Verwaltung kann festgehalten werden, dass grundsätzlich Übereinstimmung besteht, dass Evaluationen nicht das alles lösende Instrument oder nicht die einzige relevante Grundlage für Entscheidungen sein sollen und sein können. Am Schluss entscheidet die Politik. Die Nutzung von Evaluationen ist somit über die genannten Nutzungstypen hinausgehend mitunter ein politischer Prozess. Auf der anderen Seite sind Evaluationen die unabdingbare Voraussetzung für das Handeln in der Verwaltung und Grundlage für politische Entscheidungen. Darum ist es sehr wichtig, dass das Bewusstsein für dieses Instrument noch weiter verstärkt wird. Evaluationen und deren Nutzen sollen nicht überbewertet aber auch nicht vernachlässigt werden. Schliesslich ist es Aufgabe der

Politik, Entscheidungen zu treffen, die nicht nur auf Intuition oder programmatischen Überzeugungen basieren, sondern auch evidenzbasiert sind. Möglichst fundierte und objektive Ergebnisse ermöglichen es der Politik, auf einer breiten Entscheidungsgrundlage möglichst rational entscheiden zu können. Evaluatoren und Evaluatorinnen können und sollen ihrerseits nicht den Anspruch hegen, dass nur sie eine Sachlage richtig einschätzen und somit die einzig valable Grundlage für Entscheidungen liefern können. Auch von dieser Seite gilt es, den Stellenwert von Evaluationen richtig einzuschätzen.

Je nach Stakeholder und Perspektive stehen in der Praxis unterschiedliche Nutzenaspekte im Vordergrund.

Ein Punkt betrifft den von den verschiedenen Stakeholdern *erwartete Nutzen*. Was steht bei ihnen im Vordergrund? Geht es um Kontrolle, um Legitimation, um Rechenschaftslegung, um Entwicklungsaspekte oder um Anderes? Hier braucht es unbedingt eine Passung der Erwartungen an den Nutzen und der konkreten Aufbereitung und Form der Evaluation.

Das *Zeitmoment*, also der Zeitpunkt, bei welchem die Ergebnisse und Schlussfolgerungen einer Evaluation zurückgemeldet oder publiziert werden, ist entscheidend hinsichtlich der Nutzungsintensität. Es gibt Evaluationen, die für die Planungs- und Entscheidungsprozesse zu spät fertig werden und deren Nutzung durch (politische) Entscheide vorweggenommen wird. Solche Beispiele sind nicht selten und für viele Beteiligte unbefriedigend. Diese ungünstigen zeitlichen Konstellationen lassen sich ebenfalls durch Planung und Vereinbarungen wenn nicht vermeiden so doch mindern. Nicht selten sind jedoch Entscheidungsträger ungeduldig oder die politischen Rahmenbedingungen ändern sich, so dass Entscheide trotz guter Evaluationsplanung vorweggenommen werden. Bei einigen Evaluationsvorhaben spielt das Zeitmoment möglicherweise eine weniger wichtige Rolle. Erkenntnisse können durchaus auf andere Evaluationsgegenstände oder zumindest auf ähnliche Programme übertragen werden. Der Nutzen ist dann nicht mehr direkt mit dem Evaluationsgegenstand verknüpft und entspricht eher den Typen „wissensvermittelnd“ oder „wissensgenerierend“ (vgl. Drilling in diesem Band).

Von Interesse ist das Verhältnis zwischen *Wissenschaft*, hier namentlich die Evaluation, und *Politik*. „Politik vereinnahmt Wissenschaft“, so eine Sicht der Dinge. Vereinnahmt die Politik die Wissenschaft tatsächlich, um ihre Entscheide zu legitimieren? Die Praxis zeigt, dass es immer Studien, Gutachten und Gegengutachten gibt. Gesucht wird, was eigenen Überzeugungen entspricht (vgl. die weiter oben genannten „Filter“ von Stamm). Das kann ein unbewusster Prozess sein, der als selektive Wahrnehmung oder Suche nach konsistenter Information bezeichnet wird, wie die Sozialpsychologie meint (Stroebe et al. 1996: 621). Im Politikumfeld ist es eventuell eine bewusste Selektion. Eine andere Sicht der Dinge ist nüchterner und betont, dass Wissenschaft, resp. Studien und Evaluationen eine wichtige Grundlage für politische Prozesse und Entscheidungen – unter mehreren – sind. Weiss (1977) spricht in diesem Zusammenhang auch von der „enlightenment function of social research“, also von nicht direkt erkennbaren Prozessen, die aber durchaus ausgeprägte Effekte auf die Politik haben können. Allerdings ist die Diskussion zum Verhältnis von Wissenschaft und Politik wohl eine deutlich weiter führende. Auch stellt sich die Frage, ob Evaluation tatsächlich zur Wissenschaft oder eher zur Politik zu rechnen ist.

4 Unterstützende und hemmende Strukturen und Prozesse

Die effektive Nutzung der Evaluationsfunktion wird zu einem gewissen (vermutlich nicht unerheblichen) Teil durch die Verankerung in entsprechenden Strukturen beeinflusst. Zugleich spielen die Prozesse eine wesentliche Rolle. Schritte von der Auftragsformulierung, über die Ausführung der Evaluation bis zur Ergebnisdarstellung und -präsentation können die Nutzung der Evaluation beeinflussen. Strukturen und Prozesse bedingen sich gegenseitig. Folgend werden einige Aspekte skizziert.

4.1 Zu den Strukturen

In der Schweiz beispielsweise übt die Geschäftsprüfungskommission (GPK) die Oberaufsicht über die Geschäftsführung von Bundesrat, Bundesverwaltung und die Bundesgerichte aus. Die zentralen Kriterien in deren Arbeit sind die Rechtmässigkeit, Zweckmässigkeit und Wirksamkeit. Die Methodik ist vielfältig, unter anderem werden hier ‚Berichte mit evaluativem Charakter‘ erstellt. Das Parlament anerkennt Evaluationen somit grundsätzlich als wichtiges Instrument der Erfolgskontrolle staatlichen Handelns. Es dient der Kontrolle und Überprüfung der Wirksamkeit, des Vollzugs, der Rechenschaftslegung sowie der Förderung der Transparenz staatlichen Handelns. Schwachstellen einzelner Massnahmen werden aufgedeckt und tragen zur Optimierung bei. Seit Anfang der 1990er Jahren besteht die Parlamentarische Verwaltungskontrolle (PVK). Sie ist das eigentliche Evaluationsorgan des Bundesparlaments (Janett 2004: 139). Die PVK ist Teil der Parlamentsdienste und ist dem Sekretariat der GPK unterstellt. Sie erhält ihre Mandate in der Regel von der GPK. Sie ist somit in der wissenschaftlichen Durchführung der Projekte unabhängig, aber nicht in der Auswahl von Massnahmen der Bundespolitik, um sie aus eigener Initiative zu evaluieren (Janett 2004: 140). Die PVK bereitet ihre Ergebnisse in Berichten zu Händen der Rezipienten, namentlich der parlamentarischen Kommissionen, auf. Die Kommissionen nehmen Stellung dazu und erarbeiten gestützt auf die Evaluationen Empfehlungen zuhanden der Verwaltung oder des Bundesrats oder erarbeiten parlamentarische Vorstösse. Transparenz ist der GPK sehr wichtig, deshalb werden in der Regel die Berichte öffentlich publiziert. Die Nutzung und entsprechend der Nutzen sind für die Geschäftsprüfungskommissionen sehr gross. Sie schätzen die Arbeit der PVK positiv ein und sehen darin einen wichtigen Beitrag zu Effektivitätssteigerung, zur Systematisierung und zur Professionalisierung der parlamentarischen Oberaufsicht. Die Mittel sind begrenzt, so dass nicht alle potentiellen Untersuchungsgegenstände tatsächlich überprüft werden können.

Für die deutsche Bundesverwaltung ist in der Bundeshaushaltsordnung festgelegt, dass die Verwaltung für alle finanzwirksamen Massnahmen Wirtschaftlichkeitsuntersuchungen durchzuführen hat. Damit ist die Evaluation als Form der Erfolgskontrolle eine Regel der Verwaltung. Der Fokus der Tätigkeit des Bundesrechnungshofes liegt darin zu prüfen, ob die Verwaltung den Erfolg der einzelnen Massnahmen sachgerecht kontrolliert und die notwendigen Schlussfolgerungen aus der Evaluation zieht. Evaluationen werden dann zum Gegenstand von Prüfungen des Bundesrechnungshofes, wenn sie methodisch mangelhaft sind, wenn sie sehr stark auf die Interessen des Massnahmeträgers zugeschnitten sind, nicht aus-

reichend objektiv sind, unvollständig sind im Sinne des Paragraphen 7 BHO, was heisst, dass sie eine der dort geforderten Teiluntersuchung, namentlich Zielerreichungskontrolle, Wirtschaftlichkeitskontrolle und Wirkungskontrolle ohne plausible Begründung nicht vorgenommen haben. Der Bundesrechnungshof hat die Aufgabe, den Verwaltungen auf Grundlage der gewonnenen Prüfungserkenntnisse Empfehlungen für sachgerechtes Evaluierungsverhalten zu geben. Er entwickelt bei Bedarf Arbeitshilfen und macht Vorschläge für Änderungen bestehender Verwaltungsvorschriften oder Gesetze, wenn sich diese nicht bewährt haben. Im Unterschied zur Schweiz fehlen empirische Studien dazu, in welchem Umfang Evaluationen mit welcher professionellen Qualität tatsächlich umgesetzt werden und zu welchen Nutzungen es dabei kommt.

4.2 Zu den Prozessen

Betrachtet man die Prozesse der Evaluation und Nutzung innerhalb dieser Strukturen, zeigt sich, dass tatsächliche oder vermeintliche politische oder auch andere Sachzwänge problematisch sein können. Beispielsweise können datenschutzrechtliche Gründe Evaluationen behindern oder auch als Vorwand vorgeschoben werden. Blockieren können auch in der Planungsphase nicht festgelegte mess- und kontrollierbare Zielindikatoren. Dieser Mangel kann schliesslich dazu führen, dass die zuständigen Stellen in der Verwaltung unter Umständen Evaluationen gar nicht in Angriff nehmen.

Zu Österreich wird konstatiert, dass für Vorgehensweisen, wie Evaluierungen stattfinden, wie sie initiiert und wie sie dann abgewickelt werden, im politischen Prozess bislang keine Qualitätssicherung und keine einheitlichen Zugangsweisen bestehen. Die Gründe hierfür sind vielfältig, zum grossen Teil ist dies aber bedingt durch ein schwach ausgeprägtes Bewusstsein für die Notwendigkeit und den potentiellen Nutzen von Evaluationen.

Aus Deutschland wird berichtet, dass es viele Hinweise auf Mängel gibt, die es zu beheben gilt. Genannt wurden:

- a. Evaluierungen werden zu wenig in verwaltungsspezifische Entscheidungsprozesse einbezogen. Zum Beispiel wurde festgestellt, dass Evaluationsergebnisse im Haushaltsaufstellungsverfahren kaum eine Rolle spielen. Die Vergabe von Mitteln wird in der Regel nicht an den Erfolg einer Massnahme geknüpft.
- b. Evaluationen werden zum Teil rein zum Zwecke der legitimatorischen Funktion durchgeführt.⁴
- c. Die Bereitstellung von personellen Ressourcen für Evaluierungen geniesst teilweise nur eine nachrangige Priorität.
- d. Die Nachhaltigkeit bei grossen und komplexen Fördermassnahmen wird nicht ausreichend evaluiert. Bislang werden Massnahmen ex-post direkt nach Abschluss evaluiert, aber nicht zeitlich verzögert.
- e. Selbstevaluierungen sind die Regel, Fremdevaluierungen eher die Ausnahme.
- f. Es fehlt der Verwaltung an einer methodischen Orientierung sowie systematisch aufgebauter Evaluationskompetenz.

⁴ Was aber nicht heissen sollte, dass eine legitimatorische Funktion grundsätzlich nicht legitim sei.

Im Prozess der Nutzung von Evaluationen zeigt die Untersuchung von Balthasar (2007: 433), dass je mehr Distanz die Evaluatorinnen und Evaluatoren zu den Nutzenden haben, umso weniger die Studie dann genutzt wird. Je mehr die Massnahmeverantwortlichen einbezogen werden, umso mehr passiert nachher mit den Ergebnissen der Studie etwas. Amtsinterne Durchführungen von Evaluationen sichern den aktiven Einbezug der Massnahmeverantwortlichen als eine der wichtigsten Gruppen der Nutzenden. Konkret zeigen die Studienergebnisse, dass bei amtsinterner Durchführung sowohl instrumentelle als auch prozessbezogene Verwendungen höher sind.

Die Politik hat zunehmendes Interesse gerade an formativen Evaluierungen. Erfahrungen bei der Gesetzgebung in Österreich zeigen, dass wo Evaluierungen so angelegt wurden, dass die Evaluatorinnen und Evaluatoren während des konkreten politischen Gesetzgebungsprozesses eingebunden waren, die Gesetzestexte schliesslich in weiten Passagen wörtlich mit den Formulierungen der Evaluationsberichte übereinstimmten.

5 Fazit

In den Verwaltungen erkennt man vielerorts die Notwendigkeit an, die Evaluation als Instrument einer modernen Regierung zu stärken. Was hingegen aktuell die Intensität und Art der Nutzung von Evaluationen angeht, können teilweise grosse Unterschiede in den verschiedenen Verwaltungs- oder Politikbereichen und in den drei Ländern konstatiert werden. Die Bilanz ist somit einerseits ernüchternd, andererseits gibt sie Anlass zur Hoffnung (vgl. auch Balthasar 2007; DEZA 2006).

- Viele, aber nicht alle involvierten und betroffenen Personen kennen und anerkennen Evaluationen als wichtiges Instrument in Politik und Verwaltung. Bekanntheit und Anerkennung müssen gefördert werden.
- Eine breit geteilte Evaluationskultur, inklusive einer Verständigung über Begriffe sowie über Möglichkeiten, Ziele und Grenzen von Evaluation, fördert die Nutzung. Das Wissen, dass Evaluationen unterschiedliche Zwecke verfolgen und somit unterschiedliche Ausformungen haben können, muss gefördert werden.
- Die angemessene strukturelle Einbettung der Evaluationsfunktion sowie die Beachtung und Förderung günstiger Prozesse sind wichtige Grundlagen, die sich positiv auf die Nutzung auswirken (siehe auch Widmer/Leeuw in diesem Band).
- Die Kenntnisnahme und das Beherzigen der entsprechenden Evaluationsstandards unterstützten die Evaluationsprozesse und die Nutzung. Der Bekanntmachung der Bedeutung und die Förderung des Verständnisses der Standards ist hohe Priorität einzuräumen.
- Schon das Bewusstsein, dass etwas evaluiert wird resp. evaluiert werden könnte hat eine zumindest indirekte Wirkung auf den Evaluationsgegenstand und die involvierten Personen. Partizipative Prozesse sind förderungswürdig.

„Uses of evaluation results and processes does not come automatically. It must be planned and cultivated throughout the evaluation process“ (Horton et al. 2003: 119).

Sektorale und nationale Trends

Divergente Entwicklungspfade der Evaluation

Beobachtungen, Thesen und Perspektiven einer feldabhängigen und ungleichzeitigen Professionalisierung in Deutschland, Österreich und der Schweiz

1 Einleitung

Den Kern dieses Sammelbandes markieren die zehn Themenfelder, in denen sowohl ein Länderbeitrag je Land – Deutschland, Österreich, Schweiz – als auch ein ländervergleichender Beitrag vorliegen. In diesem abschliessenden Aufsatz bietet es sich an zu fragen, welche Aufschlüsse über Entwicklung und aktuellen Stand der Evaluation quer zu den Themenfeldern und den drei Ländern gezogen werden können. Selbstverständlich kann dies keinen umfassenden Vergleich darstellen, der die Reichhaltigkeit der Einzelbeiträge hinreichend aufgreifen würde. Vielmehr soll dieser Beitrag in selektiver Weise einige zentrale Auffälligkeiten thematisieren und in komparativer Perspektive vertiefen.

Da die Ländervergleiche innerhalb der Themenfelder geleistet sind, drängt es sich auf zu versuchen, einen zusammenfassenden Themenvergleich vorzunehmen. Auf den ersten Blick scheint dies angemessen, handelt es sich doch bei den drei benachbarten Ländern in vielerlei Hinsicht um ähnliche Gemeinwesen: Sie sind etablierte Demokratien, haben – bei gewissen Unterschieden – einen föderalen Staatsaufbau, stehen in zahlreichen Zusammenhängen in regem Austausch – sei dies in Politik, Wirtschaft, Technologie oder Kultur –; die dort lebenden Menschen können sich zu grossen Teilen über die deutsche Sprache miteinander verständigen.

Wie die Kontextbeschreibungen der Länderbeiträge zeigen, gibt es Politikfelder mit ausgeprägten länderübergreifenden Gemeinsamkeiten in Bezug auf Voraussetzungen, verfolgte Ziele oder Steuerungsmechanismen. Es bestehen aber auch bedeutsame länderspezifische Unterschiede. Als ein Beispiel genannt sei die sehr unterschiedliche Ausgangssituation für die Arbeitsmarktpolitik in den drei Ländern: Deutschland mit einer langjährig hohen, die Schweiz mit einer über lange Zeit sehr niedrigen und auch heute noch tiefen Arbeitslosenquote, sowie Österreich, das diesbezüglich eine Mittelposition einnimmt. Deutschland und Österreich sind beide Mitglieder der Europäischen Union, was für viele Bereiche der Politikgestaltung eine Homogenisierung mit sich bringt, und was auch für Evaluationen vielfach Impulse zur Etablierung und Angleichung gibt – hierzu illustrativ sind die Struktur- oder die Agrarpolitik (vgl. den Beitrag von Tissen in diesem Band). Die Schweiz verfügt in vielen Politikfeldern über bilaterale Verträge mit der Europäischen Union, was einerseits zu Adaptionen, andererseits zu weiterhin sehr eigenständigen Wegen führt, sowohl in der

Politikgestaltung als auch bei der Evaluation in diesen Politikfeldern. Die Einflüsse der Europäischen Union als Impulsgeberin der Evaluation beschränken sich hier auf verhältnismässig kleine Nischen in Politikfeldern, in denen die Schweiz an Projekten und Programmen der Europäischen Union partizipiert (etwa bei der grenzüberschreitenden Regionalpolitik oder bei den EU-Rahmenprogrammen in der Forschungsförderung; vgl. die Beiträge von Stefan Kuhlmann, Hummelbrunner/Maier, Thierstein und Gerheuser in diesem Band).

Ein zweiter Blick zeigt, dass ein länderspezifischer Vergleich des Standes der Evaluation für manche Themenfelder lediglich bedingt angebracht ist, da es auch für die Evaluation so etwas wie „internationalisierte Themenfelder“ gibt. Ein besonders herausragendes Beispiel hierfür ist die Entwicklungszusammenarbeit, für die auf beiden Ebenen – sowohl der Gestaltung der Interventionsinstrumente als auch bei der Ausgestaltung der Evaluationsfunktion – ein intensiver internationaler Austausch besteht. Deutschland, Österreich und die Schweiz als Staaten mit hoch entwickelten Volkswirtschaften der nördlichen Hemisphäre gehören traditionell zu den Geberländern, die den Empfängerländern gegenüber stehen, und zunächst Entwicklungshilfe, dann nach einem veränderten Paradigma partnerschaftliche Entwicklungszusammenarbeit betreiben und inzwischen intensiv untereinander abstimmen. Im internationalen Rahmen haben sie sich über Prinzipien zur Evaluation von Entwicklungszusammenarbeit verständigt, die in zunehmendem Masse den durch die nationalen Ministerien und Agenturen der Entwicklungszusammenarbeit beauftragten Evaluationsaktivitäten zugrunde gelegt werden. Deren Grundlagen, Zugänge, Konzepte und Methoden der Evaluation bis hin zu den Kommunikationskanälen für Evaluationsberichte haben sich zumindest teilweise von nationalen Evaluationsdiskursen entkoppelt. Wir können hier von einem „internationalisierten Evaluationsfeld“ sprechen.

Ähnlich verhält es sich mit dem Themenfeld Forschung und Technologie, das stark durch einen internationalen Diskurs und durch Beiträge international tätiger Evaluatorinnen¹ geprägt ist. Dies ist auch ablesbar an den Literaturnachweisen der Länderbeiträge in diesem Themenfeld. Forschungsförderung und der Diskurs über deren Steuerung sind stark internationalisiert. Die internationale Mobilität von Forschenden ist im Vergleich zu anderen Berufsgruppen in anderen Themenfeldern wesentlich höher. Die herausragenden Publikationsorgane von Forschenden sowie wichtige Kongresse sind durch ein weltweites Publikum charakterisiert und durch die englische Sprache geprägt. So hat sich die Evaluation auch in diesem Bereich – zumindest teilweise – vom nationalen Diskurs entkoppelt. Einen Sonderfall stellt hier die österreichische Situation dar, wo im Bereich der Forschung und Technologie eine überaus aktive Evaluationsplattform mit internationaler Ausstrahlung besteht (siehe Beitrag von Zinöcker/Dinges in diesem Band; Zinöcker et al. 2007).

Eine andere Betrachtungsweise fragt danach, ob es (abweichend vom nationalen Haupttrend) herausragende nationale Evaluationsvorreiter oder – umgekehrt – Felder mit besonderem Entwicklungsbedarf für die Evaluation gibt.

¹ In diesem Beitrag wird ausschliesslich die weibliche Form zur Bezeichnung der evaluierenden Person genutzt. Andere Akteure erscheinen dagegen in männlicher Form. In beiden Fällen sind Personen beider Geschlechter gemeint.

2 Themen- und länderspezifische Formen der Professionalisierung

Geht man – wie in vielen Vergleichsbeiträgen herausgearbeitet – davon aus, dass die Schweiz *insgesamt* seit den 1990er Jahren im Vergleich der drei Länder eine Vorreiterposition in der Evaluation einnimmt (vgl. dazu auch die Beiträge von Balthasar und von Spiel/Bergsmann in diesem Band), so finden sich *speziell* im Bereich der Evaluation von Forschung und Technologie erstaunlich wenig nachgewiesene Aktivitäten (siehe Beitrag von Farago in diesem Band)². Erklärt werden kann dies durch die Situation, dass die öffentliche gegenüber der privaten Finanzierung naturwissenschaftlicher, medizinischer oder technischer Grundlagenforschung in der Schweiz – speziell im Vergleich zu Deutschland – eine untergeordnete Rolle spielt. Während Deutschland eine grosse Zahl vollständig oder teilweise öffentlich geförderter Forschungseinrichtungen aufweist, die unter erheblichen Legitimationserfordernissen stehen und dabei auch in den letzten Jahren immer wieder unter „Evaluationsdruck“ gerieten, stellt sich die Situation in der Schweiz ganz anders dar. Die Notwendigkeit eines breiten Angebots an Evaluation öffentlich geförderter Forschung ist hier weniger ausgeprägt – Evaluation ist auf diesem Gebiet in der Schweiz daher weniger entfaltet. Dies steht in auffälligem Kontrast zu der bereits angeführten Situation in einem ähnlich kleinen Land wie Österreich.

Eine Sonderrolle kommt in Österreich auch der Nachhaltigkeits-Evaluation zu, die im Bereich der Energie- und Umweltpolitik vielfältige Aktivitäten ebenfalls mit internationaler Ausstrahlung entfaltet hat (siehe Beitrag von Martinuzzi/Leschnik in diesem Band). Im Gegensatz zum Bereich der Forschung und Technologie sind diese Evaluationsaktivitäten aber getragen durch wenige, sehr engagierte Personen. Die eigene Nachhaltigkeit dieses Evaluationszweigs kann (noch) nicht als gesichert gelten. Denn mit Österreich handelt es sich um ein Land, dessen politisch-strukturelle Rahmenbedingungen für Evaluation – ähnlich wie in Deutschland – (noch) nicht stabilisiert sind.

Wenn wir hier einschätzen, dass es in Deutschland und Österreich – im Unterschied zur Schweiz – in der Mehrzahl der Themenfelder keine stark verbreitete Evaluation auf hohem Reflexionsstand und mit ausgeprägter Institutionalisierung gibt, so behaupten wir damit nicht, dass dort weniger Evaluationen (z. B. in Relation zu den öffentlichen Haushalten) stattfinden. Teilweise finden sogar sehr viele oder – unter Einsatz beachtlicher Ressourcen – sehr grosse Evaluationen statt. Dies gilt für viele Bereiche der EU-induzierten Evaluationen, zum Beispiel im Bereich der Strukturpolitik, der Agrarpolitik oder der Arbeitsmarktpolitik. Nur ein Teil der dort tätigen Evaluatoren beteiligt sich am theoretischen und methodischen Evaluationsdiskurs. Evaluationen sind vielfach stark geprägt durch einen themenfeldspezifischen Austausch, seien dies politikfeldspezifische Diskussionen, Diskurse spezifischer wissenschaftlicher Disziplinen (z. B. der Mikroökonomie, der Arbeitspsychologie oder verschiedener gesundheitswissenschaftlicher und medizinischer Teildisziplinen), oder durch einzelne im jeweiligen Politikfeld etablierte Professionen (z. B. Schulpädagogen oder Hochschullehrer).

² Auf ein ähnliches Phänomen stossen wir auch bei der schweizerischen Berufsbildung, in der ebenfalls relativ spärliche Evaluationsaktivitäten zu beobachten sind – jedenfalls, wenn sie mit der Allgemeinbildung verglichen wird.

Der Zuschnitt von Evaluationsfragestellungen, die Entwicklung von Wirkungsmodellen, die Festlegung von Bewertungskriterien und weitere strategisch wichtige Schritte in der Planung und Durchführung von Evaluationen folgen (falls es überhaupt dazu kommt) einem im jeweiligen Themenfeld engen, als zulässig geltenden Raster, über das sich die themenfeldinterne Gemeinschaft (oft stillschweigend) verständigt hat. Dies gilt gemäss weit verbreiteter Meinung für die Schulen, wo möglichst ausgebildete Lehrer oder Schulleiter evaluieren sollen, für Universitäten, wo dies möglichst Professoren tun sollen, genauso wie für den Arbeitsmarkt, wo Ökonomen weithin als prädestiniert gelten oder für den Gesundheitsbereich, wo man oft Mediziner oder Epidemiologen oder andere Spezialisten für die bestgeeigneten hält, Evaluationen von hoher Güte durchzuführen. Diese disziplinär und/oder fachprofessionell fokussierten Diskurse erschweren oder verunmöglichen im Extremfall einen Austausch zu den spezifischen evaluationsprofessionellen Anforderungen. Die Zumutung der Evaluation wird zwar von den jeweiligen Fachkreisen (zum Teil pflichtgemäss) aufgenommen, es wird jedoch von der Grundvermutung ausgegangen, dass eine in erster Linie themenspezifische Expertise für die Konzeption und Durchführung von Evaluationen notwendig und hinreichend ist. Je robuster das Selbstverständnis und je stärker die Position der entsprechenden Disziplinen und Professionen sind, um so erfolgreicher werden Evaluationspraktiken mit Distanz zur Evaluationsfachlichkeit entwickelt, was sich auch in einem diffusen oder zumindest heterogenen Verständnis von Evaluation ausdrücken kann (vgl. für die deutschen Hochschulen Schmidt in diesem Band; weniger kritisch Kohler in diesem Band für die österreichischen Hochschulen, der jedoch ebenfalls erheblichen Entwicklungsbedarf sieht). Davon betroffen sind nicht nur Fragen im Zusammenhang mit Evaluationsdesign und -verfahren sowie zum Leistungsprofil der Evaluation und den angemessenen Erwartungen, sondern auch zu den jeweils beizuziehenden Bewertungsgrundlagen.

Auf der Ebene der Evaluation kann sich die intern differenzierte Struktur der betreffenden Professionen und Disziplinen replizieren, wie dies etwa in den Bereichen Hochschule, Forschung oder Gesundheit deutlich wird. Die Evaluation und deren Gestaltung drohen in einem solchen Fall zum Spielball der internen Machtauseinandersetzungen zu werden, in denen die jeweils dominante Fraktion auch die Gestaltungsmacht auf der Ebene der Evaluation an sich zieht. In einem derart strukturierten Machtspiel hat die Evaluationsfachlichkeit schlechte Karten und muss sich darauf beschränken, als Handlanger der jeweils dominierenden Fraktion zuzudienen. Evaluationsspezifische Anforderungen (wie sie etwa in Evaluationsstandards formuliert sind) erlangen dabei oft nur eine marginale Bedeutung, besonders wenn sie Positionen der dominanten Fraktionen und im jeweiligen Feld breit etablierte Axiome und Praktiken in Frage stellen. Um derartige Auseinandersetzungen vollends zu unterbinden, werden nach Möglichkeit Mitglieder der entsprechenden Profession oder Disziplin mit der Durchführung der Evaluationen betraut. Die auf diesem Weg zu „Evaluatorinnen“ Berufenen verfügen über eine Identität, die in erster Linie auf der Loyalität zur entsprechenden Profession oder Disziplin gründet und auf deren Werten, Interessen und Wissensbeständen aufbaut. Die Fähigkeit zur Evaluation wird diesen Personen aufgrund der Leistungen im jeweiligen Feld zugesprochen. Eine Verpflichtung auf evaluationsspezifische Kenntnisse, Fähigkeiten und Fertigkeiten ist hingegen nachrangig.

Blicken Evaluatorinnen von ausserhalb des jeweiligen Themenfeldes auf die Kriterien-sets, die dort zugrunde gelegt werden, stellen sie oft fest, dass bestimmte Bewertungskrite-

rien von vornherein gar nicht herangezogen werden. Dies unterscheidet sich je nach Themenfeld deutlich: So werden beispielsweise in einigen Themenfeldern (wie etwa in der sozialen Arbeit oder im Bildungswesen) häufig Fragen der Kosten oder der Kostengerechtigkeit der Massnahmen kaum oder sehr nachrangig berücksichtigt, während die Güte des Prozesses oder vielleicht auch die erbrachten Leistungen für ausgewählte Zielgruppen beinahe exklusiv betrachtet werden. In anderen Fällen (wie etwa verbreitet in der Arbeitsmarktpolitik) werden nahezu ausschliesslich Kostengesichtspunkte oder Ersparnisse für die öffentlichen Haushalte zur Bewertung herangezogen, hingegen Fragen der Nachhaltigkeit oder der Verteilungsgerechtigkeit gar nicht gestellt. Zuweilen ergeben sich auch innerhalb der Themenfelder wesentliche Divergenzen. Wenn zum Beispiel primär in der Jugendarbeit Tätige mit Evaluationen betraut werden, tendieren diese dazu, Werte und Interessen älterer Menschen in ihre Evaluationen von vornherein gar nicht einzubeziehen. Evaluationen an Universitäten gehen häufig davon aus, dass allein Kriterien der akademischen Güte von Lehre und vor allem Forschung zur Bewertung herangezogen werden dürfen. Bewertungsmassstäbe hingegen, welche andere Stakeholder (z. B. Steuerzahler, künftige Arbeitgeber von Universitätsabsolventen) artikulieren, werden als nicht zulässig gänzlich ignoriert oder als a priori nachrangig behandelt.

Ein Diskurs darüber, ob gerade die Tatsache gute Evaluation ausmacht, dass die Kriterien wesentlich „aus dem Themenfeld geschöpft werden“, oder dass das Themenfeld sich hinsichtlich der anzulegenden Bewertungsdimensionen gegenüber den Interessen der Gesellschaft respektive der Öffentlichkeit öffnet, und auch fremde Kriterien (z. B. die anderer Themenfelder) zulässt, werden tendenziell nicht geführt (vgl. die kritische Anmerkung zur Arbeitsmarktpolitik von Helmstädter in diesem Band). Dies kann zwar die Akzeptanz und Glaubwürdigkeit der Evaluation im entsprechenden Themenfeld stärken, trägt jedoch kaum zu einer Akzeptanz in einem weiteren Kreis von Stakeholdern oder der Öffentlichkeit bei. Zudem wird dadurch ein themenfeldübergreifender Austausch erschwert oder verunmöglicht und das damit verbundene Lernpotential nicht ausgeschöpft (vgl. die detaillierten Hinweise dazu im Fazit des Beitrages von Balzer in diesem Band).

Aufgrund der zuweilen starken internen Differenzierung der feldspezifischen Fachlichkeit erhält die Auswahl der Evaluatorinnen eine ausschlaggebende Bedeutung für die interne Glaubwürdigkeit. Die jeweilige personelle Bestimmung bevorzugt oft eine spezifische fachliche Schule (beispielsweise eine spezifische Subdisziplin) mit den ihr eigenen Bewertungsmassstäben, was die feldweite (bzw. disziplinweite) Glaubwürdigkeit und Akzeptanz gefährdet. Erschwerend kommt hinzu, dass dieser Vorgang selten in transparenter und expliziter Weise abgewickelt wird und dazu auch nur selten eine nachträgliche Explikation erfolgt. Dadurch kann auch die feldinterne Akzeptanz und Glaubwürdigkeit der konkreten Evaluation wie auch der Evaluation im Allgemeinen in Frage gestellt werden. Dies zeigt sich etwa im Gesundheitsbereich bei den Auseinandersetzungen in der Medizin zwischen der Schulmedizin und der Komplementärmedizin oder in den Sozialwissenschaften zwischen den stärker an geistes- respektive naturwissenschaftlichen Paradigmen orientierten Schulen. Neben den aus unterschiedlichen Grundlagen abgeleiteten Bewertungskriterien trifft dies auch für die favorisierten Methoden und Verfahren zu, die vielfach alleine aufgrund der Selektion der Evaluatorin bereits determiniert sind. Beispiele hierfür finden sich etwa in der Arbeitsmarktpolitik mit der Ökonometrie, in der sozialen Arbeit mit induktiv-

qualitativen Ansätzen, in der Bildung mit der Leistungsmessung (wie PISA, TIMSS etc.), in der Medizin mit der Favorisierung experimenteller Untersuchungsanordnungen (sogenannte ‚randomized controlled trials‘, RCT) oder in der akademischen Wissenschaft mit Zitationsanalysen und verwandten Zugängen einer „Wissenschaftsevaluationswissenschaft“ (Widmer 2008a: 268).

In Analogie dazu werden häufig in der Gegenstandsebene etablierte Zugänge und Praktiken als Leitlinien auch zur Gestaltung der Evaluation herangezogen. Dabei geht vergessen, dass die Zielsetzungen der Evaluation nicht kongruent zu jenen des Evaluationsgegenstandes sein müssen – und häufig auch nicht sind. So werden für deliberative Programme deliberierende Evaluationen und für Entwicklungsprojekte emergente Evaluationsdesigns eingefordert; Projekte mit starker Gewichtung der Inklusion Beteiligter & Betroffener seien auch in partizipativer Weise zu evaluieren, während in medizinischen Settings ausschliesslich experimentelle Untersuchungsanordnungen akzeptabel erscheinen. Damit wird ausser Acht gelassen, dass eine Evaluation eine spezifische Funktionalität aufweist, die nicht per se mit jener des evaluierten Gegenstandes übereinstimmt.

Folgt man dieser Argumentation, ergibt sich folgendes Bild im Ländervergleich: Man stellt fest, dass es in der Schweiz vergleichsweise viele Evaluatorinnen gibt, die themenfeldübergreifend aktiv sind. Es gibt mehrere Evaluationsbüros und an Hochschulen tätige Evaluatorinnen, die quer zu den Themenfeldern tätig sind. Sie verstehen sich primär als professionelle *Evaluatorinnen* und höchstens sekundär als Themenexperten. Diese Differenz zeigt sich auch an den inneren Strukturen der Evaluationsfachgesellschaften. In der Schweizerischen Evaluationsgesellschaft (SEVAL) mit rund 400 Mitgliedern haben sich kaum dauerhaft themenspezifische Arbeitskreise oder Arbeitsgruppen herausgebildet.³ Im Unterschied dazu ist die Gesellschaft für Evaluation (DeGEval), die ihre Mitgliedschaft, zusammengesetzt aus rund 500 Einzel- und rund neunzig institutionellen Mitgliedern, aus Deutschland und Österreich rekrutiert, durch thematische Arbeitskreise und deren Diskurse geprägt. Einige dieser Arbeitsgruppen sind ausgesprochen stabil, stark profiliert und kontinuierlich handlungsfähig, zum Beispiel der Arbeitskreis Entwicklungszusammenarbeit oder der Arbeitskreis Strukturpolitik. Andere wie zum Beispiel die Arbeitskreise Schule oder Hochschule – beides ausgesprochen grosse Evaluationsfelder – sind weniger aktiv und verfügen über einen in Relation dazu sehr kleinen Kern von Personen mit einer kontinuierlichen Beteiligung. Die Idee ist gleichwohl, über eine Stabilisierung der themenfeldbezogenen Evaluatorinnen die Mitgliedschaft zu mobilisieren und auf diesem Weg eine Professionalisierung von Evaluation zu erreichen. Dies ist deshalb in Deutschland und Österreich schwierig, da es hier oft beide Länder übergreifende, fachspezifische Organisationen gibt, zum Beispiel im Gesundheitsbereich oder in der beruflichen Bildung, in denen (Begleit-)Forscher oder auf Konzeptentwicklung spezialisierte Fachleute organisiert sind, Tagungen durchführen, Publikationen herausgeben, die unter anderem auch Evaluationen oder evaluationsähnliche Studien durchführen, und die dort ihren fachspezifischen Bezug haben. Sie sind sehr stark auf dieses Themenfeld orientiert, verfügen über hohe Feldexpertise und werden von Auf-

³ Eine Ausnahme bildet hierbei die Arbeitsgruppe Entwicklungszusammenarbeit, die seit nunmehr gut fünf Jahren aktiv ist. Die Ausrichtung dieser Gruppe mit einem starken Fokus auf internationale Zusammenhänge ist wiederum illustrativ für den Fall eines internationalisierten Evaluationsfelds, wie wir ihn oben beschrieben haben.

traggebern (so weit sie nicht selbst dort Mitarbeiter sind) bevorzugt mit Studien und – so weit diese so genannt werden – Evaluationsaufträgen betraut.

Anhand der Grösse der beiden Fachgesellschaften lässt sich weiter erkennen, dass der Organisationsgrad der mit Evaluation befassten Personen in der Schweiz deutlich höher ist als in Deutschland und Österreich. Die DeGEval verfügt zwar absolut über eine grössere Mitgliedschaft als die SEVAL, relativiert man dies jedoch durch das Einzugsgebiet der beiden Gesellschaften wird deutlich, dass sich namentlich in Deutschland neben der DeGEval Evaluationsdiskurse in disziplinär- oder themenorientierten Kreisen etabliert haben, während in der Schweiz die SEVAL die Evaluationsdiskurse stärker prägt.

Vier weitere Faktoren haben unseres Erachtens dazu beigetragen, dass es in der Schweiz zu einer themenfeldübergreifenden Professionalisierung und Organisation von Evaluation kommen konnte, die in Deutschland und Österreich in diesem Masse (noch) nicht so stark stattgefunden hat:

- Eine in Relation zur Grösse des Landes oder zu den nationalen sozialwissenschaftlichen Forschungsaktivitäten insgesamt hohe Dichte an Forschung und Reflexion über Evaluation, beginnend in den 1990er Jahre mit den durch die Arbeitsgruppe Gesetzes-evaluation (AGEVAL) ausgelösten Studien (AGEVAL 1991), dem Nationalen Forschungsprogramm „Wirksamkeit staatlicher Massnahmen“ (NFP 27) mit zahlreichen Evaluationsprojekten, und weitere Studien, die sich mit Evaluation als Gegenstand befassen, wie etwa die Habilitationsschriften von Margrit Stamm (2003) oder von Andreas Balthasar (2007). Derartige Aktivitäten sind in Deutschland und Österreich vergleichsweise selten (siehe etwa die Dissertationen von Stockbauer 2000a, Balzer 2005 oder Schwarz 2006).
- Natürlich auch – wie bereits im Teil über die institutionelle Einbettung der Evaluationsfunktion dargestellt – die starke Verankerung der Evaluationsfunktion im politischen System der Schweiz, beginnend beim entsprechenden Verfassungsartikel und zahlreichen Evaluationsklauseln in vielen Rechtsgrundlagen, über die Kontrollorgane mit Evaluationsauftrag (Parlamentarische Verwaltungskontrolle, Eidgenössische Finanzkontrolle) und zahlreiche Evaluationsstellen in Einrichtungen der Bundesverwaltung (siehe den Beitrag von Mader in diesem Band). Auch dies findet keinerlei Entsprechung in Deutschland und Österreich (siehe die Beiträge von Konzendorf und von Pichler, sowie jenen von Widmer/Leeuw in diesem Band). So hat sich mit dem „Netzwerk Evaluation in der Bundesverwaltung“ auch eine Entwicklung in der Schweiz ergeben, dass es ressortübergreifend zu einer Professionalisierung der Evaluationsfunktion gekommen ist. Hiervon ist man in Deutschland weit entfernt, wo es bislang nicht möglich war, ein Dutzend Personen aus den Bundesbehörden zu identifizieren, die sich als dauerhaft an themenfeldübergreifender Evaluation Interessierte bezeichnen würden (vgl. Beiträge von Balthasar und von Spiel/Bergsmann in diesem Band).
- Weiter zeichnet sich die Schweiz im Vergleich zu den anderen beiden Ländern dadurch aus, dass sie über eine deutlich schlankere Verwaltung verfügt und dass eine fest etablierte Tradition besteht, fachliche Aufgaben häufig auf projektorientierter Ebene extern zu beauftragen und darauf zu verzichten, die entsprechenden Kompetenzen selbst aufzubauen. Letzteres hat nicht nur zur Folge, dass die öffentlichen Stellen – auch un-

ter Berücksichtigung der Kleinheit des Landes – deutlich weniger gross sind, sondern es erleichtert auch einen ressortübergreifenden Diskurs – auch in Evaluationsfragen. Die Evaluatorinnen sind zudem, da sie nicht institutionell Teil einer Verwaltungsstelle sind, eher in der Lage (und auch gewillt), in verschiedenen Themenfeldern zu arbeiten. Die Bürokratien in Deutschland und Österreich hingegen bauen allenfalls Evaluationskompetenz intern auf, die aber primär themenspezifisch ausgerichtet ist und eine ressortübergreifende Perspektive deutlich erschwert.

- Schliesslich erhöht sich durch die schlankere Verwaltung in der Schweiz, die für die Politikumsetzung oftmals externe Stellen beauftragt, der Bedarf an evaluativen Informationen zur Tätigkeit dieser ausserhalb der Verwaltungshierarchie stehenden Umsetzungsagenten (vgl. Widmer/Frey 2006). Diese Tradition des ‚outsourcing‘, die sich in der Schweiz deutlich vor der Einführung des englischen Begriffs etabliert hatte, erhöht nicht nur den Informationsbedarf der Verwaltung hinsichtlich der Aktivitäten der Durchführungsorganisationen, sondern steigert auch deren Bereitschaft, unabhängig durchgeführte Evaluationen zu veranlassen.

Letzteres kann nicht nur zur Erklärung der Differenzen zwischen Deutschland und Österreich einerseits und der Schweiz andererseits beitragen, sondern ist aus unserer Sicht mit ein Grund für die je nach Themenfeld unterschiedliche Stellung der Evaluation. Dies zeigt sich etwa in der Aussenpolitik, die sich mit Ausnahme der Entwicklungszusammenarbeit (siehe die Beiträge von Zintl, Mair, Zürcher und von Reade in diesem Band) weitgehend als evaluative Leerstelle präsentiert (vgl. dazu Rudolf 2007, Widmer 2008b). Ähnliches lässt sich auch für andere Politikfelder postulieren, wie etwa die Verteidigungspolitik, die sich traditionell ebenfalls durch einen hohen Anteil der Leistungserbringung innerhalb der staatlichen Strukturen auszeichnet.⁴ In der Verteidigungspolitik wird der Begriff der Evaluation weitgehend mit der Produktbewertung gleichgesetzt, die bei der Beschaffung von Rüstungsgütern bei externen Anbietern eingesetzt wird, was obige These stützt.

Die Aussenpolitik und die Verteidigungspolitik sind Beispiele für Themenfelder, die im vorliegenden Band nicht abgedeckt sind; nicht zuletzt weil sie bisher in den drei Ländern kaum durch rege Evaluationsaktivitäten in einem weiteren Sinne aufgefallen wären. Dem Aspekt der Evaluationsdichte wollen wir uns im folgenden Abschnitt zuwenden.

3 Dichte, Portionierung und Distribution von Evaluation

Die Allokation der Mittel zur Durchführung von Evaluationen unterscheidet sich in auffälliger Weise von Themenfeld zu Themenfeld und von Land zu Land. Es stellen sich dazu drei Fragen: Erstens ist zu fragen, wie sich die Themenfelder und Länder bezüglich der Allokation von Evaluationsmitteln unterscheiden. Zweitens ist zu klären, welche Gründe für diese Differenzen angeführt werden können. Und drittens ist in einer evaluativen Tradi-

⁴ Bei Zoll- oder Steuerbehörden ist die Situation ähnlich, während bei der Polizei vor allem bei präventiv ausgerichteten Programmen oder Projekten Evaluationen anzutreffen sind, in anderen Tätigkeitsfeldern (Sicherheitsdienst, Ermittlungen etc.) dagegen selten.

tion die normative Frage zu diskutieren, welche Form der Zuweisung von Mitteln zur Realisierung von Evaluationsprojekten angemessen ist. Bevor wir uns diesen drei Fragen zuwenden, werden wir einige generelle Aussagen zur Thematik einbringen, die uns im Anschluss als Grundlage zur Diskussion dienen sollen.

3.1 Kosten der Evaluation

Die Frage der Mittel, die zur Realisierung von Evaluationen aufgewendet werden können, sollen oder müssen, beschäftigen Evaluatorinnen und Auftraggeber gleichermaßen, wenn auch aus anderen Blickwinkeln. Der Auftraggeber einer Evaluation – so könnte man annehmen – zielt darauf ab, eine bestimmte „Leistung“ zu einem möglichst kleinen Preis einzukaufen. Die Evaluatorin wiederum – so die Unterstellung – hat als Anbieterin die Absicht, für diese „Leistung“ einen möglichst hohen Preis zu erzielen. Diese Beschreibungen sind jedoch nicht angemessen, weil sie auf zwei Annahmen beruhen, die – betrachtet man die Evaluationspraxis – offensichtlich unzutreffend sind:

- Den Bezugspunkt einer a priori fest definierten Leistung gibt es in der Evaluation nicht. Der Auftraggeber hat (teilweise) eine Vorstellung davon, welchen Nutzen er aus der Evaluation ziehen möchte oder für Dritte erwartet. Dies kann beispielsweise die Erfüllung eines zumeist diffusen Informationsbedürfnisses sein, wie es in einer (zumeist umfangreichen) Fragestellung zum Ausdruck kommt; es kann sich aber auch um den Nutzen handeln, einer formalen Verpflichtung zur Evaluation nachzukommen. In der Regel sind diese Nutzenerwartungen multidimensional und werden je nach Perspektive unterschiedlich gewichtet. Sie sind auch einer zeitlichen Dynamik unterworfen, können sich im Verlauf der Evaluation also verändern. Mit welcher Leistung dieser recht diffuse und volatile Nutzen erzielt werden kann, ist dem Auftraggeber in der Regel nicht klar.
- Die Evaluatorin wiederum hat nur eine mehr oder weniger beschränkte Einsicht in die Nutzenerwartungen des Auftraggebers. Sie konzipiert die Evaluation – im Idealfall – in einer Weise, dass die Evaluation dazu geeignet ist, die vermuteten Nutzenerwartungen zu erfüllen. Diese sind – wie dargestellt – nicht nur ein schwer erkennbares, sondern auch ein volatiles Ziel. Aus kommerzieller Sicht wird es die Evaluatorin darauf anlegen, mit einem möglichst geringen eigenen Aufwand dem Auftraggeber einen möglichst hohen Nutzen in Aussicht zu stellen; dies mit der Absicht, das Projekt im budgetierten Rahmen sicher abzuschliessen und darüber hinaus gehend wenn möglich einen Gewinn zu erzielen. Da sich der Aufwand der Evaluatorin weitgehend durch die Leistungsumschreibung definiert, besteht die Tendenz, den zu erwartenden Nutzen zu überzeichnen (um einen hohen Preis zu erzielen) und die Leistungen (und damit den Aufwand) möglichst gering zu halten.

Mit Fragen der Evaluationskosten hat sich auch die Literatur schon längere Zeit befasst (vgl. etwa Alkin/Solmon 1983; Bickner 1980; Widmer et al. 1996). Michael Scriven (1974) formulierte das Ideal einer „cost-free evaluation“, ein Begriff der *nicht* suggerieren möchte

„that one can get it all for nothing“ (Scriven 1983: 42). Stattdessen postuliert das Modell, dass eine Evaluation einen Nutzen stiften soll, der grösser ist als die dafür notwendigen Aufwendungen. Es zirkulieren auch zahlreiche Angaben, wie gross der Anteil des Budgets für Evaluationsaktivitäten am Gesamtbudget sein soll, wobei die Spannweite sehr gross ist. Analog lassen sich in der Praxis je nach Feld und Land sehr unterschiedliche Anteile eruieren, die von wenigen Promillen am Budget für den Evaluationsgegenstand reichen, bis zu Anteilen über zehn Prozent (vgl. den Beitrag von Balthasar in diesem Band). In Ausnahmefällen – etwa riskanten Pilotprojekten, die anschliessend vielfach repliziert werden sollen – übersteigen die Evaluationskosten gar die Mittel für den zu evaluierenden Gegenstand. Zur Bemessung eines Evaluationsbudgets besteht somit kaum Konsens in Literatur oder Praxis.

Empirische Studien zeigen, dass – unter bestimmten Bedingungen – eine geringe Ressourcenausstattung einer Evaluation nicht zu einer schlechten Evaluation führen muss (Widmer et al. 1996). Als Erfolgsbedingungen für sogenannte Kurzevaluationen⁵ werden genannt (nach Widmer et al. 1996: 257):

Evaluationsgegenstände,

- ohne allzu grosse politische, soziale, ethische und finanzielle Relevanz
- zu denen bereits ausreichend Vorkenntnisse vorliegen
- mit eher geringerer Komplexität
- deren Ausgestaltung und Umsetzung unter Berücksichtigung evaluationsspezifischer Erfordernisse erfolgt
- bei denen ein hoher Zeitdruck besteht

Evaluationen,

- mit hoher Akzeptanz
- mit präziser Planung, enger Projektbegleitung und einfacher Projektorganisation
- mit formativem, nicht mit summativem Ansatz
- die sich mit Prozessen und weniger mit Wirkungen befassen
- von denen keine Vertiefung und keine Innovationen erwartet werden
- mit ausserordentlich kompetenten und flexiblen Evaluatoreninnen

Unter der Voraussetzung, dass diese Bedingungen erfüllt sind, können Kurzevaluationen eine Alternative zu umfangreicheren Evaluationen darstellen, indem sie in kurzer Zeit und mit geringen Kosten zu Ergebnissen führen, die methodisch ausreichend abgesichert sind. Sind die angeführten Bedingungen jedoch nicht erfüllt, sind umfangreichere Evaluationsstudien vorzuziehen, weil die mit Kurzevaluationen verbundenen Risiken zu hoch werden.

⁵ Als Kurzevaluationen bezeichnet werden hier Evaluationen mit geringen zeitlichen und finanziellen Ressourcen (Widmer et al. 1996: 20). Die zeitlichen und finanziellen Ressourcen einer Evaluation kovariieren in der Regel stark miteinander (Widmer et al. 1996: 238-242).

3.2 Beschreibung

Auffällig ist bei der Durchsicht der Beiträge in diesem Band, dass je nach Land und Themenfeld sehr unterschiedliche Praktiken hinsichtlich der Verteilung der zur Verfügung stehenden Mittel zu beobachten sind. Das eine Extrem bilden jene Fälle, in denen eine starke Konzentration der Mittel auf eine oder einige wenige Evaluationsprojekte stattfindet. Exemplarisch sei hierzu auf die Arbeitsmarktpolitik in Deutschland verwiesen, wo ein grosser Teil der entsprechenden Mittel in die Hartz IV-Evaluation fliesst (vgl. den Beitrag von Wagner in diesem Band). Ähnliches lässt sich auch in Österreich und in der Schweiz beobachten, wo zentral geführte Monitoringsysteme eine erweiterte Auswertungen ermöglichende informationstechnologische Grundlage für übergreifende Evaluationen darstellen, die Programm- und Projektevaluationen konkurrenzieren (vgl. die Beiträge von Wroblewski und von Bieri/Ragni in diesem Band). Auf dem anderen Extrem finden sich jene Fälle, bei denen dem Grundprinzip gefolgt wird, dass jedes einzelne Projekt und jedes einzelne Programm einer Evaluation zu unterziehen ist, also eine maximale Distribution der Evaluationsressourcen angestrebt wird. Als Beispiel sei hier auf die Entwicklungszusammenarbeit (siehe die Beiträge von Zintl, Mair, Zürcher und von Reade in diesem Band) oder auf die 2002 bis 2007 einzeln durchgeführten Evaluationen von mehr als 200 EQUAL-Entwicklungspartnerschaften in Deutschland verwiesen (www.evaluation-equal.de), in der diese Praxis weite Verbreitung findet. Zwischen diesen beiden Extremen sind verschiedene Variationen zu erkennen, bei denen eine selektive Auswahl von Evaluationsgegenständen erfolgt und gleichzeitig nicht alle Mittel auf eine oder wenige Evaluationen konzentriert werden. Exemplarisch sei auf die programmatische schweizerische Praxis in der Energiepolitik und teilweise auch in der Agrarpolitik verwiesen (siehe die Beiträge von Bisang und von Jung in diesem Band). Während eine breite Verteilung der Mittel auf mehr oder weniger alle Einzelaktivitäten keine selektiven Entscheide benötigt, erfordern die anderen beiden Modelle (Konzentration auf einen oder wenige sowie auf eine Reihe von Gegenständen) Auswahlentscheide. Bei der extremen Konzentration auf einen oder wenige Gegenstände spielen dabei politische Faktoren eine zentrale Rolle bei der Auswahl, bei der Selektion mehrerer Gegenstände liegt zumindest teilweise eine fachlich motivierte Programmatik zugrunde.

Es ergibt sich aus den zwei Grössen ‚Gesamtevaluationsbudget‘ und ‚Anzahl der Evaluationsprojekte‘ eine erhebliche Spannweite in den zur Verfügung stehenden Ressourcen zur Durchführung der einzelnen Evaluation. Angesichts der Darstellungen in diesem Band ist davon auszugehen, dass die oben genannten Bedingungen bei Evaluationen mit geringer Ressourcenausstattung nicht immer erfüllt werden. Es ist weiter anzunehmen, dass die Ressourcensituation in Fällen, in denen eine Verteilung eines beschränkten Budgets auf viele oder alle potentiellen Evaluationsgegenstände erfolgt, zu Evaluationen mit oft sehr kleiner Ressourcenausstattung führt. In Kombination können diese Begleitumstände qualitativ schlechte Evaluationen befördern.

Aus diesen Gründen erscheint es vielfach angezeigt, eine selektive Vorgehensweise anzustreben. Soll die Selektion nicht nur auf politischen sondern auch auf fachlichen Kriterien

beruhen, muss der Auftraggeber⁶ über evaluationsfachliche Kompetenz verfügen, die es ihm erlaubt, eine begründete Auswahl der Evaluationsgegenstände vorzunehmen und für deren Evaluation eine angemessene Ausrichtung und ein angemessenes Budget vorzusehen. Beispielhaft verdeutlicht sei dies für die beiden Leistungsarten der summativen und der formativen Evaluation:

- Summative Evaluationen lohnen sich dann aus Sicht von Auftraggebern, wenn Richtungsentscheidungen über die Fortführung oder Einstellung, die Ausweitung oder die Reduktion eines (kostenintensiven) Programms durch Evaluationsergebnisse tatsächlich beeinflussbar sind: Ist zu erwarten, dass die Evaluation umstrittene Sachverhalte klärt, dass sie bislang nicht vorhandene Informationen erzeugt und dass diese, wenn in geeigneter Weise in Berichtsform aufbereitet und den Entscheidungsträgern kommuniziert, mit hoher Wahrscheinlichkeit den Entscheidungsprozess zumindest graduell beeinflussen? Auftraggebende sollten in der Lage sein, dies abzuschätzen und Mittel für Evaluationen so zuzuweisen, dass sie besonders für solche Entscheidungen eingesetzt werden, für die Spielräume bestehen und für die Bereitschaft bei den Entscheidern besteht, Evaluationsergebnisse zu nutzen. Sollte hingegen von vornherein feststehen, dass es lediglich darum geht, ein ohnehin fortzuführendes Programm zu legitimieren beziehungsweise ein Programm, zu dem es bereits einen Einstellungsentscheid gibt, zu delegitimieren, sollten Evaluationsmittel – wenn überhaupt – möglichst sparsam eingesetzt werden.
- Formative Evaluationen sind auf den ersten Blick weniger riskant und treffen auf weniger Widerstände bei den Beteiligten & Betroffenen als dies für summative Evaluationen der Fall ist. Zunächst scheint weniger auf dem Spiel zu stehen, denn es geht nicht um Fortführung oder Beendigung von Programmen. Doch auch bei formativen Evaluationen muss auf Seiten der Beteiligten – dies sind oft die Programmverantwortlichen, in Politik und Administration diejenigen, welche für die Steuerung der Programme verantwortlich sind beziehungsweise hierauf Einfluss nehmen – die Bereitschaft bestehen, nicht nur die Stärken der Programme zur Kenntnis zu nehmen sondern auch die Schwächen zu akzeptieren und Verbesserungsvorschläge zu diskutieren und – auch gegen Widerstände – durchzusetzen. Hierfür sind Engagement und eine gewisse Risikobereitschaft derjenigen erforderlich, welche auf der jeweils obersten Ebene für die Programminhalte und deren Umsetzung zuständig sind. Evaluationsmittel sollten in solche Felder gesteuert werden, wo sich solche Programmverantwortlichen mit einem ‚Commitment‘ für das Programm und seine Optimierung finden, die offen sind auch für schlechte Nachrichten aus dem Programm und die bereit sind, Veränderungen einzuleiten, so weit dies aus den Evaluationsergebnissen, Folgerungen und Bewertungen hervorgeht.

⁶ Wir gehen hier von einer Konzentration der Zuständigkeit zur Veranlassung von Evaluationen auf einen Auftraggeber aus. Häufig sind jedoch diese Zuständigkeiten auf mehrere Akteure verteilt. Solche Konstellationen erschweren die Festlegung einer umfassenden und systematischen Evaluationsprogrammatik, weil Evaluationsaktivitäten häufig durch Machtüberlegungen geprägt sind (Palumbo 1987: 12; Weiss 1987; Karlsson 1996; Widmer et al. 2001: 84-85; Widmer/Neuenschwander 2004: 404) und es eher zu negativer denn positiver Koordination (Scharpf 1993) zwischen den zuständigen Stellen kommt.

Derartige komplexe Stakeholderprozesse zu steuern verlangt auch und gerade auf Seiten der Auftraggeber Erfahrung und Kompetenz im Umgang mit Evaluationen – was auch ein Mindestmass an Institutionalisierung von Evaluationskompetenz in den entsprechenden politischen und administrativen Behörden erfordert.

Dieses Evaluationsvermögen („evaluation capacity“) ist – was ein zentrales Ergebnis der Länder- und Themenvergleiche darstellt – sehr unterschiedlich ausgeprägt. Nutzung und Nützlichkeit der Evaluation sind auf ausgeprägte Evaluationsfachlichkeit und angemessene Institutionalisierung der Evaluationsfunktion angewiesen (vgl. resümierend den Beitrag von Fabian in diesem Band). Um die Chancen einer gegenseitigen Befruchtung durch Intensivierung des evaluationsbezogenen Austausches zwischen den Ländern und Themenfeldern genauer abschätzen zu können, bedarf es als weiteren Schritt einer Erklärung für die vorgefundenen divergenten Ausprägungen.

3.3 Erklärung

Stellen wir uns die Frage, wie es zu den oben dargelegten Differenzen der Evaluationsdurchdringung kommt, bieten sich verschiedene Erklärungsansätze an.

Die Erklärungsfaktoren können sich in den verschiedenen Ländern oder für verschiedene Themenfelder oder Unterthemen gegenseitig verstärken oder auch abschwächen. Es fehlt bislang an genaueren Erkenntnissen oder Wirkungsmodellen, wie es letztlich zu stärker oder schwächer ausgeprägten Evaluationsaktivitäten kommt. Nachfolgend führen wir thesenartige Überlegungen illustriert mit konkreten Beispielen an:

In Themenfeldern, die einem hohen *Legitimationsdruck* ausgesetzt sind, wird mehr evaluiert, als in unbestrittenen Feldern.

Beispielhaft hierfür ist die Entwicklungszusammenarbeit, die in allen drei Ländern auf eine lange Tradition zurückblickt und die insgesamt einen „hohen Institutionalisierungs- und Professionalisierungsgrad“ erreicht hat (vgl. Reade in diesem Band). Obwohl der Anteil der Ausgaben für die Entwicklungszusammenarbeit am jeweiligen öffentlichen Haushalt relativ gering ist, wird in allen drei Ländern zyklisch immer wieder eine entwicklungspolitische Debatte über die Wirksamkeit der Entwicklungspolitik geführt und ihre genauere Überprüfung verlangt, womit Evaluation kontinuierlich gefordert ist. Da diese Debatte für die Mehrzahl der Geberländer ein dauerhafter Begleiter der Innenpolitik ist, hat sich gerade hier eine Internationalisierung der Evaluationspraxis und Evaluationskultur herausgebildet mit der Folge einer gewissen Stabilisierung für die Methodologie der Evaluation und evaluationsfachliche Standards.

Themenfelder, die dauerhaft unter hohem Legitimationsdruck stehen, da starke und dauerhafte *Wertkonflikte* bestehen, sind zum Beispiel der Bildungs- und Sozialbereich.

Im Schul- und Hochschulwesen führte bereits in den 1960er Jahren der so genannte Sputnik-Schock zu zahlreichen Reformaktivitäten, die in den drei Ländern – zum Teil zeitlich stark versetzt – Überlegungen auslösten, den Hochschulzugang zu erleichtern und die Hochschulausbildung insgesamt auszuweiten. Debatten, wie zunächst um die Verlängerung, später um die Verkürzung der Schulzeiten, zur Schulstruktur, zur Einführung von Ganztagsangeboten, zum verstärkten Angebot von Kindergarten- und Hortplätzen et

cetera, markieren bildungspolitische Konfliktfelder, die von starken Wertdebatten begleitet sind. Im Zusammenhang mit all diesen Reformüberlegungen und Entwicklungen wurden häufig Evaluationen durchgeführt, teils unter anderen Bezeichnungen, wie zum Beispiel wissenschaftliche Begleitung oder Qualitätsentwicklung (vgl. die Beiträge zur Evaluation im Bildungsbereich in diesem Band). Ähnliche Reformüberlegungen und Innovationsschübe bei grossen Unsicherheiten über die jeweils „richtigen“ Wege gibt es zum Beispiel sowohl in der Sozialhilfe, in der Jugendhilfe (vgl. die Beiträge von Haubrich, Loidl-Keil und von Drilling in diesem Band) als auch in der Arbeitsmarktpolitik, die in den letzten 25 Jahren in den beteiligten Ländern verschiedene Paradigmenwechsel erfahren hat (vgl. die Beiträge von Wagner, Wroblewski, Ragni/Bieri und von Helmstädter in diesem Band). Charakteristisch war beispielsweise in der Arbeitsmarktpolitik – ausgeprägt für Österreich in den 1990er Jahren, teils auch für Deutschland und die Schweiz –, dass Modellvorhaben durchgeführt und evaluiert wurden. Dies ist einerseits der Unsicherheit über die Wirkfähigkeit alternativer Interventionsansätze geschuldet, andererseits der Tatsache, dass den Alternativen sehr unterschiedliche, teils konträre Wertvorstellungen zugrunde liegen: marktwirtschaftliche Steuerung versus staatliche Regulierung, individuelle Selbstverantwortung versus kollektive Versorgung, Ansporn versus Sicherheit und so weiter. Analog lässt sich auch im Falle der Reformen institutioneller Politiken argumentieren (vgl. die Beiträge von Sager, Sabine Kuhlmann, Pleschberger und von Bussmann in diesem Band.)

Für alle genannten Bereiche ist also kennzeichnend, dass bei den beteiligten Akteuren – Politik, Verbänden, Fachexperten – unterschiedliche Leitprinzipien und Wertvorstellungen darüber bestehen, welches grundsätzlich angemessene Leitziele und Vorgehensweisen sind, um zu guten Lösungen für die betroffenen Zielgruppen und die Gesellschaft insgesamt zu kommen. Neben das Legitimationsprinzip treten als Auslöser für Evaluationsaktivitäten tiefgreifende Wertkonflikte.

Zusätzlich für Evaluationen spricht, wenn der *Kostendruck* im jeweiligen Themenfeld besonders gross ist. Veranschaulicht sei dies für einen Bereich der Sozialen Arbeit. Hier haben die kostenintensiven stationären Massnahmen der erzieherischen Hilfe immer wieder Evaluationen ausgelöst: Einerseits wurde versucht, die mittel- und längerfristigen Wirkungen von Heimerziehung zu überprüfen und in Relation zu den verschiedenen Kostenarten zu setzen, andererseits wurden die Alternativen zur Heimerziehung auf ihre Wirksamkeit untersucht und vergleichende Kosten-Nutzen-Abschätzungen durchgeführt (vgl. den Abschnitt 6 im Beitrag von Haubrich in diesem Band)⁷. Loidl-Keil (in diesem Band) postuliert, dass gerade im Zusammenhang mit der direkteren Bezahlung von sozialarbeiterischen Leistungen zunehmend Evaluation eingefordert wird, die grundsätzlich nicht-marktfähige Produkte sozialer Arbeit auf ihre Qualität hin untersucht. Drilling (in diesem Band) findet hierfür insofern einen Indikator, als er für die Schweiz überwiegend „wissensverwertende“ Evaluationen vorfindet, die in Bezug auf derartige unmittelbar praktisch-nützliche Fragen instrumentelles Wissen zur Verfügung stellen sollen.

⁷ Diese Konstellation einer „kostendruckgetriebenen“ Evaluation bildet eine Ausnahme zum allgemeinen Befund, dass in der Evaluation sozialer Arbeit generell Kostenüberlegungen eine nachrangige Bedeutung zukommt (siehe dazu oben, Abschnitt 2.).

Legitimationsbedarf, Wertkonflikte und Kostendruck reichen jedoch nicht in jedem Falle aus, um Evaluationsaktivitäten auszulösen. Ein Beispiel hierfür ist der Bereich der Rentenversicherung respektive Alters- und Hinterbliebenenversorgung. Auch diese Politiken stehen mit dem demographischen Wandel unter erheblichem Kostendruck, sie werden öffentlich diskutiert und es gibt zahlreiche Diskussionen über Alternativen in Form von Grundsicherungsrenten und ähnlicher Modellen. Ein zunehmend grosser Teil der Gesellschaft ist zudem von diesem Themenfeld betroffen. Es finden sich jedoch kaum Evaluationen, die sich mit diesem Themenfeld beschäftigen. Auch in benachbarten Feldern, etwa der Arbeitsmarktintegration von Menschen über 65 Jahre oder anderer Programme, die sich an spezifische Zielgruppen richten, sind Evaluationen eher selten, bilden zumindest kein eigenes Sub-Themenfeld. Es müssen also noch andere Faktoren wirken, um intensive Evaluationsaktivitäten in einem bestimmten Themenfeld auszulösen. Hierzu gehört mit Sicherheit eine Akteurskonstellation, die nicht zu konfliktreich aber auch nicht zu konfliktarm ist, um Evaluationen anzufordern, aber auch unerwartete Ergebnisse verarbeiten zu können.

Ein Themenfeld mit ausgeprägt volatiler Themenkonjunktur ist das der Energie- und Umweltpolitik. Gab es für die Energie- und Umweltpolitik in Österreich und der Schweiz durchgängig auf einem gewissen Niveau bestimmte Evaluationsaktivitäten (vgl. die Beiträge von Martinuzzi/Leschnik, Bisang sowie Hirschi in diesem Band), brach diese Entwicklung für Deutschland Ende der 1990er Jahre ab (vgl. den Beitrag von Duscha/Klemisch/Meyer in diesem Band). Erst Ende 2007 – mit der Intensivierung der Debatte um den Klimawandel und seiner Folgen für Natur und Gesellschaft – wurden auch wieder verstärkt Evaluationsaufträge durch die öffentliche Hand in Deutschland vergeben. In Deutschland fristet zudem der Verkehrsbereich evaluationsbezogen weiterhin ein stiefmütterliches Dasein, obwohl er in den Jahren 2007/2008 in der öffentlichen Debatte immer wieder grosse Aufmerksamkeit fand (Tempolimit, Feinstaub-Emissionen, Fahrverbote für Innenstädte, Innenstadt-Maut-Debatten u. v. m.). Für Evaluationen bleibt er in Deutschland – im Gegensatz zu Österreich und der Schweiz – weithin verschlossen, wobei uns Erklärungsansätze hierfür fehlen.

In verschiedenen rechtlichen oder anderweitig verpflichtenden Grundlagen bestehen Vorgaben, Projekte, Programme, Massnahmen et cetera ausnahmslos zu evaluieren. Diese „*rechtliche Evaluationspflicht*“ ist besonders ausgeprägt für Themenfelder, die in Deutschland und Österreich durch Mittel der Europäischen Gemeinschaft ko-finanziert werden. Wie dies den Beiträgen zu den entsprechenden Themenfeldern zu entnehmen ist, gilt dies besonders für die Agrarpolitik, die Arbeitsmarktpolitik und die Raumentwicklungspolitik, in abgeschwächten Mass auch für andere Politikfelder. Durch die Vorgaben der Europäischen Kommission werden hier auf nationaler Ebene Evaluationen ausgelöst, die von dort teilweise bis hinunter auf die Ebene der Programme beziehungsweise der einzelnen Projekte weitergegeben werden. Dies führt zu zahlreichen Evaluationen, zur Verausgabung von Evaluationsmitteln in einem bestimmten Umfang – der teilweise sogar durch Richtlinien der Europäischen Union vorgegeben ist. Ohne diesen Anreiz käme es, wie für manche Themenfelder festgestellt, in Österreich und Deutschland zu weit weniger Evaluationen (vgl. die Beiträge von Ortner und Schwab in diesem Band). Wie einige Beiträge (z. B. jener von Grajewski in diesem Band) thematisieren, folgt daraus teils eine pflichtgemässe Evaluierung. Evaluation droht zu einem Bestandteil des routinemässigen Verwaltungshandelns zu werden, wird von

vornherein so konzipiert – eventuell auch an Auftragnehmer vergeben –, dass möglichst wenig unerwartete Ergebnisse zutage gefördert werden, um die Kapazitäten der Verwaltungsabläufe nicht über Gebühr zu beanspruchen. Insofern zu der Vielzahl der so ausgelösten Evaluationen keine unabhängige Forschung beziehungsweise Meta-Evaluationen vorliegen, können keine Aussagen über deren Nutzen und Einwirkung gemacht werden; es steht die Vermutung im Raum, dass hier Optimierungsmöglichkeiten bestehen.

In anderen Fällen führen knappe Ressourcen entweder zu einer einseitigen Zentralisierung oder zu einer einseitigen Dezentralisierung von Evaluationen. Ein Beispiel hierfür bietet der Schulbereich in Österreich und Deutschland, während in der Schweiz bislang eine gegenläufige Entwicklung zu beobachten ist:

In Deutschland und Österreich dominierten lange, bis über das Jahr 2000 hinaus, dezentrale Ansätze der Schulevaluation, oft verbunden mit Konzepten der „Selbstevaluation“ oder der „Schulentwicklung“ (vgl. die Beiträge von Maag Merki und von Specht in diesem Band). Evaluation wurde als Bestandteil des didaktischen Handelns der Lehrpersonen in der Schule angelegt, verknüpft mit Konzepten der kollegialen Beratung und (unter der Bezeichnung „critical friends“) Ansätzen des Peer-Review. Schulleute steuerten die Selbstevaluationen. Charakteristisch dafür war der Ansatz des Instituts für Schulentwicklungsforschung in Dortmund (Kempfert/Rolff 2005; vgl. auch Nevo 1995; Gonon et al. 1998; Altrichter/Posch 1999; Strittmatter 1999; Büeler 2006).

Mit dem schlechten Abschneiden – besonders von Deutschland – in den internationalen Schulvergleichsstudien PISA und TIMSS verschoben sich die Gewichte und es entstand die Vision, auf der Makroebene ein Evaluationssystem zu implementieren, zu dem Elemente wie Schulinspektion, zentrale Abschlussprüfungen, nationale Bildungsstandards, Bildungsmonitoring und so weiter gehören. Auf die Spannungsfelder und „offenen Baustellen“, die sich aus einer weitgehend evaluationskompetenzfreien Schule und einem hochprofessionellen Messsystem auf der Makroebene ergeben, weist Maag Merki (in diesem Band) in ihrer Bilanzierung hin. Specht (in diesem Band) spricht davon, dass Evaluation im Schulbereich „meist noch punktuell und anlassbezogen“ erfolgt.

Hingegen verfolgt die Schweiz im Bildungsbereich die zentrale wie die dezentrale Evaluationsstrategie parallel: Bildungsstandards und Bildungsberichterstattung auf der nationalen Ebene, weiterer Ausbau der kantonalen Fachstellen für externe Schulevaluation mit ausgebildeten Schulevaluatorinnen sowie Evaluation als Regelbestandteil der Schulleiterfortbildung auf der anderen Seite (vgl. den Beitrag von Rhyn in diesem Band).

Das Entweder-Oder in Deutschland und Österreich auf der einen Seite, das Vorgehen im Gegenstromverfahren in der Schweiz auf der anderen Seite mag auch damit zusammenhängen, dass, wie Widmer und Leeuw (in diesem Band) aufzeigen, die Schweiz über kleinere Verwaltungsapparate verfügt und grössere Mittel für externe Beratungen, Evaluationen oder auch zur Weiterbildung einsetzen kann, als dies in Deutschland oder Österreich der Fall ist. Hier sind gerade im Bildungsbereich praktisch alle Mittel als Personalmittel (sei es für das Lehr- oder für das Verwaltungspersonal) gebunden, sodass äusserst knappe Bewegungsspielräume für Evaluation oder für Weiterbildung in Evaluation bestehen. Die Chancen, aktuell neben dem eingeleiteten Ausbau von Evaluationssystemen auf der Makroebene auch auf der dezentralen Ebene Evaluationsvermögen zu schaffen, sind von daher ausgesprochen gering. Die Wirkungsmöglichkeiten einer Evaluation, welche auf das Gegen-

stromprinzip angewiesen ist, sind damit grundsätzlich eingeschränkt. Wie weit der – ausserordentlich wichtige – Bereich Schule hier einen Ausnahmefall darstellt oder auch andere Themenfelder von dieser Spannung von Zentralität und Dezentralität stark betroffen sind, gilt es in weiteren Untersuchungen zu klären.

Der Hochschulbereich – ganz besonders in Deutschland – ist Anfang der 1990er Jahre durch die damals veröffentlichten Rankings in grossen deutschen Publikumszeitschriften unter erheblichen öffentlichen Legitimationsdruck geraten. Im Gefolge der zunehmenden administrativen Autonomie der Hochschulen (vgl. Widmer 2008a: 268-270) haben diese in allen drei Ländern – in auffälligem Gleichschritt – Verfahren der Qualitätssicherung und des Qualitätsmanagements eingeführt. Dabei sind für den Forschungsbereich wie für den Bereich Lehre auch Evaluationsverfahren zur Anwendung gekommen (vgl. die entsprechenden Beiträge im Themenfeld Forschung und Technologie sowie Bildung in diesem Band). Die Forschungsevaluation – bei der es nicht ausschliesslich um hochschulgetragene Forschung, sondern auch um (speziell in Deutschland) ebenfalls öffentlich finanzierte hochschulnahe Forschung sowie um Forschung ausserhalb der Hochschulen geht – hat sich international eine starke Professionalisierung und methodologische Standardisierung der Evaluation entwickelt (vgl. die Beiträge von Stefan Kuhlmann, Zinöcker/Dinges sowie Polt in diesem Band). Hingegen beschränkt sich die Evaluation der Lehrfunktion von Hochschulen ganz überwiegend auf so genannte Feedbackbefragungen bei Studierenden beziehungsweise ist untergeordneter Bestandteil von Peer-Reviews, die überwiegend der Analyse der Forschungsleistungen von Universitätseinheiten dienen. Hier hat also die gesetzliche Auflage, Qualitätssicherung zu betreiben, dazu geführt, dass die Hochschulen Evaluationsressourcen letztlich ungleichgewichtig investiert haben: Den weitaus grössten Teil der Entwicklung von Evaluationsvermögen in den Bereich der Forschungsaktivitäten, den geringeren Teil zur Abdeckung der gesetzlichen Verpflichtungen in den Bereich der Lehre. Bei Letzterem zeigen sich also Parallelen zu den Reaktionen auf Evaluationsverpflichtungen, wie sie im Bereich der EU-induzierten Evaluationsauflagen, etwa für den Bereich der Agrar- und Arbeitsmarktpolitik, in Deutschland und Österreich festzustellen sind.

Weiter erschwerend wirkt in föderalistischen Systemen, wie sie die drei Vergleichsländer verkörpern, dass die oftmals anzutreffenden Mehrebenenkonstellationen (vgl. Widmer/Frey 2006) ohne wirkungsvolle Mechanismen der vertikalen Koordination noch zu einer Verschärfung der Problematik einer unausgewogenen Ressourcenverteilung führen. Illustrativ sei hierbei wiederum auf die Hochschulsysteme, aber auch auf den Gesundheitsbereich (vgl. die Beiträge von Mayring, Spiel/Petscharnig, Uchtenhagen sowie Spycher in diesem Band) und die – vor allem in Deutschland und Österreich – durch EU-Evaluationsanforderungen geprägten Evaluationsfelder wie Agrar-, Raumentwicklungs- oder Strukturpolitik verwiesen, die aufzeigen, dass Evaluationsaktivitäten einen Gegenstand in unkoordinierter Weise multipel abzudecken vermögen.

In diesen Fällen ist eine unzureichende Ressourcenausstattung für die einzelne Evaluation zu erwarten, wenn nicht eine äusserst kompetente Evaluationskonzipierung kombiniert mit einer üppigen Ressourcenausstattung gegeben ist.

3.4 Bewertung

Zusammenfassend seien stichwortartig einige Risiken aufgezeigt, die sich für das weitere Gedeihen der Evaluation aus den beobachteten Entwicklungen in den Themenfeldern über die drei Länder hinweg oder für einige der Länder zeigen:

Falls rechtliche Evaluationsverpflichtungen nicht durch die umsetzenden Stellen aktiv getragen und mit fundierter Evaluationsexpertise realisiert werden, wenn kein ‚Commitment‘ gegenüber der Evaluation besteht und keine Chancen gesehen werden, dass sich ein Mehrwert ergeben könnte (etwa eine substantielle und nachhaltige Verbesserung des evaluierten Programms), kann dies für die Evaluation negative Konsequenzen haben: Es droht ihr die Unter- und Einordnung in Verwaltungsroutinen, Aufsichtsfunktionen und reine Rechtmässigkeitskontrollen. Derartige „Evaluation“ genannten Routinen (seien sie selbst durchgeführt oder beauftragt) werden weithin so konzipiert, dass sie keine überraschenden, unerwarteten oder unerwünschten Ergebnisse erbringen (können). Sollten Berichte dennoch kritische oder negative Befunde enthalten, werden sie nicht veröffentlicht.

Unabhängige Evaluatorinnen, welche diese Spielregeln nicht verstehen oder nicht einhalten, laufen Gefahr, nicht (wieder) beauftragt zu werden. Derartige in Verwaltungsroutinen eingepasste Evaluationen würden Edward Suchman (1967: 143) oder Daniel Stufflebeam (2000) wohl den Pseudo-Evaluationen zuordnen und sie stellen aus der Sicht des Steuerzahlers eine Geldverschwendung dar, insofern sie keinen Unterschied machen: Ob sie stattfinden oder nicht, trägt nicht zur Verbesserung der Entscheidungsfindung oder zu deren besserer Absicherung bei.

Es besteht die Gefahr, dass Evaluationen von bestimmten Organisationen immer wieder und ausschliesslich feldgebunden durchgeführt werden. Das heisst, es werden vorrangig oder ausschliesslich „Evaluatorinnen“ beauftragt, welche über grosse Feldexpertise verfügen und welche die Werte der jeweiligen Experten und Professionellen des Feldes teilen: Ärzte evaluieren medizinische Leistungen, Lehrer evaluieren pädagogische Leistungen und Sozialarbeiter evaluieren sozialarbeiterische Leistungen und so weiter. Damit drohen die der Bewertung der Programme und Politiken zugrundegelegten Bewertungsmaßstäbe ebenso wie die überhaupt gestellten Fragestellungen a priori auf einen engen Bereich begrenzt zu werden, der fach- und themengebunden ist. Die hohe Akzeptanz der Insider wird dadurch erkaufte, dass – zumindest auf längere Sicht – die Glaubwürdigkeit der Evaluation aus Sicht feldexterner Stakeholder leidet.

Bei Themenbereichen mit erheblichem Finanzierungsdruck besteht die Gefahr, dass Evaluationsressourcen einseitig dezentral oder einseitig zentral zugewiesen werden, dass damit Evaluationskapazitäten im vernachlässigten Bereich fehlen und die systemische Wirkfähigkeit von Evaluation insgesamt in Frage steht. Darauf zu vertrauen, dass das Pendel nach einigen Jahren rein zentraler, durch Methodenexperten betriebener Evaluation auf der Makroebene wieder umschlagen und dann ein heilsamer Schwung angemessene Evaluationen mit hoher Nützlichkeit auch auf der Meso- und Mikroebene auslösen würde, ist sicher eine naturalistische Fehlannahme. Auch die Evaluation muss sich der Frage stellen, in welchem Zeitraum und über welche Mechanismen ihre Ergebnisse mit hoher Sicherheit Nutzen entfalten können, auch sie muss Wirkungsmodelle vorweisen. Dabei müssen in der Regel die verschiedenen Beteiligtebenen angemessen repräsentiert und involviert sein,

Evaluationsvermögen muss in allen Bereichen angemessen vorhanden sein; eine Konzentration in einem exklusiven Bereich bedeutet, Ressourcen suboptimal zu allozieren.

4 Schluss

Eine zentrale, zukünftige Herausforderung besteht in der Etablierung von zweckmässigen Vorgehensweisen bei der Auswahl geeigneter Evaluationsgegenstände. Ohne derartige Grundlagen ist damit zu rechnen, dass entweder eine flächendeckende Evaluation mit zu kleiner Ressourcenausstattung und erheblichen Qualitätsdefiziten Einzug hält oder dass eine rein politisch determinierte Selektion der Evaluationsgegenstände erfolgt. Themenfachlich geprägten Evaluationszugängen mangelt es an Fokussierung respektive Diskriminierung, was zur flächendeckenden und ständigen Realisierung von Kleinstevaluationen führt. Politisch geprägte Evaluation wiederum bringt mit sich, dass Evaluationsaktivitäten auf einzelne, politisch brennende Gegenstände fokussiert werden, ohne den Einsatz der Evaluation hinsichtlich seines Nutzenpotentials kritisch zu reflektieren.

Für eine zweckmässige Verteilung der Evaluationsaktivitäten auf Themenfelder und Gegenstände ist eine Evaluationsplanung notwendig, die sich auf fachliche Grundlagen der Evaluation abstützt. Dies gilt für die Festlegung einer Verteilung der Evaluationsaktivitäten auf verschiedene Aktivitätsfelder, für die Zuweisung von Evaluationszuständigkeiten, für die Auswahl spezifischer Evaluationsgegenstände, für eine angemessene Konzipierung der einzelnen Evaluation wie auch für die entsprechende Mittelausstattung.

Um den Gefahren einer Fehlallokation von Evaluationsressourcen entgegen zu wirken, braucht es für die Evaluation im politischen System – und bei den relevanten professionellen und fachpolitischen Akteuren – ein Mindestmass an Institutionalisierung der Evaluationsfunktion. Im Vergleich der drei Länder und der zehn Themenfelder lassen sich hier verschiedene Optionen feststellen und Tendenzen identifizieren. Darüber hinaus ist eine gewisse gesetzliche Absicherung von Evaluationen über Evaluationsklauseln grundsätzlich wünschenswert, wobei diese den Bedarf nach evaluationsfachlicher Expertise nicht ersetzen können. Vielmehr droht bei einer nicht evaluationsfachlich abgestützten Rechtsverpflichtung zur Evaluation die bereits angesprochene Routinisierung in Form von Pseudo-Evaluationen. Weiter ist darauf zu achten, dass Evaluationspflichten nicht in einer Form abgefasst werden, welche eine Routinisierung befördern, indem sie etwa flächendeckende und regelmässige Evaluationen einfordern.

Es wird weiterhin vorrangig eine Aufgabe der Fachorganisationen der Evaluation und anderer Fachgesellschaften sein, die Relevanz der Evaluationsfachlichkeit aufzuzeigen und deren Akzeptanz – namentlich bei den Organisationen der Themenfachlichkeit sowie in Politik und Verwaltung – zu fördern, um so die Qualität und den Nutzen der Evaluation zu steigern.

Herausgeber

PD Dr. **Thomas Widmer** ist Leiter des Forschungsbereichs ‚Policy-Analyse & Evaluation‘ sowie Privatdozent am Institut für Politikwissenschaft, Universität Zürich. Forschungs- und Lehrtätigkeit an den Universitäten St. Gallen, Bern, Konstanz und an der Harvard University. Mitglied des Vorstandes der Schweizerischen Evaluationsgesellschaft (SEVAL; seit 1998), des Executive Board der European Evaluation Society (EES; 2002-2005) sowie des Ethics Committee der American Evaluation Association (AEA; seit 2008). Forschungs-, Evaluations-, Lehr- und Beratungstätigkeit in diversen Politikfeldern.

Dr. **Wolfgang Beywl** ist Leiter des Weiterbildungsprogramms ‚Evaluation‘ an der Koordinationsstelle für Weiterbildung (KWB) der Universität Bern, Gründer und wissenschaftlicher Leiter der Univation GmbH – Institut für Evaluation Köln. Lehrtätigkeit am Institut für Höhere Studien, Wien sowie in Weiterbildungsstudiengängen der Universität und der Fachhochschule Bern. 1997 bis 2005 Vorstandsmitglied der Gesellschaft für Evaluation (DeGEval). Evaluations- und Beratungstätigkeit in diversen Politikfeldern.

Prof. **Carlo Fabian** ist Dozent an der Fachhochschule Nordwestschweiz, Hochschule für Soziale Arbeit in Basel. Forschungs- und Lehrtätigkeiten an der Universität Zürich, am Institut für Suchtforschung Zürich, an der Fachhochschule München, sowie im interuniversitären Masterstudiengang Public Health (CH). Forschungs-, Evaluations-, Lehr- und Beratungstätigkeit in diversen Bereichen der Gesundheitsförderung, Prävention und Sozialen Arbeit. Seit 2008 Leiter RADIX Nordwestschweiz, Kompetenzzentrum für Gesundheitsförderung und Prävention.

Autorinnen und Autoren

Die Autorinnen und Autoren äussern in ihren Beiträgen in diesem Band ihre persönlichen Auffassungen, die nicht mit jenen der entsprechenden Institutionen übereinstimmen müssen.

PD Dr. Andreas Balthasar, Interface – Institut für Politikstudien, Luzern, Schweiz

Dr. Lars Balzer, Eidgenössisches Hochschulinstitut für Berufsbildung, Zollikofen, Schweiz

Evelyn Bergsmann, Fakultät für Psychologie, Universität Wien, Österreich

Dr. Wolfgang Beywl, Koordinationsstelle für Weiterbildung, Universität Bern, Schweiz

Dr. Oliver Bieri, Interface – Institut für Politikstudien, Luzern, Schweiz

Dr. Kurt Bisang, Energiepolitik, Bundesamt für Energie, Ittigen, Schweiz

Dr. Werner Bussmann, Rechtsetzungsprojekte und -methodik, Bundesamt für Justiz, Bern, Schweiz

Michael Dinges, Institut für Technologie- und Regionalpolitik, Joanneum Research Forschungsgesellschaft mbH, Wien, Österreich

Prof. Dr. Matthias Drilling, Institut Sozialplanung und Stadtentwicklung, FHNW – Hochschule für Soziale Arbeit, Basel, Schweiz

Markus Duscha, ifeu – Institut für Energie- und Umweltforschung Heidelberg GmbH, Deutschland

Prof. Carlo Fabian, Institut Kinder- und Jugendhilfe, FHNW – Hochschule für Soziale Arbeit, Basel, Schweiz

Dr. Peter Farago, Swiss Foundation for Research in Social Sciences FORS, Lausanne, Schweiz

Frohmut Gerheuser, POLIS Politikberatung und Sozialforschung, Brugg AG, Schweiz

Regina Grajewski, Institut für Ländliche Räume, Johann Heinrich von Thünen-Institut (vTI), Braunschweig, Deutschland

Karin Haubrich, Abteilung Jugend und Jugendhilfe, Deutsches Jugendinstitut, München, Deutschland

Wolfgang Helmstädter, Referat NL 14, Bundesministerium für Verkehr, Bau und Stadtentwicklung, Berlin, Deutschland

Dr. Christian Hirschi, Institut für Umweltentscheidungen, Eidgenössische Technische Hochschule, Zürich, Schweiz

Richard Hummelbrunner, ÖAR Regionalberatung – Büro Steiermark, Graz, Österreich

Vinzenz Jung, Direktionsbereich Strategie und Evaluation, Bundesamt für Landwirtschaft, Bern, Schweiz

Dr. Herbert Klemisch, Klaus Novy Institut, Köln, Deutschland

Alexander Kohler, Österreichische Qualitätssicherungsagentur (AQA), Wien, Österreich

PD Dr. Götz Konzendorf, Bundesministerium des Innern, Berlin, Deutschland

PD Dr. Sabine Kuhlmann, Institut für Sozialwissenschaften, Humboldt-Universität zu Berlin, Deutschland

Prof. Dr. Stefan Kuhlmann, School of Management and Governance, University of Twente, Enschede, Niederlande

Prof. Frans L. Leeuw, Maastricht University, Niederlande

Melanie Leschnik, Research Institute for Managing Sustainability, Wirtschaftsuniversität Wien, Österreich

Prof. Dr. Rainer Loidl-Keil, Sozialarbeit und Sozialmanagement, FH Joanneum, Graz, Österreich

Prof. Dr. Katharina Maag Merki, Pädagogische Hochschule Freiburg, Deutschland

Prof. Dr. Luzius Mader, Direktionsbereich Öffentliches Recht, Bundesamt für Justiz, Bern, Schweiz

Andreas Maier, Österreichische Raumordnungskonferenz, Wien, Österreich

Anton Mair, Bundesministerium für europäische und internationale Angelegenheiten, Wien, Österreich

Doz. Dr. André Martinuzzi, Research Institute for Managing Sustainability, Wirtschaftsuniversität Wien, Österreich

Prof. Dr. Philipp Mayring, Institut für Psychologie, Universität Klagenfurt, Österreich

Dr. Wolfgang Meyer, Centrum für Evaluation, Universität des Saarlandes, Saarbrücken, Deutschland

Karl M. Ortner, Bundesanstalt für Agrarwirtschaft, Wien, Österreich

Joachim Petscharnig, Pro Mente: Kinder Jugend Familie, Klagenfurt, Österreich

Dr. Rupert Pichler, Forschungs- & Technologieförderung, Bundesministerium für Verkehr, Innovation und Technologie, Wien, Österreich

Prof. Dr. Werner Pleschberger, Wirtschafts- und Sozialwissenschaften, Universität für Bodenkultur, Wien, Österreich

Wolfgang Polt, Institut für Technologie- und Regionalpolitik, Joanneum Research Forschungsgesellschaft mbH, Wien, Österreich

Thomas Ragni, Staatssekretariat für Wirtschaft, Bern, Schweiz

Nicolà Reade, Centrum für Evaluation, Universität des Saarlandes, Saarbrücken, Deutschland

Dr. Heinz Rhy, Schweizerische Konferenz der kantonalen Erziehungsdirektoren (EDK), Bern, Schweiz

Prof. Dr. Fritz Sager, Institut für Politikwissenschaft, Universität Bern, Schweiz

Dr. Uwe Schmidt, Zentrum für Qualitätssicherung und -entwicklung (ZQ), Universität Mainz, Deutschland

Oliver Schwab, IfS Institut für Stadtforschung und Strukturpolitik, Berlin, Deutschland

Werner Specht, Bundesinstitut für Bildungsforschung, Innovation und Entwicklung des Bildungswesens (bifie), Graz, Österreich

Prof. DDr. Christiane Spiel, Fakultät für Psychologie, Universität Wien, Österreich

Dr. Georg Spiel, Neurologie und Psychiatrie des Kindes- und Jugendalters, Landeskrankenhaus Klagenfurt, Österreich

Dr. Stefan Spycher, Schweizerisches Gesundheitsobservatorium OBSAN, Neuchâtel, Schweiz

Prof. Dr. Alain Thierstein, Lehrstuhl für Raumentwicklung, Technische Universität München, Deutschland

Günter Tissen, Bundesministerium für Ernährung, Landwirtschaft und Verbraucherschutz (BMELV), Bonn, Deutschland

Prof. DDr. Ambros Uchtenhagen, Institut für Sucht- und Gesundheitsforschung, Zürich, Schweiz

Gerald Wagner, isw Institut, Halle (S.), Deutschland

PD Dr. Thomas Widmer, Institut für Politikwissenschaft, Universität Zürich, Schweiz

Dr. Angela Wroblewski, Institut für höhere Studien IHS, Wien, Österreich

Klaus Zinöcker, Wiener Wissenschafts- Forschungs- und Technologiefonds WWTF, Wien, Österreich

Michaela Zintl, Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung BMZ, Bonn, Deutschland

Dieter Zürcher, KEK-CDC Consultants, Zürich, Schweiz

Personenverzeichnis

(Seitenzahlen in Klammern verweisen auf Ko-Autoren und Mitherausgeberschaften)

- Abs, Hermann Josef 161, 197
Achermann, Christin 348
Achtenhagen, Frank (194)
Ackermann, Heike 193
Aeppli, Daniel 140, 141
Ahmad, Sharjeel (265)
Aichholzer, Georg 130
Airaghi, Angelo 288, 292
Albisser, Gregor (105)
Alexaki, Anastasia (500, 505)
Aliesch, Beat 428
Alkin, Marvin 13, 70, 477, 479, 480, 490, 517
Alliata, Roberta (187, 194)
Allraum, Jürgen 187, 194
Altrichter, Herbert 172, 173, 175, 444, 524
Angele, Hans-Christian 233
Arends, Lena (194)
Argyris, Chris 158
Arnold, Benedikt 141
Arnold, Erik 45, 296, 299, 300, 301
Arnoldt, Bettina 161, 199
Atria, Moira 485
Avenarius, Hermann (160, 161)
- Babic, Edwin (194)
Baccini, Peter 434
Bacher, Joachim (296)
Bachmann, Gaudenz (348)
Bachmann, Helmut 172
Bachmann, Nicole (197, 219, 312, 344, 349, 431, 444, 471, 498)
Bachmann, Ruth 141
Bachtler, John 430,
Bachtlor, John (410)
Badura, Bernhard 327
Baethge, Martin 194
Balthasar, Andreas 12, 16, 60, 68, 69, (141), 231, 232, 233, 238, 368, 479, 480, 481, 482, 483, 486, 488, 491, 494, 495, 498, 499, 505, 511, 515
Balzer, Lars 193, 199, 513, 515
Barth, Ingola (194)
Barz, Andreas 164
Basler, Ernst 264
Battaglia, Markus 346
Bättig, Christoph (368), 427
Bauer, Helfried 389
Bauer, Helmut (208, 238)
Bauer, Michael 377
Bauer, Tobias 139
Baumann, Beat (139)
Baumert, Jürgen 159, 194
Baumgartner, Edgar 348, (464, 468), 469
Baur, Martin (366)
Baur, Priska 107
Baur, Roland (464)
Beck, John (346)
Beck, Ulrich 203
Becker, Franz J.E. 481
Becker, Nadine (83)
Behrendt, Heiko 423, 425, 426, 427
Behrens, Johann 327
Bellenberg, Gabriele 159, 160
Bellmann, Lutz 408
Belser, Katharina 309
Bennett-Lartey, Samuel (500, 505)
Benz, Arthur 372, 433, 438
Berger, Christa 346, (347)
Berger, Fritz (265)
Berger-Künzli, Daniela 468
Bergmann, Andreas 187
Bergs, Rolf (83)
Bergschmidt, Angela 83
Bergsmann, Evelyn 477, 498, 511, 515
Bernhard, Josef (390)
Bertelsmeier, Markus 77
Berto, Daniele (348)
Besson, Jaques (349)
Beule, Wolfgang 348
Beyme, Klaus von 381
Beywl, Wolfgang 11, 13, 158, 445, 446, 448, 471, 509
Bickman, Leonard 222
Bickner, Robert E. 517
Bieger, Thomas 426, 427
Bieling, Alfons (83)
Bieri, Oliver 137, (141), 519, 522

- Binder, Bruno 42
 Binder, Michael 296
 Bindeus, Heinz 390
 Birkeland, Sarah 480
 Bisang, Kurt 238, 240, (342, 348), 519, 523
 Bischof, Ueli (233)
 Bisig, Brigitte 348
 Biwald, Peter (382), 388, (389, 390)
 Blaser, Cornelia (473)
 Blättler, Richard (347)
 Blaudszun, Annette (348)
 Block, Hans-Jürgen 286
 Blum, Werner (160, 161)
 Bober, Marcie J. (14)
 Bochsler, Daniell 362
 Bock, Karin 441
 Böde, Ulla 207
 Bogner, Daniel 95
 Bogumil, Jörg 361, 377, 378
 Böhnke, Jan (252)
 Bohnsack, Ralf 471
 Bohr, Nils (427)
 Böhret, Carl 28, 31
 Boillat, Patrick (428)
 Bolliger-Salzmänn, Heinz (346)
 Bonfadelli, Heinz 346
 Borges, Peter 337
 Bornett, Walter 300
 Borrmann, Axel 247, 272, 274
 Bortz, Jürgen 324
 Bos, Wilfried (159), 160, (194)
 Bouckaert, Geert (70)
 Boulianne, Louis Marie 425, (428)
 Bourguignon, Jean-Pierre 302
 Boza, Albina Maestrey (500, 505)
 Brand, Karl-Werner 203
 Brand-Sassen, Henning (80)
 Brandt, Tasso 500
 Brandt, Willy 203
 Braun, Dietmar 66, 394
 Brauns, Dieter 14
 Brice, Kim Noële (500, 505)
 Brinkmann, Henrik (43)
 Bruderer, Ernst (187)
 Brügger, Urs (348)
 Brugnoli, Gianni (428)
 Brun, Stephan (140)
 Brunsson, Nils 374, 379
 Bucek, Milan (416)
 Buchen, Herbert (159)
 Büchler, Peter 263
 Büeler, Xaver 183, 524
 Buhren, Claus G. 175
 Buliard, Jean-Luc (348)
 Burkard, Christoph 185, 193
 Burla, Sophie 310
 Burnand, Bernard (348, 349)
 Bürsch, Michael 376
 Busch, Nils E. (288, 292)
 Bussmann, Kristina (366)
 Bussmann, Werner 16, 58, 65, 68, 69, 231, 361, (367), 392, 397, 399, 489, 522
 Bütler, Mirjam (104)
 Bützer, Michael 68, 69
 Caduff, Gabriel 233
 Callon, Michel 284
 Campbell, David F.J. 45
 Campbell, Donald 495
 Campilan, Dindo (500, 505)
 Caracelli, Valerie 479
 Carden, Fred (500, 505)
 Caspari, Alexandra 277
 Castelli, Francesco 464, 465
 Castelli, Stefano (428)
 Cattacin, Sandro (348)
 Cavegn, Gion 107
 Cecon, Franziska (383, 387, 389, 390)
 Cemerin, Michael (364, 366, 370)
 Xhang, Mathilde (500, 505)
 Chelimsky, Eleanor 67
 Chen, Huey-Tsyh 222
 Chimienti, Milena 348
 Christ, Paschal (428)
 Clausnitzer, Klaus-Dieter 210
 Clerici, Remo (428)
 Conen, Dieter (348)
 Conrad, Susanna 448
 Cook, Thomas (444)
 Cornuz, Jaques 348
 Cousins, Bradley J. 477, 479, 486, 494
 Czerwanski, Annette 159, 193
 Dahler-Larsen, Peter 70, 381
 Daintith, Terence 15
 Dalin, Peter 159
 Dallinger, Alfred 129
 Daniel, Hans-Dieter 285

- Daub, Regina (80)
 Davies, Huw T.O. 495
 Davies, Philipp 65
 de Coulon, Anne (141)
 de la Teja, Ileana (14)
 De Tommasi, Roberto 233, (428)
 Dearing, Elisabeth 382, 383, 384, 385, 386, 387, 390
 Debrunner, Anneliese 348
 Dederling, Kathrin (160)
 Delley, Jean-Daniel 59, 68
 Denny, Evelyne (425)
 Derlien, Hans-Ulrich 15, 16, 69, 373, 379, 396, 398, 443, 469
 Deschenaux, Chantal 427
 Dettli, Reto 233
 Devouassoux, Christophe 425
 Dewe, Bernd 449
 Dietler, Urs (232)
 Dietzel, Gottfried T.W. 442
 Dinges, Michael 20, 295, (296), 299, 300, 313, 314, 510, 525
 Dinkel, Rolf (348)
 Dirksmeyer, Walter 80, (83)
 Ditton, Hartmut 161, 199
 Dittrich, Gisela 448
 Divorski, Stan (67)
 Döbert, Hans 159
 Dobler-Mikola, Anja (347)
 Döbrich, Peter (160, 161)
 Doll, Helmut (80)
 Dombrowski, Daniela 346
 Donaldso, Steward I. 222
 Döring, Nicola 70, 324
 Downes, Ruth (416)
 Dozio, Edo (187, 194)
 Draxl, Petra 416
 Dresing, Thorsten (471)
 Drilling, Matthias (141), 463, 464, (468), 471, 473, 502, 522
 Dubas, Daniel (346, 348)
 Dünnhoff, Elke 207
 Duong, Le Thanh (500, 505)
 Duscha, Markus 203, (207), 208, 238, 241, 523
 Duve, Thomas 383, 386
 Dyllick, Thomas 208, 209
 Eastus, Christopher (349)
 Eberhard, Sybille 347
 Ebinger, Falk 377
 Ebner, Heinz (348)
 Eckey, Hans-Friedrich 403
 Eder, Ferdinand 173
 Edler, Jakob 292, 300, (301), 304
 Effionayi-Mäder, Denise 348
 Efken, Josef (80, 83, 84)
 Eggenberger, Christina (349)
 Egger, Mathias (346)
 Eggimann, Jean-Claude (348)
 Egli, Regula 427
 Ehrler, Melchior (232)
 Eichler, Klaus (348)
 Eigler, Gunther 481
 Eikenbusch, Gerhard 193
 Eisenring, Claudia 348
 Eisinger, Angelus 434
 Elbe, Sebastian 78
 Elias, Jiri 144
 Elkeles, Thomas 325, 328
 Elte, Ronald (66, 67)
 Emmerich, Marcus 157
 Endres, Heinz 326
 Engel, George 342
 Engelniederhammer, Stefan 376
 Engleitner, Johann 174
 Erlanger, Jaques 427
 Erni, Rita 141
 Ernst, Tobias (43)
 Eser Davolio, Miryam 464, 468, 473
 Eser, Thiemo W. 79, 86, (425)
 Esterling, Kevin M. 70
 Etzkowitz, Henry 284
 Eugster, Patrick (350)
 Evans, Alison (249)
 Fabian, Carlo 11, 347, 462, 463, 464, 468, 469, 498, 521,
 Fahrenkrog, Gustavo (292, 318)
 Falk, Armin 139
 Falk, Chantal 473
 Falk, Rahel 299
 Falkner, Gerda 51
 Faltermaier, Toni 322
 Farago, Peter 305, 313, 315, 316, 511
 Fasbender, Karl (274)
 Federspiel, Barbara (348)
 Fegeler, Marie (232)
 Fehr-Duda, Helga (129, 130)
 Feld, Lars P. 368

- Felder, Stefan 107
Feldmann, Ursula (470)
Fend, Helmut 158, 172, 193
Ferro Luzzi, Giovanni 140
Fertig, Michael 123
Fetterman, David 485
Filtzinger, Julia 498
Finz, Alfred 384, 385
Fischedick, Manfred 206
Fischer, Rudolf 265
Fischl, Iris (296, 304)
Flanders, Dana (348)
Fleck, Dieter 415
Fleischmann, Eduard 382
Flick, Uwe 471
Florida, Richard 427
Flösser, Gabi 448
Flückiger, Alexandre 231
Flückiger, Martin (141)
Flückiger, Yves 140, 141, 232
Flury, Christian 103, 104, 105
Foresti, Marta 249
Fornerod, Serge 16, 69
Forstner, Bernhard 78, 80, 82, 83, 84
Foxon, Marguerite 14
Freeman, Howard E. 14, 486
Frei, Andreas 348
Freiburghaus, Dieter 138
Freimann, Jürgen 208
Freitag, Markus 368, 369
Freudenthaler, Harald H. 175
Frey, Bruno S. 15, 66, 70, 197, 368, 369, 516
Frey, Kathrin 70, 231, 516
Frey, Rahel (231)
Friboulet, Jean-Jacques 141
Frick, Andres 139
Frick, Frank 43
Frick, Roman (231)
Frick, Uli (348)
Friemel, Thomas 346
Frölich, Markus 140
Fullan, Michael 161
Furrer, Cornelia 346, 366
Fürst, Dietrich 438
Furubo, Jan-Eric 16, 69, 70

Gabi, Simone (423, 428, 432)
Gabriel-Schärer, Pia (462, 463)
Gadient, Brigitta M. 498

Gamper, Andreas (187)
Gänsfuß, Rüdiger (194)
Ganzaroli, Dennis 139
Gärtner, Ludwig 140
Gassner, Erich 205
Gaus, Hansjörg (205, 238)
Gautschi, Ursula (105)
Georgiou, Luke (288), 292
Gerfin, Michael 139, 140
Gerheuser, Frohmut (232), 233, 368, 369, 433, 510
Germann, Raimund E. 68, 362
Gerster, Richard 264
Gerwig, Claude (105)
Geser, Hans 362
Gessler, Rahel (231)
Gesù, Remo 263
Giacomazzi, Fabio 428
Gianola, Giacomo (428)
Gibbons, Michael 284
Gilbert, Allison (348)
Gindl, Michaela (302)
Glänzel, Wolfgang (285, 286)
Glanzmann, Lars (423, 432)
Gleich, Albrecht von (274)
Gocht, Alexander (77)
Gonon, Philipp 184
Good, Barbara 69, 309
Gormley, William T. jr. 70
Gottweis, Herbert 47, 51
Götz, Konrad (232)
Grabensberger, Eva 174
Grajewski, Regina 75, (83, 84), 523
Grawe, Klaus 347
Gredig, Daniel 461
Greenhalgh, Trisha 326
Greil, Leopold (386)
Greiling, Michael 327
Greiner, Roger-Axel (248)
Greiwe, Stephanie (348)
Greve, Carsten 381, 384
Grillon, Nathalie (423, 432)
Grin, François (187, 194)
Grogger, Günther 173
Grohmann, Romano 444
Grohs, Stephan 378
Groß Ophoff, Jana 161, 198
Gruber, Edelgard 207
Gruber, Hans (160, 161)
Gruber, Heinz 172

- Grunwald, Armin 205
 Grüter, Karin 468
 Grüter, Kurt 56
 Geschwend, Patrick (347)
 Gsponer, Gabriele (232, 366)
 Gstetter, Peter 444
 Gude, Stefanie (302)
 Guggenbühl, Lisa 346, 347
 Guhl, Jutta (464)
 Guidotti, Sabrina (141)
 Guillaume-Gentil, Sylvain (428)
 Güller, Peter 428
 Gunten, Ursula von 468
 Gürtler, Leo (330)
 Guston, David 285
 Gutmann, Renate 464
 Güttinger, Franziska 347
 Gutzwiller, Felix (347), 348
- Haake, Michael (326)
 Haari, Roland 348
 Haari-Oberg, Ilse (122)
 Hackenberg, Helga 122
 Haefeli, Ueli (231)
 Haering, Barbara (366)
 Hafner, Sabine (301)
 Hagen, Tobias 123
 Haider, Günter 173
 Hambruch, Josef 94
 Hammer, Andreas (346)
 Hammer, Stephan 232, (346), 348, (426)
 Hammerschmid, Gerhard 381, 382, 383, 390
 Hämmig, Robert (349)
 Hamschmidt, Jost 208, 209
 Hanft, Anke 181
 Hangartner, Yvo 56
 Hansen, Hanne Fossen 381, 383
 Hansen, Jane 351
 Hanser, Christian 232
 Harmelink, Mirjam (234)
 Harmsen, Robert (234)
 Hartmann, Peter (428)
 Haubrich, Karin 20, 441, 442, 444, 446, 461, 470, 471, 522
 Hauff, Volker 203
 Hauke, Eugen 336
 Häuser, Simon 167
 Heck, Anne (448)
 Heckman, James 303
- Hefti, Helena 103
 Heil, Karolus 470
 Heiner, Maja 441, 444, 445, 470
 Heintel, Martin 85
 Heinze, Thomas 283, 286, 288, 292, 294
 Hellstern, Gerd-Michael 15, 443, 470
 Helmke, Andreas 194, 478, 481
 Helmstädter, Wolfgang 148, 522
 Hendricks, Michael 479
 Henneberger, Fred 140
 Hennen, Manfred 167
 Hennicke, Peter 206
 Henrichsmeyer, Wilhelm 109
 Henry, Gary T. 480, 493
 Herbeck, Gabriele 383, 386
 Hering, Sabine 441
 Herweg, Karl 271
 Hess, Kurt 348
 Hewstone, Miles (502)
 Heyer, Gerd 124
 Hirschi, Christian 523
 Hirsing, Peter (104)
 Hochreutener, Marc Anton (348)
 Hödl, Josef 456
 Hofer, Eduard 106
 Hofer, Helmut (134), 149
 Hoffmann, Brigitte (141)
 Hoffmann, Karl-Heinz 302
 Hofmarcher, Maria M. 331
 Hofreither, Markus 92, 93
 Hofstetter, Carolyn H. 477, 479, 480
 Höhn, Martina (451)
 Holland, Doris 286, 290
 Höllinger, Andreas (425)
 Holtappels, Heinz Günter (159, 160)
 Holthus, Manfred (274)
 Holthusen, Bernd (461)
 Holzer-Möstl, Elke (336)
 Holzinger, Elisabeth 50, 216, 419, 421, 434
 Homfeld, Hans Günther 441
 Hood, Christopher 15
 Höpflinger, François 362
 Hoppe, Andrea 405
 Horber-Papazian, Katia 16, 68, 69, 363
 Hornbostel, Stefan 15, 197, 285
 Hornung, Rainer (464)
 Horton, Douglas 500, 505
 Hoschka, Alexander 444
 Hosenfeld, Ingmar 161, 198

- Hössl, Alfred (444)
 Hotz, Cora (141)
 House, Ernest R. 13
 Hovorka, Gerhard 95
 Huber, Günther L. 229, 238
 Hübler, Karl-Hermann 407
 Hudson, Alan (249)
 Hugentobler, Valerie (141)
 Hügli, Ernst (184, 524)
 Hummelbrunner, Richard 403, 413, 419, 510
 Hunold, Claude 139
 Hunt, Neil (348)
 Hurrelmann, Klaus 327
 Hüsler, Gebhard 346
- Iby, Manfred (172, 173)
 Inderbitzin, Jürg 141
 Inquart, Ernst (390)
 Ipfling, Heinz-Jürgen 158, 193
 Irrek, Wolfgang (234)
 Irresberger, Karl 41, 42
 Isermeyer, Folkhard 77
 Iten, Rolf (231, 233), 234, (348)
- Jacob, Steve 16, 58, 69
 Jacoby, Klaus-Peter 205
 Jäger, Reinold S. 194
 Janetschek, Hubert 95
 Janett, Daniel 56, 496, 503
 Jänicke, Martin 238
 Jann, Werner 361, 372, 374, 375, 378
 Jans, Armin 141
 Jetzinger, Daniela 41
 Johnson, R. Burke 365, 366, 480
 Jordan, Erwin 442, 443
 Jörg, Leonhard 295, (296), 299, 300, (301), 302
 Jouval, Norbert (428)
 Jud, Thomas (300)
 Jung, Vinzenz 99, 519
 Junker, Christoph 346
 Jürgenssen, Olaf (339)
- Kaegi, Urs (461, 464, 468)
 Kägi, Wolfram 348
 Kälín, Roli (140)
 Kaluza, Gert 328
 Kanaan, Sami (231, 427)
 Kapella, Olaf 134
 Karagiannis, Giannis (95)
- Karbach Manfred 14
 Karger, Thomas 346
 Karlsson, Ove 520
 Kaufmann, Philippe 458
 Kehr, Jochen (46)
 Kellaghan, Thomas (18)
 Keller, Odile 263
 Keller, Peter 427
 Keller, Roger 464
 Keller-Ebert, Cornelia 481
 Kempfert, Guy 524
 Keppelmüller, Joachim (174)
 Kern, Augustin (172)
 Kerschl, Viktoria (348)
 Kerseboom, Jeroen S. (66, 67)
 Kersting, Norbert (376)
 Kessler, Stefan (233)
 Kessler, Stephan (231)
 Kettiger, Daniel 16, 59, 69
 Keusch, Gérald (348)
 Khadar, Ibrahim (500, 505)
 Khan, Jamil 235
 Kiegelmann, Mechtild (330)
 Kiendl-Wendner, Doris 451
 Kierkegaard, Sören Aabye 148
 Kilius, Dagmar (159)
 Kirkpatrick, Donald 352
 Kirner, Leopold 94
 Kißler, Leo 376, 378
 Kissling-Näf, Ingrid 369, 489
 Kleinhanß, Werner 77
 Klemenz, Regina (464, 468)
 Klemisch, Herbert 205, 206, 208, 209, 238, 241, 523
 Klepp, Cornelia 189
 Klepper, Reiner (80)
 Kletzan, Daniela 92
 Klieme, Eckhard 159, 160, 161, 197
 Klöti, Ulrich 237, 368, 486
 Kluve, Jochen (123)
 Knapp, Gerald 454
 Knill, Christoph (377)
 Knoepfel, Peter 227, 231, 238, 361, 367, (368), 369, 489
 Kobi, Sylvie 463
 Koch, Benedikt 104
 Koch, Birgit 80, 82
 Koch, Peter 366
 Kodama, Fumio 284
 Koeck, Christian (348)

- Koehler, Jan (252)
 Koehne, Godehard (277)
 Kohler, Alexander 177, 181, 512
 Kohler, Britta 198, 478, 481
 Kohlmann, Uwe (205, 238)
 Koller, Christophe (362)
 Koller, Martin 408
 Kolly, Doris (486)
 König, Joachim 328
 Konle-Seidl, Regina 126, 153
 Konzendorf, Götz 27, 28, 31, 64, 69, 376, 515
 Köpp, Bodo (376)
 Kopp, Ursula 216
 Köppl, Franz 130
 Kosfelder, Joachim (347)
 Köster, Ulrich (78)
 Koszalka, Tiffany A. (14)
 Kottmann, Steffen 377
 Krah, Volker (80)
 Kranz-Plote, Jutta (277)
 Krapf, Sandra (377)
 Krapp, Stefanie (277)
 Krappmann, Lothar (448)
 Kreß, Stefan 329
 Krieg, Oliver (207)
 Kriesi, Hanspeter (368)
 Kriz, Jürgen 326
 Kromrey, Helmut 14, 159, 164, 329, 453
 Krull, Wilhelm 286
 Kruse, Christian (423, 432)
 Kübler, Daniel 370
 Kuckartz, Udo 471
 Kuffner, Astrid 216, 220
 Kuhlmann, Sabine 371, 373, 377, (378), 380, 392, 394, 395, 397, 445, 522
 Kuhlmann, Stefan 167, 196, 282, 283, 284, 285, 286, 288, 290, 292, 293, 294, 313, 316, 317, 510, 525
 Kuhn, Carsten (208, 237)
 Kundert, Sonja 464
 Künzli, Kilian (139)
 Küpker, Bernd (77)
 Kurz Gygax, Rachel (368)
 Kuster, Jürg (231, 428)
- Laat, Bastiaan de (79, 480, 481)
 Labance, Nicola (234)
 Labudde, Peter (187, 194)
 Ladner, Andreas 364, 366, 370
- Ladowski, Wolfgang 327
 Lalive d'Epinay, Rafael (139), 145
 LaLonde, Robert J. (303)
 Lammers, Konrad 408, 409, 411
 Lampart, Daniel 139
 Landert, Charles 183, (197, 219, 312, 344, 349, 431, 444), 469, 470, (471), 473
 Landis, Flurina (231)
 Landsteiner, Günther 434
 Landwehr, Norbert (184), 193, 194
 Lang, Jochen 406
 Lange, Hans-Jürgen (376)
 Langer, Gero 327
 Lanz, Andreas 347
 Larédo, Philippe 283
 Laser, Ulrich 327
 Laskowski, Wolfgang 451
 Lassnigg, Lorenz (130), 134, (135)
 Latzer, Michael 47, 51
 Lauber, Volkmar 213
 Läubli Loud, Marlène M. 344
 Lauterbach, Karl 327
 Lechner, Ferdinand 129, 130, (134, 135)
 Lechner, Michael 134 139, 140, 151
 Ledermann, Simone 346, 348, 492
 Ledoux, Marc J. (288, 292)
 Lee, Barbara 486
 Leeuw, Frans L. 66, 67, 69, 70, 505, 515, 524
 Legewie, Heiner 326
 Lehmann, Rainer H. (159), 194
 Lehmruch, Gerhard 382
 Leinter, Andrea (302)
 Leithwood, Kenneth A. 479, 486, 494
 Leitl, Barbara (42)
 Leitner, Andrea 134, (135)
 Leiner, Karl-Heinz 301
 Lemaire, Donald (67)
 Lemne, Marja 497
 Lenkeit, Jenny 194
 Lenz, Peter (185)
 Leresche, Jean-Philippe (370)
 Leschnik, Melanie 216, 239, 240, 511, 523
 Leu, Robert (140)
 Leutenegger, Franziska 346
 Leutner, Detlev 175, 197
 Leviton, Laura C (444)
 Leydesdorff, Loet 284
 Limoges, Camille (284)
 Lincoln, Yvonna S. 70

- Linder, Wolf 368, 487
 Lippuner, Christoph (232)
 Lipsey, Mark W. (14, 486)
 Lischer, Petra (448)
 Lo, Vivien (300)
 Lohaus, Arnold 328
 Loidl-Keil, Rainer 327, 328, 450, 451, 452, 455, 457, 522
 Löning, Kathrin (425)
 Losa, Stefano (348)
 Loy, Jens-Peter (78)
 Lucyshyn, Josef 173, 194
 Lüders, Christian (442), 446, 470
 Luedi, Christoph 263
 Luks, Fred 203
 Lüthi, Jean Christophe 348
 Lüthi, Stefan 422
 Luthiger, Philip (348)
 Lutkat, Sabine (194)
 Lutz, Hedwig 134, 135
 Lutz, Wolfgang 347
- Maag Merki, Katharina 157, (161, 197), 524
 Mass, Stephanie (329)
 Mack, Gabriele 105
 Madaus, George F. (18)
 Mader, Luzius 52, 55, 64, 68, 69, (367), 425, 487, 496, 515
 Mäding, Heinrich 406
 Maeder, Christoph 140
 Maffli, Etienne 347
 Magnin, Chantal 140
 Mahon, Pascal (486)
 Mahringer, Helmut 130, 134, (135)
 Maibach, Markus (429)
 Maier, Andreas 413, 510
 Mair, Anton 255, 276, 516, 519
 Mair, Peter 64
 Mandl, Irene (301)
 Manegold, Dirk (77)
 Mangold, Max 186
 Mann, Stefan 105
 Maradan, Olivier 186
 March, James G. 361
 Margarian, Anne (80)
 Margraf, Jürgen 347
 Mark, Melvin M. 480
 Marti, Michael 140
 Marti, Peter (428)
- Martinovits-Wiesendanger, Alex 139
 Martinuzzi, André 216, 221, 222, 238, 239, 240, 511, 523
 Mastronardi, Philippe 55
 Mathey, Marie Claire (347)
 Mathiasen, David 67
 Matschek, Markus 389
 Mattes, Christoph 463
 Matti, Daniel (231, 427)
 Matzdorf, Bettina 83
 Mauch, Corine 231, 232
 Maugué, Marc (232)
 Mayer, Kurt (134)
 Mayne, John 67
 Mayntz, Renate 444, 469
 Mayr, Johannes (175)
 Mayring, Philipp 321, 328, 330, 525
 McDonald, Paul (348)
 McLellan, William (348)
 McSweeney, Tim (348)
 Mecklenburg, Roland (448)
 Mehl, Peter 76, 77, 78, (83)
 Mehnert, Angelika (348)
 Meier, Hans-Rudolf 428
 Meier, Peter 263
 Menegale, Sarah (426)
 Meneghin, Franco (28)
 Merton, Robert K. 71
 Merz, Annemarie 141
 Messner, Elgrid (173)
 Mettler, Monika (185)
 Meusburger, Martina 469
 Meyer, Leo (107)
 Meyer, Lutz (277)
 Meyer, Renate E. 382, 383
 Meyer, Wolfgang 203, 204, (205), (216), 237, 238, 240, (277), 523
 Meyer-Krahmer, Frieder 284, 285
 Meylan, Jean-Pierre 193
 Michaelowa, Katharina 272, 274
 Minder, Christoph (346)
 Miquel, Ruth (134)
 Moed, Henk F. 285, 286
 Morales, Daniel 141
 Morand, Charles-Albert 55
 Moser, Christian (342, 348)
 Moser, Heinz 444
 Mücke, Stephan (194)
 Müller, André (140)

Müller, Axel 228, 237
 Müller, Brigitte 376
 Müller, C. Wolfgang 441, 469
 Müller, Caroline (463)
 Müller, Christoph (368, 369)
 Müller, Marianne 23
 Müller, Max (348)
 Müller, Sabine (159, 160), 193
 Müller, Silke (464, 468)
 Müller, Stephan (187, 194)
 Müller, Verena (348)
 Müller-Kohlenberg, Hildegard 445, 470
 Münch, Richard 286
 Münchmeier, Richard 441
 Munier, Gerald (206)
 Muniruzzaman, Imrul Kayes (500, 505)
 Münstermann, Klaus 470
 Murhpy-Graham, Erin 480
 Mustar, Philippe 283

Nadai, Eva 140
 Nahrath, Stéphane 227, 238
 Naschold, Frieder 381, 390
 Neeser, Kurt 348
 Neidhardt, Friedhelm 285
 Nemecek, Nicole (134, 135)
 Nenninger, Peter 481
 Nentwig-Gesemann, Iris 471
 Neuenschwander, Peter 16, 60, 68, 69, 71, (470),
 479, 480, 481, 486, 498, 501, 520
 Neuhäusler, Irène (462, 463)
 Neurath, Wolfgang 295
 Nevo, David 524
 Newburn, Tim 238
 Newcomer, Kathryn (65)
 Nideröst, Bruno (183)
 Niebuhr, Annekatrin 408, 409, 411
 Niederkofler, Carmen (390)
 Nitsch, Heike 100
 Nölle, Joachim (80)
 Nones, Brigitte (301)
 Novak, Rudolf 296
 Novotny, Birgit (134, 135)
 Nowotny, Helga (284)
 Nuspliger, Kurt 59
 Nutley, Sandra M. 495

O'Neil, Tammie 249
 Oberender, Peter O. 327

Obermair, Anna 384
 Obernberger, Ingwald (92)
 Oechler, Melanie 448
 Oedl-Wieser, Theresia 95
 Oetterli, Dominik (428)
 Oeuvray, Kerrie (348)
 Oevreveit, John 330
 Offermann, Frank (77)
 Ohler, Fritz (300)
 Ohm, Anna 378
 Olsen, Johan P. 361
 Ortner, Karl M. 87, 93, 94, 95, 523
 Osinger, Dietmar (172)
 Osterburg, Bernhard 100, (77)
 Osterwald, Stephan (140)
 Oswald, Franz 434
 Ott, Walter (140)
 Otto, Hans-Uwe 449
 Owen, John M. 478, 479

Paccaud, Fred (348)
 Page, Alan 15
 Pagnossin, Elisabetta (187, 195)
 Paliokas, Kathleen L. 67
 Palumbo, Dennis J. 520
 Papagiannis, Yannis 479
 Paschon, Andreas 174
 Passardi Gianola, Lorenza (428)
 Patry, Jean-Luc 175
 Patton, Michael Q. 14, 477, 478, 479, 485, 495, 499
 Paulischin, Herbert 469
 Pechar, Hans 45
 Peek, Rainer (194)
 Pelinka, Anton 170
 Pellert, Ada (302)
 Perez, Jocelyn (500, 505)
 Pesendorfer, Dieter 213, 238
 Peter, Kathrin 104
 Peter, Martin 103, 429
 Peters, Guy B. (70)
 Peters, Matthias (140), 231, 346, 348, (366)
 Petri, Gottfried 171, 172, 193
 Petscharnig, Joachim 331, 525
 Pfeifer, Sylvie (347)
 Pfeiffer, Hermann 193
 Pfingstner, Hubert 93, 94
 Pfister Giauque, Barbara 425, 426, (427, 429)
 Pfusterschmid, Sophie 95
 Pichl, Claudia 92

- Pichler, Rupert 64, 68, 69, 295, 498, 515
 Piller, Edith Maud (464, 468)
 Pinget, Christophe (348)
 Pinson, Gaëlle (427)
 Pinter, Karin 129, 130
 Pitts, Stephen (348)
 Plankl, Reiner 76, 77, 80
 Platzter, Reinhard 388, 391
 Pleschberger, Werner 373, 381, 382, 389, 390, 392, 395, 397, 398, 522
 Pluto, Liane 447
 Pofperl, Angelika 203
 Pohl, Christian (80)
 Pohl, Wolfgang (206)
 Pointner, Wolfgang 296, 303
 Pollitt, Christopher 386
 Polt, Wolfgang 292, 296, (300), 303, 313, 317, 318, 525
 Polutta, Andreas 447
 Polverari, Laura (410)
 Popp, Reinhold 454
 Posch, Klaus 455, 457
 Posch, Peter 172, (173), 524
 Pöschl, Andrea (134, 135)
 Potter, Philip 209
 Power, Michael 15
 Preissing, Christa (448)
 Prenner, Peter (134)
 Prenzel, Manfred (160, 161)
 Preskill, Hallie 479
 Preußler, Annabell 197
 Prittwitz, Volker von 203
 Proeller, Isabella 365, 390
 Promberger, Kurt 383, 386, 387, 388, 389, 390
 Pulli, Raffael (348)
 Puppo, Irene (348)
 Püspök, Rudolf (339)
 Puwein, Wilfried (92)

 Raab, Erich (444)
 Rack, Herta M. 331
 Radauer, Alfred (296, 301, 304)
 Rädiker, Stefan (471)
 Radnitzky, Edwin (172), 173, 193
 Ragni, Thomas 137, 144, 519, 522
 Rahmstorf, Stefan 206
 Raich, Traudlinde 44
 Ramirez, José V. (140)
 Ramseier, Erich 187, 194

 Raschert, Jürgen 158
 Rässler, Susanne (122)
 Rathkolb, Oliver 51
 Reade, Nicolà 516, 519, 521
 Rees, William 204
 Reger, Guido 292
 Rehm, Jürgen (346, 347), 349
 Reichard, Christoph 375, 376, 378, 390
 Reiche, Danyel 237
 Reichl, Bettina (274)
 Reimann, Ralph (485)
 Rein, Martin 286
 Reinhard, Mark (428, 429)
 Reiss, Kristina (160, 161)
 Reissert, Bernd (433)
 Reißig, Kristin (377)
 Reiter, Karin 83
 Reiter, Walter (129, 130)
 Rentsch, Hans 232
 Reuther, Iris 434
 Reutter, Michaela (83)
 Rey, Alfred 433
 Rey, Michel 425, (426, 427)
 Rheinhard, Mark (426, 427)
 Rhomberg, Wolfram 48
 Rhyn, Heinz 182, 183, 184, 524
 Richard, Jörg (448)
 Ricka, Regula (184, 524)
 Riedel, Jochim (77)
 Rieder, Peter 104, 107
 Rieder, Stefan (231), 232, 237, 363, 364, 366, 392
 Riedli, Anna Maria 463
 Riesenfelder, Andreas 134, (135)
 Riess-Passer, Susanne 384
 Riffert, Franz 174
 Rigassi, Reto (233)
 Rigby, John 300, 304
 Rille-Pfeiffer, Christine 134
 Riquarts, Kurt (160, 161)
 Rist, Ray C. 67, 396, 398, 495
 Ritz, Adrian 361, 363, (364), 365, 366, 369, (370), 372
 Röber, Manfred (376)
 Rog, Debra J. 493
 Rogers, Patricia J. 478, 479
 Roggendorf, Wolfgang 83
 Rojo, Jaime (292, 318)
 Rolff, Hans-Günter 158, 159, 172, 524
 Roos, Markus (183)

- Rösner, Ernst (159, 160)
 Rossi, Peter H. 14, 222, 486
 Rossier-Affolter, Roger 348
 Rossmann, Bruno 384
 Rost, Jürgen (160, 161)
 Rothmayr, Christine (70, 517, 518)
 Rudel, Roman (428)
 Rudolf, Peter 516
 Rudow, Katja (80)
 Rüefli, Christian (231, 346, 348)
 Rüegg, Erwin (16), 60, (232, 470, 486, 520)
 Ruegg, Rosalie (296, 304)
 Ruskala, Iris (383, 387, 389, 390)
 Russ-Eft, Darlene F. 14
 Rüst, Thomas 348
- Sager, Fritz 232, (346, 348), 361, (364), 366, 368,
 369, 370, 371, 392, 522
 Salamon, Petra (77)
 Saleemi, Razzaq Abdul (265)
 Salhofer, Klaus (95)
 Salomon, Alice 441
 Salto, Esteve (348)
 Sandahl, Rolf 16, 69
 Sandoz, Menga (348)
 Santa Maria, Alberto (348)
 Sardi, Massimo (232)
 Sassi, Claudia (141)
 Sauter, Joseph (428)
 Savioz, Marcel R. 368
 Sawillion, Martin 208
 Schaaf, Susanne (347), 348
 Schäfer, Martin (80)
 Schäffer, Heinz 16, 42, 388
 Schaffler, Roland (336)
 Schaffner, Dorothee 463
 Schäfli, Martin (348)
 Schalk, Hans-Joachim 408
 Scharpf, Fritz W. 405, 433, 444, 469, 520
 Schedler, Kuno 365, (427)
 Scheffler, Monika (206)
 Scheiblhofer, Elisabeth (134, 135)
 Schellnhuber, Hans-Joachim 206
 Schelske, Oliver (232)
 Schenkel, Walter 231, (370, 428)
 Scherer, Roland 425, (426)
 Scheuer, Markus (123)
 Schewe, Torben (122)
 Schibany, Andreas (296, 299), 301, 302
- Schienstock, Gerhard (130)
 Schiller-Dickhut, Reiner (206)
 Schilling, Karl (348)
 Schindler, Julia (302)
 Schleicher-Tappeser, Ruggero 425
 Schmidmayer, Julia (301, 302, 510)
 Schmid, Günther 151
 Schmidt, Boris 198, 337
 Schmidt, Nicolas (346, 348, 427)
 Schmidt, Ralf 337
 Schmidt, Uwe 164, 168, 448, 512
 Schmidt-Eichstaedt, Gerd 379
 Schmoch, Ulrich 284, (285, 286)
 Schnabel, Fritz (433)
 Schneid, Theo (448)
 Schneider, Christian 233
 Schneider, Friedrich (92)
 Schneider, Gerald (486)
 Schneider, Vera (205)
 Schneidewind, Peter (416)
 Schnell, Klaus-Dieter (425), 426
 Schnorr, Valentin (463)
 Schöberl, Marianne (134, 135)
 Schönenberger, Alain (427)
 Scholl, Bernd 427
 Schömann, Klaus (134)
 Schön, Donald 158, 286
 Schötz, Dirk 207
 Schrader, Friedrich-Wilhelm 198, 478, 481
 Schramek, Jörg 83
 Schrappe, Matthias 327
 Schratz, Michael 173, 193
 Schremmer, Christoph (427)
 Schreyer, Christoph (233)
 Schröder, Jan 446
 Schroeder, Marco (425)
 Schubert, Hans-Joachim 328
 Schug, Liselotte 348
 Schuh, Andreas 134
 Schuh, Bernd 416
 Schulte, Bernd (347)
 Schulze-Krüdener, Jörgen 441
 Schwab, Brigitte (370)
 Schwab, Oliver 84, 403, 409, 523
 Schwappach, David 348
 Schwar, Beatrix 451
 Schwarb, Thomas (348)
 Schwartzman, Simon (284)
 Schwarz, Christine 15, 515

- Schwarz, Wolfgang T. (174)
 Schwengler, Barbara 408
 Sciarini, Pascal (362, 368)
 Scott, Peter (284)
 Scriven, Michael 368, 517, 518
 Seckinger, Mike (447)
 Sedlacek, Peter 407
 Seel, Andrea 175
 Seiler, Benno (366)
 Seitz, Helmut 377
 Serdült, Uwe (70, 517, 518)
 Sfar, Doris (426, 427)
 Shadish, William R. Jr. 444
 Shams, Rasul (274)
 Shapira, Philip 292
 Sharma, Bavna (249)
 Sheikh, Sonja 295, (296), 300, 301
 Sheldon, George 139
 Shulha, Lyn M. 477
 Siegfried, Gerhard 263
 Siegrist, Johannes 327
 Siegrist, Stefan (348)
 Sieverts, Thomas (427)
 Silva, José de Souza (500, 505)
 Simma, Angelika 41, 44
 Simon, Christian 310
 Simon, Dieter 15, 70
 Simon, Martha-Maria (386)
 Simonet, Jean-Claude (141)
 Sinabell, Franz 95
 Sinz, Manfred 433
 Skinner, Denise 381
 Slama, Karen (348)
 Smeral, Egon 427
 Smeral, Kristin (134, 135)
 Smith, Jeffrey A. (303)
 Smith, Peter C. (495)
 Sohlman, Åsa 497
 Solmon, Lewis C. 70, 517
 Sommer, Ulrich (77)
 Sommerfeld, Peter (461), 464, 465, 468
 Soratori, Robert 454
 Sorger, Claudia (134)
 Soulet, Marc-Henry (141)
 Sousa-Poza, Alfonso (140)
 Soydan, Haluk (65)
 Specht, Werner 170, 172, 173, 174, 175, 524
 Speck, Raphael (231)
 Speer, Sandra (446)
 Spermann, Alexander 123
 Spicher, Benjamin (348)
 Spiegel, Hiltrud von 445
 Spiel, Christiane 12, (173), 477, 485, 498, 499, 500, 515
 Spiel, Georg 331, 525
 Spinatsch, Markus 16, 69, 397
 Spycher, Stefan 353, 525
 Stalder, Ursula (187, 194)
 Stamm, Margrit 192, 477, 478, 493, 501
 Stampfer, Michael 295
 Stanley, Julian C. 495
 Staubli, Nicole 23
 Stecher, Brian 490
 Stefer, Claus (471)
 Steffen, Thomas (347)
 Stegmann, Susanne (83)
 Steiger, Heidi (140)
 Steimel, Kirsten (83)
 Steindl, Claudia (48)
 Steinemann, Myriam (426, 427)
 Steiner, Isabelle (232)
 Steiner, Kurt 271
 Steiner, Mario (134, 135)
 Steiner, Olivier 463, (464), 468
 Steiner, Peter (134, 135, 184), 193, 194, (524)
 Steiner, Roald 296, (304)
 Steiner, Wolfgang 44
 Steininger, Karl (92)
 Stekel, Herbert (336)
 Stephan, Gesine 122
 Stephenson, Geoffrey M. (502)
 Stern, Elliot 477
 Steurer, Reinhard 216, 221
 Stevens, Alex 348
 Stockbauer, Uta 478, 499, 515
 Stocker, Eugen (187, 194)
 Stockmann, Reinhard 15, 164, 205, 238, 247, 277, 278, 470
 Storni, Marco (141)
 Straubhaar, Thomas 149
 Strauf, Simone (425)
 Strebel, Nicole (426, 427)
 Strehl, Franz (42)
 Streicher, Gerhard 296, 299
 Streitberger, Konrad (326)
 Striewe, Ludwig 78
 Strittmatter, Anton 524
 Stroebe, Wolfgang 502

Strohmeier, Gerhard 50, 419, 421
 Strubelt, Wendelin 407
 Struhkamp, Gerlinde (442, 461)
 Stübi, Beat (464)
 Stuck, Andreas 346
 Studer, Thomas 185
 Stufflebeam, Daniel 18, 526
 Stutzer, Alois 368, 369
 Suchman, Edward A. 453, 526
 Summa, Hilka 66
 Surry, Felix von 263
 Sutter, Daniel (104, 429)
 Szucs, Thomas 348

Tálos, Emmerich 51
 Tamme, Oliver 95
 Tanquerel, Thierry (231)
 Tatzer, Ernst (339)
 Tauber, Magdalena 451
 Taut, Sandy 14
 Taylor, Sandra 410
 Telser, Harry (350)
 Tempelmaier, Brigitte 295
 Tenorth, Heinz-Elmar (160, 161), 194
 Theiss, Roland (141)
 Thévoz, Laurent 16
 Thiel, Thomas (448)
 Thierau, Heike 453, 470
 Thierstein, Alain 422, 423, (425), 426, 427, 428,
 432, 510
 Thole, Werner 441
 Tholen, Sven (347)
 Thonhauser, Josef 175
 Tiemann, Silja (83)
 Tissen, Günter 27, 109, 509
 Toepel, Kathleen 27, 84, 409
 Topolnik, Michaela 295
 Torisi, Marco (348)
 Tötschel, Anja 454
 Trageser, Judith (346, 348)
 Traimond, Sylvie (362)
 Trefflich, Anette (80)
 Tribl, Christoph 93
 Trinkl, Barbara (348)
 Trippolini, Ivar (362)
 Trow, Martin (284)
 Tschemer, Marlies (452)
 Tschöpe, Andreas (140)
 Turowski, Gerd 403

Uchtenhagen, Ambros 342, 346, 347, (348), 349,
 525
 Uetrecht, Inge (80, 83)
 Uffer, Sabina (231)
 Umbach, Eberhard 70
 Untiedt, Gerhard 408
 Urbahn, Julia (205, 238)
 Vader, Jean-Paul 349
 Valovirta, Ville 486
 Van Ackeren, Isabell 159, 160
 Van den Berg, Max (427)
 Van der Knaap, Peter 486
 Van Raan, Anthony F.J. (288, 292)
 Van Santen, Eric (447)
 Vaney, Claude 347
 Varone, Frédéric 16, 58, 69
 Vassiliev, Anatoli (140, 141)
 Vaterlaus, Stefan 350
 Vatter, Adrian 362, 368, 369
 Vedung, Evert 486, 493, 501
 Verhoest, Koen 70
 Vernooy, Ronnie (500, 505)
 Verschuere, Bram (70)
 Vettori, Anna (346, 348, 427)
 Viana Baptista, Jorge (288, 292)
 Victor, Norbert (326)
 Viechtbaur, Karin 451, (452)
 Vodoz, Luc 429
 Voelzkow, Helmut 405
 Vogt Gehri, Franziska (187)
 Voillat, Françoise (428)
 Vollmer, Helmut J. (160, 161)
 von Fischer, Christoph 428
 von Stokar, Thomas 426, 427
 von Troschke, Jürgen 442
 von Wartburg, Sara (428)
 Vonortas, Nicholas 317, 318
 Wacker, Hans Rudolf (347)
 Wackernagel, Mathis 204
 Wagner, Gerald 117, 519, 522
 Waldhauser, Franz 339
 Walser, Fredi 141
 Walser, Manfred (425, 426, 427)
 Wälti, Sonja 369
 Wälty, Samuel 263
 Wapf, Bettina 346
 Warta, Katharina (296, 301)

- Wasem, Jürgen (329)
 Watson, Nancy 161
 Watts, Jamie (500, 505)
 Watts, Ronald 394, 396
 Weber, Andrea 149
 Weber, Matthias (48)
 Weber, Max 456
 Weber, Susanne (194)
 Weis, Katja 362
 Weiß, Christoph (92)
 Weiß, Gerhard 92
 Weiss, Carol H. 14, 238, 366, 369, 480, 492, 493, 494, 501, 502, 520
 Wez, Dietmar (83)
 Wendt, Heinz 80
 Werlen, Egon 346
 Werner, Heinz 149
 Werner, Matthias 302
 Werner, Stephan (134)
 Wettach, Ralph 348
 Wetzel, Petra (134)
 Wicke, Lutz 205
 Widmer, Alex 231
 Widmer, Céline 23, 231
 Widmer, Thomas 11, 13, 14, 15, 16, 17, 58, 60, 64, 65, 66, 68, 69, 70, 71, (183), 7, 219, 231 286, 312, 342, 344, 348, 351, 362, 363, 364, 371, 392, 397, 431, 444, 445, 446, 469, 470, 471, 473, 486, 498, 501, 505, 509, 514, 515, 516, 517, 518, 520, 525
 Wiehler, Stephan (134)
 Wiesner, Reinhard 442
 Willi, Bruno 462, 463
 Williams, Kevin 79, 477, 480, 481
 Willimann, Ivo (187, 194)
 Willisegger, Jonas (231)
 Willsberger, Barara (134)
 Wimmer, Manfred (173)
 Winkelbrandt, Arndt 205
 Witte, Steffen (326)
 Witzke, Heinz Peter 109
 Woitech, Birgit (302)
 Wolf, Bernhard (448)
 Wollmann, Hellmut 15, 28, 371, 372, 373, 375, 377, 378, 380, 408, 411, 443, 470
 Wonik, Martin (448)
 Wortmann, Klaus (207)
 Wottawa, Heinrich 453, 470, 485
 Wren, Colin 430
 Wroblewski, Angela 129, 131, (134, 135), 302, 519, 522
 Wronski, Siddy 441
 Zarth, Michael (408)
 Zehnder, Tanja (145)
 Zellweger, Erich (231)
 Zenz, Michael (326)
 Zetter, John (427)
 Zeyen Bernasconi, Petra (346, 348)
 Ziegler, Alexandre (140)
 Zimmer, Jürgen 448
 Zimmer, Yelto (80)
 Zimmermann, Carmen 388, 389
 Zimmermann, Hubert (231)
 Zimmermann, Willi (368), 486, 487
 Zinöcker, Klaus 20, (292), 295, 296, 300, 302, 304, 313, 314, (318), 510, 525
 Zintl, Michaela 245, 246, 276, 516, 519
 Zucker, Judith 86
 Zürcher, Christoph 252
 Zürcher, Dieter 263, 276, 516, 519
 Zürcher, Markus 364
 Zweifel, Peter (350)
 Zweimüller, Josef (139, 145)

Sachverzeichnis

- 2Q, siehe Qualität und Qualifizierung
Aargau 59, 347, 365
Abfall 230, 231
Abfallpolitik 231
Abfallwirtschaft 208
ABM, siehe Arbeitsbeschaffungsmassnahme
Abrüstung 273
Absolventenbefragung 164, 180
Accountability 381
ADA, siehe Austrian Development Agency
Adoption, internationale 60
Afghanistan 252
Afrika 270, 279
AGEG, siehe Arbeitsgemeinschaft
entwicklungspolitischer Gutachter
Agency-Modell 375
AGES, siehe Österreichische Agentur für
Gesundheit und Ernährungssicherheit
AGEVAL, siehe Arbeitsgruppe
Gesetzesevaluation
Agglomerationspolitik 425, 428
AGJ, siehe Arbeitsgemeinschaft für Kinder- und
Jugendhilfe
AGMEMOD, siehe Agricultural Member States
Modelling
Agrarhandelspolitik 75
Agrarökonomie 102, 230
Agrarpolitik 21, 65, 215, 509, 511, 519, 523, 525
Agrarsozialpolitik 75, 77, 110
Agrarstrukturpolitik 110, 111, 112
Agrarumweltpolitik 75, 110, 111, 112
Agrarwissenschaft 111
Agricultural Member States Modelling
(AGMEMOD) 93
Agronomie 102
Agroscope Reckenholz-Tänikon (ART) 105, 106
AHS, siehe Allgemeinbildende Höhere Schule
AHV, siehe Alters- und
Hinterlassenenversicherung
Aidsprävention 346
AiF, siehe Arbeitsgemeinschaft industrieller
Forschungsvereinigungen
Akademie für Raumforschung und
Landesplanung 407, 410
Akademien der Wissenschaften Schweiz 311
Akkreditierung 45, 164, 166, 169, 189, 190, 195,
196, 198
Akkreditierungsrat 166, 178
Akkreditierungsverfahren 45
Aktion 8000 129
Aktionsforschung 18
Akupunktur 326
Akzeptanzforschung 453
Albanien 256, 259, 277
Alibi-Evaluation 501
Alkohol 349, 467
Allgemeinbildende Höhere Schule (AHS) 170,
171
Allgemeine Rahmenrichtlinien zur Gewährung
von Förderungen aus Bundesmitteln (ARR)
42
Altenheim 14
Altenhilfe 441, 470
Alter 144
Alternativmedizin 328
Alters- und Hinterlassenenversicherung (AHV)
139, 489, 523
AMA-Gesetz 88
AMM, siehe arbeitsmarktliche Massnahmen
AMS, siehe Arbeitsmarktservice
AMSG, siehe Bundesgesetz über den
Arbeitsmarktservice
Analyse, makroökonomische 124, 410, 412
Analyse, ökonomische 229
Analyse, robuste 269
Angebot, sozialpädagogisches 441
Ansatz, systemischer 418, 419
Anwendungsdefizit 34
AQA, siehe Österreichische
Qualitätssicherungsagentur
ARAMIS 58
Arbeiterkammer 90
Arbeitsbeschaffungsmassnahme (ABM) 124
Arbeitsförderung 117, 118, 119
Arbeitsförderungsgesetz 151

- Arbeitsgemeinschaft entwicklungspolitischer Gutachter (AGEG) 250, 275
 Arbeitsgemeinschaft Externe Evaluation von Schulen (ARGEV) 185, 195
 Arbeitsgemeinschaft für Kinder- und Jugendhilfe (AGJ) 471
 Arbeitsgemeinschaft industrieller Forschungsvereinigungen (AiF) 284
 Arbeitsgruppe Gesetzesevaluation (AGEVAL) 487, 515
 Arbeitskreis Evaluationen im Umweltbereich 204
 Arbeitskreis Hochschule 514
 Arbeitskreis Schule 514
 Arbeitskreis Soziale Dienstleistungen 471
 Arbeitskreis Sozialwissenschaftliche Arbeitsmarktforschung (SAMF) 126
 Arbeitskreis Stadt- und Regionalentwicklung 412
 Arbeitskreis Strukturpolitik 412, 514
 Arbeitslosenhilfe 120, 121
 Arbeitslosenversicherung 138, 139, 140, 141, 142, 150
 Arbeitslosenversicherungsgesetz (AVIG) 138, 143, 145, 146, 150
 Arbeitslosigkeit 60, 117, 118, 130, 131, 141, 142, 147, 148, 149, 234
 Arbeitsmarkt 28, 191, 442, 455, 512
 Arbeitsmarktförderungsgesetz 129, 131
 Arbeitsmarktforschung 30, 119, 122, 123, 470
 Arbeitsmarktliche Massnahme (AMM) 138, 139, 140, 141, 142, 143, 144, 145, 146, 151
 Arbeitsmarktpolitik 21, 28, 303, 404, 460, 466, 509, 511, 513, 519, 522, 523, 525
 Arbeitsmarktservice (AMS) 130, 131, 132, 135
 ARC, siehe Forschungszentrum Seibersdorf
 ARE, siehe Bundesamt für Raumentwicklung
 ARGEV, siehe Arbeitsgemeinschaft Externe Evaluation von Schulen
 ARR, siehe Allgemeine Rahmenrichtlinien zur Gewährung von Förderungen aus Bundesmitteln
 ART, siehe Agroscope Reckenholz-Tänikon
 Arzt/Ärztin 14, 15
 Ärztesgesetz 338
 Ärztekammer 331, 333, 338
 Arztpraxis 326
 Asien 267, 270
 Äthiopien 256
 Audit 15, 180, 277, 500
 Audit Commission 373
 Audit Society 15
 Audit, internes 266
 Auditisierung 15
 Aufsicht 182
 Aufsichtsdelegation 55
 Aufsichtsfunktion 60, 61, 63
 Aufsichtskommission 55
 Auftraggebende 11, 12, 20, 22, 82
 Aus- und Weiterbildung, berufliche 28, 256, 259
 Aussenministerium 255, 256, 257, 258, 260, 261
 Aussenpolitik 19, 516
 Austrian Agency for Quality Assurance, siehe Österreichische Qualitätssicherungsagentur
 Austrian Development Agency (ADA) 255, 256, 258, 260, 261, 275
 Austrian Research Center 299, 301
 Autobahntunnel Vue-des-Alpes 428
 Avenir Suisse 107, 231
 AVIG, siehe Arbeitslosenversicherungsgesetz
 AWI, siehe Bundesanstalt für Agrarwirtschaft
 BA, siehe Bundesanstalt für Arbeit
 Baar 141
 BABF, siehe Bundesanstalt für Bergbauernfragen
 Bachelorstudiengang 163, 195, 451
 Baden-Württemberg (Bundesland) 166, 290, 375, 376
 BAFU, siehe Bundesamt für Umwelt
 BAG, siehe Bundesamt für Gesundheit
 Balanced Scorecard (BSC) 327, 337, 389
 Barcelona-Ziele 295
 Basel 11, 463
 Basel-Landschaft (Kanton) 365
 Basel-Stadt (Kanton) 12, 140, 228, 463
 Basisstufe 185
 BASPO, siehe Bundesamt für Sport
 Bauministerium 206
 Bauordnung 214
 BAW, siehe Bundesamt für Wasserwirtschaft
 Bayern (Bundesland) 166, 290, 376
 BBG, siehe Berufsbildungsgesetz
 BBI, siehe Bundesblatt
 BBT, siehe Bundesamt für Berufsbildung und Technologie
 Befragung 139, 389, 390, 391, 408, 410, 418, 463
 Begleitforschung 409, 410, 411, 443, 444, 469
 Belgien 251
 Benchmarking 29, 125, 131, 142, 152, 187, 337, 371, 385, 391, 416

- Beratungsbüro 429
Beratungsinstitut 388
Beratungsunternehmen 82, 83, 111, 145, 230, 233, 267, 335, 385, 409, 417, 452
Bericht 53, 79, 80, 82, 86, 92, 95, 97, 112, 164, 251, 258, 259, 287, 288, 292, 303, 339, 348, 387, 437, 479, 499, 503
Berichterstattung 383
Berichtsintervall 35
Berichtspflicht 30, 50
Berichtswesen 371, 388
Berlin (Bundesland) 375, 376
Bern (Kanton) 141, 197, 356
Bern (Stadt) 59, 354, 366, 37
Bertelsmann Stiftung 36, 122
Berufsbildung 60, 96, 186, 188, 194, 277, 489, 511
Berufsbildung, höhere 188, 189
Berufsbildungsgesetz (BBG) 186, 194
Beschäftigung 42, 140, 150
Best Practice 370, 394
Beteiligte & Betroffene 16, 70, 220, 321, 479, 498, 514, 520
Betriebswirtschaft 237, 367, 395, 445, 450, 455
Bewertung, ex-ante, siehe auch ex-ante-Evaluation 79
Bewertung, ex-post, siehe auch ex-post-Evaluation 79
BfArM, siehe Bundesinstitut für Arzneimittel und Medizinprodukte
BFE, siehe Bundesamt für Energie
BFS, siehe Bundesamt für Statistik
BFW, siehe Bundesforschungs- und Ausbildungszentrum für Wald, Naturgefahren und Landschaft
BGA, siehe Bundesgesundheitsamt
BHG, siehe Bundeshaushaltsgesetz
BHO, siehe Bundeshaushaltsordnung
Bhutan 256
BIBB, siehe Bundesinstitut für Berufsbildung
BIGA, siehe Bundesamt für Industrie, Gewerbe und Arbeit
Bildung 13, 15, 19, 21, 48, 140, 144, 256, 259, 273, 310, 349, 442, 443, 455, 471, 513
Bildung, berufliche 19, 157, 514
Bildung, frühe 19
Bildung, tertiäre 48, 49
Bildungsabschluss 182
Bildungsbericht 160, 174
Bildungsberichterstattung 160
Bildungsdirektorenkonferenz Zentralschweiz (BKZ) 185
Bildungsforschung 190, 191, 444
Bildungsministerium 173
Bildungsmonitoring 190, 191, 524
Bildungsplanung 190, 191
Bildungspolitik 137, 138, 171, 173, 191, 295
Bildungsrat 443
Bildungssektor 153
Bildungsstandard 159, 160, 161, 173, 175, 185, 194, 524
Bildungsstatistik 190, 191
Bildungswesen, siehe Bildung
Biodiversität 97, 103, 104, 106, 229
BIQG, siehe Bundesinstitut für Qualität im Gesundheitswesen
Birdlife Österreich 92
BJ, siehe Bundesamt für Justiz
BK, siehe Bundeskanzlei
BKA, siehe Bundeskanzleramt
BKZ, siehe Bildungsdirektorenkonferenz Zentralschweiz
Block 286
Blockade-Evaluation 501
Blumau (Steiermark) 173
BLW, siehe Bundesamt für Landwirtschaft
BMBF, siehe Bundesministerium für Bildung und Forschung
BMBWK-Programm proVition 215
BMEIA, siehe Bundesministerium für europäische und internationale Angelegenheiten
BMELV, siehe Bundesministerium für Ernährung, Landwirtschaft und Verbraucherschutz
BMG, siehe Bundesministerium für Gesundheit
BMLFUW, siehe Bundesministerium für Land- und Forstwirtschaft, Umwelt und Wasserwirtschaft
BMVIT, siehe Bundesministerium für Verkehr, Innovation und Technologie
BMW A, siehe Bundesministerium für Wirtschaft und Arbeit
BMZ, siehe Bundesministerium für wirtschaftliche Entwicklung und Zusammenarbeit
BÖL, siehe Bundesprogramm Ökologischer Landbau
Bolivien 270

- Bologna-Deklaration 195
 Bologna-Prozess 166, 178, 181
 Bologna-Reform 64
 Bonn 197
 Botschaft (des Bundesrats) 57
 Botschaftsleitfaden 57
 Brandenburg (Bundesland) 406, 435
 Bregenz 383
 Bremen (Bundesland) 166
 BRH, siehe Bundesrechnungshof
 BSC, siehe Balanced Scorecard
 BSV, Bundesamt für Sozialversicherungen
 Budgethilfe 251, 262, 263, 266, 269, 270
 Bundesamt für Bauwesen und Raumordnung 407, 410
 Bundesamt für Berufsbildung und Technologie (BBT) 144
 Bundesamt für Energie (BFE) 12, 229, 231, 232, 233, 238, 487
 Bundesamt für Gesundheit (BAG) 12, 341, 344, 345, 346, 348, 350, 352, 458, 460, 466, 467, 487, 489, 490
 Bundesamt für Industrie, Gewerbe und Arbeit (BIGA) 138, 139
 Bundesamt für Justiz (BJ) 12, 56, 58, 458, 467, 489
 Bundesamt für Kommunikation 487
 Bundesamt für Landwirtschaft (BLW) 99, 111, 102, 103, 104, 106, 107, 108, 487
 Bundesamt für Polizei 487
 Bundesamt für Raumentwicklung (ARE) 227, 229, 424, 428, 429
 Bundesamt für Sozialversicherungen (BSV) 144, 351, 466, 467, 487
 Bundesamt für Sport (BASPO) 346
 Bundesamt für Statistik (BFS) 106, 107, 344
 Bundesamt für Umwelt (BAFU) 12, 103, 229, 231, 234, 238, 241
 Bundesamt für Umwelt, Wald und Landschaft (BUWAL) 103
 Bundesamt für Verkehr, Bau und Stadtentwicklung 407
 Bundesamt für Veterinärwesen (BVET) 103
 Bundesamt für Wasserwirtschaft (BAW) 92
 Bundesamt für Wirtschaft und Arbeit (BWA) 139
 Bundesamt für Wirtschaft und Ausfuhrkontrolle 408
 Bundesanstalt für Agrarwirtschaft (AWI) 91, 97, 112
 Bundesanstalt für Arbeit (BA) 28, 30, 118, 119, 120, 121, 123, 125, 151, 375, 393
 Bundesanstalt für Bergbauernfragen (BABF) 91
 Bundesblatt (BBI) 56
 Bundesforschungs- und Ausbildungszentrum für Wald, Naturgefahren und Landschaft (BFW) 92, 97
 Bundesforschungsanstalt für Landwirtschaft (FAL) 78, 82
 Bundesforschungsministerium 286
 Bundesgericht 503
 Bundesgesetz für die Gewährung von Bürgschaften im Berggebiet 422
 Bundesgesetz gegen die Schwarzarbeit 57
 Bundesgesetz über den Arbeitsmarktservice (AMSG) 150
 Bundesgesetz zur Verbesserung der Wohnverhältnisse im Berggebiet 422
 Bundesgesundheitsagentur 334
 Bundesgesundheitsamt (BGA) 28
 Bundesgesundheitskommission 334
 Bundesgesundheitsministerium 325
 Bundeshaushalt 29
 Bundeshaushaltsgesetz (BHG) 41, 42, 43
 Bundeshaushaltsordnung (BHO) 77, 78, 85, 500
 Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM) 29, 323
 Bundesinstitut für Berufsbildung (BIBB) 28, 121, 122
 Bundesinstitut für gesundheitlichen Verbraucherschutz und Veterinärmedizin 29
 Bundesinstitut für Qualität im Gesundheitswesen (BIQG) 334, 340
 Bundeskanzlei (BK) 57, 58, 487
 Bundeskanzler 50, 413
 Bundeskanzleramt (BKA) 37, 38, 241, 385, 413, 416, 417
 Bundesminister/in 50
 Bundesministerium für Arbeit und Soziales 30
 Bundesministerium für Bildung und Forschung (BMBF) 157, 159, 287, 290, 291
 Bundesministerium für Ernährung, Landwirtschaft und Verbraucherschutz (BMELV) 77, 111
 Bundesministerium für europäische und internationale Angelegenheiten (BMEIA) 255, 275
 Bundesministerium für Familie, Senioren, Frauen und Jugend 472

- Bundesministerium für Finanzen 255, 385
Bundesministerium für Gesundheit (BMG) 29, 30
Bundesministerium für Land- und
Forstwirtschaft, Umwelt und
Wasserwirtschaft (BMLFUW) 88, 97, 112, 214,
223, 224, 225
Bundesministerium für Unterricht und Kunst
171, 194
Bundesministerium für Verkehr, Innovation und
Technologie (BMVIT) 214, 223, 416, 417, 498
Bundesministerium für Wirtschaft und Arbeit
(BMWA) 124, 130, 135, 214, 224, 225, 300, 417
Bundesministerium für wirtschaftliche
Entwicklung und Zusammenarbeit (BMZ)
245, 246, 247, 248, 250, 251, 252, 253, 259, 275,
278
Bundesministerium für Wissenschaft und
Forschung 48
Bundesprogramm Ökologischer Landbau (BÖL)
78
Bundesrat 56, 57, 58, 78, 104, 106, 108, 495, 496,
503
Bundesrechnungshof (BRH) 29, 376, 393, 498,
503, 504
Bundesrecht 37
Bundesregierung 27, 35, 36, 40, 41, 47, 49, 51, 78,
111, 121, 122, 215, 298, 301, 302, 500
Bundesstatistik 50
Bundesstatistikgesetz 344
Bundestag 291
Bundesverfassung 55, 57, 99, 106, 109, 110, 112,
188, 332, 340, 343, 353, 354, 413, 424, 437, 458,
477, 484, 496
Bundesverfassung, schweizerische (BV) 14, 52,
54, 55, 65, 101, 102, 182
Bundes-Verfassungsgesetz (B-VG) 40, 44
Bundesvergabegesetz (BVerG) 50, 451
Bundesversammlung 56, 106
Bundesverwaltung 27, 38, 41, 56, 60, 67, 106, 233,
307, 385, 386, 486, 487, 489, 497, 503, 515
Bundeswirtschaftsministerium 408
Bundeszentrale für gesundheitliche Aufklärung
329, 354
Bund-Länder-Gemeinschaftsaufgabe zur
Verbesserung der Agrarstruktur und
Küstenschutz (GAK) 76, 78, 82
Bündnis für Arbeit 152
Burgenland (Bundesland) 89, 336, 383
Burkina Faso 256, 259, 270, 277
Büro DI Tischler 415
Büro für soziale und kulturelle Planung (SCP) 66
Büro für Technikfolgenabschätzung (TAB) 29
Büro Vatter 231
Bürokratieabbau 36, 37
Bürokratiekosten 33, 35, 37
Bürokratisierung 70, 378, 382
BUWAL, siehe Bundesamt für Umwelt, Wald
und Landschaft
BV, siehe Bundesverfassung, Schweizerische
BVerG, siehe Bundesvergabegesetz
BVET, siehe Bundesamt für Veterinärwesen
B-VG, siehe Bundes-Verfassungsgesetz
BWA, siehe Bundesamt für Wirtschaft und Arbeit
C.E.A.T. 429
Campbell libraries 350
Caritas Schweiz 263
CCE, siehe Kompetenzzentrum für Evaluation
Centre for Sustainable Management (CSM) 206
Centrum für Evaluation (CEval) 205, 247, 250,
275, 278
Centrum für Hochschulentwicklung (CHE) 195,
290
CEPP, siehe Commission externe d'évaluation
des politiques publiques
CEST, siehe Zentrum für Wissenschafts- und
Technologiestudien
CEval, siehe Centrum für Evaluation
Chancengleichheit 419, 464
CHE, siehe Centrum für Hochschulentwicklung
Checkpoint EVA 216
CHE-Hochschulranking 179
Christian Doppler Forschungsgesellschaft 301
Clearinghouse 292
Cochrane Review 349
Cochrane-Institute 324
Commission externe d'évaluation des politiques
publiques (CEPP) 68, 141
Consulting-Unternehmen 218, 278
Conterganskandal 323
Controlling 28, 58, 121, 130, 131, 136, 208, 229,
230, 232, 245, 265, 266, 309, 341, 363, 371, 373,
385, 386, 389, 390, 391, 430, 431, 447, 454, 500
CPB, siehe Nationales Büro für
Wirtschaftsforschung
CSM, siehe Centre for Sustainable Management
DAC, siehe Development Assistance Committee

- DACH-Gruppe 251
 Dachverband der Schweizer Lehrerinnen und Lehrer (LCH) 184
 Dänemark 149, 250, 410
 DANIDA, siehe Dänische Internationale Entwicklungsagentur
 Dänische Internationale Entwicklungsagentur (DANIDA) 27
 Datawarehouse (DWH) 130, 133, 134, 135, 151, 452
 Datenbank 287, 452, 458, 459, 466
 Datenzugang 83, 127, 150, 303, 317, 427
 Dauersubvention 78
 DED, siehe Deutscher Entwicklungsdienst
 DeGEval, siehe (Deutsche) Gesellschaft für Evaluation
 DeGEval-Standard 11, 82, 209, 219, 251, 253, 298, 329, 332, 453
 DELBI, siehe Deutsches Instrument zur methodischen Leitlinienbewertung
 Demokratie, direkte 16, 64, 101, 361, 368, 369
 Departement 57, 60, 61
 DESI-Studie 194
 Design, randomisiertes 347
 Deutsche Bundesstiftung Umwelt 205
 Deutsche Forschungsgemeinschaft (DFG) 167, 283, 285, 287, 289, 314, 377
 Deutsche Gesellschaft für Evaluation (DeGEval), siehe Gesellschaft für Evaluation
 Deutscher Bundestag 29, 67, 122, 151
 Deutscher Entwicklungsdienst (DED) 247
 Deutsches Institut für Entwicklungspolitik (DIE) 250, 276
 Deutsches Instrument zur methodischen Leitlinienbewertung (DELBI) 327
 Deutsches Jugendinstitut (DJI) 446
 Development Assistance Committee (DAC) 245, 247, 251, 253, 261, 265, 270, 271, 272, 274
 DEZA, siehe Direktion für Entwicklung und Zusammenarbeit
 Dezentralisierung 16, 64, 151, 263, 377
 DFG, siehe Deutsche Forschungsgemeinschaft
 DFG-Ranking 287, 289
 DFID, UK Department for international Development
 DG Agri, siehe Generaldirektion für Landwirtschaft und ländliche Entwicklung
 DIE, siehe Deutsches Institut für Entwicklungspolitik
 Direktion für Entwicklung und Zusammenarbeit DEZA 12, 263, 264, 265, 266, 269, 271, 272, 275, 470, 487, 489
 Direktzahlung 75, 100, 101, 102, 105, 107, 108, 226
 DJI, siehe Deutsches Jugendinstitut
 Doha-Runde 93
 Dokumentenanalyse 459, 463, 466
 Doppelblindstudie 324
 DORE 308, 460, 466, 467
 Dortmund 524
 Drogen 348, 349, 455, 467, 489
 Durchführbarkeit 209, 222, 269, 276, 499
 DWH, siehe Datawarehouse

 E + C, siehe Sektion Evaluation und Controlling
 EAGFL, Europäischer Ausrichtungs- und Garantiefonds der Landwirtschaft
 EAK, siehe Eidgenössische Arzneimittelkommission
 EASY-ECO, siehe Evaluation of Sustainability 222
 Econcept 230
 E-Control GmbH 224
 Ecoplan 230
 EDA, siehe Eidgenössisches Department für auswärtige Angelegenheiten
 EDK-Ost, siehe Erziehungsdirektorenkonferenz der Ostschweizer Kantone und des Fürstentums Liechtenstein 185
 EEG, siehe Erneuerbare-Energien-Gesetz
 EES, siehe Europäische Evaluationsgesellschaft
 Effektivität 54, 103, 104, 158, 263, 268, 275, 292, 312, 333, 334, 377, 381, 450
 Effektstärkenschatzung 324
 Effizienz 54, 66, 78, 103, 104, 137, 139, 140, 210, 230, 231, 268, 275, 292, 312, 333, 365, 377, 381, 383, 450
 EFK, siehe Eidgenössische Finanzkontrolle
 EFQM, siehe European Foundation for Quality Management
 EFRE, siehe Europäischer Fonds für Regionale Entwicklung
 EGZ, siehe Eingliederungszuschüsse
 Eidgenössische Arzneimittelkommission EAK 343
 Eidgenössische Finanzkontrolle (EFK) 12, 56, 58, 66, 231, 234, 489, 515
 Eidgenössische Kommission für Allgemeine Leistungen (ELK) 343

- Eidgenössischen Kommission für
Grundsatzfragen der Krankenversicherung
(GLK) 343
- Eidgenössische Technische Hochschule (ETH)
230, 305, 308, 429
- Eidgenössischer Datenschutz- und
Öffentlichkeitsbeauftragter 57
- Eidgenössisches Büro für die Gleichstellung von
Frau und Mann 490
- Eidgenössisches Department des Innern (EDI)
344
- Eidgenössisches Department für auswärtige
Angelegenheiten (EDA) 264
- Eidgenössisches Justiz- und Polizeidepartement
(EJPD) 487
- Eidgenössisches Volkswirtschaftsdepartement
(EVD) 57, 99, 111, 142, 190
- Einbettung, institutionelle 21, 23, 38, 46, 52, 64,
65, 515
- Eingliederungszuschüsse (EGZ) 124
- Einkommenspolitik 75
- Einkommenssicherung 109, 110
- Einzelschule 158, 159, 161, 172, 173, 194, 198
- Eisenbahntunnel Vereine 428
- EJPD, siehe Eidgenössisches Justiz- und
Polizeidepartement
- ELK, siehe Eidgenössische Kommission für
Allgemeine Leistungen
- EMAS, siehe Environmental management und
audit scheme
- Emissionsstandard 213, 237
- Emissionszertifikat 214
- Empfehlung 219, 258, 265, 272, 302, 420, 479, 494,
497, 499, 503, 504
- Empowerment 248, 270
- Energie 106, 234, 240, 256, 306, 489
- Energie Schweiz 227, 228, 230, 232, 233, 234, 240
- Energie, erneuerbare 204, 206, 213, 227, 228, 232,
234, 237, 238
- Energie2000, siehe auch Energie Schweiz 227,
228, 231, 489
- Energieeffizienz 204, 206, 228, 237, 238
- Energieforschung 58, 287
- Energiepolitik 21, 201, 231, 232, 489, 511, 519, 523
- Energieversorgung, konventionelle 206, 229
- Entwicklung, ländliche 88, 112
- Entwicklung, räumliche 424, 433, 434
- Entwicklungsfonds der Europäischen Union 245
- Entwicklungshilfe 15, 58
- Entwicklungsparadigma 330
- Entwicklungspolitik 27, 521
- Entwicklungspolitik, ländliche 85
- Entwicklungsprogramm (EPLR) 91
- Entwicklungszusammenarbeit 21, 243, 303, 306,
487, 510, 516, 519, 521
- Entwicklungszusammenarbeitsgesetz (EZA-
Gesetz) 255, 256
- Environmental management and audit scheme
(EMAS) 208, 209, 214
- ENWISS, siehe Evaluationsnetzwerk
- EPLR, siehe Entwicklungsprogramm
- EPRC, siehe European Policy Research Center
- EQUAL 133, 454, 519
- Erasm 231
- Erfolg 65
- Erfolgskontrolle 18, 77, 78, 110, 111, 141, 371, 373,
406, 411, 412, 503
- Ergebnisevaluation 327, 328, 350
- Ergebnisorientierung 175, 269
- Ergebnisqualität 168, 333, 334, 335, 336, 337
- Erhebungsbogen 452
- Erhebungsinstrument 14, 349, 482
- Ernährung 349
- Erneuerbare-Energien-Gesetz (EEG) 75
- Erwachsenenbezogene Soziale Hilfe 441
- Erziehung 157, 471
- Erziehungsdirektorenkonferenz der Ostschweizer
Kantone und des Fürstentums Liechtenstein
(EDK-Ost) 185
- Erziehungswissenschaft 158
- ESF, siehe European Science Foundation 120,
126, 133
- ETH, siehe Eidgenössische Technische
Hochschule
- Ethik 344
- Ethikkommission UREK 344
- EU, siehe Europäische Union
- EU-Beitritt 47
- EU-KOM, siehe Europäische Kommission
- EU-Kommission 49, 79, 84
- EU-Mitgliedstaat 46, 64, 79, 88
- EU-Rahmenprogramm 45
- EU-Ratsverordnung 46
- EU-Regionalpolitik 414, 417
- Europäische Evaluationsgesellschaft (EES) 351
- Europäische Kommission (EU-KOM) 34, 38, 65,
80, 82, 85, 91, 94, 97, 126, 246, 323, 418, 477,
497, 523

- Europäische Union (EU) 34, 49, 51, 64, 65, 79, 86, 91, 109, 110, 111, 112, 113, 130, 131, 133, 144, 152, 153, 211, 215, 251, 271, 313, 323, 382, 405, 414, 416, 420, 422, 423, 431, 433, 435, 436, 437, 438, 455, 467, 477, 490, 509, 525
- Europäischer Ausrichtungs- und Garantiefonds der Landwirtschaft (EAGFL) 76
- Europäischer Entwicklungsfonds 255
- Europäischer Fonds für Regionale Entwicklung (EFRE) 405, 409, 417
- Europäischer Gerichtshof 179
- Europäischer Sozialfonds 152
- European Association for Quality Assurance in Higher Education 168, 178
- European Foundation for Quality Management (EFQM) 189
- European Policy Research Center (EPRC) 420
- European Science Foundation (ESF) 126, 133
- EUROSTAT 289
- EU-Strukturfonds 46, 49, 215, 406, 409, 414, 416, 430, 433, 437
- EU-Verordnung 77
- evalag, siehe Evaluationsagentur Baden-Württemberg 166
- evaluanda 231
- Evaluation capacity 521
- Evaluation community 351, 417, 421, 430
- Evaluation der Maturitätsreform (EVAMAR) 187, 194
- Evaluation influence 480
- Evaluation of Sustainability, siehe EASY-ECO 222
- Evaluation, begleitende 218, 219, 221, 239, 418, 419, 420, 421, 436, 437, 438
- Evaluation, externe 28, 164, 165, 172, 185, 260, 265, 266, 268, 269, 270, 271, 275, 276, 277, 285, 371, 373, 375, 376, 379, 386, 387, 388, 390, 391, 393, 395, 417, 425, 445, 446, 447
- Evaluation, formative 302, 337, 350, 364, 385, 389, 390, 431, 447, 505, 518, 520
- Evaluation, institutionelle 164, 286
- Evaluation, interne 164, 165, 172, 183, 234, 247, 260, 268, 275, 276, 277, 285, 372, 375, 377, 379, 393, 446
- Evaluation, modellgestützte 168
- Evaluation, ökonomische 345, 350
- Evaluation, partizipative 252
- Evaluation, strategische 258, 260, 261, 276, 277, 278
- Evaluation, summative 302, 364, 385, 447, 518, 520
- Evaluation, systemische 317, 425
- Evaluation, theoriegestützte 168
- Evaluation, unabhängige 222, 245, 246, 248, 265, 266, 269, 272, 277
- Evaluationsabteilung 50
- Evaluationsagentur 66
- Evaluationsagentur Baden-Württemberg (evalag) 165
- Evaluationsauftrag 515
- Evaluationsbericht 15, 20, 79, 82, 83, 216, 217, 232
- Evaluationbestimmung 42
- Evaluationsbudget 261, 489, 490, 518
- Evaluationsdefizit 373, 390
- Evaluationsdesign 126, 134, 147, 447, 452, 512
- Evaluationsdruck 511
- Evaluationsergebnis 79, 86, 215, 266, 268, 275, 279, 339, 353, 364, 366, 427, 437, 447, 477, 478, 479, 482, 484, 493, 504
- Evaluationsfachlichkeit 17
- Evaluationsforschung 158, 171, 355, 373, 453, 454, 456, 458, 471, 486
- Evaluationsfrage 268
- Evaluationsfunktion 20, 21, 22, 28, 38, 48, 49, 52, 67
- Evaluationsgegenstand 66
- Evaluationsgesellschaft 17, 19, 20
- Evaluationsinstanz 66, 262
- Evaluationskapazität 59, 62, 279, 526
- Evaluationsklausel 56, 66
- Evaluationskommission 59, 60
- Evaluationskompetenz 59, 66, 274, 504, 521
- Evaluationskonzept 269, 298, 322
- Evaluationskosten 81, 517
- Evaluationskriterium 268
- Evaluationskultur 14, 42, 60, 65, 182, 192, 303, 313, 331, 332, 339, 340, 352, 353, 355, 431, 453, 484, 485, 497, 505
- Evaluationslücke 373
- Evaluationsmethode 207, 212, 18, 221, 223
- Evaluationsmodell, ökonomisches 205
- Evaluationsmüdigkeit 70, 304
- Evaluationsnetzwerk 20, 84, 165, 166
- Evaluationsorgan 61
- Evaluationspflicht 30, 56, 57, 58, 409, 433, 437, 523, 527
- Evaluationspraxis 46
- Evaluationsrhetorik 384, 398

- Evaluationsroman 13
Evaluationsskepsis 388, 393, 395
Evaluationsstandard, siehe auch SEVAL-Standard, DeGEval-Standard 14, 17, 161, 197, 209, 210, 211, 222, 246, 261, 312, 321, 335, 338, 340, 344, 354, 412, 418, 431, 444, 445, 452, 455, 471, 474, 484, 485, 498, 505, 512
Evaluationsstrategie 58
Evaluationsstudie, externe 174
Evaluationssynthese 20
Evaluationstheorie 15, 444
Evaluationstradition 27
Evaluationstyp 53
Evaluationsverfahren 27, 31, 38, 44, 48
Evaluationsverpflichtung 50, 65, 66, 526
Evaluationsverwendung 480, 486, 494, 497
Evaluationszweck 479
Evaluierbarkeit 434
Evaluierung, siehe Evaluation
Evaluierungsplattform, siehe auch Plattform
Forschungs- und Technologieevaluierung 47
Evaluitis 15, 70, 197
EVAMAR, siehe Evaluation der Mauritätsreform
EVD, siehe Eidgenössisches Volkswirtschaftsdepartment
Evidence-based 268, 446, 456
Evidence-based medicine 326, 330, 342, 356
Evidenztriangulation 330
ex-ante-Evaluation 92, 96, 97, 105, 121, 218, 229, 247, 248, 297, 302, 309, 314, 364, 377, 410, 417, 420, 424, 425, 433, 436, 437, 438
Existenzsicherung 140
Experiment 330, 443, 444
experimentell 151, 152, 284, 324, 325, 328, 329, 330, 412, 433, 444, 514
Expertenpanel 45
Export-Basis-Theorie 405
Exportsubvention 75
ex-post-Evaluation 95, 218, 228, 234, 240, 246, 247, 249, 252, 265, 276, 278, 290, 297, 300, 303, 304, 309, 314, 410, 424, 425, 426, 436, 437, 438
Exzellenzinitiative 285, 286
EZA-Gesetz, siehe Entwicklungszusammenarbeitsgesetz
Fachhochschule 45, 48, 49, 67, 141, 146, 163, 177, 178, 189, 190, 195, 307, 308, 312, 339, 432, 452, 453, 455, 460, 470, 473
Fachhochschule Nordwestschweiz (FHNW) 11, 12
Fachhochschulkonferenz 48
Fachhochschulrat 45, 49, 178
Fachhochschul-Studiengesetz (FHStG) 45, 177, 195
Fachkommission Qualitätsmanagement (FQM) 189, 196
Fairness 209, 222, 321, 499
FAL, siehe Bundesforschungsanstalt für Landwirtschaft
Fallstudie 83, 103, 428
Familienhilfe 455
Familienpolitik 191
FAT, siehe Forschungsanstalt für Agrarwirtschaft und Landtechnik Tänikon
FbW, siehe Förderung der beruflichen Weiterbildung
FDZ, siehe Forschungsdatenzentrum
FE EZ, siehe Fortbildungsprogramm Evaluation in der Entwicklungszusammenarbeit
FER, siehe Förderungsaktion Eigenständige Regionalentwicklung
FFF, siehe Forschungsförderfonds für die gewerbliche Wirtschaft
FFG, siehe Forschungsförderungsgesellschaft
FFG-G, siehe Gesetz zur Errichtung der Österreichischen Forschungsförderungsgesellschaft
FGÖ, siehe Fonds gesundes Österreich 334
FhG, siehe Fraunhofer-Gesellschaft
FHNW, siehe Fachhochschule Nordwestschweiz
FHStG, siehe Fachhochschul-Studiengesetz
Finanzaufsicht 66
Finanzausgleich 62, 363, 392, 404, 423, 424
Finanzkontrolle 260
Finanzkontrollgesetz 56
Finanzministerium 384, 386
Finanzpolitik 137, 454
FLAG, siehe Führen mit Leistungsauftrag und Globalbudget
Flexibilisierungsklausel 386
FMH, siehe Verband der Ärzt/innen
Focus Group, siehe Fokusgruppe
Föderalismus 16, 44, 64, 76, 138, 157, 182, 206, 236, 307, 313, 331, 332, 354, 361, 362, 368, 369, 371, 381, 392, 394, 509, 525
FODOS, siehe Forschungsdokumentation Sozialwissenschaften

- FOG, siehe Forschungsorganisationsgesetz 44
 Fokusgruppe 296, 349, 350
 Folgenabschätzung, siehe
 Gesetzesfolgenabschätzung (GFA)
 Fonds gesundes Österreich (FGÖ) 334
 Fonds zur Förderung der wissenschaftlichen
 Forschung (FWF) 45, 296, 299, 300, 304, 314
 Förderorientiertes Qualitätsevaluations-System
 (FQS) 184
 Förderung der beruflichen Weiterbildung (FbW)
 124
 Förderungsaktion Eigenständige
 Regionalentwicklung (FER) 416, 418
 FORS, siehe Schweizer Stiftung für
 sozialwissenschaftliche Forschung 308
 Forschung 21, 44, 45, 46, 47, 48, 49, 68, 157, 164,
 167, 289, 341, 344, 512
 Forschung und Entwicklung (FuE) 283, 384
 Forschung und Technologie 510, 511
 Forschungs- und Technologieförderungsgesetz
 (FTF-G) 44
 Forschungs-, Technologie- und
 Innovationspolitik (FTI) 295, 315
 Forschungsanstalt für Agrarwirtschaft und
 Landtechnik Tänikon (FAT) 105
 Forschungsdatenzentrum (FDZ) 127
 Forschungsdokumentation Sozialwissenschaften
 (FODOS) 17
 Forschungseinrichtung 295, 298, 315, 317, 371,
 417, 511
 Forschungsförderfonds für die gewerbliche
 Wirtschaft (FFF) 45, 296, 299, 300, 304
 Forschungsförderungsgesellschaft (FFG) 45, 296,
 299, 300
 Forschungsgesetz 307, 309
 Forschungsgesetz (FG) 343
 Forschungsinstitut 218, 298, 410, 472
 Forschungsinstitut zur Zukunft der Arbeit (IZA)
 122
 Forschungskonsortium 124
 Forschungsleistung 285
 Forschungsorganisationsgesetz (FOG) 44
 Forschungsparadigma 329
 Forschungspolitik 51, 236, 281
 Forschungszentrum Seibersdorf (ARC) 452
 Fortbildungsprogramm Evaluation in der
 Entwicklungszusammenarbeit (FEEZ) 250,
 275
 FQM, siehe Fachkommission
 Qualitätsmanagement
 FQS, siehe Förderorientiertes
 Qualitätsevaluations-System
 Fragebogen 338, 464
 Frankfurt am Main 29
 Fraunhofer-Gesellschaft (FhG) 167, 284, 288, 289
 Fraunhofer-Institut für Systemtechnik und
 Innovationsforschung 287
 Fraunhofer-Institut für Systemtechnik und
 Innovationsforschung (ISI) 290
 Freiburg 324
 Freiburg (Kanton) 141
 Freie Universität Berlin (FU Berlin) 252
 Freizeit und Sport 346
 Fremdevaluation 447, 504
 Friedensförderung 251, 273
 FTE-Richtlinien 297
 FTF-G, siehe Forschungs- und
 Technologieförderungsgesetz
 FTI, siehe Forschungs-, Technologie- und
 Innovationspolitik
 FuE, siehe Forschung und Entwicklung
 Führen mit Leistungsauftrag und Globalbudget
 (FLAG) 398
 Funktionalreform 377
 Funktionendiagramm 31
 Fürstentum Liechtenstein 463
 FWF, siehe Fonds zur Förderung der
 wissenschaftlichen Forschung
 GAK, siehe Bund-Länder-Gemeinschaftsaufgabe
 zur Verbesserung der Agrarstruktur und
 Küstenschutz
 Ganztageesschule 193
 Ganztageesschule 158
 GAO, siehe United States Government
 Accountability Office
 GAP, siehe Gemeinsame Agrarpolitik
 GATT, siehe General Agreement on Tariffs and
 Trade
 Gebarungsüberprüfung 41, 50
 Gebietsreform 361, 363, 365, 377
 Geisteswissenschaft 284, 513
 Gemeindebund 413
 Gemeindeexekutive 395
 Gemeindefusion 362, 363, 392
 Gemeindeparlament 395

- Gemeinsame Agrarpolitik (GAP) 75, 76, 77, 79, 83, 88, 93, 100, 109, 112
- Gemeinsame Geschäftsordnung der Bundesministerien (GGO) 32, 33
- Gemeinsamer Unterricht behinderter und nichtbehinderter Schüler 172
- Gemeinschaftliches Förderkonzept (GFK) 80
- Gemeinschaftsaufgabe Verbesserung der regionalen Wirtschaftsstruktur (GRW) 405, 406, 408, 409
- Gemeinschaftsevaluation 248, 250, 251, 259, 262, 274, 276, 278, 279
- Genauigkeit 209, 220, 222, 499
- General Agreement on Tariffs and Trade (GATT) 92, 99
- Generaldirektion 34
- Generaldirektion für Landwirtschaft und ländliche Entwicklung (DG Agri) 79, 82, 84
- Generaldirektorat Forschung der Europäischen Kommission 292
- Generalsekretariat 34, 35
- Genf 59, 68, 354
- Genf (Kanton) 140, 141
- Genomforschung 302
- GEPRIS 289
- Gesamtschule 158, 171, 172, 193
- Geschäftsprüfungskommission 264, 309, 312, 495, 503
- Gesellschaft für Evaluation (DeGEval) 11, 12, 17, 23, 30, 126, 135, 153, 167, 197, 204, 206, 210, 222, 239, 250, 253, 278, 286, 298, 303, 313, 321, 335, 354, 356, 412, 420, 455, 471, 479, 482, 483, 484, 498, 514, 515
- Gesellschaft für Technische Zusammenarbeit (GTZ) 246, 247, 253, 259, 279
- Gesellschaft Sozialwissenschaftlicher Infrastruktureinrichtungen e.V. (GESIS) 17, 18
- Gesetz über die Gemeinschaftsaufgabe „Verbesserung der regionalen Wirtschaftsstruktur“ (WiStruktG) 404
- Gesetz zur Errichtung der Österreichischen Forschungsförderungsgesellschaft (FFG-G) 44
- Gesetz zur Verbesserung der Zusammenarbeit von Arbeitsämtern 119
- Gesetzesevaluation 52, 53, 62, 63
- Gesetzesfolge 41, 42
- Gesetzesfolgenabschätzung (GFA) 27, 31, 32, 34, 35, 38, 42, 69, 77, 111, 218
- Gesetzgebung 52, 53, 62, 66
- Gesetzgebung, evidenzgestützte 63
- Gesetzgebungsmethodik 62
- GESIS, siehe Gesellschaft Sozialwissenschaftlicher Infrastruktureinrichtungen e.V.
- Gesundheit 14, 15, 21, 29, 66, 144, 308, 319, 349, 442, 455, 460, 467, 470, 512
- Gesundheit Österreich GmbH (GÖG) 334
- Gesundheit, psychische 345
- Gesundheitsbereich 58, 460, 512, 513, 514, 525
- Gesundheitsförderung 344, 345, 464
- Gesundheitsförderung Schweiz 345
- Gesundheitsministerium 334
- Gesundheitsobservatorium (OBSAN) 344, 347, 350, 351, 352
- Gesundheitsplattform 334
- Gesundheitspolitik 137
- Gesundheitsproblem 348
- Gesundheitsqualitätsgesetz (GQG) 333
- Gesundheitsreform 333, 335, 336, 340
- Gesundheitsreformgesetz 333
- Gesundheitssystemevaluation 329, 354, 355
- Gesundheits telematikgesetz (GtelG) 333
- Gesundheitswesen, siehe Gesundheit
- Gewaltprävention 464
- Gewässerschutz 104, 230
- GFA, siehe Gesetzesfolgenabschätzung
- GFK, siehe Gemeinschaftliches Förderkonzept
- GGO, siehe Gemeinsame Geschäftsordnung der Bundesministerien
- Glaubwürdigkeit 63, 234, 235, 275, 479, 492, 495, 513, 526
- Gleichstellung 19, 60, 144, 273, 275
- Gleichstellungsgesetz 60
- Gleichstellungspolitik 490
- Gliedstaatsvertrag 332
- GLK, siehe Eidgenössische Kommission für Grundsatzfragen der Krankenversicherung
- Globalbudget 363
- Glückskette 266, 466
- GÖG, siehe Gesundheit Österreich GmbH
- Good Practice 272, 418, 429
- Governance 47, 254, 256, 424, 428
- Government Accountability Office (GAO), siehe United States Government Accountability Office
- GQG, siehe Gesundheitsqualitätsgesetz
- Graue Literatur 453
- Graz 223

- Grundgesetz für die Bundesrepublik Deutschland 29
- Grundlagenforschung 284, 308, 345, 369, 370, 392
- Grundlagenverbesserung 102
- Grundsicherung 151, 153
- Grundsicherung für Arbeitslose 120
- Grüner Bericht 90, 111
- GRW, siehe Gemeinschaftsaufgabe Verbesserung der regionalen Wirtschaftsstruktur
- GtelG, siehe Gesundheitstelematikgesetz
- GTZ, siehe Gesellschaft für Technische Zusammenarbeit
- Guidelines for Quality Assurance in the European Higher Education Area 168
- Halbzeitbewertung 79, 82, 410, 417, 418, 419
- Halbzeitevaluation 95, 433
- Hamburg (Bundesland) 166
- Hamburgisches Welt-Wirtschafts Archiv (HWWA) 274, 278
- Hamburgisches WeltWirtschaftsInstitut (HWWI) 247
- Handlungsforschung 444
- Hannover 208
- Hans-Böckler-Stiftung 329, 377, 378, 394
- Hanser und Partner AG (BHAP) 230
- Happy-Sheets 14
- HarmoS-Konkordat, Interkantonale Vereinbarung über die Harmonisierung der obligatorischen Schule 185
- Hartz IV-Evaluation 122, 519
- Hartz-Gesetz 118, 120, 152
- Hartz-Kommission 120, 122
- Hauptschule 170, 171
- HBLFA Raumberg-Gumpenstein (RG) 92
- Health Technology Assessment Zentrum 356
- Heilmittelgesetz HMG 344
- Heilverfahren 325
- Heimministerium 386
- Helmholtz-Gemeinschaft (HGF) 167, 283
- Helvetas 263, 271
- Hessen (Bundesland) 30, 81, 166
- HGF, siehe Helmholtz-Gemeinschaft 288, 289, 290
- Hilfsorganisation 19
- HIS, siehe Hochschulinformationssystem
- HIV/Aids 347, 349, 460, 467
- HMG, siehe Heilmittelgesetz
- Hochschulautonomie 178, 180, 195, 196
- Hochschule 13, 14, 15, 18, 19, 20, 22, 45, 67, 157, 161, 163, 174, 177, 188, 190, 194, 231, 267, 283, 285, 287, 288, 301, 305, 307, 309, 310, 311, 314, 328, 429, 431, 455, 459, 460, 470, 512, 514, 521, 525
- Hochschuleevaluation 163, 164, 167, 448
- Hochschulevaluierungsverbund 165
- Hochschulgesetz 178
- Hochschulinformationssystem (HIS) 290
- Hochschullehrende 14, 15
- Hochschulrahmengesetz (HRG) 163, 289
- Hochschulrektorenkonferenz (HRK) 166
- HoF, siehe Institut für Hochschulforschung Wittenberg
- Hospitation 183
- Hotelkreditgesetz 422
- HRG, siehe Hochschulrahmengesetz
- HRK, siehe Hochschulrektorenkonferenz
- Humanitäre Hilfe 263
- Humboldt-Universität Berlin (HU Berlin) 250, 377, 378
- HWWA, siehe Hamburgische Welt-Wirtschafts Archiv
- HWWI, siehe Hamburgisches WeltWirtschaftsInstitut
- IAB, siehe Institut für Arbeitsmarkt- und Berufsforschung
- ICD, siehe International Statistical Classification of Diseases
- Ich-AG 120, 124, 125
- IDT-HSG, siehe Institut für Öffentliche Dienstleitungen und Tourismus, Universität St. Gallen
- IFES, siehe Interkantonale Fachstelle für Externe Schulevaluation auf der Sekundarstufe II
- IFEU, Institut für Energie- und Umweltforschung Heidelberg 205
- IFQ, siehe Institut für Forschungsinformation und Qualitätssicherung
- IG-L, siehe Immissionsschutzgesetz-Luft
- IHG, siehe Investitionshilfegesetz
- IIZ, siehe interinstitutionelle Zusammenarbeit
- IKRK, siehe Internationales Komitee vom Roten Kreuz
- IKT, siehe Institut für Kulturtechnik und Bodenwasserhaushalt
- Immissionsschutzgesetz-Luft (IG-L) 43
- Impact 103, 139, 145, 199, 265, 268, 317, 429

- Impact Analyse 272, 286
Impact Assessment 34
Impact Evaluation 274, 277
Implementationsforschung 175
ImPlus 385
Indicators of Educational Systems (INES) 191
Indien 270
Indikator 314, 317
Individualfeedback 183
Industrie 283, 456
INES, siehe Indicators of Educational Systems 191
Infektionskrankheit 349
Infras 230
InnoRegio 409, 411
InnoTour 423
Innovation 36, 47, 285
Innovations-Evaluation 501
Innovationspolitik 283, 416, 417
Input 158, 208, 270, 272, 287, 291, 375, 387, 429, 436
Input-Steuerung 158, 170
Inspektorat 184
Institut für Alpine Naturgefahren und Forstliches Ingenieurwesen der Universität für Bodenkultur, Wien 92
Institut für Arbeitsmarkt- und Berufsforschung (IAB) 28, 121, 122, 150, 150, 408, 410
Institut für Forschungsinformation und Qualitätssicherung (IFQ) 167, 289, 294, 315
Institut für Hochschulforschung Wittenberg (HoF) 163
Institut für Höhere Studien (IHS) 452
Institut für Kulturtechnik und Bodenwasserhaushalt (IKT) 92
Institut für Ländliche Räume der Bundesforschungsanstalt für Landwirtschaft (FAL) 78
Institut für Nachhaltige Wirtschaftsentwicklung der Universität für Bodenkultur (NWE) 91
Institut für Öffentliche Dienstleitungen und Tourismus, Universität St. Gallen (IDT_HSG) 429
Institut für Ökologie und Naturschutz der Universität, Universität Wien 92
Institut für Ökologie und Umweltplanung 92
Institut für ökologische Wirtschaftsforschung (IÖW) 205
Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen 356
Institut für Qualitätsentwicklung im Bildungswesen (IQB) 160
Institut für Schulentwicklungsforschung 524
Institut für Sozial- und Präventivmedizin 459
Institut für sozialökologische Forschung (ISOE) 205
Institut für Wissenschafts- und Technikforschung (IWT) 290
Institut ORL der ETH 230
Institution 64, 122, 361, 466, 467, 485
Institution, formale 368
Institution, formelle 65
Institution, informelle 65
Institutionalisierung der Evaluation 22, 27, 28, 29, 32, 35, 36, 37, 38, 52, 57, 58, 59, 60, 61, 64, 279, 315, 317, 375, 469, 486, 487, 494, 511, 521, 527
Institutionelle Politik, siehe Politik, institutionelle
Institutionenforschung 368
Institutionenpolitik 361, 371, 372, 392
Institutionenreform 371
Integration 464
Integrationspolitik 426
Integrationsprozess, europäischer 64
Intercooperation 263, 271
Interface, Institut für Politikstudien, Luzern 231
Interimevaluation 302
Interinstitutionelle Zusammenarbeit (IIZ) 141
Interkantonale Fachstelle für Externe Schulevaluation auf der Sekundarstufe II (IFES) 187
Interkantonalen Vereinbarung über die Harmonisierung der obligatorischen Schule (HarmoS-Konkordat) 185
Interkantonaler Verband für Arbeitnehmerschutz IVA 346
International Organization for Standardization (ISO) 186, 208, 209, 214, 228
International Quality Indicator Project (QIP) 337
International Statistical Classification of Diseases (ICD) 333
Internationale Vernetzung 312
Internationale Weiterbildung und Entwicklung gGmbH (InWEnt) 247
Internationales Komitee vom Roten Kreuz – IKRK 263, 271
Internet 13, 69, 82, 97, 484
Interreg 409, 417, 423, 425, 426, 429, 433, 437

- Interview 141, 146, 349, 427, 435, 453, 485
 Investitionsförderung 80, 83, 89, 412
 Investitionshilfegesetz (IHG) 422, 423, 425, 426
 InWent, siehe Internationale Weiterbildung und Entwicklung gGmbH 279
 IÖW, siehe Institut für ökologische Wirtschaftsforschung 206
 IQB, siehe Institut für Qualitätsentwicklung im Bildungswesen 160
 ISI, siehe Fraunhofer-Institut für Systemtechnik und Innovationsforschung
 ISO, siehe International Organization for Standardization
 ISOE, siehe Institut für sozialökologische Forschung
 ISO-Norm 186
 IVA, siehe Interkantonaler Verband für Arbeitnehmerschutz
 IWT, siehe Institut für Wissenschafts- und Technikforschung
 IZA, Forschungsinstitut zur Zukunft der Arbeit 122

 Joanneum Research 97, 295, 299, 301
 Job-AQTIV-Gesetz 119
 Johnson 365, 366, 480
 Joint Committee on Standards for Educational Evaluation 168, 471, 479, 498
 Joint Evaluation 266, 268, 269, 270, 272
 Jugendamt 442, 447, 466
 Jugendarbeit 442, 447, 451, 463, 513
 Jugendarbeitslosigkeit 148
 Jugendfürsorge 451
 Jugendhilfe 463, 522
 Jugendhilfeausschuss 442
 Jugendschutz 60, 442
 Jugendsozialarbeit 442
 Jurisprudenz 459
 Justiz 66, 392
 Justizministerium 66

 Kantonales Sozialamt Zug 141
 Kantonsverwaltung 466, 467
 Kap Verde 256
 KAP-EVA, siehe Koordinations- und Arbeitsplattform Evaluierung
 Karlsruhe 290
 Kärnten (Bundesland) 336, 383, 388, 416, 417
 Kartellgesetz 495
 Katastrophenschutz 19
 KFH, siehe Konferenz der Fachhochschulen der Schweiz
 KfW Entwicklungsbank 246, 247, 253
 KgSt, siehe Kommunale Gemeinschaftsstelle für Verwaltungsmanagement
 Kinder- und Jugendhilfegesetz 445, 451
 Kinder-, Jugend- und Familienhilfe 441, 470
 Kindertagesbetreuung 443, 447, 448
 Kindertagesstätte 473
 Klassenscockpit 183
 Klaus Novy Institut 205
 Kleine und mittlere Unternehmen (KMU) 57
 Klima 106
 Klima:aktiv 215, 222
 Klimapolitik 205
 Klimaschutz 203, 237
 KMK, siehe Ständige Konferenz der Kultusminister der Länder (KMK)
 KMU Forschung Österreich 298
 KMU, siehe Kleine und mittlere Unternehmen
 Knowledge Utilization Research 479
 Koalitionsregierung 90, 91
 Koalitionsvertrag 30, 38
 KOF, siehe Konjunkturforschungsstelle der ETH Zürich
 Kohortengruppe 347, 349
 Kohortenstudie 325
 Kombilohn-Modell 119, 121, 149
 Kommission für Technologie und Innovation (KTI) 308, 309, 310, 466, 467
 Kommunale Gemeinschaftsstelle für Verwaltungsmanagement (KGSt) 394, 445
 Kommunalverband 37
 Kompetenzzentrum für Evaluation (CCE) 351, 352, 354
 Komplementärmedizin 347, 513
 Konferenz der Fachhochschulen der Schweiz (KFH) 189, 196
 Konjunkturforschungsstelle der ETH Zürich (KOF) 139
 Konkordanzdemokratie 379
 Konkordanzsystem 62, 64, 368, 392, 397
 Konkurrenzsystem 67, 397
 Konsensus Konferenz 349
 Konstruktivismus 418
 Kontrolle 157, 159, 164, 279, 296, 338, 412, 444, 445, 502, 503
 Kontrollfunktion 61, 63

- Kontrollgruppe 269, 272, 277, 347, 408, 418
Kontrollgruppe, experimentelle 139
Kontrollparadigma 329
Koordinations- und Arbeitsplattform
 Evaluierung („KAP-EVA“) 50, 241, 419, 421, 438
Korrektheit 220
Kosovo 256
Kosten 31, 33, 35, 54, 79, 141, 151, 205, 209, 229, 268, 347, 348, 355, 376, 380, 388, 427, 431, 481, 483, 489, 494, 497, 500, 513, 517, 522
Kostendruck 522, 523
Kosten-Leistungsrechnung 28, 385, 386
Kosten-Nutzen-Analyse 135, 141
Krankenanstalt 331, 332, 337
Krankenanstalten- und Kuranstaltengesetz 333
Krankenhaus 14, 326, 332, 353
Krankenkasse 327, 343
Krankenversicherungsgesetz (KVG) 355, 344, 348, 351
Kriminalität 66
Kritische Theorie 444
Kritischer Rationalismus 444
KTI, siehe Kommission für Technologie und Innovation
Kultur 259, 308, 509
Kulturerbe 259
Kulturpolitik 489
Kundenbindung 19
Kundengewinnung 19
Kunden-Zufriedenheitsmessung 14
Kurzevaluation 518
KVG, siehe Krankenversicherungsgesetz
Kyoto-Protokoll 203, 227, 229
Kyoto-Ziel 213

Label 214, 228
Land- und Forstwirtschaft (LWF) 87
Landesbauordnung 36
Landesgesetzgebung 44
Landesgesundheitsfonds 334
Landeskrankenhaus (LKH) 337
Landesplanung 411
Landesprogramm Albanien 259
Landesrechnungshof 29, 91, 98, 376, 379, 406
Landesrecht 36
Landesregierung (LReg) 30, 36, 40, 49, 78, 90, 91, 98, 224, 225
Landesregierung, Steiermärkische 415
Landesregierung, Vorarlberger 415
Landkreistag 394
Ländliche Entwicklung 256
Ländliches Entwicklungsprogramm (LEP) 88
Landsteiner 434
Landtag 30, 40
Landwirtschaft 15, 75, 87, 101, 102, 106, 109, 228, 305, 489
Landwirtschaftsgesetz 88, 90, 104, 109
Landwirtschaftskammer 90, 91
Landwirtschaftsministerium 214
Langzeitarbeitslosigkeit 119
Lärm 346
Lateinamerika 267, 270
LCH, siehe Dachverband der Schweizer Lehrerinnen und Lehrer
LEADER 95, 423
Lebensministerium 91, 94, 214, 217, 218, 238
Legalitätsprinzip 52
Legislative 66, 67
Legislativkommission 55
Legitimation 137, 159, 164, 171, 219, 272, 296, 365, 366, 381, 423, 450, 469, 489, 493, 502, 504, 511, 521, 522, 523, 525
Legitimierung 28
Lehre 164, 167, 289
Lehrerdienstrecht 175
Lehrevaluation 167, 198
Lehrmittel 182
Lehrplan 182, 183
Lehrveranstaltungsbefragung 164, 165
Lehrveranstaltungsevaluation 179, 198
Lehrveranstaltungsfeedback 13
Leibniz-Gemeinschaft 167
Leistung 274, 376
Leistungsauftrag 363
Leistungsfähigkeit 286, 288
Leistungsindikator 395, 429
Leistungsmessung 272, 371, 373, 514
Leistungsrechnung 500
Leistungsvergleichsstudie 159, 196
Leistungsvertrag 454
Leitfaden 349
Leitfadeninterview 463
LEP, siehe Ländliches Entwicklungsprogramm 88, 91, 94, 95, 96
Lernende Region 409, 411
Linz 383
Lissabon-Prozess 152, 295

- Lissabon-Strategie für Wachstum und Beschäftigung 51
 LKH, siehe Landeskrankenhaus
 Logic Charts 296
 Logical Framework Model 419
 Logit-Probit-Analyse 296
 Lokale Agenda21 227, 407
 LReg, siehe Landesregierung
 Ludwig Boltzmann Institut für Health Technology Assessment 356
 Luftreinhaltung 230
 Luzern 59
 Luzern (Kanton) 141

 Machbarkeit 234
 Machbarkeitsstudie 302
 Magadinoebene 428
 Mainzer Modell 119, 121
 Majorz 368
 MAMAC, siehe medizinisch-arbeitsmarktliches Assessment mit Case Management
 Management 445, 489
 Managementforschung 206
 Managerialismus 373
 MAR, siehe Reglement der EDK über die Anerkennung von gymnasialen Maturitätsausweisen
 Marketing 456
 Marktmassnahmen 102
 Marktordnung 89, 92, 94
 Marktordnungsgesetz 88
 Marktpolitik 75
 Marktregulierung 109, 110
 Marktstützung 100, 104, 105
 Marktunterstützung 75
 Masterstudiengang 163, 195, 451
 Matched Pairs 296, 303
 Matching 150, 408
 Materialforschung 287
 Max-Planck-Gesellschaft (MPG) 167, 284, 288, 289
 Mazedonien 256
 MDG, siehe Millenium Development Goals
 Mecklenburg-Vorpommern 166
 Medikament 323, 354, 355
 Medizin 29, 511
 Medizinisch-arbeitsmarktliches Assessment mit Case Management (MAMAC) 141
 Mehrebenensystem 70
 Menschenrechte 273

 Meta-Analyse 222, 349
 Meta-Evaluation (Evaluation von Evaluation/en) 20, 83, 171, 222, 350, 351, 524
 Methode, partizipative 328, 408, 409, 410, 418, 514
 Methodenhandbuch 35
 Metropolregion 432
 Mid-Term Review 265, 266
 Mid-Term-Evaluation 133
 MigrantInnen 455
 Migration 144, 348, 349
 Migrationspolitik 191, 489
 Mikroevaluation 355
 Mikroökonomie 511
 Milizparlament 67
 Milizverwaltung 62, 68
 Millenium Development Goals 259, 267, 273, 275, 276
 Millenniumserklärung 251, 273
 Mindeststandard 33, 34, 35
 Minijob 120
 Ministerialverwaltung 27, 33
 Ministerium für Ernährung, Landwirtschaft und Verbraucherschutz 206
 Ministerium für Verkehr, Innovation und Technologie 299, 300
 Ministerium für Wirtschaft und Arbeit 299
 Ministerium für Wissenschaft und Forschung 299
 Ministerrat 387
 Ministerratsbeschluss 42, 43
 Modellprojekte zur Verbesserung der Zusammenarbeit von Arbeitsämtern und Trägern der Sozialhilfe (MoZArT) 121
 Moldau 256
 Monitoring 53, 97, 103, 105, 106, 107, 112, 130, 131, 133, 196, 229, 265, 266, 272, 274, 292, 317, 341, 407, 418, 419, 430, 431, 447, 452, 500, 519
 Montenegro 256
 Mosambik 256
 MoZArT, siehe zur Verbesserung der Zusammenarbeit von Arbeitsämtern und Trägern der Sozialhilfe 121
 MPG, siehe Max-Planck-Gesellschaft
 München 208
 Münster 208
 Musikerziehung 60

 Nachdiplomstudiengang Entwicklungsländer (NADEL) 276

- Nachhaltige Entwicklung 203, 204, 216, 227, 229, 234, 275
- Nachhaltigkeit 68, 106, 203, 211, 212, 238, 240, 249, 268, 275, 277, 334, 340, 347, 407, 433, 504, 511, 513
- Nachhaltigkeitsbeurteilung 229
- Nachhaltigkeitsevaluation 211, 220, 238, 511
- Nachhaltigkeitspolitik 223
- Nachhaltigkeitsprüfung 215
- Nachhaltigkeitsstrategie 215, 216, 217, 221, 237, 241
- NADEL, siehe Nachdiplomstudiengang
Entwicklungsländer
- Nanotechnologie 302
- Narrative Based Medicine 326
- Nationale Umweltplanung (NUP) 214
- Nationaler Normenkontrollrat (NKR) 37
- Nationales Büro für Wirtschaftsforschung (CPB) 66
- Nationales Forschungsprogramm (NFP) 140, 144, 146, 230, 308, 345, 462, 466, 467, 487, 515
- Nationalrat 40, 41, 42
- Naturschutz 214
- Naturschutzgesetz 214
- Naturwissenschaft 204, 211, 214, 217, 233, 237, 238, 239, 284, 511, 513
- Network of Networks Impact Evaluation
Initiative (NONIE) 274
- Netzwerk Evaluation in der Bundesverwaltung 61, 233, 515
- Netzwerkanalyse 408
- Neue Regionalpolitik 424, 427, 430, 431
- Neuer Finanzausgleich 433
- Neues Steuerungsinstrument (NSI) 376
- Neues Steuerungsmodell (NSM) 376, 395, 470
- Neugestaltung des Finanzausgleichs (NFA) 62, 424
- New Public Management (NPM) 59, 62, 142, 146, 317, 363, 364, 365, 370, 375, 381, 382, 383, 386, 387, 389, 390, 391, 392, 393, 395, 397, 398, 405
- NFA, siehe Neugestaltung des Finanzausgleichs
- NFP, siehe Nationales Forschungsprogramm
- Niederlande 35, 66, 67, 149, 251
- Niederösterreich (Bundesland) 336, 416
- Niedersachsen (Bundesland) 290
- Nikaragua 256
- NKR, Nationaler Normenkontrollrat
- NONIE, siehe Network of Networks Impact
Evaluation Initiative
- Nordrhein-Westfalen (NRW) 30, 81, 160, 166, 290, 375, 448
- Nordwestschweizerische
Erziehungsdirektorenkonferenz (NW EDK)
184, 185, 187
- Normenbegutachtung 41
- Normenkontrollrat 37
- Normprüfungsstelle 33
- NPM, siehe New Public Management
- NRP, siehe Neue Regionalpolitik
- NSI, siehe Neues Steuerungsinstrument
- NSM, siehe Neues Steuerungsmodell 377, 378, 394
- NUP, siehe Nationale Umweltplanung
- Nutzen 23, 31, 54, 70, 141, 151, 209, 229, 268, 347, 348, 380, 390, 409, 419, 478, 479, 482, 484, 494, 497, 518, 522
- Nutzenorientierung 16
- Nützlichkeit 209, 211, 220, 222, 247, 275, 349, 418, 478, 479, 482, 484, 496, 498, 521, 526
- Nutzung 22, 28, 85, 86, 199, 212, 275, 279, 365, 369, 370, 412, 444, 478, 479, 482, 484, 521
- Nutzung der Evaluation 475
- Nutzung, instrumentelle 500
- Nutzung, konzeptionelle 501
- Nutzung, taktische 501
- Nutzwertanalyse 31
- NW EDK, siehe Nordwestschweizerische
Erziehungsdirektorenkonferenz
- NWE, siehe Institut für Nachhaltige
Wirtschaftsentwicklung der Universität für
Bodenkultur
- OAQ, siehe Organ für Akkreditierung und
Qualitätssicherung der Schweizerischen
Hochschulen
- Oberaufsicht 55, 56, 67
- Oberösterreich (Bundesland) 299, 336, 337, 417
- Oberstufe, gymnasiale 171, 193
- ÖBIG, siehe Österreichisches Bundesinstitut für
Gesundheitswesen
- OBSAN, siehe Gesundheitsobservatorium
- OECD, siehe Organisation für wirtschaftliche
Zusammenarbeit und Entwicklung 32, 51, 93, 108, 144, 148, 153, 160, 191, 245, 251, 257, 261, 265, 274, 275, 289, 329, 423, 429
- OEZA, siehe Ost- und
Entwicklungszusammenarbeit
- Öffentlichkeit 312

- Öffentlichkeitsgesetz 53, 56
 Öffentlichkeitsverordnung 57
 Official Development Assistance (ODA) 256
 ÖkoBusinessPlan 222
 Öko-Effizienz 208, 209
 Öko-Institut 205
 Ökologie 102, 106, 108, 211
 Ökologischer Leistungsnachweis (ÖLN) 100, 101, 104
 Ökonometrie 144, 408, 436, 513
 Ökonomie 106, 139, 142, 191, 211, 230
 ÖLN, siehe Ökologischer Leistungsnachweis
 Opferhilfegesetz 60, 496
 ÖPUL, siehe Österreichisches Programm für eine umweltgerechte und extensive Landwirtschaft
 Öqmed 338
 Ordnungsmäßigkeit 29, 56, 65
 Organ für Akkreditierung und Qualitätssicherung der Schweizerischen Hochschulen (OAQ) 188, 190, 315
 Organisation for Economic Cooperation and Development (OECD) 160, 191
 Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD) 32, 51, 93, 108, 144, 148, 153, 160, 191, 245, 251, 257, 261, 265, 274, 275, 289, 329, 423, 429
 Organisationsanalyse 14
 ÖROK, siehe Österreichische Raumordnungskonferenz
 Ost- und Entwicklungszusammenarbeit (OEZA) 255, 256
 Ostdeutschland 286
 Österreichische Agentur für Gesundheit und Ernährungssicherheit (AGES) 92
 Österreichische Bundesverfassung 40
 Österreichische Hochschülerschaft 48
 Österreichische Qualitätssicherungsagentur (AQA) 48, 179, 180, 181, 195, 196, 301, 315
 Österreichische Raumordnungskonferenz (ÖROK) 50, 414, 418, 433, 438
 Österreichische Rektorenkonferenz 48
 Österreichischen Volkspartei (ÖVP) 90
 Österreichischer Staatsschuldenausschuß 381
 Österreichischer Städtebund 389
 Österreichisches Bundesinstitut für Gesundheitswesen (ÖBIG) 334
 Österreichisches Institut für Wirtschaftsforschung (WIFO) 91, 299, 302
 Österreichisches Kuratorium für Landtechnik und Landentwicklung 92, 215
 Österreichisches Programm für eine umweltgerechte und extensive Landwirtschaft (ÖPUL) 94, 95, 97
 Outcome 103, 199, 317, 334, 340, 366, 372, 375, 429
 Output 199, 208, 270, 272, 288, 291, 317, 365, 366, 367, 369, 375, 429, 436
 Outputindikator 289
 ÖVP, siehe Österreichische Volkspartei
 Pädagogische Hochschule 177, 178, 460
 Pädagogisches Institut Oberösterreich 174
 Pariser Erklärung 251, 259, 261, 267, 272, 273, 279
 Parlament 11, 13, 21, 23, 52, 55, 59, 61, 65, 67, 90, 101, 105, 107, 108, 264, 265, 365, 366, 371, 373, 387, 392, 495, 499, 503
 Parlamentarische Verwaltungskontrolle (PVK) 12, 56, 58, 67, 231, 269, 309, 495, 503, 515
 Parlamentsgesetz 55, 57
 Parlamentsreform 62, 361, 363
 Parlamentsverwaltungsverordnung 56
 Partizipation 219, 275, 368, 443, 447, 505
 Patientenorientierung 334
 Patientenzufriedenheit 343
 Peer evaluation 309, 477
 Peer Review 45, 165, 80, 179, 190, 247, 261, 265, 285, 287, 299, 300, 314, 316, 534, 525
 Performance, siehe Performanz
 Performance Measurement 500
 Performanz 263, 270, 376, 377, 378, 386, 387
 Performanzeffekt 375
 Performanzevaluation 372, 375, 377, 378, 380
 Personalgesetz 183
 Personal-Service-Agentur (PSA) 120, 124
 Perspektive 50 plus 121
 Pflegebedürftigkeit älterer Menschen 346
 Pflegeperson 15, 29
 Pflegeversicherung 30
 Pharmakotherapie 347
 Phosphor 103, 106
 PISA, siehe Programme for International Student Assessment
 Placebogruppe 324, 326
 Planspiel 31
 Planung 30, 58, 86, 131, 276, 279, 405, 407, 434, 443, 453, 479, 497, 498

- Plattform Forschungs- und
Technologieevaluierung (FTEval) 47, 48, 215,
295, 296, 298, 313
- Plattform FTEval, siehe Plattform Forschungs-
und Technologieevaluierung
- Policy-Evaluation 294
- Politik 38
- Politik, institutionelle 21, 359, 361, 522
- Politik, materielle 361, 367
- Politik, substanzielle 361, 366, 367
- Politikanalyse 14
- Politikevaluation 60, 103, 220
- Politikfolgenabschätzung 77, 112
- Politikoption 34
- Politikverflechtungsfalle 405, 433
- Politikwissenschaft 231, 310, 368
- Politikzyklus 40, 47, 51
- Polizei 516
- Positivismusstreit 444
- Poweranalyse 324
- Praktikabilität 31, 38
- Prävention 323, 328, 329, 345, 355
- Präventionsprogramm 354
- Praxistest 31
- Preußische Oberrechnungskammer 29
- Primarschulbereich 193
- Prinzipal-Agent-Problem 301
- Privatuniversität 177, 178, 195
- Privatwirtschaftsverwaltung 42
- Produktivitätsindikator 286
- Produktpolitik, integrierte 209
- Profession 450, 451, 453, 454, 456, 511, 512
- Professionalisierung 185, 196, 197, 233, 234, 240,
241, 279, 301, 312, 351, 362, 370, 445, 449, 451,
453, 469, 485, 499, 503, 509
- Programme for International Student Assessment
(PISA) 157, 159, 161, 173, 191, 194, 195, 481,
514, 524
- Programm-Evaluation 164, 276, 277, 278, 290,
300, 309, 315, 366, 437, 448, 519
- Projekt eXe 20, 473
- Projekt Qualität in Schulen (Q.I.S.) 194
- Projektevaluation 164, 220, 276, 277, 278, 308, 309,
314, 384, 437, 519
- Projektzyklusphilosophie 270
- Proporz 368
- Prozess-Erfolgsstudie 326
- Prozessevaluation 327, 328, 330, 350, 355
- Prozess-Nutzung 501
- Prozessorientierung 175, 340
- Prozessqualität 168, 334, 336, 338
- Prüfung 343
- Prüfverfahren 31
- PSA, siehe Personal-Service-Agentur
- Pseudo-Evaluation 526, 527
- Psychologie 455, 459
- Psychotherapie 326, 347
- Psychotherapiegesetz 326
- Public policies 52
- Public Relations 456
- Public-Private-Partnerships 434
- PVK, siehe Parlamentarische
Verwaltungskontrolle
- Q.I.S., siehe Projekt Qualität in Schulen
- Q2E, siehe Qualität durch Evaluation und
Entwicklung
- QIBB, siehe Qualität im berufsbildenden
Schulwesen 175
- QIP, siehe Quality Indicator Project
- QS-Initiative 472
- Qualität 210, 219
- Qualität durch Evaluation und Entwicklung
(Q2E) 184, 186
- Qualität im berufsbildenden Schulwesen (QIBB)
175
- Qualität und Qualifizierung (2Q) 184
- Qualitativ-partizipatorisch 310
- Qualitätsentwicklung 182, 183, 184, 186, 189, 191,
192, 333, 352, 445, 446, 447, 448, 449, 470, 473,
522
- Qualitätskontrolle 289
- Qualitätskonzept 183
- Qualitätskriterium 312
- Qualitätsmanagement 159, 164, 172, 178, 181, 184,
185, 186, 187, 189, 194, 245, 264, 270, 272, 279,
297, 376, 445, 500, 525
- Qualitätssicherung 29, 33, 35, 46, 48, 82, 157, 163,
165, 166, 172, 177, 178, 180, 181, 182, 183, 184,
185, 188, 189, 191, 192, 195, 196, 198, 211, 215,
228, 292, 311, 326, 451, 452, 453, 469, 477, 504,
525
- Qualitätsstandard 274, 333
- Quality Audit 188
- Quality Indicator Project (QIP) 336
- Quasi-experimentell 140, 277, 328
- RAE, siehe Research Assessment Exercise

- Randomisierung 324, 325, 328, 349
 Randomized Controlled Trial (RCT) 324, 514
 Ranking 164, 165, 190, 196
 Rat für Forschung und Technologieentwicklung 47, 298, 302, 314
 Raumentwicklungspolitik 21, 401, 403, 523, 525
 Raumluft 346
 Raumordnung 101, 413
 Raumordnungsgesetz (ROG) 404, 406, 415
 Raumordnungskonferenz (ÖROK) 413
 Raumordnungspolitik 406, 408, 410, 422, 490
 Raumplanung 230, 413
 Raumplanungsgesetz (RPG) 425, 427, 428
 Raumplanungspolitik 403, 407, 414, 415
 RAV, siehe regionale Arbeitsvermittlungszentren
 RCT, siehe Randomized Controlled Trial
 Reaktions-Evaluation 501
 Rechenschaftslegung 18, 157, 269, 502, 503
 Rechnungshof 40, 41, 43, 44, 50, 66, 223, 371, 373, 377, 385, 390, 391, 393, 415
 Rechnungshofgesetz (RHG) 41
 Rechnungskontrolle 40
 Rechtmässigkeit 56, 65, 503
 Rechtsextremismus 473
 Rechtsoptimierung 35
 Referat 120 "Evaluierung der Entwicklungszusammenarbeit" 246
 Referentenentwurf 31
 Referenzgruppe 140
 Reformdruck 379
 Reformpolitik 28
 Reformrhetorik 379
 Regelungsalternative 31, 34
 Regierung 21, 23, 55, 56, 61, 392
 Regierungs- und Verwaltungsorganisationsgesetz (RVOG) 363
 Regierungsrat des Kantons Basel-Stadt 12
 Regierungsreform 361, 363
 Regierungssystem 20, 361
 Regionale Arbeitsvermittlungszentren (RAV) 138, 143, 144, 146, 150, 151
 Regionale Impulsförderung 416
 Regionale Innovationszentren (RIZ) 416
 Regionalentwicklung 46, 49
 Regionales Arbeitsvermittlungszentrum (RAV) 137, 138, 139, 142, 145, 466
 Regionalförderung 46, 49, 50
 Regionalmanagement 416, 419, 427
 Regionalplanung 411
 Regionalpolitik 65, 101, 216, 295, 403, 406, 408, 411, 412, 414, 417, 421, 422, 510
 Regionen Aktiv 78
 RegioPlus 423, 427
 Reglement der EDK über die Anerkennung von gymnasialen Maturitätsausweisen (MAR) 186
 Regulierung 522
 Regulierungsfolgenabschätzung 229
 Reichsgesundheitsamt 28
 Reintegrationsprogramm 249
 Rentenversicherung 523
 Reporting 58
 Repräsentativität 139
 Research Assessment Exercise (RAE) 45
 Ressortforschung 229, 341
 Ressortforschungseinrichtung 111, 112
 Review 265, 271, 272, 277, 350
 Review, externe 265
 Review, interne 266
 Rezeptionsforschung 198, 478, 481
 Rheinland-Pfalz 32, 81, 166
 RHG, siehe Rechnungshofgesetz
 Riehen 370
 Ritual, symbolisches 70
 RIZ, siehe Regionale Innovationszentren
 ROG, siehe Raumordnungsgesetz
 RPG, siehe Raumplanungsgesetz
 Ruhr-Universität Bochum 377
 RVOG, siehe Regierungs- und Verwaltungsorganisationsgesetz
 Saarbrücken 197
 Saar-Gemeinschaftsinitiative (SGI) 119, 121
 Saarland (Bundesland) 160, 166
 Sachsen (Bundesland) 32, 81, 376
 Sachsen-Anhalt (Bundesland) 32, 81
 SAGW, siehe Schweizerische Akademie der Geistes- und Sozialwissenschaften
 Salutogenese 351
 Salzburg (Bundesland) 91, 383, 388, 417
 SAMF, Arbeitskreis Sozialwissenschaftliche Arbeitsmarktforschung
 S-Bahn Zürich 428
 SBF, siehe Staatssekretariat für Bildung und Forschung
 SBV, siehe Schweizerischer Bauernverband
 Schneeballverfahren 350
 Schnittstellenanalyse 31
 Schulaufsicht 184, 185, 195, 196

- Schulautonomie 172, 193
Schule 18, 19, 157, 170, 193, 346, 467, 524
Schule BEWUSST 174
Schuleffektivitätsforschung 161
Schulentwicklung 158, 185
Schulentwicklungsforschung 161
Schulevaluation 161, 183, 184, 185, 524
Schulevaluation, externe 174, 198
Schulevaluation, interne 194
Schulgesetz 182, 183
Schulgesetzgebung 170
Schulinspektion 160, 161, 195
Schulleistungsmessung 13
Schulleitung 162, 183, 184
Schulsozialarbeit 466
Schulsystem 184
Schulvergleich 13
Schulversuch 171, 172
Schweden 149
Schweizer Bundesverfassung (BV), siehe Bundesverfassung, Schweizerische
Schweizer Stiftung für sozialwissenschaftliche Forschung (FORS) 308
Schweizerische Akademie der Geistes- und Sozialwissenschaften (SAGW) 309, 312
Schweizerische Evaluationsgesellschaft (SEVAL) 11, 12, 17, 23, 153, 197, 278, 233, 239, 312, 344, 350, 351, 352, 354, 356, 395, 430, 459, 471, 498, 514, 515
Schweizerische Friedensstiftung swisspeace 309
Schweizerische Gesundheitsstiftung 467
Schweizerische Maturitätskommission (SMK) 186
Schweizerische Studiengesellschaft für Raumordnung und Regionalpolitik 430
Schweizerische Unfallversicherungsanstalt (SUVA) 346
Schweizerische Universitätskonferenz (SUK) 190
Schweizerische Vereinigung für Qualitäts- und Management-Systeme (SQS) 186
Schweizerischer Bauernverband (SBV) 101, 106, 107
Schweizerische Informations- und Datenarchivdienst für die Sozialwissenschaften (SIDOS) 17, 308, 351, 459
Schweizerischer Nationalfonds (SNF) 12, 307, 308, 309, 310, 311, 314, 345, 350, 459, 460, 462, 465, 466, 470, 473
Schweizerischer Wissenschafts- und Technologierat (SWTR) 307, 308, 309, 310, 311, 312, 315
Schweizerischer Wissenschaftsrat (SWR) 309
Schweizerisches Haushalts-Panel (SHP) 308
Schweizerisches Heilmittelinstitut 350
Schweizerisches Sozialarchiv 309
SCP, siehe Büro für soziale und kulturelle Planung 66
SECO, siehe Staatsekretariat für Wirtschaft
Sektion Evaluation und Controlling (E + C) 265
Sektion Umweltforschung 229
Sektoralpolitik 372, 424, 428, 433, 434
Sekundärprävention 346
Sekundarstufe II 186, 187
Selbstbeurteilung 183
Selbstevaluation 159, 161, 171, 172, 184, 188, 193, 245, 247, 260, 265, 267, 269, 270, 271, 276, 277, 314, 328, 332, 338, 350, 354, 355, 385, 386, 398, 445, 446, 447, 452, 471, 485, 504, 524
Selbstevaluation, interne 174, 371, 373
Selbstverwaltungskörper 40
Selden & Partner 224
Seminar für Ländliche Entwicklung (SLE) 250
SEVAL, siehe Schweizerische Evaluationsgesellschaft
SEVAL-Standard 11, 219, 344, 431
SGB, siehe Sozialgesetzbuch
SGI, siehe Saar-Gemeinschaftsinitiative
Shanghai-Ranking 195
SHP, siehe Schweizerisches Haushalts-Panel
SIDOS, siehe Schweizerische Informations- und Datenarchivdienst für die Sozialwissenschaften
Siedlungsentwicklung 428
Siedlungshygiene 256, 259
Siedlungspolitik 425
SKM, siehe Standardkosten-Modell
SLE, siehe Seminar für ländliche Entwicklung
SMK, siehe Schweizerische Maturitätskommission
SNF, Schweizerischer Nationalfonds
Sockelarbeitslosigkeit 119
SOFIS, siehe sozialwissenschaftliches Forschungsinformationssystem
Soliwork 141
Soll-Ist-Vergleich 368
Solothurn (Kanton) 59, 141, 365
Sonderforschungsbereich 700 252

- Sozialarbeit, siehe Soziale Arbeit
 Sozialarbeitender 15
 Sozialarbeitsausbildung 450, 451
 Soziale Arbeit 20, 21, 439, 471, 522
 Soziales 211
 Sozialforschung 444, 451
 Sozialgesetzbuch (SGB) 30, 118, 119, 125, 126, 149, 325, 326, 442
 Sozialhilfe 120, 121, 141, 466
 Sozialhilfegesetz 341, 451
 Sozialpädagogik 441, 443, 455, 459, 469, 471
 Sozialpartner 51
 Sozialpolitik 130, 140, 191, 459
 Sozialpsychologie 502
 Sozialstrukturanalyse 408
 Sozialwesen 308, 450
 Sozialwissenschaft 197, 203, 204, 205, 230, 233, 237, 238, 239, 294, 309, 350, 351, 429, 444, 461, 513, 515
 Sozialwissenschaftliches
 Forschungsinformationssystem (SOFIS) 17, 18
 Soziologie 106, 455
 Sparsamkeit 415, 500
 Sportwissenschaft 346
 Sprachregion 182, 191
 SQS, siehe Schweizerische Vereinigung für
 Qualitäts- und Management-Systeme
 SR, siehe Systematische Rechtssammlung 56, 57
 St. Gallen (Kanton) 365
 Staatssekretär 37
 Staatssekretariat für Bildung und Forschung (SBF) 308
 Staatssekretariat für Wirtschaft (Seco) 12, 143, 144, 145, 146, 147, 151, 259, 263, 265, 266, 268, 269, 270, 272, 275, 346, 424, 426, 430, 487
 Staatssekretärsausschuss 37
 Staatsvertrag 30
 Städtebau 407, 410
 Städtebund 413
 Stadtentwicklung 407
 Städtetag 394
 Stadtplanung 465
 Stadtsoziologie 465
 Stakeholder, siehe auch Beteiligte und Betroffene
 12, 16, 420, 437, 438, 478, 479, 481, 482, 499, 500, 502, 513, 521, 526
 Stammzellenforschung 314
 Standard 63, 207, 211, 215, 233, 234, 251, 275, 298
 Standardisierung 211
 Standardkostenmodell 27, 31, 35, 38, 43, 69
 Ständemehr 369
 Ständige Konferenz der Kultusminister der
 Länder (KMK) 157
 Standortentwicklung 432
 Standortförderung 424
 Statistik 229, 287, 427
 Statistisches Bundesamt (StBA) 37, 163, 205
 StBA, siehe Statistisches Bundesamt
 Steiermark (Bundesland) 174, 336, 383, 416
 Steiner 44, 193, 194, 271, 296, 463
 Stellwerk 183
 Steuer 60
 Steuerpolitik 77, 137
 Steuerung 30, 221, 361, 376, 411, 453, 510, 522
 Steuerung, ergebnisorientierte 173
 Steuerung, wirkungsbasierte 363, 366, 446
 Stichprobe 158, 160, 268, 326, 338, 350
 Stickstoff 103, 106
 Stiftung 19
 Stiftung Gesundheitsförderung Schweiz 348
 Stiftung Klimarappen 229, 233
 Stipendium 60
 Strafvollzug 460, 473
 Strahlung 349
 STRAT.AT 420
 Strategische Umweltprüfung (SUP) 97, 415
 Strukturevaluation 327
 Strukturfonds 46, 84, 222, 411, 417, 420, 425
 Strukturfondsmittel 46
 Strukturpolitik 65, 75, 422, 509, 511, 525
 Studium 164
 Subsidiaritätsprinzip 182, 443
 Subvention 232, 454
 Subventionsbericht 56
 Subventionsgesetz 56
 Sucht 455, 460, 467
 Suchtforschung 460
 Suchtpräventionsstelle 347
 SUK, siehe Schweizerische Universitätskonferenz
 Sunset legislation 30, 38, 68
 SUP, Stategische Umweltprüfung
 SUVA, siehe Schweizerische
 Unfallversicherungsanstalt
 SWI, siehe Spotwissenschaftliches Institut
 Swissmedic 343, 344, 348
 SWR, siehe Schweizerischer Wissenschaftsrat
 SWTR, siehe Schweizerischer Wissenschafts- und
 Technologierat

- Systematische Rechtssammlung (SR) 56, 57
Systemevaluation 158, 184, 185, 286, 288, 289, 290, 291, 342
Systemsteuerung 182
- TAB, siehe Büro für Technikfolgenabschätzung
Tabak 349
Tabakpräventionsfonds (TPFV) 344
Tansania 250, 277
TA-Swiss 310, 312
Technikfolgenabschätzung 29, 67, 205
Technische Hilfe 257
Technischer Überwachungsverein (TÜV) 204
Technologie 15, 20, 21, 44, 45, 46, 47, 48, 49, 68, 509
Technologie- und Innovationspolitik (FTI-Politik) 295
Technologiefolgenabschätzung 307, 314
Technologiepolitik 214, 236, 281
Technologietransfer 232
Technopolis 299
Telekommunikationsgesetz (TKG) 42
Terms of Reference (ToR) 258
Tertiärstufe 186, 188
Tessin (Kanton) 149
Testmethode 31
Theoretical Sampling 350
Thurgau (Kanton) 365
Thüringen (Bundesland) 32, 81
Tierhaltung 103
TIMSS, siehe Trends in Mathematic and Science
Tirol (Bundesland) 299, 336, 383, 416
TKG, siehe Telekommunikationsgesetz
ToR, siehe Terms of Reference
TPFV, siehe Tabakpräventionsfonds
Transplantation 349
Trends in Mathematic and Science Stud (TIMSS) 159, 173, 194, 514, 524
Triangulation 464
Triple C Evaluierung 251
Triple Helix Modell 284
Tschernobyl 213
TÜV, siehe Technischer Überwachungsverein
- UBA, siehe Umweltbundesamt
Überzeugungs-Nutzung 501
UG, siehe Universitätsgesetz 42, 45
Uganda 256
- UK Department for International Development (DFID) 249
UKG, siehe Umweltkontrollgesetz
UMS, siehe Umweltmanagementsystem
Umwelt 57, 101, 256, 273, 419, 489
Umweltbewußtseinsforschung 205
Umweltbildung 231, 238
Umweltbildungsforschung 205
Umweltbundesamt 43
Umweltbundesamt (UBA) 92, 97, 224
Umweltcontrolling 205
Umweltevaluation 203, 222
Umweltforschung 231, 287
Umweltgesetz 226
Umweltkontrollgesetz (UKG) 43
Umweltmanagement 209
Umweltmanagementsystem (UMS) 208
Umweltministerium 75, 206, 207, 214, 238
Umweltpolitik 21, 43, 201, 236, 295, 511, 523
Umweltschutz 203, 229, 273
Umweltverträglichkeitsprüfung (UVP) 204, 228, 231, 237, 415
Umweltverträglichkeitsprüfungsgesetz (UVP_G) 43
Umweltwissenschaft 102, 111, 230
UN, siehe Vereinte Nationen
Unabhängigkeit 17, 234, 247, 261, 272, 275, 301, 495
UNESCO, siehe United Nations Educational, Scientific and Cultural Organization
UNFCCC, siehe United Nations Framework Convention on Climate Change
Unipark 12
United Nations Educational, Scientific and Cultural Organization (UNESCO)
United Nations Framework Convention on Climate Change (UNFCCC) 233
United States Government Accountability Office (GAO) 67
Univation, Köln 11
Universität 42, 48, 145, 163, 174, 177, 178, 188, 190, 195, 196, 218, 263, 265, 283, 285, 287, 289, 291, 297, 298, 299, 301, 305, 308, 312, 314, 315, 354, 411, 417, 436, 455, 459, 460, 512, 513, 525
Universität Bern 11
Universität Bielefeld 166, 290
Universität Bonn 166
Universität des Saarlandes 205, 275
Universität Duisburg-Essen 166

- Universität Halle 166
 Universität Jena 166
 Universität Kassel 290
 Universität Konstanz 378
 Universität Krems 177
 Universität Leipzig 166
 Universität Lüneburg 206
 Universität Mainz 168
 Universität Marburg 378
 Universität Potsdam 378
 Universität Salzburg 174
 Universität St. Gallen 388, 429
 Universität Wien 301
 Universität Zürich 11, 188
 Universitäts-Akkreditierungsgesetz 177
 Universitätsgesetz (UG) 42, 45, 48, 177, 179, 297
 Universitätsorganisationsgesetz (UOG) 297
 Unparteilichkeit 275
 Unter-Evaluation 304
 Unternehmensberatung 231
 Unternehmenssteuerreform 61
 UOG, siehe Universitätsorganisationsgesetz
 Update-Evaluierung 95
 Urban 423
 Uruguay-Runde 92
 USA 158, 329, 471, 477, 479, 486
 UVP, siehe Umweltverträglichkeitsprüfung
 UVP-G, siehe
 Umweltverträglichkeitsprüfungsgesetz

 Validität 338
 Valorisierung 62
 Value for money 48
 VBS, siehe Eidgenössisches Department für
 Verteidigung, Bevölkerungsschutz und Sport
 Verankerung 33, 48, 61
 Verankerung der Evaluationsfunktion 515
 Verankerung, institutionelle 20, 49, 50, 64, 478,
 485
 Verband der Ärzt/innen (FMH) 356
 Verbleibsanalyse 452
 Verbraucher/in 33
 Verbreitung 275
 Verbund Norddeutscher Universitäten 165
 Verbund Technischer Universitäten 165
 Verein EOS 173
 Vereinte Nationen (UN) 245, 255, 267, 273
 Verfassung 108, 342, 515
 Verfassungsreform 62

 Vergleichsgruppe 418
 Vergleichsgruppenuntersuchung 18, 135, 410,
 412
 Vergleichsring 29
 Verhütbare Krankheiten 342
 Verkehr 15, 217, 218, 231, 240, 523
 Verkehr, öffentlicher 60
 Verkehrsinfrastruktur 428
 Verkehrspolitik 213, 214, 232, 425
 Verlässlichkeit 456
 Vernehmlassungsverfahren 99, 369
 Verrechtlichung 382
 Versorgung, psychosoziale 323, 328
 Versorgung, psychotherapeutische 323
 Verteidigung 489
 Verteidigungspolitik 516
 Verwaltung 21, 23, 27, 29, 38, 61, 69, 82, 85, 86, 98,
 231, 305, 310, 361, 362, 411, 431, 433, 435, 437,
 460, 462, 466, 495, 496, 497, 500, 501, 503, 516
 Verwaltungsführung, leistungsorientierte 382,
 387
 Verwaltungsführung, wirkungsorientierte 62,
 229, 487
 Verwaltungskontrolle Parlamentarische, siehe
 Parlamentarische Verwaltungskontrolle
 Verwaltungskontrolle, wirkungsorientierte 62
 Verwaltungsmodernisierung 371, 373, 382, 383,
 384, 389, 392, 394, 395, 397
 Verwaltungspolitik 361, 371, 372
 Verwaltungsreform 28, 361, 362, 372, 382, 384,
 392, 393, 395
 Verwaltungswissenschaft 395
 Verwendung, instrumentelle 492, 494
 Verwendung, konzeptionelle 492, 494
 Verwendung, prozessbezogene 492, 494, 495
 Verwendung, symbolische 493, 494, 495
 Verwendungsforschung 486, 487
 Vetopunkt 369
 VirilistInnen 44
 Vollzugsaufwand 33
 Vollzugsevaluation 104, 231
 Vollzugskontrolle 53, 408
 Vorarlberg (Bundesland) 336, 383, 388, 415
 Vorarlberger Naturschutzrat 223
 Vorblatt 33

 Wahlsystem 361
 Wald 230
 WaP, siehe Wirkungsanalyseprogramm 388, 389

- Wasser 106, 256, 259
Wasserkraft-Kleinkraftwerke 259
Weiterbildung 19, 60, 157, 186, 354, 356
Weiterbildungsrecht 36
Weltbank 251, 266
Weltgesundheitsorganisation (WHO) 322, 329
Weltkommission für Umwelt und Entwicklung 203
Westschweiz 149
Wettbewerbsfähigkeit 42
WGL, siehe Wissenschaftsgemeinschaft G.W. Leibniz
WHO, siehe Weltgesundheitsorganisation
WiDuT, siehe Wissenschaftliche Dokumentations- und Transferstelle für Verwaltungsmodernisierung in den Ländern 376
Wiedereingliederung 149, 153
Wiedereingliederungspolitik 138, 139
Wien 218, 224, 225, 257, 259, 299, 336, 383
Wiener Institut für Sozialwissenschaftliche Dokumentation und Methodik (WISDOM) 17
Wiener Universität für Bodenkultur 301
Wiener Wissenschafts- und Technologiefonds (WWTF) 295
WIFO, siehe Österreichisches Institut für Wirtschaftsforschung 91, 299, 302
Wirksamkeit 14, 54, 55, 65, 78, 79, 106, 107, 137, 145, 150, 208, 210, 230, 237, 239, 240, 250, 263, 264, 267, 268, 312, 334, 343, 344, 348, 350, 365, 427, 428, 434, 436, 459, 503, 521
Wirksamkeitsdialog 448
Wirksamkeitsevaluation 138
Wirksamkeitsprüfung 349
Wirksamkeitsüberprüfung 55, 57, 58, 101, 340, 341, 342, 345, 425, 431
Wirkung 31, 65, 79, 97, 103, 130, 140, 142, 145, 147, 152, 153, 199, 204, 208, 219, 221, 223, 230, 231, 232, 239, 240, 251, 265, 267, 268, 270, 271, 272, 274, 275, 277, 279, 291, 397, 416, 419, 426, 427, 429, 436, 446, 449, 459, 463, 498
Wirkungsanalyse 78, 112, 134, 138, 141, 153, 269, 270, 271, 302, 303, 368, 371, 388, 406, 408, 416, 496
Wirkungsanalysen 272
Wirkungsanalyseprogramm (WaP) 388, 389
Wirkungsbeurteilung 53, 57
Wirkungsermittlung 53
Wirkungsreichung 274
Wirkungsforschung 18, 30, 286, 485
Wirkungshypothese 54
Wirkungsindikator 142, 143, 152, 271, 286, 288, 289, 429
Wirkungskette 97, 147, 153, 221, 229, 232, 251, 270, 367, 434
Wirkungskontrolle 53, 57, 78, 138, 408, 504
Wirkungsmessung 272, 278
Wirkungsmodell 428
Wirkungsorientierte Jugendhilfe 447
Wirkungsprognose 53, 55
Wirkungsüberprüfung 53
Wirkungsvereinbarung 142
Wirkungszusammenhang 63, 146, 147
Wirtschaft 38, 57, 101, 433, 435, 456, 509
Wirtschaftlichkeit 29, 56, 65, 66, 78, 98, 103, 230, 325, 350, 391, 415, 500, 504
Wirtschaftlichkeitsanalyse 110
Wirtschaftlichkeitsprüfung 56, 406
Wirtschaftsförderung 433
Wirtschaftsforschung 287, 288
Wirtschaftsforschungsinstitut (WIFO) 452
Wirtschaftskammer Österreichs (WKÖ) 90, 94, 225
Wirtschaftsministerium 206, 207
Wirtschaftspolitik 236, 295, 404, 426
Wirtschaftsstrukturgesetz 405
Wirtschaftsverband 37
Wirtschaftswissenschaft 205, 408, 410, 412
WISDOM, siehe Wiener Institut für Sozialwissenschaftliche Dokumentation und Methodik
Wissenschaftliche Dokumentations- und Transferstelle für Verwaltungsmodernisierung in den Ländern (WiDuT) 376
Wissenschaftliche Kommission für Niedersachsen 166
Wissenschaftliches Institut der Allgemeinen Ortskrankenkassen 356
Wissenschaftsgemeinschaft G.W. Leibniz (WGL) 284, 288, 289, 290
Wissenschaftsmanagement 283
Wissenschaftsministerium 301
Wissenschaftspolitik 295
Wissenschaftsrat 167, 285, 286, 287, 290, 291
Wissenschaftszentrum Berlin (WZB) 290
Wissensmanagement 279
WKÖ, siehe Wirtschaftskammer Österreichs

- WODC, siehe Zentrum für wissenschaftliche
 Forschung und Dokumentation
 Wohlfahrt, soziale 66
 Wohlfahrtsverband 19
 Wohnbauförderung 60
 Wohnen 460
 Wohnungsbau 467
 Workshop 418
 World Food 249
 World Trade Organisation (WTO) 93, 100
 WOV, siehe Wirkungsorientierte Verwaltung
 WR, Wissenschaftsrat
 WTO, siehe World Trade Organisation
 Wuppertal-Institut 205
 WWTF, siehe Wiener Wissenschafts- und
 Technologiefonds
 WZB, siehe Wissenschaftszentrum Berlin

 Zeitreihenanalyse 428
 Zentralamerika 259, 279
 Zentrale Evaluations- und
 Akkreditierungsagentur (ZEvA) 165, 166
 Zentralismus 170
 Zentralregierung 66
 Zentrum für empirische Bildungsforschung 158
 Zentrum für Schulentwicklung 174, 195

 Zentrum für Schulversuche und
 Schulentwicklung 171
 Zentrum für Technologiefolgen-Abschätzung 356
 Zentrum für Wissenschafts- und
 Technologiestudien (CEST) 310, 311, 315
 Zentrum für wissenschaftliche Forschung und
 Dokumentation (WODC) 66
 Zertifizierung 186, 229, 237, 452, 456, 458
 ZEvA, siehe Zentrale Evaluations- und
 Akkreditierungsagentur
 Zielerreichung 65, 78, 187, 229, 268, 270, 274, 337,
 408, 411, 427, 504
 Zielerreichungsgrad 54
 Zielerreichungskontrolle 78
 Zielformulierung 365
 Zielgruppe 268, 462
 Zitationsanalyse 514
 Zoll- oder Steuerbehörden 516
 Zufallsstichprobe 272, 277, 279
 Zug (Kanton) 141
 Zürich (Kanton) 140, 141, 347, 365, 429
 Zweckmäßigkeit 415, 503
 Zweistufenplan 120
 Zweite Republik 51
 Zwei-Wellen-Theorie 396
 Zwischenevaluation 364, 425, 427

Literatur

- Abs, Herman Josef/Klieme, Eckhard (2005): Standards für schulbezogene Evaluation. *Zeitschrift für Erziehungswissenschaft* Beiheft 4: 46-64.
- Abs, Herman Josef/Maag Merki, Katharina/Klieme, Eckhard (2006): Grundlegende Gütekriterien für Schulevaluation. In: Böttcher, Wolfgang/Holtappels, Heinz Günter/Brohm, Michaela (Hrsg.): *Evaluation im Bildungswesen. Eine Einführung in Grundlagen und Praxisbeispiele*. Weinheim: Juventa, S. 97-108.
- Achermann, Christin/Chimienti, Milena (2006): Migration, Prekarität und Gesundheit. Ressourcen und Risiken von vorläufig Aufgenommenen und Sans Papiers in Genf und Zürich. Neuenburg: Schweizerisches Forum für Migration.
- Ackermann, Heike (2001): Möglichkeiten und Grenzen einer Evaluation von Schule und Unterricht. *Salzburger Beiträge zur Erziehungswissenschaft* 5(1): 87-95.
- Aeppli, Daniel (2000): Die Ausgesteuerten. Situationsbericht – Dritte Studie. Bern: Haupt.
- Aeppli, Daniel/Hoffmann, Brigitte/Theiss, Roland (1998): Ausgesteuerte in der Schweiz. Ein Situationsbericht. Bern: Haupt.
- Aeppli, Daniel/Hotz, Cora/Hugentobler, Valerie/Theiss, Roland (1996): Die Situation der Ausgesteuerten. Bern: Haupt.
- Aeppli, Daniel/Kälin, Roli/Ott, Walter/Peters, Matthias (2004): Wirkungen von Beschäftigungsprogrammen für ausgesteuerte Arbeitslose. Zürich: Rüeegger.
- AGEZ/FORUM/ÖKOBÜRO (2006): Trendumkehr in Richtung Nachhaltigkeit – jetzt! Forderungspapier für Regierungsverhandlungen 2006. Arbeitsgemeinschaft Entwicklungszusammenarbeit (AGEZ), Forum Nachhaltiges Österreich (FORUM) und ÖKOBÜRO (Koordinationsstelle österreichischer Umweltorganisationen). (www.nachhaltigkeit.at/strategie/forum/pdf/Positionspapier-060913.pdf [22.8.2007])
- Aichholzer, Georg/Lasnigg, Lorenz/Schienstock, Gerhard (1984): Arbeitsmarktpolitische Qualifizierungsmaßnahmen für Jugendliche. In: Bundesministerium für Soziale Verwaltung (Hrsg.): *Jugendliche auf dem österreichischen Arbeitsmarkt*, Forschungsberichte aus Sozial- und Arbeitsmarktpolitik Nr. 12. Wien: Eigenverlag, S. 115-142.
- AID-EE (2007): Success and failure in energy efficiency policies. Ex-post evaluations of 20 instruments to improve energy efficiency across Europe. Report prepared within the framework of the AID-EE project. Utrecht: Ecofys.
- Airaghi, Angelo/Busch, Niels E./Georgiou, Luke/Kuhlmann, Stefan/Ledoux, Marc J./van Raan, Anthony F.J./Viana Baptista, Jorge (1999): Options and limits for assessing the socio-economic impact of European RTD programmes. Report to the European Commission, DG XII, Evaluation Unit. Brussels/Luxembourg: Office for Official Publications of the European Communities.
- Aliesch, Beat/Sauter, Joseph/Hartmann, Peter/Kuster, Jürg/Meier, Hans-Rudolf (2006): Räumliche Auswirkungen des Vereinatunnels – eine ex-post Analyse. Zusammenfassung. Bern: ARE.
- Alkin, Marvin C. (2004): Evaluation roots. Tracing theorists' views and influences. Thousand Oaks: Sage.
- Alkin, Marvin C. (2005): Evaluation use. In: Mathison, Sandra (ed.): *Encyclopedia of evaluation*. Thousand Oaks: Sage, p. 145.
- Alkin, Marvin C./Solmon, Lewis C. (eds) (1983): *The costs of evaluation*. Beverly Hills: Sage.
- Alkin, Marvin C./Stecher, Brian (1983): A study of evaluation costs. In: Alkin, Marvin C./Solmon, Lewis C. (eds): *The costs of evaluation*. Beverly Hills: Sage, pp. 119-132.

- Altrichter, Herbert (1999a): Evaluation als Alltäglichkeit, als Profession und als Interaktion. In: Thonhauser Josef/Patry Jean-Luc (Hrsg.): Evaluation im Bildungsbereich. Wissenschaft und Praxis im Dialog. Innsbruck: Studienverlag, S. 103-120.
- Altrichter, Herbert (1999b): Einige Anmerkungen zur Diskussion mit Detlev Leutner. In: Thonhauser Josef/Patry Jean-Luc (Hrsg.): Evaluation im Bildungsbereich. Wissenschaft und Praxis im Dialog. Innsbruck: Studienverlag, S. 133-136.
- Altrichter, Herbert/Buhren, Claus G. (1997): Schulen vermessen oder entwickeln? Zur Bedeutung von Evaluation in Schulentwicklungsprozessen. *Journal für Schulentwicklung* 1(3): 4-21.
- Altrichter, Herbert/Gstetter, Peter (1993): Aktionsforschung – ein abgeschlossenes Kapitel in der Geschichte der deutschen Sozialwissenschaft? *Sozialwissenschaftliche Literatur Rundschau* 16(26): 67-83.
- Altrichter, Herbert/Messner, Elgrid/Posch, Peter (2004): Schulen evaluieren sich selbst. Ein Leitfaden. Seelze: Kallmeyer.
- Altrichter, Herbert/Posch, Peter (1999): Wege zur Schulqualität. Innsbruck: Studienverlag.
- Angele, Hans-Christian/Gerheuser, Frohmüt/Rüegg, Erwin/Schelske, Oliver/Lippuner, Christoph (2003): Einbettung von EnergieSchweiz in die Verkehrspolitik des Bundes. Bern: Bundesamt für Energie.
- Arbeitsgruppe Gesetzesevaluation (1991): Die Wirksamkeit staatlichen Handelns besser ermitteln: Probleme, Möglichkeiten, Vorschläge. Schlussbericht. Bern: Eidgenössisches Justiz- und Polizeidepartement.
- Arbeitskreis „Evaluation von Entwicklungspolitik“ in der DeGEval (Hrsg.) (2005): Lernen aus Evaluationen. Leitfaden für Geberinstitutionen und GutachterInnen. Erarbeitet von der AG „Lernen aus Evaluationen“. Arbeitspapiere Nr. 2. (www.uni-saarland.de/fak5/stockmann/akepol/download/AK_ARBEITSPAPIER2.pdf [29.4.2008])
- ARC systems research/ÖAR Regionalberatung (2004): Evaluierung der Inkubatorfunktion und der regionalen Innovationsfunktion der RIZ-NÖ Holding GmbH. Wien: ARC systems research.
- ARE – Bundesamt für Raumentwicklung (2004): Nachhaltigkeitsbeurteilung: Rahmenkonzept und methodische Grundlagen. Bern: ARE.
- ARE – Bundesamt für Raumentwicklung /seco – Staatssekretariat für Wirtschaft (2006): Agglomerationspolitik des Bundes. Zwischenbericht 2006. Bern: ARE/seco.
- ARGEV – Interkantonale Arbeitsgemeinschaft Externe Evaluation von Schulen (2005): Jahresbericht über die Aktivitäten der ARGEV 2005. Frauenfeld/Winterthur: ARGEV. (www.argev.ch/darkstar_system/system_daten/argev/doc_attachments/rw_argev_page_ueberuns_8.pdf [25.8.2007]).
- Argyris, Chris/Schön, Donald (1999): Die lernende Organisation. Grundlagen, Methode, Praxis. Stuttgart: Klett-Cotta.
- Arnold, Benedikt/Drilling, Matthias/Flückiger, Martin/Storni, Marco (2003): Kurzfassung Evaluation ‚Projekt Integration‘. Basel: Fachhochschule für Soziale Arbeit beider Basel.
- Arnold, Erik (ed.) (2004): Evaluation of the Austrian Industrial Research Promotion Fund (FFF) and the Austrian Science Fund (FWF). Synthesis Report. Brighton: Technopolis. (www.fteval.at/files/evstudien/FFF_FWF_Synthesis_Report.pdf [Stand: 28. 3. 2008])
- ART (diverse Jahrgänge): Hauptbericht. Ettenhausen: Agroscope Reckenholz-Tänikon.
- Atria, Moira/Reimann, Ralph/Spiel, Christiane (2006): Qualitätssicherung durch Evaluation. Die Bedeutung von Zielexplication und evaluativer Haltung. In: Steinebach, Christoph (Hrsg.): Handbuch Psychologische Beratung. Stuttgart: Klett-Cotta, S. 574-586.
- Avenir Suisse (2006): Der befreite Bauer. Zürich: Verlag Neue Zürcher Zeitung.
- Baccini, Peter/Oswald, Franz (2003): Netzstadt. Einführung zum Stadtentwerfen. Basel: Birkhäuser.
- Bachmann, Helmut/Iby, Manfred/Kern, Augustin/Osinger, Dietmar/Radnitzky, Edwin/Spacht, Werner (1996): Auf dem Weg zu einer besseren Schule. Evaluation der Schulautonomie in Österreich, Auswirkungen der 14. SchOG-Novelle. Bildungsforschung des Bundesministeriums für Unterricht und kulturelle Angelegenheiten, Band 11. Innsbruck: Studienverlag.

- Bachmann, Ruth/Bieri, Oliver/Balthasar, Andreas (1998): Die Situation der ausgesteuerten Stellenlosen im Kanton Luzern. Luzern: Interface Politikstudien.
- Bachtler, John/Wren, Colin (2006): Evaluation of European Union cohesion policy: Research questions and policy challenges. *Regional Studies* 40(2): 143-153.
- Badura, Bernhard/Siegrist, Johannes (Hrsg.) (1999): Evaluation im Gesundheitswesen. Ansätze und Ergebnisse. Weinheim: Juventa.
- Baethge, Martin/Achtenhagen, Frank/Weber, Susanne/Arends, Lena/Babic, Edwin (2006): Berufsbildungs-PISA. Machbarkeitsstudie. Stuttgart: Franz Steiner Verlag.
- BAG – Bundesamt für Gesundheit (Hrsg.) (1997): Leitfaden für die Planung von Projekt- und Programmevaluation. Bern: BAG.
- BAG – Bundesamt für Gesundheit (Hrsg.) (2002): Evaluationen des BAG im Suchtbereich 1990-2000. Bern: BAG.
- BAG – Bundesamt für Gesundheit (Hrsg.) (2004): Prävention im Sport. Was haben wir gelernt? Bern: BAG.
- BAG – Bundesamt für Gesundheit (Hrsg.) (2005a): Glossar von Evaluationsbegriffen. Bern: BAG.
- BAG – Bundesamt für Gesundheit (Hrsg.) (2005b): Assessing the effects of public health policies and actions. Evaluation strategic framework. Bern: BAG.
- BAG – Bundesamt für Gesundheit (Hrsg.) (2005c): Evaluation der Stiftung Gesundheitsförderung Schweiz. Bern: BAG.
- BAG – Bundesamt für Gesundheit (Hrsg.) (2005d): Programm Evaluation Komplementärmedizin – Schlussbericht. Bern: BAG.
- BAG – Bundesamt für Gesundheit (Hrsg.) (2006a): Gesetzliche Grundlagen von (Ressort-)Forschung und Evaluation im BAG. Bern: BAG. (www.bag.admin.ch/dokumentation/gesetzgebung [10.3.2008])
- BAG – Bundesamt für Gesundheit (Hrsg.) (2006b): Forschung Migration und Gesundheit 2002-2007. Sammelband. Bern: BAG.
- Balthasar, Andreas (1998): Nutzen von Evaluationen. Luzern: Interface. (www.interface-politikstudien.ch/downloads/deutsch/BI_Nutzen_von_Evaluationen_2_d.pdf [15.2.2007])
- Balthasar, Andreas (2000): Energie 2000. Programmwirkungen und Folgerungen aus der Evaluation. Chur: Rüegger.
- Balthasar, Andreas (2002): Was und wie wird evaluiert? Eine Übersicht über die Evaluationspraxis in der Energiepolitik. *LeGes – Gesetzgebung & Evaluation*, 13(1): 23-38.
- Balthasar, Andreas (2006): The effects of the institutional design on the utilization of evaluation. *Evaluation* 12(3): 353-371.
- Balthasar, Andreas (2007): Institutionelle Verankerung und Verwendung von Evaluationen. Praxis und Verwendung von Evaluationen in der schweizerischen Bundesverwaltung. Zürich: Rüegger.
- Balthasar, Andreas/Bättig, Christoph/Zimmermann, Willi/Kurz Gygax, Rachel (1995): Der Vollzug der Bundespolitik durch die Kantone – eine Metaevaluation. Studie zuhanden der Parlamentarischen Verwaltungskontrollstelle. Bern: EDMZ.
- Balthasar, Andreas/Rieder, Stefan (2000): Learning from evaluations. Effects of the evaluation of the Swiss Energy 2000 programme. *Evaluation* 6(3): 245-260.
- Balzer, Lars (2005): Wie werden Evaluationsprojekte erfolgreich? Ein integrierender theoretischer Ansatz und eine empirische Studie zum Evaluationsprozess. Landau: Verlag Empirische Pädagogik.
- Balzer, Lars (2006): "Wie werden Evaluationsprojekte erfolgreich?" – Ergebnisse einer Delphistudie. In: Böttcher, Wolfgang/ Holtappels, Heinz G./Brohm, Michaela (Hrsg.): Evaluation im Bildungswesen – Eine Einführung in Grundlagen und Praxisbeispiele. Weinheim: Juventa, S. 123-135.
- Barz, Andreas (1998): Evaluation im deutschen Hochschulsystem: Ziele, Instrumente, Erfahrungen, Trends. (www.degeval.de/koeln1998/barz.htm [31.08.2007])

- Basler, Ernst (1984): Bericht zu „Sinn und Erfolg von Projekten der Schweizerischen Entwicklungshilfe in Nepal“ zuhanden der Geschäftsprüfungskommission des Nationalrates. Bern: EDMZ.
- Battaglia, Markus/Junker, Christoph (2001): Auswirkungen der Aufnahme von präventivmedizinischen Leistungen in den Pflichtleistungskatalog. Teilstudie Impfungen im Schulalter. Bern: Bundesamt für Sozialversicherung.
- Bättig, Christoph/Kanaan, Sami/Matti, Daniel (2002): Zwischenevaluation Regio Plus und Landwirtschaftliche Absatzförderung. Luzern: ARGE Interface/evaluanda.
- Bauer, Helfried/Biwald, Peter/Dearing, Elisabeth (Hrsg.) (2003): Öffentliches Management in Österreich. Realisierungen und Perspektiven. Wien: KDZ Managementberatungs- und Weiterbildungs GmbH.
- Bauer, Helfried/Biwald, Peter/Dearing, Elisabeth (Hrsg.) (2005): Public Governance. Öffentliche Aufgaben gemeinsam erfüllen und effektiv steuern. Wien: Neuer Wissenschaftlicher Verlag.
- Bauer, Michael/Bogumil, Jörg/Knill, Christoph/Ebinger, Falk/Krapf, Sandra/Reißig, Kristin (2007): Modernisierung der Umweltverwaltung. Reformstrategien und Effekte in den Bundesländern. Berlin: Edition Sigma.
- Bauer, Tobias/Baumann, Beat/Künzli, Kilian (1999): Evaluation der Regelung des Zwischenverdienstes in der Schweiz. Beiträge zur Arbeitsmarktpolitik Nr. 1. Bern: Seco.
- Baumert, Jürgen/Bos, Wilfried/Lehmann, Rainer H. (2000): TIMSS/III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie. Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn. 2 Bände. Opladen: Leske+Budrich.
- Baumgartner, Edgar (2000): Evaluationen für wen? Meta-evaluative Anmerkungen zu Evaluationssettings in der Sozialen Arbeit. In: Müller-Kohlenberg, Hildegard/Münstermann, Klaus (Hrsg.): Qualität von Humandienstleistungen. Opladen: Leske+Budrich, S. 241-247.
- Baumgartner, Edgar (2003): Der Nutzen der betrieblichen Sozialarbeit. Eine Kosten-Nutzen-Analyse in zwei Unternehmen. DoRe-Abschlussbericht. Olten: FHNW (unveröffentlicht).
- Baumgartner, Edgar/Greiwe, Stephanie/Schwarb, Thomas (2004): Die berufliche Integration von behinderten Personen in der Schweiz. Forschungsbericht 04/04. Bern: Bundesamt für Sozialversicherung.
- Baur, Priska/Meyer, Leo/Rieder, Peter (1994): Politökonomische Analyse der Getreidepolitik. In: Nationales Forschungsprogramm 27 (Hrsg.): Evaluationen in der Agrarpolitik II, Reihe Schlussberichte. Bern: SNF.
- Beck, Ulrich (1986): Risikogesellschaft. Frankfurt: Suhrkamp.
- Becker, Franz J.E. (1983): Adressatenbezogene Analyse und didaktische Entscheidung bei der Planung von Lehrerfortbildungsveranstaltungen. *Unterrichtswissenschaft* 11(3): 264-275.
- Behrendt, Heiko/Eser, Thimo/Scherer, Roland/Schnell, Klaus-Dieter/Strauf, Simone (2003): Bericht zur Halbzeitbewertung INTERREG III A. Alpenrhein-Bodensee-Hochrhein. St. Gallen: IDT-HSG.
- Behrens, Johann/Langer, Gero (2004): Evidence-based Nursing. Vertrauensbildende Entzauberung der „Wissenschaft“. Qualitative und quantitative Methoden bei täglichen Pflegeentscheidungen. Bern: Huber.
- Bellmann, Lutz/Schwengler, Barbara (2000): Einzelbetriebliche Erfolgskontrolle – Weg zwei. Auswertung des IAB-Betriebspanels zur Inanspruchnahme von Mitteln der GA. Gutachten im Rahmen der Gemeinschaftsaufgabe "Verbesserung der regionalen Wirtschaftsstruktur" (GA). Nürnberg: Institut für Arbeitsmarkt- und Berufsforschung.
- Belser, Katharina (2006): Das Marie Heim-Vögtlin-Programm in den Jahren 1991-2002. Evaluationsbericht. Bern: Schweizerischer Nationalfonds.
- Benz, Arthur (2004): Institutionentheorie und Institutionenpolitik. In: Benz, Arthur/Siedentopf, Heinrich/Sommermann, Karl-Peter (Hrsg.): Institutionenwandel in Regierung und Verwaltung. Festschrift für Klaus König zum 70. Geburtstag. Berlin: Duncker & Humblot, S. 19-31.

- Benz, Arthur (2005): Verwaltung als Mehrebenensystem. In: Blanke, Bernhard/von Bandemer, Stephan/Nullmeier, Frank/Wewer, Götz (Hrsg.): Handbuch zur Verwaltungsreform. 3. Auflage. Wiesbaden: VS Verlag, S. 18-26.
- Benz, Arthur/Fürst, Dietrich (2002): Policy learning in regional networks. *European Urban and Regional Studies* 9(1): 21-35.
- Berger, Christa/Uchtenhagen, Ambros (2000): Infektionsprophylaxe und Gesundheitsförderung im Strafvollzug. Forschungsbericht. Zürich: ISGF.
- Bergmann, Andreas/Bruderer, Ernst/Vogt Gehri, Franziska/Gamper, Andreas (2006): Benchmarking Schulen Sekundarstufe II: Berichterstattung Herbst 2006, Teilprojekte 1, 2 und 3. Zürich: ZHW.
- Bergschmidt, Angela/Dirksmeyer, Walter/Efken, Josef/Forstner, Bernhard/Utrecht, Inge (2006): Proceedings of the European workshop on the evaluation of farm investment support, investment support for improvement of processing and marketing of agricultural products. Braunschweig: Herausgegeben vom Institut für Betriebswirtschaft der FAL und dem Institut für Marktanalyse und Agrarhandelspolitik der FAL. (literatur.fal.de/fallitdok_extern/bitv/zi040499.pdf [1.8.2007])
- Bericht des Bundesrates (2006): Bericht des Bundesrates vom 15. Februar 2006 über die Wirksamkeit des Gleichstellungsgesetzes in Erfüllung der Motion Vreni Hubmann 02.3142. *Bundesblatt* 12: 3161.
- Bericht von den Staatspolitischen Kommissionen der eidgenössischen Räte (1996): Kompetenzverteilung zwischen Bundesversammlung und Bundesrat. *Bundesblatt* 2: 428-534.
- Bertelsmeier, Markus/Kleinhanß, Werner/Offermann, Frank (2004): Rahmenbedingungen und Folgen unterschiedlicher Milchmarktpolitiken in Nordrhein-Westfalen. Arbeitsbericht des Instituts für Betriebswirtschaft, H. 04/2004. Braunschweig: Bundesforschungsanstalt für Landwirtschaft. (literatur.fal.de/fallitdok_extern/bitv/zi033821.pdf [20.2.2008])
- Beule, Wolfgang/Müller, Max/Spicher, Benjamin (2004): Patientenzufriedenheit: Die Helsana Studie. Eine Befragung über die Qualität von Arztpraxen. Schriftenreihe Schweizerische Gesellschaft für Gesundheitspolitik Bd. 77. Bern: Schweizerische Gesellschaft für Gesundheitspolitik.
- Beyme, Klaus von (2007): Föderalismus und regionales Bewusstsein. Ein internationaler Vergleich. München: Beck.
- Beywl, Wolfgang (1988): Zur Weiterentwicklung der Evaluationsmethodologie. Grundlegung, Konzeption und Anwendung eines Modells der responsiven Evaluation. Frankfurt: Lang.
- Beywl, Wolfgang (1999): Standards für Evaluation. *DeGEval-News* 1999(1): 1-4.
- Beywl, Wolfgang/Mecklenburg, Roland/Richard, Jörg/Schneid, Theo/Wonik, Martin (Hrsg.) (2001): Evaluation im Alltag. Jugendverbände untersuchen ihre Wirkungen. Münster: Votum.
- Beywl, Wolfgang/Speer, Sandra/Kehr, Jochen (2004): Wirkungsorientierte Evaluation im Rahmen der Armuts- und Reichtumsberichterstattung. Forschungsprojekte-Lebenslagen des Bundesministeriums für Gesundheit, A-232. Bonn: BMGS (Bundesministerium für Gesundheit und soziale Sicherung). (www.univation.org/download/Evaluation_der_Armuts_und_Reichtumsberichterstattung.pdf [31.05.2007])
- BFS – Bundesamt für Statistik (2006): F+E der Schweiz 2004. Neuchâtel: BFS.
- Bickman, Leonard (1987): The function of program theory. In: Bickman, Leonard (ed.): Using program theory in evaluation. *New Directions for Program Evaluation* 33: 5-18.
- Bickner, Robert E. (1980): Pitfalls in the analysis of costs. In: Majone, Giandomenico/Quade, Edward S. (eds): Pitfalls of analysis. Chichester: John Wiley, pp. 57-69.
- Bieger, Thomas/Rey, Michel/Scherer, Roland/Schnell, Klaus-Dieter/Sfar, Doris/Strebel, Nicole/Rheinhard, Mark (2004): Evaluation der Investitionshilfe für Berggebiete (IHG). Schlussbericht. St. Gallen/Lausanne: IDT-HSG/C.E.A.T.
- BIGA – Bundesamt für Industrie, Gewerbe und Arbeit (1993): Reform der öffentlichen Arbeitsvermittlung in der Schweiz. Beiträge zur Arbeitsmarktpolitik Nr. 3. Bern: BIGA.

- Binder, Bruno/Strehl, Franz/Leitl, Barbara (1999): Studie zur Berechnung und Abschätzung der Folgekosten von Gesetzen in Österreich unter Berücksichtigung ausländischer Maßnahmen und Erfahrungen. Linz: Johannes Kepler Universität.
- Binder, Michael (2005): The Austrian Research Promotion Agency (FFG). *Plattform ffeval Newsletter* 25: 39-40.
- Bindeus, Heinz/Inquart, Ernst/Biwald, Peter (2005): Interkommunaler Vergleich Oberösterreich – Benchmarking zwischen 10 Städten – Ergebnisse, Erfahrungen, Erfolgsfaktoren. *Österreichische Gemeindezeitung* 3: 32-38.
- Bisig, Brigitte/Gutzwiller, Felix (Hrsg.) (2004): Gesundheitswesen Schweiz: Gibt es Unter- oder Überversorgung? Band 1 Gesamtübersicht, Band 2 Detailresultate. Zürich: Rüegger Verlag.
- Biwald, Peter (2005): Voneinander und miteinander lernen – Benchmarking in Bezirksverwaltungsbehörden. *Österreichische Gemeindezeitung* 4: 45.
- BJ – Bundesamt für Justiz (2004): Wirksamkeit von Bundesmassnahmen, Vorschläge zur Umsetzung von Artikel 170 der Bundesverfassung bei Bundesrat und Bundesverwaltung, Bericht der Interdepartementalen Kontaktgruppe „Wirkungsprüfungen“ an die Generalsekretärenkonferenz der Schweizerischen Eidgenossenschaft. Bern: BJ.
- BJ – Bundesamt für Justiz (2006): Entscheide des Bundesrats vom 3. November 2004 zur Verstärkung der Wirksamkeitsüberprüfungen bei Bundesrat und Bundesverwaltung (ergänzt durch die Empfehlungen der Generalsekretärenkonferenz (GSK) vom 19. Dezember 2005 über Qualitätsstandards). Konsolidierte Übersicht über die Entscheide. Empfehlungen zum Zeitpunkt der Umsetzung. Bern: BJ. (www.bj.admin.ch/etc/medialib/data/staat_buerger/evaluation.Par.0002.File.tmp/konsolidierte-uebersicht-d.pdf [5.3.2008])
- Block, Hans-Jürgen/Krull, Wilhelm (1990): What are the consequences? Reflections on the impact of evaluations conducted by a science policy advisory body. *Scientometrics* 19(5-6): 427-437.
- BLW – Bundesamt für Landwirtschaft (1999): Konzeptbericht Evaluation der Ökomassnahmen und Tierhaltungsprogramme. Bern: BLW.
- BLW – Bundesamt für Landwirtschaft (2003): Evaluation im BLW: Evaluationsleitfaden und Evaluationsstrategie 2004/07. Bern: BLW.
- BLW – Bundesamt für Landwirtschaft (diverse Jahrgänge): Agrarbericht. Bern: BLW.
- BMBF – Bundesministerium für Bildung und Forschung (2005): Das BMBF-Förderprogramm InnoRegio – Ergebnisse der Begleitforschung. Berlin: BMBF.
- BMEIA (2008): Leitlinien für Evaluierung in der österreichischen Entwicklungszusammenarbeit. Wien: OEZA. (www.bmeia.gv.at/fileadmin/user_upload/bmeia/media/2-Aussenpolitik_Zentrale/EZA/1_leitlinien_dt_.pdf [4.4.2008])
- BMFSFJ (Hrsg.) (1998): Leistungen und Grenzen der Heimerziehung. Ergebnisse einer Evaluationsstudie stationärer und teilstationärer Erziehungshilfen. Forschungsprojekt Jule. Schriftenreihe des Bundesministeriums für Familie, Senioren, Frauen und Jugend, Band 170. Stuttgart: Kohlhammer.
- BMFSFJ (Hrsg.) (2002a): Elfter Kinder- und Jugendbericht. Bericht über die Lebenslagen junger Menschen und die Leistungen der Kinder- und Jugendhilfe in Deutschland. Berlin: BMFSFJ.
- BMFSFJ (Hrsg.) (2002b): Effekte erzieherischer Hilfen und ihre Hintergründe. Schriftenreihe des Bundesministeriums für Familie, Senioren, Frauen und Jugend, Band 219. Stuttgart: Kohlhammer.
- BMLFUW (2002): Österreichs Zukunft nachhaltig gestalten – Die österreichische Strategie zur nachhaltigen Entwicklung. Wien: Bundesministerium für Land- und Forstwirtschaft, Umwelt und Wasserwirtschaft.
- BMSG – Bundesministerium für soziale Sicherheit und Generationen (Hrsg.) (2001): Messung der Ergebnisqualität im Krankenhaus. Wien: BMSG.
- BMSG – Bundesministerium für soziale Sicherheit und Generationen (ed.) (2003): Health Care Outcomes: The Path to Understanding Processes. Conference Report. Wien: BMSG.
- BMWA (2005): Arbeitsmarkt und Arbeitsmarktpolitik in Österreich. Wien: BMWA.

- BMWA (2006): Zielvorgaben des Bundesministers für Wirtschaft und Arbeit zur Durchführung der Arbeitsmarktpolitik an das Arbeitsmarktservice. Wien: BMWA. (www.bmwa.gv.at/NR/rdonlyres/74B2E6E8-F62D-4936-A7EA-08D9411D5A17/0/ZielvorgabenFassung0806.pdf [6.6.2007])
- BMZ (2005): Mehr Wirkung erzielen: Die Ausrichtung der deutschen Entwicklungszusammenarbeit auf die Millenniums-Entwicklungsziele. Die Umsetzung der Paris Declaration on Aid Effectiveness. Bonn: BMZ. (www.bmz.de/de/service/infothek/fach/spezial/spezial130pdf.pdf [29.4.2008])
- BMZ (2006): Evaluierungskriterien für die deutsche bilaterale Entwicklungszusammenarbeit. Bonn: BMZ. (www.bmz.de/de/ErfolgKontrolle/ziele/infos/starterkit/evaluierungskriterien.pdf [4.4.2008])
- BMZ (2007): Leitlinien für die bilaterale Finanzielle und Technische Zusammenarbeit mit Kooperationspartnern der deutschen Entwicklungszusammenarbeit vom 1. März 2007. Bonn: BMZ.
- Bochsler, Daniel/Koller, Christophe/Sciarini, Pascal/Traimond, Sylvie/Trippolini, Ivar (2004): Die Schweizer Kantone unter der Lupe. Behörden, Personal, Finanzen. Bern: Haupt.
- Bock, Karin (2002): Die Kinder- und Jugendhilfe. In: Thole, Werner (Hrsg.): Grundriss Soziale Arbeit. Ein einführendes Handbuch. Opladen: Leske & Budrich, S. 299-315.
- Bogner, Daniel (2003): Evaluierung der Wirkung des ÖPUL auf die österreichische Kulturlandschaft in ausgewählten Regionen mit Hilfe eines Landschaftsmodells. Klagenfurt: Bogner & Golob OEG Büro für Ökologie und Landwirtschaft.
- Bogumil, Jörg/Ebinger, Falk (2005): Die Große Verwaltungsstrukturreform in Baden-Württemberg. Erste Umsetzungsanalyse und Überlegungen zur Übertragbarkeit der Ergebnisse auf NRW. Schriftenreihe der Stiftung Westfalen-Initiative, Band 9. Ibbenbüren: Ibbenbürener Vereinsdruckerei GmbH.
- Bogumil, Jörg/Grohs, Stephan/Kuhlmann, Sabine (2006): Ergebnisse und Wirkungen kommunaler Verwaltungsmodernisierung in Deutschland – Eine Evaluation nach zehn Jahren Praxiserfahrung. In: Bogumil, Jörg/Jann, Werner/Nullmeier, Frank (Hrsg.): Politik und Verwaltung. *PVS Sonderheft 37*: 151-184.
- Bogumil, Jörg/Grohs, Stephan/Kuhlmann, Sabine/Ohm, Anna (2007): Zehn Jahre Neues Steuerungsmodell. Eine Bilanz kommunaler Verwaltungsmodernisierung. Berlin: Edition Sigma.
- Bogumil, Jörg/Jann, Werner (2005): Verwaltung und Verwaltungswissenschaft in Deutschland. Einführung in die Verwaltungswissenschaft. Wiesbaden: VS Verlag.
- Bogumil, Jörg/Kottmann, Steffen (2006): Verwaltungsstrukturreform – die Abschaffung der Bezirksregierungen in Niedersachsen. Schriftenreihe der Stiftung Westfalen-Initiative, Band 11. Ibbenbüren: Ibbenbürener Vereinsdruckerei GmbH.
- Bohnsack, Ralf/Nentwig-Gesemann, Iris (Hrsg.) (2008): Dokumentarische Evaluationsforschung. Theoretische Grundlagen und Beispiele aus der Praxis. Opladen: B. Budrich (in Vorbereitung).
- Böhret, Carl/Konzendorf, Götz (1997): Ko-Evolution von Gesellschaft und funktionalem Staat – Ein Beitrag zur Theorie der Politik. Wiesbaden: Westdeutscher Verlag.
- Böhret, Carl/Konzendorf, Götz (2001): Handbuch Gesetzesfolgenabschätzung. Baden-Baden: Nomos.
- Bonfadelli, Heinz/Friemel, Thomas (2006): Kommunikationskampagnen im Gesundheitsbereich. Grundlagen und Anwendungen. Konstanz: UVK Verlagsgesellschaft.
- Borges, Peter/Schmidt, Ralf (2002): Die Balanced Scorecard als Steuerungsinstrument im Krankenhaus. *Betriebswirtschaftliche Forschung und Praxis* 2/2002: 102-117.
- Borrmann, Axel/Fasbender, Karl/Holthus, Manfred/Gleich, Albrecht von/Reichl, Bettina/Shams, Rasul (1999): Erfolgskontrolle in der deutschen Entwicklungszusammenarbeit. Analyse, Bewertung, Reformen. Baden-Baden: Nomos.
- Borrmann, Axel/Gleich, Albrecht von/Holthus, Manfred/Shams, Rasul (2001): Reform der Erfolgskontrolle in der deutschen Entwicklungszusammenarbeit. Eine Zwischenbilanz. Baden-Baden: Nomos.
- Borrmann, Axel/Stockmann, Reinhard (2008): Evaluation in der deutschen Entwicklungszusammenarbeit (Arbeitstitel). Münster: Waxmann Verlag (in Vorbereitung).

- Bortz, Jürgen/Döring, Nicola (2006): Forschungsmethoden und Evaluation. 4. Auflage. Berlin: Springer.
- Bos, Wilfried/Dedering, Kathrin/Holtappels, Heinz Günter/Müller, Sabine/Rösner, Ernst (2007): Schulinspektion in Deutschland – eine kritische Bestandsaufnahme. In: Van Buer, Jürgen/Wagner, Cornelia (Hrsg.): Qualität von Schule. Ein kritisches Handbuch. Frankfurt: Lang, S. 241-258.
- Bos, Wilfried/Holtappels, Heinz Günter/Rösner, Ernst (2006): Schulinspektion in den deutschen Bundesländern – eine Baustellenbeschreibung. In: Bos, Wilfried/Holtappels, Heinz Günter/Preiffer, Hermann/Rolff, Hans-Günter/Schulz-Zander, Renate (Hrsg.): Jahrbuch der Schulentwicklung. Daten, Beispiele und Perspektiven. Band 14. Weinheim: Juventa, S. 81-123.
- Boulianne, Louis Marie/Rey, Michel (1992): Arrêté fédéral du 6 octobre 1978 instituant une aide financière en ses régions dont l'économie est menacée. Son application, son apport. Lausanne: C.E.A.T.
- Brand, Karl-Werner (Hrsg.) (1997): Nachhaltige Entwicklung. Eine Herausforderung an die Soziologie. Opladen: Leske + Budrich.
- Brand, Karl-Werner/Poferl, Angelika (1997): Ökologische Kommunikation in Deutschland. Opladen: Westdeutscher Verlag.
- Brandt, Tasso (2007): Sozialer Kontext der Evaluation. In: Reinhard Stockmann (Hrsg.): Handbuch zur Evaluation – Eine praktische Handlungsanleitung. Münster: Waxmann, S. 164-194.
- Braun, Dietmar (1999): Changing governance models in higher education: the case of the new managerialism. *Swiss Political Science Review* 5(3): 1-24.
- Braun, Dietmar (2003): Dezentraler und unitarischer Föderalismus. Die Schweiz und Deutschland im Vergleich. *Schweizerische Zeitschrift für Politikwissenschaft*, 9(1): 57-89.
- Braun, Dietmar (2008): Evaluation und unintendierte Effekte – eine theoretische Reflexion. In: Matthies, Hildegard/Simon, Dagmar (Hrsg.): Wissenschaft unter Beobachtung. *Leviathan Sonderheft* 24: 103-124.
- Brunsson, Nils (1989): The Organization of Hypocrisy: Talk, Decision and Actions in Organizations. Chichester: John Wiley & Sons.
- BSV – Bundesamt für Sozialversicherung (Hrsg.) (2001): Analyse der Auswirkungen des Krankenversicherungsgesetzes. Synthesebericht. Forschungsbericht 16/01. Bern: BSV.
- Büeler, Xaver (2006): Qualitätsevaluation und Schulentwicklung. In: Stockmann, Reinhard (Hrsg.): Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder. 3. Auflage. Münster: Waxmann, S. 260-288.
- Buhren, Claus G./Killus, Dagmar/Müller, Sabine (1999): Wege und Methoden der Selbstevaluation. Ein praktischer Leitfaden für Schulen. 3. Auflage. Dortmund: IFS-Verlag.
- Bundesblatt (2006): Ressortforschung des Bundes: Evaluation des Behördenarrangements sowie der Forschungskonzepte und deren Umsetzung. Expertenbericht von Landert Farago Partner im Auftrag der Parlamentarischen Verwaltungskontrolle vom 3. April 2006. Bern: Bundesblatt.
- Bundeshaushaltsordnung vom 19. August 1969 (BGBl. I S. 1284), geändert durch Artikel 15 des Gesetzes vom 14. August 2006. BGBl. I S. 1911.
- Bundeskanzleramt (2006): Das Verwaltungsinnovationsprogramm der Bundesregierung. Ziele-Maßnahmen-Ergebnisse. Bilanz 2006. Wien: Bundeskanzleramt.
- Bundeskanzleramt (2007): Regierungsprogramm für die XXIII. Gesetzgebungsperiode. Wien: BKA.
- Bundesministerium des Innern – BMI (2000): Aktivitäten zur Staats- und Verwaltungsmodernisierung in Bund und Ländern 2000. Berlin: BMI.
- Bundesministerium des Innern – BMI (2002): Stabstelle Moderner Staat – Moderne Verwaltung. Bilanz 2002. Berlin: BMI.
- Bundesministerium des Innern – BMI (2004): Initiative Bürokratieabbau. Zwischenbericht nach zwölf Monaten. Berlin: BMI.
- Bundesministerium für Arbeit und Sozialordnung (Hrsg.) (2002): Kommission „Moderne Dienstleistungen am Arbeitsmarkt“. Bericht der Kommission, Broschüre Nr. A 306. Berlin: BMAS.

- Bundesministerium für Bildung und Forschung (BMBF) (2006): Bundesbericht Forschung 2006. Berlin/Bonn: BMBF.
- Bundesministerium für Bildung und Forschung (BMBF) (2007): Bericht zur technologischen Leistungsfähigkeit Deutschlands 2007. Berlin/Bonn: BMBF.
- Bundesministerium für Bildung und Forschung (BMBF)/Medizinischer Fakultätentag (Hrsg.) (2002): Forschungslandkarte Hochschulmedizin. (Bearbeitung: Susanne Bühner, Fraunhofer ISI). Stuttgart: Fraunhofer IRB-Verlag.
- Bundesministerium für Bildung, Wissenschaft und Kultur (bm:bwk) (2002): Universitätsgesetz 2002, Stand 9. August 2002. Wien: BM:BWK.
- Bundesministerium für Bildung, Wissenschaft und Kultur (bm:bwk)/Bundesministerium für Verkehr, Innovation und Technologie (bmvit)/Bundesministerium für Wirtschaft und Arbeit (bmwa) (2006): Österreichischer Forschungs- und Technologiebericht 2006. Bericht der Bundesregierung an den Nationalrat gem. § 8 (2) FOG über die Lage und Bedürfnisse von Forschung, Technologie und Innovation in Österreich. Wien: BM:BWK/BMVIT/BMWA.
- Bundesministerium für Land- und Forstwirtschaft, Umwelt und Wasserwirtschaft (2005): Update-Evaluierung des Österreichischen Programms für die Entwicklung des ländlichen Raums. Wien: BMLFUW. (land.lebensministerium.at/filemanager/download/14162/ [2.7.2007])
- Bundesministerium für Land- und Forstwirtschaft, Umwelt und Wasserwirtschaft (2006): Grüner Bericht 2006. Wien: BMLFUW. (www.gruenerbericht.at/cms/index.php [2.7.2007])
- Bundesministerium für Land- und Forstwirtschaft, Umwelt und Wasserwirtschaft (2007): Entwurf des Österreichischen Programms für die Entwicklung des ländlichen Raums 2007 – 2013. Anlage II: Ex-ante Evaluierung. Wien: BMLFUW. (www.landnet.at/article/archive/5127 [2.7.2007])
- Bundesministerium für Soziale Verwaltung (BMSV) (1982): Soziale Problemgruppen. Forschungsberichte aus Sozial- und Arbeitsmarktpolitik Nr. 1. Wien: Eigenverlag.
- Bundesministerium für Soziale Verwaltung (BMSV) (1984): Lokale Beschäftigungsinitiativen. Forschungsberichte aus Sozial- und Arbeitsmarktpolitik Nr. 6. Wien: Eigenverlag.
- Bundesministerium für Umwelt, Naturschutz und Reaktorsicherheit (Hrsg.) (2001): Handbuch Umweltcontrolling. 2. Auflage. München: Vahlen.
- Bundesministerium für Unterricht und Kunst (1999): Aufgabenprofil der Schulaufsicht (Allgemeine Weisung gemäß § 18 Abs. 3 Bundes-Schulaufsichtsgesetz) – Kundmachung im Ministerialverordnungsblatt (Rundschreiben Nr. 64/1999 vom 17.12.1999). Wien: BMUK. (www.bewegung.ac.at/download/22/sub0/22_1132_4091.pdf [21.5.2008])
- Bundesministerium für Wissenschaft und Forschung (2007): Evaluierung der Auswirkungen des § 124b des Universitätsgesetzes 2002, Bericht an den Nationalrat. Wien. (images.derstandard.at/20070308/Evaluierungsbericht.pdf [20.08.2007])
- Bundesrat (1993): Oeko-Beitragsverordnung vom 26. April 1993, AS 1993 1581.
- Bundesrat (1996): Botschaft zur Reform der Agrarpolitik: Zweite Etappe (Agrarpolitik 2002) vom 26. Juni 1996. *Bundesblatt* 1996 IV 1.
- Bundesrat (1998): Verordnung über die Beurteilung der Nachhaltigkeit vom 7. Dezember 1998, SR 919.118.
- Bundesrat (1999): Bericht über die Kernstädte. Bern: ARE/seco.
- Bundesrat (2001): Verordnung vom 4. April 2001 über die regionale Förderung der Qualität und der Vernetzung von ökologischen Ausgleichsflächen in der Landwirtschaft (Öko-Qualitätsverordnung, ÖQV); SR 910.14.
- Bundesrat (2002): Botschaft zur Weiterentwicklung der Agrarpolitik (Agrarpolitik 2007) vom 29. Mai 2002, *Bundesblatt* 2002 4721.
- Bundesrat (2004): Beschluss vom 3. November 2004 zur Umsetzung von Artikel 170 BV / Verstärkung der Wirksamkeitsüberprüfung. Bern: Bundesamt für Justiz.

- Bundesrat (2006): Botschaft zur Weiterentwicklung der Agrarpolitik (Agrarpolitik 2011) vom 17. Mai 2006, *Bundesblatt* 2006 6337.
- Bundesrechnungshof (2005): Zusammenfassung der Ergebnisse zur Peer Review der Eidgenössischen Finanzkontrolle, 27. Mai 2005. Bonn: Bundesrechnungshof.
- Bundesrechnungshof (2006): Bericht nach § 99 Bundeshaushaltsordnung über die Modernisierung des staatlichen Haushalts- und Rechnungswesens. Berlin: Bundesrechnungshof.
- Bundesregierung (2006): Programm Bürokratieabbau und bessere Rechtssetzung – Einführung des Standardkosten-Modells, Methodenhandbuch der Bundesregierung. Berlin: Bundesregierung.
- Bundesumweltministerium/Umweltbundesamt (Hrsg.) (2000): Umweltmanagementsysteme. Fortschritt oder heiße Luft? Frankfurt: Frankfurter Allgemeine Buch.
- Bundesversammlung (1991): Bundesgesetz vom 24. Januar 1991 über den Schutz der Gewässer (Gewässerschutzgesetz, GSchG), SR 814.20.
- Bundesversammlung (1998): Landwirtschaftsgesetz, SR 910.1.
- Bundesversammlung (1999): Bundesverfassung, SR 101.
- Burkard, Christoph (2005): Ergebnisorientierte Systemsteuerung: Konsequenzen für die externe Evaluation. In: Brägger, Gerold/Bucher, Beat/Landwehr, Norbert (Hrsg.): Schlüsselfragen zur externen Schulevaluation. Bern: hep-Verlag, S. 79-109.
- Burkard, Christoph/Eikenbusch, Gerhard (2000): Praxishandbuch Evaluation in der Schule. Berlin: Cornelsen Scriptor.
- Burkard, Christoph/Pfeiffer, Hermann (1995): Evaluation von Einzelschulen - Entwicklungslinien und aktuelle Trends. *Zeitschrift für Sozialisationsforschung und Erziehungssoziologie* 15(4): 294-312.
- Bürsch, Michael/Müller, Brigitte (1999): Verwaltungsreformen in den deutschen Bundesländern. Bonn: Friedrich-Ebert-Stiftung. (library.fes.de/fulltext/stabsabteilung/00927toc.htm [19.3.2008])
- Bussmann, Werner (1981): Gewässerschutz und kooperativer Föderalismus in der Schweiz. Bern: Haupt Verlag.
- Bussmann, Werner (1995): Evaluations and grassroots politics: The case of Switzerland. *Knowledge and Policy* 8(3): 85-98.
- Bussmann, Werner (1996): Democracy and evaluation's contribution to negotiation, empowerment and information. *Evaluation* 2(3): 307-19.
- Bussmann, Werner (1997): Evaluationen in der Schweiz. In: Bussmann, Werner/Klöti, Ulrich/Knoepfel, Peter (Hrsg.): Einführung in die Politikevaluation. Basel: Helbing & Lichtenhahn, S. 13-35.
- Bussmann, Werner (2004): Ungeliebt aber unverzichtbar: Evaluation im Kontext der Politik. In: Beywl, Wolfgang (Hrsg.): Evaluation im Kontext. Arbeitsbericht Nr. 31. Bern: Koordinationsstelle für Weiterbildung, Universität Bern, S. 14-24.
- Bussmann, Werner (2007): Institutionalisierung der Evaluation in der Schweiz: Verfassungsauftrag, Konkretisierungsspielräume und Umsetzungsstrategie. In: Schäffer Heinz (Hrsg.): Evaluierung der Gesetze / Gesetzesfolgenabschätzung (II). Wien: Manz, S. 1-20.
- BWV, Bundesbeauftragte für Wirtschaftlichkeit in der Verwaltung (1998): Erfolgskontrolle finanzwirksamer Maßnahmen in der öffentlichen Verwaltung, Gutachten der Präsidentin des Bundesrechnungshofes als Bundesbeauftragte für Wirtschaftlichkeit in der Verwaltung (Band 2 der Schriftenreihe der Bundesbeauftragten für Wirtschaftlichkeit in der Verwaltung). (www.bundesrechnungshof.de/bundesbeauftragter-bwv/bwv-baende/bwv-band2.pdf. [11.5.2007])
- BWV, Bundesbeauftragter für Wirtschaftlichkeit in der Verwaltung (2002): Bericht zu den Finanzbeziehungen zwischen Bund und Ländern, Mischfinanzierungen nach Art. 91a, 91b und 104a Abs. 4 Grundgesetz (Band 9 der Schriftenreihe des Bundesbeauftragten für Wirtschaftlichkeit in der Verwaltung). (www.bundesrechnungshof.de/bundesbeauftragter-bwv/bwv-baende/bwv-band9.pdf. [11.5.2007])

- Callon, Michel (1992): The dynamics of techno-economic networks. In: Coombs, Rod/Saviotti, Paolo/Walsh, Vivian (eds): Technological change and company strategies: Economic and sociological perspectives. London: Academic Press, pp. 72-102.
- Campbell, David F.J. (2006): The Research Assessment Exercise 2008 in the United Kingdom. In: Plattform Forschungs- und Technologieevaluierung (Hrsg.): New Frontiers in Evaluation. Newsletter Special Edition. Wien: Plattform FTEval, pp. 32-47.
- Campbell, Donald T./Stanley, Julian C. (1963): Experimental and quasi-experimental designs for research. Chicago: Rand McNally.
- Caspari, Alexandra (2004): Evaluation der Nachhaltigkeit von Entwicklungszusammenarbeit – Zur Notwendigkeit angemessener Konzepte und Methoden. Wiesbaden: VS Verlag.
- Caspari, Alexandra/Kranz-Plote, Jutta/Meyer, Lutz (2000): Langfristige Wirkungen deutscher Entwicklungszusammenarbeit und ihre Erfolgsbedingungen. Eine Ex-post-Evaluierung von 32 abgeschlossenen Projekten“. BMZ-Spezial Nr. 19; Bonn: BMZ.
- Cavegn, Gion/Meyer, Leo/Rieder, Peter (1994): Analyse des Angebotsverhaltens der Getreideproduzenten. In: Nationales Forschungsprogramm 27 (Hrsg.): Evaluationen in der Agrarpolitik II, Reihe Schlussberichte. Bern: SNF.
- CEST (2003a): Evaluation des Schweizerischen Sozialarchivs. CEST 2003/5. Bern: Center for Science and Technology Studies.
- CEST (2003b): Evaluation der Schweizerischen Friedensstiftung „swisspeace“. CEST 2003/6. Bern: Center for Science and Technology Studies.
- Chelimsky, Eleanor (1990): Evaluation und Politik. Die Umsetzung von Evaluationsergebnissen in exekutiven und legislativen Bundesorganen der USA. In: Koch, Uwe/Wittmann, Werner W. (Hrsg.): Evaluationsforschung. Bewertungsgrundlage von Sozial- und Gesundheitsprogrammen. Berlin: Springer, S. 249-280.
- Chelimsky, Eleanor (2006): The purposes of evaluation in a democratic society. In: Shaw, Ian F./Greene, Jennifer C./Mark, Melvin M. (eds): Handbook of evaluation. Policies, programs and practices. London: Sage, pp. 33-55.
- Chen, Huey-Tsyh (2005): Practical program evaluation: Assessing and improving planning, implementation, and effectiveness. Thousand Oaks: Sage.
- Chen, Huey-Tsyh/Rossi, Peter H. (1980): The multi-goal, theory-driven approach to evaluation: A model linking basic and applied social science. *Social Forces* 59(1): 106-122.
- Clausnitzer, Klaus-Dieter (2000): Situation der Evaluation deutscher Energiepolitik. Bremen: Bremer Energie-Institut.
- Commission externe d'évaluation des politiques publiques (CEPP) (1998): Chômeurs en fin de droit. Evaluation de la politique cantonale d'emploi temporaire. Genève: CEPP.
- Commission externe d'évaluation des politiques publiques (CEPP) (2002): Politique cantonale de lutte contre le chômage de longue durée. Evaluation des mesures cantonales. Genève: CEPP.
- Conrad, Susanna/Lischer, Petra/Wolf, Bernhard (1997): Erhebungsmethoden der externen empirischen Evaluation des Modellvorhabens Kindersituationen. Zentrum für Empirische Pädagogische Forschung, Universität Koblenz-Landau, Berichte, 19. Landau: Verlag Empirische Pädagogik.
- Convelop – kooperative knowledge design GmbH/Technopolis Forschungs- und Beratungsgesellschaft mbH (2007): Ex-post Evaluierung des EFRE Programms Innovative Maßnahmen Salzburg 2003 – 2006. Wien: Convelop.
- Cornuz, Jacques/Gilbert, Allison/Pinget, Christophe/McDonald, Paul/Slama, Karen/Salto, Estève/Paccaud, Fred (2006): Cost-effectiveness of pharmacotherapies for nicotine dependence in primary care settings: a multi-national comparison. *Tobacco Control* 15(3): 152-159.
- Cousins, Bradley J./Leithwood, Kenneth A. (1986): Current empirical research on evaluation utilization. *Review of Educational Research* 56(3): 331-364.

- Cousins, Bradley J./Leithwood, Kenneth A. (1993): Enhancing knowledge utilization as a strategy for school improvement. *Knowledge: Creation, Diffusion, Utilization* 14(3): 305-333.
- Czerwanski, Annette (Hrsg.) (2003): Schulentwicklung durch Netzwerkarbeit: Erfahrungen aus den Lernnetzwerken im "Netzwerk innovativer Schulen in Deutschland". Gütersloh: Verlag Bertelsmann Stiftung.
- DAC (1998): Review of DAC principles for evaluation of development assistance. Paris: OECD. (www.oecd.org/dataoecd/63/50/2065863.pdf [4.4.2008])
- DAC (2000): DAC peer review Switzerland. Paris: OECD.
- DAC (2004): Peer review Austria. Paris: OECD.
- DAC (2005a): Prüfbericht über die Entwicklungszusammenarbeit – Deutschland. Paris: OECD. (www.oecd.org/dataoecd/10/22/36770168.pdf [10.3.2008])
- DAC (2005b): Paris Declaration on Aid Effectiveness: Ownership, harmonisation, alignment, results and mutual accountability. High level forum Paris 28.2-2.3.2005. Paris: OECD. (www.oecd.org/dataoecd/11/41/34428351.pdf [5.3.2007])
- DAC (2005c): DAC peer review Switzerland. Paris: OECD. (www.oecd.org/dataoecd/9/59/35297586.pdf [5.3.2007]).
- DAC (2006a): DAC evaluation quality standards. Paris: OECD. (www.oecd.org/dataoecd/34/21/38686856.pdf ([25.08.2007])
- DAC (2006b): Guidance for managing joint evaluations. Paris: OECD. (www.oecd.org/dataoecd/29/28/37512030.pdf ([25.08.2007])
- DAC (2006c): Erklärung von Paris über die Wirksamkeit der Entwicklungszusammenarbeit. Deutsche Übersetzung der Paris Declaration on Aid Effectiveness. Paris: OECD. (www.oecd.org/dataoecd/37/39/35023537.pdf [4.4.2008])
- DAC (2006d): Promoting pro-poor growth. Harmonizing ex-ante poverty impact assessment. Paris: OECD.
- DAC (2008): DAC criteria for evaluating development assistance. Paris: OECD. (www.oecd.org/document/22/0,3343,en_2649_34435_2086550_1_1_1_1,00.html [4.4.2008])
- Dahler-Larsen, Peter (2005): Evaluation and public management. In: Ferlie, Ewan/Lynn, Laurence E./Pollitt, Christopher (eds): The Oxford handbook of public management. Oxford: University Press, pp. 614-639.
- Dahler-Larsen, Peter (2006): Evaluation after disenchantment? Five issues shaping the role of evaluation in society. In: Shaw, Ian F./Greene, Jennifer C./Mark, Melvin M. (eds): Handbook of evaluation. London: Sage, pp. 141-60.
- Daintith, Terence/Page, Alan (1999): The executive in the constitution: Structure, autonomy, and internal control. Oxford: Oxford University Press.
- Dalin, Per/Rolff, Hans-Günter/Buchen, Herbert (1996): Institutioneller Schulentwicklungsprozess. Ein Handbuch. Soest: Verlag für Schule und Weiterbildung.
- Daniel, Hans-Dieter (1993): Guardians of science. Fairness and reliability of peer review. Weinheim: VCH.
- Davies, Huw T.O./Nutley, Sandra M./Smith, Peter C. (2000): What works? Evidence-based policy and practice in public services. Bristol: The Policy Press.
- Davies, Philipp/Newcomer, Kathryn/Soydan, Haluk (2006): Government as structural context for evaluation. In: Shaw, Ian F./Greene, Jennifer C./Mark, Melvin M. (eds): Handbook of evaluation. London: Sage, pp. 163-183.
- De Tommasi, Roberto/Caduff, Gabriel (2006): Mobilitätsmanagement in Unternehmen: Schlussevaluations der Pilotphase. Bern: Bundesamt für Energie.
- Dearing, Elisabeth (1992): Das Projekt „Verwaltungsmanagement“ – Chancen einer Verwaltungsreform in Österreich. *Die öffentliche Verwaltung* 7: 297-305.

- Dearing, Elisabeth (2003): New Public Management in der österreichischen Bundesverwaltung. In: Bauer, Helfried/Biwald, Peter/Dearing, Elisabeth (Hrsg.): Öffentliches Management in Österreich. Realisierungen und Perspektiven. Wien: KDZ Managementberatungs- u. Weiterbildungs GmbH, S. 83-106.
- DeGEval – Deutsche Gesellschaft für Evaluation (2002): Standards für Evaluation. Köln: DeGEval. (www.degeval.de/calimero/tools/proxy.php?id=227 [10.4.2008])
- DeGEval – Deutsche Gesellschaft für Evaluation (2004a): Empfehlungen für die Aus- und Weiterbildung in der Evaluation. Alfter: DeGEval.
- DeGEval – Deutsche Gesellschaft für Evaluation (2004b): Empfehlungen zur Anwendung der Standards für Evaluation im Handlungsfeld der Selbstevaluation. Alfter: DeGEval.
- DeGEval – Deutsche Gesellschaft für Evaluation (Hrsg.) (2001): Evaluation – Reformmotor oder Reformbremse. Köln: DeGEval.
- DeGEval – Gesellschaft für Evaluation (2007): Empfehlungen für Auftraggebende von Evaluationen. Mainz: DeGEval.
- Delley, Jean-Daniel (2001): L'évaluation législative entre rigueur scientifique et utilité pratique – L'expérience de la Commission externe d'évaluation des politiques publiques du canton de Genève. In: Auer, Andreas/Delley, Jean-Daniel/Hottelier, Michel/Malinverni, Giorgio (éds.): Aux confins du droit: essais en l'honneur du Professeur Charles-Albert Morand. Genève: Helbing & Lichtenhahn, p. 51-64.
- Delley, Jean-Daniel (2007): En quête de la Sainte Trinité – La commission externe d'évaluation des politiques publiques du canton de Genève. In: Jacob, Steve/Varone, Frédéric/Genard, Jean-Louis (éds.): L'évaluation des politiques au niveau régional. Frankfurt: Lang, p. 113-122.
- Derlien, Hans-Ulrich (1990a): Program evaluation in the Federal Republic of Germany. In: Rist, Ray C. (ed.): Program evaluation and the management of government. New Brunswick: Transaction, pp. 37-51.
- Derlien, Hans-Ulrich (1990b): Genesis and structure of evaluation efforts in comparative perspective. In: Rist, Ray C. (ed.): Program evaluation and the management of government. New Brunswick: Transaction, pp. 147-175.
- Derlien, Hans-Ulrich (1997): Die Entwicklung von Evaluationen im internationalen Kontext. In: Bussmann, Werner/Klöti, Ulrich/Knoepfel, Peter (Hrsg.): Einführung in die Politikevaluation. Basel: Helbing & Lichtenhahn, S. 4-12.
- Derlien, Hans-Ulrich (2002): Policy evaluation in Germany: Institutional continuation and sectoral activation. In: Furubo, Jan-Eric/Rist, Ray C./Sandahl, Rolf (eds) (2002): International atlas of evaluation. New Brunswick: Transaction, pp. 77-91.
- Derlien, Hans-Ulrich/Rist, Ray C. (2002): Policy evaluation in international comparison. In: Furubo, Jan-Eric/Rist, Ray C./Sandahl, Rolf (eds): International atlas of evaluation. New Brunswick: Transaction, pp. 439-455.
- Deschenaux, Chantal/Erlanger, Jacques (1998): Loi fédérale concernant l'amélioration du logement dans les régions de montagne (WS). Rapport d'évaluation provisoire. Lausanne: C.E.A.T.
- DESI-Konsortium (2008): Unterricht und Kompetenzerwerb zu Deutsch und Englisch: Ergebnisse der DESI-Studie. Weinheim: Beltz.
- Detli, Reto/Gsponer, Gabriele/Fegeler, Marie/Steiner, Isabelle/Sardi, Massimo (2004): Evaluation energo. Bern: Bundesamt für Energie.
- Deutscher Bundestag (2002): Entschließungsantrag der Fraktionen SPD und BÜNDNIS 90/DIE GRÜNEN zu der dritten Beratung des Gesetzentwurfs der Fraktionen SPD und BÜNDNIS 90/DIE GRÜNEN – Drucksachen 15/25, 15/77 – Entwurf eines Ersten Gesetzes für moderne Dienstleistungen am Arbeitsmarkt. Drucksache 15/98. Berlin: Deutscher Bundestag.
- Deutscher Bundestag (2005): Raumordnungsbericht 2005. Drucksache 15/5500 vom 9. Mai. 2005. Berlin: Deutscher Bundestag.

- Deutscher Bundestag (2006a): Bericht der Bundesregierung über die Entwicklung der Finanzhilfen des Bundes und der Steuervergünstigungen für die Jahre 2003 bis 2006 (20. Subventionsbericht). Drucksache 16/1020 vom 22. März 2006. Berlin: Deutscher Bundestag.
- Deutscher Bundestag (2006b): Fünfunddreißigster Rahmenplan der Gemeinschaftsaufgabe „Verbesserung der regionalen Wirtschaftsstruktur“ für den Zeitraum 2006 bis 2009. Drucksache 16/1790 vom 7. Juni 2006. Berlin: Deutscher Bundestag.
- Deutscher Bundestag (2006c): Bericht der Bundesregierung über den Stand des Ausbaus für ein bedarfsgerechtes Angebot an Kindertagesbetreuung für Kinder unter drei Jahren 2006. Drucksache 16/2250. Berlin: Deutscher Bundestag.
- Deutscher Bundestag (2006d): Unterrichtung durch die Bundesregierung: Bericht 2006 der Bundesregierung zur Wirksamkeit moderner Dienstleistungen am Arbeitsmarkt. Drucksache 16/3982 vom 21.12.2006. Berlin: Deutscher Bundestag.
- Deutsches Institut für Erwachsenenbildung/Deutsches Institut für Wirtschaftsforschung/Universität Duisburg-Essen/Ludwigs-Maximilians-Universität (2004): Ausgewählte Ergebnisse der wissenschaftlichen Begleitung des Programms „Lernende Regionen – Förderung von Netzwerken“. Bonn: DIE. (www.lernende-regionen.info/dlr/download/Lernende_Regionen_-_Empfehlungen_wiss_Begleitung_0704.pdf [13.5.2008])
- Deutsches PISA-Konstortium (2001): PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich. Opladen: Leske+Budrich.
- Devouassoux, Christophe/Pfister Giauque, Barbara (1998): Evaluation intermédiaire du programme INTERREG II France (Franche-Comté) – Suisse. Lausanne/Annemasse: C.E.A.T./Transversal.
- Dewe, Bernd/Otto, Hans-Uwe (2001): Profession. In: Otto, Hans-Uwe/Thiersch, Hans (Hrsg.): Handbuch Sozialarbeit/Sozialpädagogik. 2. Auflage. Neuwied: Luchterhand, S. 1399-1423.
- DEZA – Direktion für Entwicklung und Zusammenarbeit (2002): Leitlinien Evaluation und Controlling. Bern: DEZA. (www.deza.ch/ressources/resource_de_23568.pdf [4.4.2008])
- DEZA – Direktion für Entwicklung und Zusammenarbeit [2006]: Evaluationen erfolgreich nutzen: 10 Leitsätze. (www.deza.admin.ch/ressources/resource_de_91087.pdf [29.4.2008])
- DEZA – Direktion für Entwicklung und Zusammenarbeit (2007): SDC's performance towards empowerment of stakeholders from the recipients' perspective. (www.deza.ch/de/Home/Aktivitaeten/Evaluation/Abgeschlossene_Evaluationen/2007 [5.6.2007])
- Dietzel, Gottfried T.W./von Troschke, Jürgen (Hrsg.) (1988): Begleitforschung bei staatlich geförderten Modellprojekten – strukturelle und methodische Probleme. Schriftenreihe des Bundesministers für Jugend, Familie, Frauen und Gesundheit, Band 216. Stuttgart: Kohlhammer.
- Dinges, Michael (2006): Pressure, Deception and Peer Review. *Plattform fteval Newsletter* 28: 11-22.
- Dirksmeyer, Walter/Forstner, Bernhard/Margarian, Anne/Zimmer, Yelto (2006): Aktualisierung der Zwischenbewertung des Agrarinvestitionsförderungsprogramms (AFP) in Deutschland für den Förderzeitraum 2000 bis 2004, länderübergreifender Bericht. Braunschweig: Eigenverlag.
- Ditton, Hartmut/Arnoldt, Bettina (2004): Schülerbefragungen zum Fachunterricht – Feedback an Lehrkräfte. *Empirische Pädagogik* 18(1): 115-139.
- Dittrich, Gisela (2007): Querauswertung zum Thema Qualität. Datenbank ProkiTa. München: Deutsches Jugendinstitut. (www.dji.de/prokita [31.05.2007])
- Döbert, Hans (2003): Neue Steuerungsmodelle von Schulsystemen in Europa? In: Döbert, Hans/von Kopp, Botho/Martini, Renate/Weiss, Manfred (Hrsg.): Bildung vor neuen Herausforderungen. Historische Bezüge – Rechtliche Aspekte – Steuerungsfragen – Internationale Perspektiven. Neuwied: Luchterhand, S. 287-303.
- Donaldson, Stewart I. (2003): Theory-driven program evaluation in the new millennium. In: Donaldson, Stewart I./Scriven, Michael (eds): Evaluating social programs and problems: Visions for the new millennium. Mahwah: Erlbaum, pp. 109-141.

- Döring, Nicola (2005). Für Evaluation und gegen Evaluitis. Warum und wie Lehrevaluation an deutschen Hochschulen verbessert werden sollte. In: Berendt, Brigitte/Voss, Hans-Peter/Wildt, Johannes (Hrsg.): Neues Handbuch Hochschullehre (Ergänzungslieferung Juli 2005). Berlin: Raabe.
- Draxl, Petra/Schneidewind, Peter/Downes, Ruth/Bucek, Milan (2004): Systemische Evaluierung des Regionalmanagements in Österreich. Wien: Bundeskanzleramt/ÖSB Consulting. (www.bundeskanzleramt.at/DocView.axd?CobId=8434 [13.5.2008])
- Draxl, Petra/Schuh, Bernd (2005): Evaluierung der Regionalmanagementstellen in der Steiermark. Kurzfassung. Wien: Österreichisches Institut für Raumplanung (ÖIR). (www.oir.at/iddb/archiv6288/89_archiv6288_25820.pdf [13.5.2008])
- Drilling, Matthias/Müller, Caroline/Fabian, Carlo (2006): Schulsozialarbeit im Fürstentum Liechtenstein. Evaluationsbericht. Basel: FHNW (unveröffentlicht).
- Dünnhoff, Elke/Duscha, Markus/Krieg, Oliver (2005): Evaluation der stationären Energieberatung der Verbraucherzentralen, des Deutschen Hausfrauenbundes Niedersachsen und des Verbraucherservice Bayern. Heidelberg: ifeu – Institut für Energie- und Umweltforschung.
- Duscha, Markus/Bauer, Helmut/Kuhn, Carsten (2007): Evaluation des Förderprogramms Klimaschutz-Plus. Heidelberg: ifeu – Institut für Energie- und Umweltforschung.
- Duve, Thomas (2006): Verwaltungsmodernisierung in Österreich. In: Hill, Hermann (Hrsg.): Verwaltungsmodernisierung in den Staaten Europas – Länderberichte II, Speyrer Arbeitshefte 185 (II). Speyer: Deutsche Hochschule für Verwaltungswissenschaften, S. 397-429.
- Dyllick, Thomas/Hamschmidt, Jost (2000): Wirksamkeit und Leistung von Umweltmanagementsystemen. Zürich: vdf.
- Eberhard, Sybille/Lanz, Andreas (1997): Evaluation der Suchtpräventionsstelle Zürcher Oberland. Forschungsbericht. Zürich: ISGF.
- EBP (2007): Machbarkeitsstudie: Indikatoren für die Neue Regionalpolitik. Im Auftrag des Staatssekretariats für Wirtschaft (seco). Zürich: Ernst Basler + Partner AG.
- Eckey, Hans-Friedrich (2005): Regionale Strukturpolitik. In: Akademie für Raumforschung und Landesplanung (Hrsg.): Handwörterbuch der Raumordnung. 4. Auflage. Hannover: Akademie für Raumforschung und Landesplanung, S. 933-940.
- Eder, Ferdinand (Hrsg.) (2002): Qualitätsentwicklung und Qualitätssicherung im österreichischen Schulwesen. Bildungsforschung des Bundesministeriums für Bildung, Wissenschaft und Kultur, Band 17. Innsbruck: Studienverlag.
- EDK – Schweizerische Konferenz der kantonalen Erziehungsdirektoren (2006): Interkantonale Vereinbarung über die Harmonisierung der obligatorischen Schule (HarmoS-Konkordat). Bericht zur Vernehmlassung. Bern: EDK.
- EDK – Schweizerische Konferenz der kantonalen Erziehungsdirektoren (2008): Evaluation und Schulqualität. Orientierungsrahmen zuhanden von Bildungsbehörden, Aufsichtsorganen und Schulleitungen. Bern: EDK (im Druck).
- Edler, Jakob/Bührer, Susanne/Lo, Vivien/Sheikh, Sonja (2003): Assessment „Zukunft der Kompetenzzentrenprogramme und Zukunft der Kompetenzzentren“. Karlsruhe/Vienna: Fraunhofer ISI/Austrian Institute for SME Research.
- Edler, Jakob/Rigby, John (2004): Research network programmes evaluation for the Austrian Science Fund (FWF). Manchester/Karlsruhe: PREST/Fraunhofer ISI.
- Effionayi-Mäder, Denise/Chimienti, Milena/Losa, Stefano/Cattacin, Sandro (2001): Evaluation des Gesamtprogrammes Migration und Gesundheit des Bundesamts für Gesundheit Evaluation. Forschungsbericht 17/2001. Neuenburg: Forum für Migrationsstudien.
- Egli, Regula (2006): Zwischenbericht Pilotprojekte NRP. Bern: seco.
- Eidgenössische Finanzkontrolle (2003): Vollzugs- und Wirkungsprüfung bei den Investitionskrediten an die Forstwirtschaft. Bern: EFK.

- Eidgenössisches Justiz- und Polizeidepartement (1991): Die Wirkungen staatlichen Handelns besser ermitteln: Probleme, Möglichkeiten, Vorschläge. Schlussbericht der Arbeitsgruppe „Gesetzesevaluation“. Bern: EJPD.
- Eigler, Gunther/Nenninger, Peter (1983): Möglichkeiten und Grenzen einer interessen-orientierten Lehrerfortbildung. *Unterrichtswissenschaft* 11(3): 227-243.
- Eisenring, Claudia/Hess Kurt (2004): Wirtschaftlichkeit und Qualität der Arztpraxis. Systeme-Modelle-Methoden. Neuenburg: Schweizerisches Gesundheitsobservatorium.
- Eisinger, Angelus/Reuther, Iris (2007): Zürich baut. Konzeptioneller Städtebau. Building Zürich. Conceptual Urbanism. Basel: Birkhäuser.
- Elbe, Sebastian (2006): Ergebnisse der Begleitforschung Regionen Aktiv in der Diskussion. Das Konzept der Begleitforschung und der Steuerungsansatz von Regionen Aktiv (Vortrag). Darmstadt: Regionenaktiv. (www.regionenaktiv.de/Vortrag_Euregia_Elbe.pdf [10.5.2007])
- Elias, Jiri/Ragni, Thomas (2007): Anwendungsorientierte Forschung des Seco – Rückblick und Ausblick. *Soziale Sicherheit CHSS* 3/2007: 136-139.
- Elkeles, Thomas (2006): Evaluation von Gesundheitsförderung und die Forderung nach Evidenzbasierung. *Zeitschrift für Evaluation* 2006/1: 39-70.
- Emmerich, Marcus/Maag Merki, Katharina (2007): Schulsystementwicklung und Reformstrategien in den Ländern der Bundesrepublik Deutschland. Gutachten für den 10. Gesprächskreis Schulentwicklung im Auftrag der Bertelsmann Stiftung. Abschlussbericht. Freiburg: Pädagogische Hochschule Freiburg.
- Endres, Heinz/Victor, Norbert/Haake, Michael/Witte, Steffen/Streitberger, Konrad/Zenz, Michael (2007): Akupunktur bei chronischen Knie- und Rückenschmerzen. *Deutsches Ärzteblatt* 104(3): A-123.
- Engel, George (1977): The need for a new medical model. A challenge for biomedicine. *Science* 196:129-136.
- Engelniederhammer, Stefan/Köpp, Bodo/Reichard, Christoph/Röber, Manfred/Wollmann, Hellmut (1999): Berliner Verwaltung auf Modernisierungskurs. Berlin: Edition Sigma.
- Engleitner, Johann/Keppelmüller, Joachim/Schwarz, Wolfgang T. (1998): EVIST - Evaluation und IST-Analyse. *Erziehung und Unterricht* 148(7/8): 648-663.
- EQUAL EntwicklungspartnerInnenschaft Donau – Quality in Inclusion (Hrsg.) (2007): Sozialer Sektor im Wandel. Zur Qualitätsdebatte und Beauftragung von Sozialer Arbeit. Linz: edition pro mente.
- Eser Davolio, Miryam/Drilling, Matthias (2007): Rechtsextremismus und Soziale Arbeit. Evaluation von Interventionsansätzen und Entwicklung von Guidelines. Basel: FHNW (unveröffentlicht).
- Eser, Thiemo W. (2001): Evaluation und Qualitätsmanagement – Anforderungen und Konsequenzen für die EU-Strukturpolitik. *Informationen zur Raumentwicklung* 28(6/7): 327-339.
- Esterling, Kevin M. (2004): The political economy of expertise. Information and efficiency in American national politics. Ann Arbor: University of Michigan Press.
- Etzkowitz, Henry/Leydesdorff, Loet (2000): The dynamics of innovation: From national systems and "mode2" to a triple helix of university-industry-government relations. *Research Policy* 29, 109-123.
- EUHES – The group of Heads of the EU Member States' development cooperation evaluation services and the European Commission (2007): Evaluating Coordination, Complementarity and Coherence in EU development policy: a synthesis. Studies in European Development Co-operation Evaluation No 8. Amsterdam: Aksant Academic Publishers. (www.three-cs.net/content/download/548/5660/file/CCC_8.pdf [10.3.2008])
- EU-KOM, Europäische Kommission Generaldirektion Landwirtschaft (1999a): Bewertung von Programmen zur Entwicklung des ländlichen Raums im Zeitraum 2000-2006 mit Unterstützung des Europäischen Ausrichtungs- und Garantiefonds für die Landwirtschaft, Leitfaden (VI/8865/99). Brüssel.

- EU-KOM, Europäische Kommission Generaldirektion Regionalpolitik (1999b): Die Ex-ante-Bewertung der Strukturfondsinterventionen. Der neue Programmplanungszeitraum 2000-2006: methodische Arbeitspapiere, H. 2. Brüssel.
- EU-KOM, Europäische Kommission Generaldirektion Regionalpolitik (1999c): Indikatoren für die Begleitung und Bewertung: Eine indikative Methode. Der neue Programmplanungszeitraum 2000-2006: methodische Arbeitspapiere, H. 3. Brüssel.
- EU-KOM, Europäische Kommission Generaldirektion Regionalpolitik (2000a): Die Halbzeitbewertung der Strukturfondsinterventionen. Der neue Programmplanungszeitraum 2000-2006: methodische Arbeitspapiere, H. 8. Brüssel.
- EU-KOM, Kommission der Europäischen Gemeinschaften (2000b): Gemeinsame Bewertungsfragen mit Kriterien und Indikatoren - Bewertung von Programmen zur Entwicklung des ländlichen Raums, die von 2000 bis 2006 durchgeführt und durch den Europäischen Ausrichtungs- und Garantiefonds gefördert werden (Dokument VI/12004/00 Endg.). Brüssel.
- Europäische Kommission (2007): Politik zur Entwicklung des ländlichen Raums 2007-2013. Gemeinsamer Begleitungs- und Bewertungsrahmen. Leitfaden. Brüssel: EK. (ec.europa.eu/agriculture/rurdev/eval/index_de.htm [2.7.2007])
- Europäische Kommission, Generaldirektion Regionalpolitik (2006a): Der neue Programmplanungszeitraum 2007-2013: Indikative Leitlinien zu Bewertungsverfahren: Ex-ante Bewertungen. Arbeitsdokument 1, August 2006. Brüssel: Europäische Kommission. (ec.europa.eu/regional_policy/sources/docoffic/2007/working/wd1_exante_de.pdf [29.4.2008])
- Europäische Kommission, Generaldirektion Regionalpolitik (2006b): Der neue Programmplanungszeitraum 2007-2013: Indikative Leitlinien zu Bewertungsverfahren: Indikatoren für Begleitung und Bewertung. Arbeitsdokument 2, August 2006. Brüssel: Europäische Kommission. (ec.europa.eu/regional_policy/sources/docoffic/2007/working/wd2indic_082006_de.pdf [29.4.2008])
- Europäische Kommission, Generaldirektion Regionalpolitik (2006c): Methodologisches Papier der Kommission betreffend Leitlinien zur Berechnung öffentlicher oder gleichgestellter Strukturausgaben eines Mitgliedstaates zum Zwecke der Zusätzlichkeit. Arbeitsdokument 3, September 2006. Brüssel: Europäische Kommission. (ec.europa.eu/regional_policy/sources/docoffic/2007/working/wd3_additionality_de.pdf [29.4.2008])
- Europäische Kommission, Generaldirektion Regionalpolitik (2006d): Methodologische Leitlinien der Kommission zur Durchführung der Kosten-Nutzen-Analyse für Großprojekte und zu Einnahmen schaffenden Projekten. Arbeitsdokument 4, August 2006. Brüssel: Europäische Kommission. (ec.europa.eu/regional_policy/sources/docoffic/2007/working/wd4_cost_de.pdf [29.4.2008])
- Europäische Kommission, Generaldirektion Regionalpolitik (2007a): Indikative Leitlinien zur Bewertungsverfahren : Bewertung während des Programmplanungszeitraums. Arbeitsdokument 5, April 2007. Brüssel: Europäische Kommission. (ec.europa.eu/regional_policy/sources/docoffic/2007/working/wd5_ongoing_de.pdf [29.4.2008])
- Europäische Kommission, Generaldirektion Regionalpolitik (2007b): Messung der Beschäftigungseffekte von Strukturfonds. Arbeitsdokument 6, März 2007. Brüssel: Europäische Kommission. (ec.europa.eu/regional_policy/sources/docoffic/2007/working/wd6_employ_de.pdf [29.4.2008])
- Europäische Union (EU) (2000): Schlussfolgerungen des Vorsitzes des Europäischen Rates, Lissabon 23. und 24. März 2000. (www.consilium.europa.eu/ueDocs/cms_Data/docs/pressdata/de/ec/00100-r1.d0.htm [17.10.2007])
- European Association for Quality Assurance in Higher Education (ed.) (2005): Standards and guidelines for quality assurance in the European Higher Education Area. Helsinki: European Association for Quality Assurance in Higher Education. (www.enqa.eu/files/ENQA%20Bergen%20Report.pdf [19.5.2008])

- European Commission (2000): Towards a European research area. Communication from the Commission to the Council, the European Parliament, the Economic and Social Committee and the Committee of the Regions, Brussels, 18 January 2000, COM (2000) 6. Brussels: European Commission.
- European Commission (2004): The Common Agricultural Policy - A policy evolving with the times. (ec.europa.eu/agriculture/publi/capleaflet/cap_en.htm. [9.5.2007])
- Evaluiierungskommission (Hrsg.) (1998): Systemevaluierung der Fraunhofer-Gesellschaft. Bericht der Evaluierungskommission. München: Evaluierungskommission.
- Expertenkommission „Überprüfung und Neukonzeption der Regionalpolitik“ (2003): Neue Regionalpolitik (NRP). Schlussbericht. Zürich: BHP – Brugger und Partner AG.
- Fabian, Carlo (2002): Evaluation der Mobilen Jugendarbeit Basel MJAB. Ein Projekt im Unteren Kleinbasel. Basel: HPSABB (unveröffentlicht).
- Fabian, Carlo (2004): Evaluation Kinderbüro. Schlussbericht. Basel: HPSABB (unveröffentlicht).
- Fabian, Carlo/Guggenbühl, Lisa (2000): Suchtprävention mit Risikogruppen. Zürich: ISGF.
- Fabian, Carlo/Gutmann, Renate (2004): CRASH! Ein Forum Theater gegen Rassismus und zu Fragen des Respekts. Evaluationsbericht. Basel: HPSABB (unveröffentlicht).
- Fabian, Carlo/Kaegi, Urs/Stübi, Beat (2004): Erfolgskontrolle der Staatsbeiträge an Institutionen der stationären Kinder- und Jugendhilfe im Kanton Bern. Basel: HPSABB (unveröffentlicht).
- Fabian, Carlo/Steiner, Olivier (2004): Evaluation der Projekte ‚Aufsuchen, Vertrauensbildung, Information‘ und ‚Abklärung, Vermittlung‘. Basel: HPSABB (unveröffentlicht).
- Fabian, Carlo/Steiner, Olivier/Guhl, Jutta (2006): Schule und Cannabis. Regeln, Massnahmen und Früherfassung. Basel: FHNW (unveröffentlicht).
- FAL, Bundesforschungsanstalt für Landwirtschaft/ARUM, Arbeitsgemeinschaft Umwelt und Stadtplanung/BFH, Bundesforschungsanstalt für Forst- und Holzwirtschaft (2005): Aktualisierung der Halbzeitbewertung des NRW-Programms ländlicher Raum gem. VO (EG) 1257/1999. Im Auftrag des Ministeriums für Umwelt und Naturschutz, Landwirtschaft und Verbraucherschutz des Landes Nordrhein-Westfalen. Braunschweig/Hannover/Hamburg: FAL/ARUM/BFH.
- Falk, Armin/Lalive, Rafael/Zweimüller, Josef (2002): The success of job applications: a new approach to program evaluation. IEW Working Paper, 131. Zürich: Institut für empirische Wirtschaftsforschung, Universität Zürich. (www.iew.uzh.ch/wp/iewwp131.pdf [28.5.2008])
- Falkner, Gerda (2006): Zur „Europäisierung“ des österreichischen politischen Systems. In: Dachs, Herbert/Gerlich, Peter/Gottweis, Herbert/Kramer, Helmut/Lauber, Volkmar/Müller, Wolfgang C./Tálos, Emmerich (Hrsg.): Politik in Österreich. Das Handbuch. Wien: Manz, S. 82-94.
- Faltermaier, Toni (2005): Gesundheitspsychologie. Stuttgart: Kohlhammer.
- Feld, Lars P./Savioz, Marcel R. (1997): Direct democracy matters for economic performance: An empirical investigation. *Kyklos* 50(4): 507-538.
- Felder, Stefan/Meyer, Leo/Rieder, Peter (1994): Vorzeitige Stilllegung von Höchsttierbeständen in der Fleisch- und Eierproduktion 1980-1984. In: Nationales Forschungsprogramm 27 (Hrsg.): Evaluationen in der Agrarpolitik II, Reihe Schlussberichte. Bern: SNF.
- Felder, Stefan/Rieder, Peter (1994): Analyse des Angebotsverhaltens und der Produktionsstruktur im Agrarsektor mit Hilfe eines LP Strukturmodells. Evaluationen in der Agrarpolitik I, Nationales Forschungsprogramm 27, Reihe Schlussberichte. Bern: SNF.
- Fend, Helmut (1982): Gesamtschule im Vergleich. Bilanz der Ergebnisse des Gesamtschulversuchs. Weinheim: Beltz.
- Fend, Helmut (1986): "Gute Schulen – Schlechte Schulen". Die einzelne Schule als pädagogische Handlungseinheit. *Deutsche Schule* 78(3): 275-293.
- Ferro Luzzi, Giovanni/Flückiger, Yves/Ramirez, José V./Vassiliev, Anatoli (2004): Swiss Unemployment Policy: An Evaluation of the Public Employment Service. *Swiss Journal of Sociology* 30(3): 319-337.
- Fertig, Michael/Kluve, Jochen/ Scheuer, Markus (2004): Die Hartz-Gesetze zur Arbeitsmarktpolitik – Ein umfassendes Evaluationskonzept. RWI-Schriften Bd. 74. Berlin: Duncker & Humblot.

- Fetterman, David M. (1994): Empowerment evaluation. *Evaluation Practice* 15(1): 1-15.
- Finz, Alfred (2002): Die Hauptanliegen der Verwaltungsreform und ihre Auswirkungen auf die Haushaltsführung. *Das öffentliche Haushaltswesen in Österreich* 43(1-2): 4-10.
- Fischer, Rudolf/Saleemi, Razzaq Abdul/Ahmad Sharjeel/Berger, Fritz (2005): Tracing impacts of the Kalam integrated development project. Bern: Intercooperation.
- Fleck, Dieter (2006): Evaluierung ROG – Handelsbetriebe <800m². Graz: Regionalentwicklung – DI Tischler ZT GmbH. (www.raumplanung.steiermark.at/cms/dokumente/10470288_265384/0456c76d/Studie_Handelsbetriebe_800m%C2%B2.pdf [13.5.2008])
- Fleischmann, Eduard (2005): Öffentliche Haushalte in Österreich im Überblick. In: Steger, Gerhard (Hrsg.): Öffentliche Haushalte in Österreich, 2. Auflage. Wien: Verlag Österreich, S. 7-35.
- Flick, Uwe (Hrsg.) (2006): Qualitative Evaluationsforschung. Konzepte, Methoden, Umsetzungen. Reinbek: Rowohlt.
- Florida, Richard (1995): Toward the learning region. *Futures* 27(5): 527-536.
- Flösser, Gaby/Oechler, Melanie (2004): Chancen und Risiken von Qualitätsmanagement in der Sozialen Arbeit. In: Beckmann, Christoph/Otto, Hans-Uwe/Richter, Martina/Schrödter Mark (Hrsg.): Qualität in der Sozialen Arbeit. Zwischen Nutzerinteressen und Kostenkontrolle. Wiesbaden: VS Verlag, S. 175-183.
- Flückiger, Alexandre/Morand, Charles-Albert/Tanquerel, Thierry (2000): Evaluation du droit de recours des organisations de protection de l'environnement. Cahier de l'environnement n°314. Berne: Office fédéral de l'environnement, des forêts et du paysage.
- Flückiger, Yves/de Coulon, Anne/Vassiliev, Anatoli (2002): Les raisons de la différence entre les taux de chômage genevois et suisse. Rapport No 11 de l'OUE. Série de publication du LEA No 24. Genève: Université de Genève.
- Flückiger, Yves/Morales, Daniel (1994): Analyse des causes de l'augmentation actuelle du chômage en Suisse et à Genève. Rapport No 1 de l'Observatoire Universitaire de l'Emploi. Série de publications du LEA No 5. Genève: Université de Genève.
- Flury, Christian (2005): Bericht Agrarökologie und Tierwohl 1994-2005. Zürich: Bundesamt für Landwirtschaft.
- Flury, Christian/Gautschi, Ursula/Albisser, Gregor/Gerwig, Claude/Mack, Gabriele (2003): Effekte einer Aufhebung der Milchkontingentierung und einer Umlagerung der Milchpreisstützung, Synthese aus vier Studien des Institutes für Agrarwirtschaft der ETH Zürich, IAW, und der Eidgenössischen Forschungsanstalt für Agrarwirtschaft und Landtechnik, FAT. Zürich: Bundesamt für Landwirtschaft.
- Foresti, Marta/Sharma, Bavna/O'Neil, Tammie/Evans, Alison (2007): Evaluation of citizen's voice and accountability: Evaluation framework. London: Overseas Development Institute.
- Fornerod, Serge (2001): A quoi et a qui servent les évaluations? Cahier de l'IDHEAP 199. Chavannes-près-Renens: IDHEAP.
- Forscherguppe Governance der Forschung (2007): Rahmenbedingungen für eine leistungsfähige öffentlich finanzierte Forschung. Forschungspolitische Thesen der Forschergruppe „Governance der Forschung“ (Deutsche Forschungsgemeinschaft, FOR 517), anlässlich der Tagung „Neue Governance für die Forschung“, Berlin-Brandenburgische Akademie der Wissenschaften 14. und 15. März 2007. Speyer: Deutsches Forschungsinstitut für öffentliche Verwaltung Speyer. (www.foev-speyer.de/governance1/projektbilder/BBAW-Conference/070301_Forschungspolitische%20Thesen%20wie%20Heft.pdf [8.1.2008])
- Forstner, Bernhard (2000): Erfolgskontrolle der einzelbetrieblichen Investitionsförderung in der Landwirtschaft. In: GeWiSoLa, Gesellschaft für Wirtschafts und Sozialwissenschaften des Landbaus (Hrsg.): Wettbewerbsfähigkeit und Unternehmertum in der Land- und Ernährungswirtschaft. Schriften der Gesellschaft für Wirtschafts- und Sozialwissenschaften des Landbaus, H. 36. Münster: Landwirtschaftsverlag Münster-Hiltrup, S. 151-158.

- Forstner, Bernhard/Grajewski, Regina/Efken, Josef (2004): Arbeitsbericht zum Workshop Zwischenbewertung der Programme zur Entwicklung des ländlichen Raumes nach der Verordnung (EG) Nr. 1257/1999 - Erfahrungsaustausch und Verbesserungsansätze, durchgeführt am 27./28. Januar 2004 in Braunschweig. Braunschweig: Selbstverlag.
- Forstner, Bernhard/Grajewski, Regina/Mehl, Peter (2002): Evaluation von Politikmaßnahmen zur Förderung ländlicher Räume: Theorie und Praxis. Braunschweig: Bundesforschungsanstalt für Landwirtschaft. (literatur.fal.de/fallitdok_extern/zi028835.pdf. [3.4.2008])
- Forstner, Bernhard/Koch, Birgit (2005): Die Halbzeitbewertung der Entwicklungsprogramme für den ländlichen Raum: Erfahrungen unterschiedlicher organisatorischer Bewertungsansätze. *Zeitschrift für Evaluation* 2005(1): 109-127.
- Frei, Andreas/Greiner, Roger-Axel/Mehnert, Angelika/Dinkel, Rolf (2000): Socioeconomic evaluation of heroin maintenance treatment. In: Gutzwiller, Felix/Steffen, Thomas (eds) *Cost-benefit analysis of heroin maintenance treatment*. Basel: Karger, pp. 37-128.
- Freiburghaus, Dieter (1987): Präventivmassnahmen gegen die Arbeitslosigkeit in der Schweiz. Methoden der Wirkungsanalyse und erste Ergebnisse. Bern: Paul Haupt.
- Freimann, Jürgen (1997): Ergebnisse der Begleitforschung zu Öko-Audits in Hessen. In: Klemisch, Herbert (Hrsg.): *Öko-Audit und Partizipation. Die betriebliche Umsetzung von Umweltinformationssystemen in kleinen und mittelständischen Unternehmen*. Köln: kni-Verlag, S. 23-36.
- Freitag, Markus/Vatter, Adrian (2000): Direkte Demokratie, Konkordanz und wirtschaftliche Leistungskraft. Ein Vergleich der Schweizer Kantone. *Schweizerische Zeitschrift für Volkswirtschaft und Statistik* 136: 579-606.
- Freitag, Markus/Vatter, Adrian/Müller, Christoph (2003): Bremse oder Gaspedal? Eine empirische Untersuchung zur Wirkung der direkten Demokratie auf den Steuerstaat. *Politische Vierteljahresschrift* 44(3): 348-369.
- Freudenthaler, Harald H./Specht, Werner (2006): Bildungsstandards: Der Implementationsprozess aus der Sicht der Praxis. Ergebnisse einer Fragebogen-Studie nach dem ersten Jahr der Pilotphase II. ZSE-Report Nr. 71. Graz: Zentrum für Schulentwicklung.
- Frey, Bruno S. (2008): Evaluitis – eine neue Krankheit. In: Matthies, Hildegard/Simon, Dagmar (Hrsg.): *Wissenschaft unter Beobachtung. Leviathan Sonderheft* 24: 125-140.
- Friboulet, Jean-Jacques/Guidotti, Sabrina/Sassi, Claudia/Simonet, Jean-Claude/Soulet, Marc-Henry (1997): Scénarios pour une politique en faveur des chômeurs en fin de droit : Etat des lieux et analyse prospective à partir de l'exemple fribourgeois. Fribourg: Edition Universitaires.
- Frick, Andres/Lampart, Daniel (2005): Wirksamkeit der Kurzarbeitsregelung in der Rezession 2001-2003. Arbeitsmarktpolitik Nr. 13. Bern: Seco.
- Frick, Frank/Brinkmann, Henrik/Ernst, Tobias (2007): Das Standard-Kosten-Modell. Ein neuer Ansatz für effektiven Bürokratieabbau. In: Schäffer, Heinz (Hrsg.): *Evaluierung der Gesetze/Gesetzesfolgenabschätzung (II). Weitere Erfahrungsberichte aus dem In- und Ausland/Staatspraxis und Wirtschaft*. Wien: Manz, S. 99-116.
- Fullan, Michael/Watson, Nancy (2000): School-Based Management: Reconceptualizing to Improve Learning Outcomes. *School Effectiveness and School Improvement* 11(4): 453-473.
- Furubo, Jan-Eric/Sandahl, Rolf (2002): Introduction. A diffusion perspective on global developments in evaluation. In: Furubo, Jan-Eric/Rist, Ray, C./Sandahl, Rolf (eds): *International atlas of evaluation*. New Brunswick: Transaction Press, pp. 1-23.
- Gärtner, Ludwig/Flückiger, Yves (2005): Probleme des Sozialstaats: Ursachen, Hintergründe, Perspektiven. Synthesebericht. Zürich: Rüegger.
- Gassner, Erich/Winkelbrandt, Arndt (2004): UVP. Umweltverträglichkeitsprüfung in der Praxis. 4. Auflage. Heidelberg: C.F. Müller.
- GEFRA – Gesellschaft für Finanz- und Regionalanalysen mbH/MR Gesellschaft für Regionalberatung/TRAST Transformationsprozesse und Strukturpolitik (2004): Evaluierung des Einsatzes von

- Fördermitteln im Rahmen der Gemeinschaftsaufgabe „Verbesserung der regionalen Wirtschaftsstruktur“ (GA) in den Jahren 1997 – 2003 in Thüringen. Münster: GEFRÄ. (www.thueringen-online.de/imperia/md/content/tmwta/wirtschaft/foerderung/evaluierung/evaluierung_ga_kurzfasung.pdf [13.5.2008])
- Gemeinsamer Bundesausschuss (2005): Verfahrensordnung des Gemeinsamen Bundesausschusses. vom 20. September 2005, veröffentlicht im Bundesanzeiger 2005, S. 16 998, in Kraft getreten am 1. Oktober 2005; zuletzt geändert am 18. April 2006, veröffentlicht im Bundesanzeiger 2006, S. 4876, in Kraft getreten am 7. Juli 2006. Siegburg: Gemeinsamer Bundesausschuss. (www.g-ba.de/downloads/62-492-83/VerfO_2006-04-18.pdf [10.4.2008])
- Georgiou, Luke (1995): Research evaluation in European national science and technology systems. *Research Evaluation* 5(1): 3-10.
- Gerfin, Michael/Lechner, Michael (1999): Ökonometrische Evaluation der arbeitsmarktlichen Massnahmen in der Schweiz. Bern: Seco.
- Gerfin, Michael/Leu, Robert/Brun, Stephan/Tschöpe, Andreas (2002): Armut unter Erwerbstätigen in der Schweiz: Eine Beurteilung alternativer wirtschaftspolitischer Lösungsansätze. Bern: Seco.
- Gerheuser, Frohmüt (2006): Evaluation der Kampagne solarbegeistert. Bern: Bundesamt für Energie.
- Gerheuser, Frohmüt/Vatter, Adrian/Sager, Fritz (1997): Die Berücksichtigung von Stellungnahmen der Kantone im Vernehmlassungsverfahren des Bundes. Studie zuhanden der Parlamentarischen Verwaltungskontrollstelle. Bern: EDMZ.
- Germann, Raimund E. (1981): Ausserparlamentarische Kommissionen. Die Milizverwaltung des Bundes. Bern: Haupt.
- Germann, Raimund E./Weis, Katja (1995): Die Kantonsverwaltungen im Vergleich. Haupt: Bern.
- Gerster, Richard (2006): Zusammenarbeit mit Musterschülern im Vordergrund. Die staatliche Entwicklungszusammenarbeit der Schweiz vor grossen Herausforderungen. *EINS – Entwicklungspolitik Information Nord-Süd* 13-14/2006: 45-48.
- Geschäftsbericht des Bundesrates (2008): Bericht des Bundesrates über seine Geschäftsführung und die Schwerpunkte der Verwaltungsführung im Jahre 2007. Bern: Schweizerische Bundeskanzlei.
- Geser, Hans (1981): Bevölkerungsgrösse und Staatsorganisation. Bern: Haupt.
- Geser, Hans/Höpflinger, François (1977): Staatsverwaltung und Sozialstruktur: Ein Vergleich zwischen vier Kantonen. *Schweizerisches Jahrbuch für Politische Wissenschaft* 17: 111-135.
- Giacomazzi, Fabio/Clerici, Remo/Marti, Peter/Rudel, Roman/Brugnoli, Gianni/Passardi Gianola, Lorenza/Gianola, Giacomo/Castelli, Stefano/Meneghin, Franco/Von Wartburg, Sara (2004): Räumliche Auswirkungen der Verkehrsinfrastrukturen in der Magadinoebene – eine ex post-Analyse. Zusammenfassung. Bern: ARE.
- GIB – Gesellschaft für Innovationsforschung und Beratung mbH (2003): Protokoll Reflexionsworkshop zum Bundesprogramm Ökologischer Landbau am 4. und 5. Dezember in Bonn, erstellt im Rahmen der Evaluation des Bundesprogramms Ökologischer Landbau. (www.bundesprogramm-oekolandbau.de/fileadmin/sites/default/files/evaluation_boel_reflexionsworkshop_041011.pdf [16.7.2007])
- GIB – Gesellschaft für Innovationsforschung und Beratung mbH (2004): Abschlussbericht der Evaluation des Bundesprogramms Ökologischer Landbau, im Auftrag der Bundesanstalt für Landwirtschaft und Ernährung. Internetseite Geschäftsstelle Bundesprogramm Ökologischer Landbau in der Bundesanstalt für Landwirtschaft und Ernährung (BLE). (www.bundesprogramm-oekolandbau.de/fileadmin/sites/default/files/evaluation_boel_bericht_041011.pdf. [10.5.2007])
- Gibbons, Micheal/Limoges, Camille/Nowotny, Helga/Schwartzman, Simon/Scott, Peter/Trow, Martin (1994): The new production of knowledge. The dynamics of science and research in contemporary societies. London: Sage.

- Gonon, Philipp/Hügli, Ernst/Landwehr, Norbert/Ricka, Regula/Steiner, Peter (1998): Qualitätssysteme auf dem Prüfstand. Die neue Qualitätsdiskussion in Schule und Bildung – Analyse und Perspektiven. Aarau: Sauerländer.
- Good, Barbara (2006): Technologie zwischen Staat und Markt. Die Kommission für Technologie und Innovation und die Wirksamkeit ihrer Förderung. Zürich: Rüegger.
- Gormley, William T. jr. (2007): Public policy analysis: Ideas and impacts. *Annual Review of Political Science* 10: 297-313.
- Gottweis, Herbert/Latzer, Michael (2006): Forschungs- und Technologiepolitik. In: Dachs, Herbert/Gerlich, Peter/Gottweis, Herbert/Kramer, Helmut/Lauber, Volkmar/Müller, Wolfgang C./Tálos, Emmerich (Hrsg.): Politik in Österreich. Das Handbuch. Wien: Manz, S. 711-725.
- GPk-S (2006): Kohärenz und strategische Führung der Aktivitäten der DEZA. Bericht der Geschäftsprüfungskommission des Ständerates. Bern: BBL.
- Gredig, Daniel/Sommerfeld, Peter/Kaegi, Urs (2003): Evidence-based intervention development. Antrag zur Unterstützung eines Forschungsverbundes. Basel/Olten (unveröffentlicht).
- Greenhalgh, Trisha (1999): Narrative based medicine in an evidence based world. *British Medical Journal* 318: 323-325.
- Greiling, Michael (Hrsg.) (2004): Pfade durch das klinische Prozessmanagement. Stuttgart: Kohlhammer.
- Greve, Carsten (2003): Public sector reform in Denmark: Organizational transformation and evaluation. *Public Organization Review* 3(3): 269-280.
- Grogger, Günter/Specht, Werner (Hrsg.) (1999): Evaluation und Qualität im Bildungswesen. Problem-analyse und Lösungsansätze am Schnittpunkt von Wissenschaft und Bildungspolitik. Dokumentation eines internationalen Workshops, Blumau/Steiermark 18.-21. Feber 1999. Graz: Zentrum für Schulentwicklung.
- Grohmann, Romano (1997): Das Problem der Evaluation in der Sozialpädagogik. Europäische Hochschulschriften. Frankfurt: Lang.
- Gruber, Edelgard/Böde, Ulla (2000): REN-Impuls-Programm „RAVEL NRW“. Begleitende Bewertung. Karlsruhe (unveröffentlicht).
- Gruber, Heinz/Petri, Gottfried (1989): Schulversuche zum gemeinsamen Unterricht behinderter und nichtbehinderter Kinder: Rahmenkonzepte und Ergebnisse einer ersten Befragung von Versuchslehrern. Graz: Zentrum für Schulentwicklung.
- Grunwald, Armin (2002): Technikfolgenabschätzung. Eine Einführung. Berlin: Edition Sigma.
- Grüter, Karin/Schaffner, Dorothee (2006): Mittagstisch mit Nachmittagsbetreuung St. Johann Schällemätteli. Begleitende Evaluation. Basel: FHNW (unveröffentlicht).
- Grüter, Kurt (1999): Die Finanzaufsicht beim Bund im Wandel. *Der Schweizer Treuhänder* 9: 834-839.
- Guggenbühl, Lisa (1999a): Weitere Auswirkungen des Projektes Euro-peers in den beteiligten Europäischen Regionen. Forschungsbericht. Zürich: ISGF.
- Guggenbühl, Lisa (1999b): Europäische Vernetzung in der Suchtprävention. Schlussbericht einer qualitativen Prozessevaluation der Netzwerkentwicklung. Forschungsbericht. Zürich: ISGF.
- Guggenbühl, Lisa (2000): Qualitätsentwicklung in den Suchtpräventionsstellen des Kantons Aargau. Zürich : ISGF.
- Güller, Peter/Schenkel, Walter/De Tommasi, Roberto/Oetterli, Dominik (2004): Räumliche Auswirkungen der Zürcher S-Bahn – eine ex-post Analyse. Zusammenfassung. Bern: ARE.
- Gunten, Ursula von (2006): Evaluation of Community Building and Local Development with Minorities, Municipalities Prizren, Dragash, Shtrpce, Kosovo. Basel: Soziologisches Institut der Universität (unveröffentlicht).
- Gunten, Ursula von/Berger-Künzli, Daniela (2006): Evaluation des Projektes Image – das Festival gegen Rassismus. Basel: FHNW (unveröffentlicht).

- Guston, David (2003): The expanding role of peer review processes in the United States. In: Shapira, Philip/Kuhlmann, Stefan (eds.): Learning from science and technology policy evaluation. Cheltenham: Edward Elgar, pp. 81-99.
- Güttinger, Franziska/Gschwend, Patrick/Schulte, Bernd/Rehm, Jürgen/Uchtenhagen, Ambros (2002): Evaluating long-term-effects of heroin-assisted treatment: The results of a 6-year follow-up. *European Addiction Research* 9: 73-79.
- Haari, Roland/Haari-Oberg, Ilse/Schilling, Karl/Torisi, Marco (2001): Kostendifferenzen zwischen den Kantonen. Forschungsbericht 15/01. Bern: Bundesamt für Sozialversicherung.
- Hackenberg, Helga (Hrsg.) (2003): Lokale Arbeitsmarktpolitik – Stand und Perspektiven. Gesamtbericht des Netzwerkprojektes »BiK - Beschäftigungsförderung in Kommunen« der Bertelsmann Stiftung. Gütersloh: Verlag Bertelsmann Stiftung.
- Hagen, Tobias/Spermann, Alexander (2004): Hartz-Gesetze – Methodische Ansätze zu einer Evaluierung. ZEW Wirtschaftsanalysen – Schriftenreihe des ZEW Bd. 74. Baden-Baden: Nomos.
- Haider, Günter/Eder, Ferdinand/Specht, Werner/Spiel, Christiane/Wimmer, Manfred (2005): Abschlussbericht der Zukunftskommission. Zusammenfassende Empfehlungen für Reformmaßnahmen. Wien: Bundesministerium für Bildung, Wissenschaft und Kultur.
- Hambrusch, Josef (2006): Österreichs Agrarmärkte nach der EU-Erweiterung. BMLFUW: Ländlicher Raum. ([www.laendlicher-raum.at/article/articleview/49513/1/14535/\[2.7.2007\]](http://www.laendlicher-raum.at/article/articleview/49513/1/14535/[2.7.2007]))
- Hambrusch, Josef/Ortner, Karl Michael (2005): Der Bananenmarkt in der EU. Einfluss einer geänderten Marktordnung auf Bananenmarkt und -handel. Agrarpolitische Arbeitsbeihilfen 21. Wien: Bundesanstalt für Agrarwirtschaft.
- Hammer, Stephan/Frick, Roman/Kessler, Stephan/Iten, Rolf (2000): Evaluation der Berner Energiepolitik. Auswirkungen auf Energie und Beschäftigung. Zürich: Infrac.
- Hammer, Stephan/Pulli, Raffael/Iten, Rolf/Eggimann, Jean-Claude (2001): Auswirkungen des KVG auf die Versicherer. Bern: Bundesamt für Sozialversicherung
- Hammerschmid, Gerhard (2006): Status der Verwaltungsreform in Österreich: eine „unendliche“ Geschichte von Stabilität und Veränderung. In: Hofmeister, Alfred (Hrsg.): Verwaltung wohin? Der öffentliche Sektor zwischen Stabilität und Veränderung. SGVW-Schriftenreihe 46. Bern: SGVW, S. 75-94.
- Hammerschmid, Gerhard/Meyer, Renate E. (2005a): New public management in Austria: Local variation on a global theme. *Public Administration* 83(3): 709-733.
- Hammerschmid, Gerhard/Meyer, Renate E. (2005b): Public management dynamics in a federalist legalistic Rechtsstaat system. Results from an executive survey in Austria. *International Journal of Public Sector Management* 38(7): 629-640.
- Hammerschmid, Gerhard/Meyer, Renate E. (2006): Public management reform: An identity project. *Public Policy and Administration* 21(1): 99-115.
- Hanft, Anke/Kohler, Alexander (2007): Qualitätssicherung im österreichischen Hochschulsystem. Positionierung der AQA und Perspektiven der externen Zertifizierung von Qualitätsmanagementprozessen. *Zeitschrift für Hochschulrecht* 6(3): 83-93.
- Hangartner, Yvo (2006): Der Rechnungshof – ein neues staatsrechtliches Institut: Bemerkungen aus Anlass der Schaffung einer Cour des comptes im Kanton Genf. *Schweizerisches Zentralblatt für Staats- und Verwaltungsrecht* 10(9): 453-470.
- Hansen, Hanne Fossen (2005): Evaluation in and of public-sector reform: The case of Denmark in a Nordic perspective. *Scandinavian Political Studies* 28(4): 323-347.
- Hansen, Joe B. (1994): Book Review: The Joint Committee on Standards for Educational Evaluations (1994): The program evaluation standards: How to assess evaluation of educational programs. Thousand Oaks: Sage. *Educational and Psychological Measurement* 54(2): 550-561.

- Hanser, Christian/Kuster, Jürg/Gessler, Rahel/Ehrler, Melchior/Dietler, Urs/Zimmermann, Hubert/Speck, Raphael (2005): Evaluation der Abfallpolitik des Bundes. Bericht im Auftrag des BUWAL. Zürich: BHP-Hanser und Partner AG/Electrowatt-Ekono AG.
- Haubrich, Karin (2009): Evaluation innovativer multizentrischer Programme: Möglichkeiten und Grenzen programmtheoriebasierter Cluster-Evaluation am Beispiel von Bundesmodellprogrammen in der Jugendhilfe. Dissertation an der Freien Universität Berlin (in Vorbereitung).
- Haubrich, Karin/Holthusen, Bernd/Struhkamp, Gerlinde (2005): Evaluation – einige Bemerkungen zu einem schillernden Begriff. *DJI Bulletin plus* 72: 1-5.
- Haubrich, Karin/Lüders, Christian (2001): Evaluation – hohe Erwartungen und ungeklärte Fragen. *Diskurs – Studien zu Kindheit, Jugend, Familie und Gesellschaft* 11(3): 69-73.
- Haubrich, Karin/Lüders, Christian (2004): Evaluation – mehr als ein Modewort? *Recht der Jugend und des Bildungswesens. Zeitschrift für Schule, Berufsbildung und Jugendberufshilfe* 57(3): 316-337.
- Haubrich, Karin/Lüders, Christian/Struhkamp, Gerlinde (2007): Wirksamkeit, Nützlichkeit, Nachhaltigkeit. Was Evaluationen von Modellprogrammen realistisch leisten können. In: Schröder, Ute/Streblow, Claudia (Hrsg.): *Evaluation konkret – Fremd- und Selbstevaluationsansätze anhand von Beispielen aus Jugendarbeit und Schule*. Opladen: Barbara Budrich, S. 183-201.
- Hauff, Volker (Hrsg.) (1987): *Unsere gemeinsame Zukunft. Der Brundtland-Bericht der Weltkommission für Umwelt und Entwicklung*. Greven: Eggenkamp.
- Hauke, Eugen/Schaffler Roland/Holzer-Möstl, Elke/Stekel, Herbert (2001): Österreichisches Zertifizierungsmodell für Gesundheitseinrichtungen – Anforderungen und Rahmenbedingungen. Wien: Ludwig Boltzmann Institut für Krankenhausorganisation.
- Heckman, James/LaLonde, Robert J./Smith, Jeffrey A. (1999): The economics and econometrics of active labour market programmes. In: Ashenfelter, Orley/Card, David E. (eds.): *Handbook of labour economics*, Vol. 3A. Amsterdam: Elsevier, pp. 1865-2097.
- Heil, Karolus/Heiner, Maja/Feldmann, Ursula (Hrsg.) (2001): *Evaluation sozialer Arbeit. Eine Arbeitshilfe mit Beispielen zur Evaluation und Selbstevaluation*. Frankfurt: Deutscher Verein.
- Heiner, Maja (1988): Von der forschungsorientierten zur praxisorientierten Selbstevaluation. Entwurf eines Konzepts. In: Heiner, Maja (Hrsg.): *Selbstevaluation in der Sozialen Arbeit. Fallbeispiele zur Dokumentation und Reflexion beruflichen Handelns*. Freiburg: Lambertus, S. 7-40.
- Heiner, Maja (2001): Evaluation. In: Otto, Hans-Uwe/Thiersch, Hans (Hrsg.): *Handbuch Sozialarbeit/Sozialpädagogik*. 2. Auflage. Neuwied: Luchterhand, S. 481-495.
- Heiner, Maja (Hrsg.) (1998): *Experimentierende Evaluation. Ansätze zur Entwicklung lernender Organisationen*. Weinheim: Juventa.
- Heintzel, Martin (2004): Probleme der Evaluation von Regionalmanagement – oder: Ist Erfolg messbar? In: Sedlacek, Peter (Hrsg.): *Evaluation in der Stadt- und Regionalentwicklung*. Wiesbaden: VS Verlag, S. 123-143.
- Hellstern, Gerd-Michael/Wollmann, Hellmut (1984a): Evaluierung und Evaluierungsforschung – ein Entwicklungsbericht. In: Hellstern, Gerd-Michael/Wollmann, Hellmut (Hrsg.): *Handbuch zur Evaluierungsforschung*, Band 1. Opladen: Westdeutscher Verlag, S. 17-93.
- Hellstern, Gerd-Michael/Wollmann, Hellmut (Hrsg.) (1984b): *Handbuch zur Evaluierungsforschung*, Band 1. Opladen: Westdeutscher Verlag.
- Hellstern, Gerd-Michael/Wollmann, Hellmut (Hrsg.) (1984c): *Evaluierung und Erfolgskontrolle in Kommunalpolitik und -verwaltung*. Basel: Birkhäuser.
- Helmke, Andreas/Jäger, Reinhold S. (2002): *Das Projekt MARKUS. Mathematik-Gesamterhebung Rheinland-Pfalz: Kompetenzen, Unterrichtsmerkmale, Schulkontext*. Landau: Verlag Empirische Pädagogik.
- Helmke, Andreas/Schrader, Friedrich-Wilhelm (2001): Von der Leistungsevaluation zur Unterrichtsentwicklung. In: Silbereisen, Rainer K./Reitzle, Matthias J. (Hrsg.): *Bericht über den 42. Kongress der Deutschen Gesellschaft für Psychologie in Jena*. Lengerich: Pabst, S. 594-606.

- Hendricks, Michael/Papagiannis, Meredith (1990): Do's and don'ts for offering effective recommendations. *Evaluation Practice* 11(2): 121-125.
- Henneberger, Fred/Sousa-Poza, Alfonso/Ziegler, Alexandre (2004): Eine empirische Analyse der Arbeit auf Abruf in der Schweiz. Arbeitsmarktpolitik Nr. 11. Bern: Seco.
- Hennen, Manfred/Häuser, Simon (2002): Evaluation und Organisationsentwicklung – ein Vergleich. Mainz: Zentrum für Qualitätssicherung und -entwicklung (ZQ) an der Johannes Gutenberg-Universität Mainz. (www.hrk-qm.de/de/download/dateien/08-2002_-_Vergleich_Evaluation_und_Organisationsentwicklung_-_Hennen_-_Haeuser.pdf [21.5.2008])
- Hennicke, Peter/Fischedick, Manfred (2007): Erneuerbare Energien. Mit Energieeffizienz zur Energiewende. München: Ch. Beck.
- Henrichsmeyer, Wilhelm/Witzke, Heinz Peter (1991): Agrarpolitik. Band 1: Agrarökonomische Grundlagen. Stuttgart: UTB Ulmer.
- Henry, Gary T./Mark, Melvin M. (2003): Beyond use: Understanding evaluation's influence on attitudes and actions. *American Journal of Evaluation* 24(3): 293-314.
- Henry, Gary T./Rog, Debra J. (1998): A realist theory and analysis of utilization. In: Henry, Gary T./Junes, George/Mark, Melvin M. (eds): Realist evaluation: An emerging theory in support of practice. *New Directions for Evaluation* 78: 89-102.
- Herbeck, Gabriele (2005): Verwaltungsreform in Österreich. In: Steger, Gerhard (Hrsg.): Öffentliche Haushalte in Österreich. 2. Auflage. Wien: Verlag Österreich, S. 241-262.
- Hering, Sabine/Münchmeier, Richard (2000): Geschichte der Sozialen Arbeit. Eine Einführung. Weinheim: Juventa.
- Herweg, Karl/Steiner, Kurt (2002): Impact monitoring and assessment. Instruments for use in rural development projects with a focus on sustainable land management. Bern/Eschborn: CDE/GTZ.
- Heyer, Gerd (2006): Zielsetzung und Struktur der „Hartz-Evaluation“. *Zeitschrift für Arbeitsmarktforschung* 39(3/4): 467-476.
- Hilfe an Opfer von Straftaten (2000): Dritter Bericht des Bundesamtes für Justiz an den Bundesrat über den Vollzug und die Wirksamkeit der Opferhilfe (1993-1998). Bern: Bundesamt für Justiz.
- Hödl, Josef (2007): Das kleine Fenster zur Welt. Ambivalenzen und Implikationen der Ökonomisierung Sozialer Arbeit. In: Hödl, Josef/Posch, Klaus/Wilhelmer, Peter (Hrsg.): Sprache und Gesellschaft. Wien: Verlag Österreich, S. 527-547.
- Hofer, Eduard (1985): Preisdifferenzierung in der Landwirtschaft, Mitteilung des Schweizerischen Bauernsekretariates Nr. 211, Brugg: Schweizerischer Bauernverband.
- Hofer, Helmut/Weber, Andrea (2006): Active labor market policy in Austria: Practice and evaluation results. *Vierteljahreshefte zur Wirtschaftsforschung* 75(3): 155-167.
- Hoffmann, Karl-Heinz/Bourguignon, Jean-Pierre (2005): Evaluation von Forschung und Lehrprogrammen an den Fachbereichen für Mathematik der österreichischen Universitäten. Studie im Auftrag der Österreichischen Mathematischen Gesellschaft und des bm:bwk. Wien: bmbwk.
- Hofmarcher, Maria M./Rack, Herta M. (2006): Gesundheitssysteme im Wandel. Österreich. Berlin: Medizinisch Wissenschaftliche Verlagsgesellschaft.
- Hofreither, Markus F. (2002): Der Reformbedarf der GAP aus österreichischer Perspektive - Endbericht. Wien: Institut für Wirtschaft, Politik und Recht der Universität für Bodenkultur.
- Hofreither, Markus F./Schneider, Friedrich/Weiß, Christoph (1987): Biosprit in Österreich. Eine volkswirtschaftliche Analyse. Wien: Arbeitsgemeinschaft Erneuerbare Energien.
- Hofstetter, Carolyn H./Alkin, Marvin C. (2003): Evaluation use revisited. In: Kellaghan, Thomas (ed.): International handbook of educational evaluation. Dordrecht: Kluwer, pp. 197-222.
- Holzinger, Elisabeth (2001): 6 Jahre Programm-Evaluation – eine Bilanz. Wien: Österreichisches Institut für Raumplanung
- Holzinger, Elisabeth/Landsteiner, Günther (1993): Programmevaluation. Theoretische Grundlagen und Anwendungsbereiche. Wien: Österreichisches Institut für Raumplanung.

- Homfeld, Hans Günther/Schulze-Krüdener, Jörgen (Hrsg.) (2003): Handlungsfelder der Sozialen Arbeit. Basiswissen Pädagogik: Pädagogische Arbeitsfelder, Band 3. Baltmannsweiler: Schneider Verlag Hohengehren.
- Hood, Christopher (2004): Conclusion: Making sense of controls over government. In: Hood, Christopher/James, Oliver/Peters, B. Guy/Scott, Colin (eds): Controlling modern government. Variety, commonality and change. Cheltenham: Edward Elgar, pp. 185-205.
- Hoppe, Andrea/Voelzkow, Helmut (1999): Raumordnungs- und Regionalpolitik: Rahmenbedingungen, Entwicklungen, Perspektiven. In: Ellwein, Thomas/Holtmann, Everhard (Hrsg.): 50 Jahre Bundesrepublik Deutschland. *Politische Vierteljahresschrift, Sonderheft 30*: 279-296.
- Horber-Papazian, Katia (2006): Les communes. In: Klöti, Ulrich/Knoepfel, Peter/Kriesi, Hanspeter/Linder, Wolf/Papadopoulos, Yannis/Sciarini, Pascal (Hrsg.): Handbuch der Schweizer Politik. 4. Auflage. Zürich: Verlag NZZ, S. 233-258.
- Horber-Papazian, Katia (éd.) (1990): Evaluation des politiques publiques en Suisse. Pourquoi? Pour Qui? Comment? Lausanne: Presses polytechniques et universitaires romandes.
- Horber-Papazian, Katia/Bützer, Michael (2008): Dissemination of evaluation reports in newspapers: The case of CEPP evaluations in Geneva, Switzerland. In: Boyle, Richard/Breul, Jonathan D./Dahler-Larsen, Peter (eds): Open to the public. Evaluation in the public arena. New Brunswick: Transaction, pp. 43-66.
- Horber-Papazian, Katia/Thévoz, Laurent (1990): Switzerland: Moving towards evaluation. In: Rist, Ray C. (ed.): Program evaluation and the management of government. New Brunswick: Transaction, pp. 133-143.
- Hornbostel, Stefan (1997): Wissenschaftsindikatoren. Bewertungen in der Wissenschaft. Opladen: Westdeutscher Verlag.
- Hornbostel, Stefan (2008): Neue Evaluationsregime? Von der Inquisition zur Evaluation. In: Matthies, Hildegard/Simon, Dagmar (Hrsg.): Wissenschaft unter Beobachtung. *Leviathan Sonderheft 24*: 59-82.
- Horton, Douglas/Alexaki, Anastasia/Bennett-Lartey, Samuel/Brice, Kim Noële/Campilan, Dindo/Carden, Fred/Silva, José de Souza/Duong, Le Thanh/Khadar, Ibrahim/Boza, Albina Maestrey/Muniruzzaman, Imrul Kayes/Perez, Jocelyn/Chang, Matilde Somarriba/Vernooy, Ronnie/Watts, Jamie (2003): Evaluating capacity development: experiences from research and development organizations around the world. The Netherland/Canada: ISNAR, IDRC, CTA.
- Hoschka, Alexander/Hössl, Alfred/Raab, Erich (1978): Die wissenschaftliche Begleitung eines überregionalen Modellversuchs. In: Müller, C. Wolfgang (Hrsg.): Begleitforschung in der Sozialpädagogik. Analysen und Berichte zur Evaluationsforschung in der Bundesrepublik. Weinheim: Beltz, S. 85-114.
- Hosenfeld, Ingmar/Groß Ophoff, Jana (2007): Nutzung und Nutzen von Evaluationsstudien in Schule und Unterricht. *Empirische Pädagogik 21*(4): 352-367.
- House, Ernest R. (2007): Regression to the mean: A novel of evaluation politics. Charlotte: Information Age.
- Hovorka, Gerhard (2004): Einkommenswirkungen des österreichischen Programms für die Entwicklung des ländlichen Raums: Mid Term Evaluierung 2003. Facts & features 29. Wien: Bundesanstalt für Bergbauernfragen.
- Hovorka, Gerhard (2005): Ergebnisse der Evaluierung der Ausgleichszulage für benachteiligte Gebiete in Österreich. BMLFUW: Ländlicher Raum. (www.laendlicher-raum.at/article/articleview/49508/1/14534/ [2.7.2007])
- HRK (Hrsg.) (1998): Evaluation. Sachstandsbericht zur Qualitätsbewertung und Qualitätsentwicklung in deutschen Hochschulen. Beiträge zur Hochschulpolitik 1/1998. Bonn: HRK.
- HRK (Hrsg.) (2006a): Standards und Leitlinien für die Qualitätssicherung im Europäischen Hochschulraum. Beiträge zur Hochschulpolitik 9/2006. Bonn: HRK.

- HRK (Hrsg.) (2006b): Von der Qualitätssicherung der Lehre zur Qualitätsentwicklung als Prinzip der Hochschulsteuerung, 2 Bde. Beiträge zur Hochschulpolitik 1/2006. Bonn: HRK.
- HRK (Hrsg.) (2007): Wegweiser 2006. Qualitätssicherung an Hochschulen. Beiträge zur Hochschulpolitik 9/2007. Bonn: HRK.
- HRK/ACQUIN (Hrsg.) (2007): Prozessqualität für Lehre und Studium. Konzeption und Implementierung eines Verfahrens der Prozessakkreditierung. Bonn: HRK.
- Huber, Joseph/Müller, Axel (2004): Zur Evaluation von Umweltschutzmassnahmen. In: Stockmann, Reinhard (Hrsg.): Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder. 2. Auflage. Opladen: Leske + Budrich, S. 343-374.
- Hübler, Karl-Hermann (2002): Erfolgskontrolle und Wirkungsanalyse in der Raumplanung – weshalb Erkenntnisse dazu in Deutschland dringlich sind. In: Keim, Karl-Dieter/Kühn, Manfred (Hrsg.): Regionale Entwicklungskonzepte – Strategien und Steuerungswirkungen. Akademie für Raumforschung und Landesplanung. ARL-Arbeitsmaterial Nr. 287, S. 10-23.
- Hummelbrunner, Richard (2004): Wirkungsmonitoring – Leitfaden für die Anwendung bei Programmen und Projekten. Graz: ÖAR Regionalberatung.
- Hunold, Claude (1998): Qualitätsbeurteilung der Regionalen Arbeitsvermittlungszentren (RAV) und der arbeitsmarktlichen Massnahmen. Telefonische Repräsentativbefragung von Stellensuchenden und Unternehmungen. Bern: BWA.
- Hurrelmann, Klaus/Laser, Ulrich (Hrsg.) (2003): Handbuch Gesundheitswissenschaften. 3. Auflage. Weinheim: Juventa.
- Hüsler, Gebhard/Werlen, Egon/Rehm, Jürgen (2005): The Action Plan – A New Instrument to Collect Data on Interventions in Secondary Prevention in Adolescents. *Substance Use & Misuse* 40: 761-777.
- Hüsler, Gebhard/Werlen, Egon (2006): Supra-f: Ein Sucht-Präventionsforschungsprogramm. Schlussbericht. Bern: Bundesamt für Gesundheit.
- IDD and Associates (2006): Evaluation of general budget support: Synthesis Report. A joint evaluation of General Budget support 1994-2004. Birmingham: International Development Department, School of Public Policy, University of Birmingham.
- ifo Institut für Wirtschaftsforschung (1999): Gemeinschaftsaufgabe „Verbesserung der regionalen Wirtschaftsstruktur in Thüringen – Evaluierung des Fördermitteleinsatzes von 1991 bis 1996. ifo Dresden Studien 21. Dresden: ifo.
- IfS Institut für Stadtforschung und Strukturpolitik/GEFRE Gesellschaft für Finanz- und Regionalanalysen/MR Gesellschaft für Regionalberatung mbH (2003): Halbzeitbewertung des EFRE in Sachsen. Berlin: IfS.
- IfS Institut für Stadtforschung und Strukturpolitik/GEFRE Gesellschaft für Finanz- und Regionalanalysen/MR Gesellschaft für Regionalberatung mbH (2005): Aktualisierung der Halbzeitbewertung des EFRE in Sachsen. Berlin: IfS.
- Independent Evaluation Group (IEG)/World Bank (2007): Sourcebook for evaluating global and regional partnership programs. Washington: World Bank. (siteresources.worldbank.org/EXTGLOREGPARPRO/Resources/sourcebook.pdf [10.3.2008])
- Inderbitzin, Jürg/Erni, Rita (2000): Erfolgskontrolle Soziallohn-Projekte im Kanton Zug. Zug: Kantonales Sozialamt Zug.
- Institut für Sozialforschung und Sozialwirtschaft e.V. – ISO (2006): Organisatorischer Umbau der Bundesagentur für Arbeit. Evaluation der Maßnahmen zur Umsetzung der Vorschläge der Hartz-Kommission. Saarbrücken: ISO.
- Internationale Kommission (1999): Forschungsförderung in Deutschland. Bericht der internationalen Kommission zur Systemevaluation der Deutschen Forschungsgemeinschaft und der Max-Planck-Gesellschaft. Hannover: Volkswagen-Stiftung.
- Ipfling, Heinz-Jürgen (1981): Modellversuche mit Ganztagschulen und anderen Formen ganztägiger Förderung. Bonn: Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung.

- Irresberger, Karl (2005): Evaluierung und Gesetzesfolgenabschätzung in der österreichischen Bundesgesetzgebung – Erfahrungen und Leistungen. In: Schäffer, Heinz (Hrsg.): Evaluierung der Gesetze/Gesetzesfolgenabschätzung in Österreich und im benachbarten Ausland. Wien: Manz, S. 75-89.
- Isermeyer, Folkhard/Gocht, Alexander/Kleinhanß, Werner/Küpker, Bernd/Offermann, Frank/Osterburg, Bernhard/Riedel, Jochim/Sommer, Ulrich (2005): Vergleichende Analyse verschiedener Vorschläge zur Reform der Zuckermarktordeung. Eine Studie im Auftrag des Bundesministeriums für Verbraucherschutz, Ernährung und Landwirtschaft. Landbauforschung Völkenrode, Sonderheft, H. 282. Braunschweig: Bundesforschungsanstalt für Landwirtschaft. (literatur.fal.de/fallitdok_extern/bitv/zi037418.pdf [21.2.2008])
- Iten, Rolf/Schneider, Christian/Schreyer, Christoph/Rigassi, Reto (2006): Wirkungsanalyse EnergieSchweiz 2005. Bern: Bundesamt für Energie.
- Jacob, Steve/Varone, Frédéric (2003): Evaluer l'action publique. Etat des lieux et perspectives en Belgique. Gent: Academia Press.
- Jacoby, Klaus-Peter/Schneider, Vera/Meyer, Wolfgang/Stockmann, Reinhard (2005): Umweltkommunikation im Handwerk. Bestandsaufnahme, vergleichende Analyse, Entwicklungsperspektive. Münster: Waxmann.
- Janetschek, Hubert (2001): Bewertungsergebnisse der einzelbetrieblichen Investitionsförderung und Niederlassungsbeihilfe der VO(EG) 950/97 in Österreich 1995-1999. Agrarpolitische Arbeitsbeihelfe 5. Wien: Bundesanstalt für Agrarwirtschaft.
- Janett, Daniel (2003). Erfolgskontrolle staatlichen Handelns durch das Parlament. Zu den Aktivitäten und Erfolgsbedingungen der parlamentarischen Verwaltungskontrollstelle der Bundesversammlung. In: Schweizer, Rainer J./Jeanrenaud, Claude/Kux, Stephan/Sitter-Liver, Beat (Hrsg.): Verwaltung im 21. Jahrhundert. Herausforderungen, Probleme, Lösungswege. Freiburg: Universitätsverlag, S. 97-109.
- Janett, Daniel (2004). Die Evaluationsfunktion des Bundesparlaments – eine Zwischenbilanz. *LeGes – Gesetzgebung & Evaluation* 15(2): 127-150.
- Jänicke, Martin (2006): Umweltpolitik – auf dem Weg zur Querschnittspolitik. In: Schmidt, Manfred G./Zohnhörer, Reimut (Hrsg.): Regieren in der Bundesrepublik Deutschland. Innen- und Aussenpolitik seit 1949. Wiesbaden: VS Verlag, S. 405-418.
- Jann, Werner (2001): Verwaltungsreform und Verwaltungspolitik: Verwaltungsmodernisierung und Policy-Forschung. In: Schröter, Eckhard (Hrsg.): Empirische Policy- und Verwaltungsforschung. Lokale, nationale und internationale Perspektiven. Opladen: Leske & Budrich, S. 321-344.
- Jann, Werner (2006): Verwaltungsmodernisierung auf Bundesebene. In: Jann, Werner/Bogumil, Jörg/Bouckaert, Geert/Budäus, Dietrich/Holtkamp, Lars/Kißler, Leo/Kuhlmann, Sabine/Mezger, Erika/Reichard, Christoph/Wollmann, Hellmut: Statusreport Verwaltungsreform – eine Zwischenbilanz nach zehn Jahren. 2. Auflage. Berlin: Edition Sigma, S. 100-111.
- Jans, Armin (1999): Evaluation der Integration von Sozialhilfeempfängerinnen und -empfängern in der Gemeinde Baar. Winterthur: Institut für Wirtschaft und Politik, Zürcher Hochschule Winterthur.
- Jetzinger, Daniela (2007): Ex-post-Evaluierung von Rechtsnormen, und die Mitwirkung von Consulting-Unternehmen im Prozess der retrospektiven Gesetzesfolgenabschätzung. In: Schäffer, Heinz (Hrsg.): Evaluierung der Gesetze/Gesetzesfolgenabschätzung (II). Weitere Erfahrungsberichte aus dem In- und Ausland/Staatspraxis und Wirtschaft. Wien: Manz, S. 21-46.
- Joanneum Research/Institut für Technologie- und Regionalpolitik InTeReg (2005): Bericht zur begleitenden Evaluierung des einzelstaatlichen strategischen Rahmenplans STRAT.AT. Wien: ÖROK. (www.oerok.gv.at/EU_Regionalpolitik_in_Oesterreich/strukturfonds_2007_2013/strat_at/STRAT_AT_Ex-Ante-Evaluierungsbericht_2005-10-24.pdf [13.5.2008])
- Joanneum Research/Institut für Technologie- und Regionalpolitik InTeReg (2000): Förderungsaktion eigenständige Regionalentwicklung (FER) – Außenseiter oder Mitspieler in Österreichs Regional-

- politik? Evaluation der FER-Projektförderung 1979 bis 1999. Wien/Graz: Joanneum Research/InTeReg.
- Johnson, R. Burke (1998): Toward a theoretical model of evaluation utilization. *Evaluation and Program Planning* 21(1): 93-110.
- Joint Committee on Standards for Educational Evaluation (1994): The program evaluation standards. 2nd ed. Thousand Oaks: Sage.
- Joint Committee on Standards for Educational Evaluation (Hrsg.) (2006): Handbuch der Evaluationsstandards. Die Standards des „Joint Committee on Standards for Educational Evaluation“. 3. Auflage. Wiesbaden: VS Verlag.
- Jordan, Erwin (2005): Kinder- und Jugendhilfe. Einführung in Geschichte und Handlungsfelder, Organisationsformen und gesellschaftliche Problemlagen. Grundagentexte Pädagogik. 2. Auflage. Weinheim: Juventa.
- Jörg, Leonhard/Falk, Rahel (2004): Evaluation of the Austrian Industrial Research Promotion Fund (FFF) and the Austrian Science Fund (FWF). Background report 3.1.2., FFF: Internal functioning and customer satisfaction. Wien: Technopolis/WIFO.
- Jörg, Leonhard/Jud, Thomas/Ohler, Fritz/Pointner, Wolfgang/Polt, Wolfgang/Zinöcker, Klaus (2000): Evaluierung des ITF-Schwerpunktprogrammes Technologietransfers. Wien: Technopolis.
- Jörg, Leonhard/Werner, Matthias (2006): Interimsevaluierung der Österreichischen NANO Initiative. Wien: Technopolis/NMTC.
- Jusline (o.J). Bundeshaushaltsordnung. ([www.jusline.de/Bundeshaushaltsordnung_\(BHO\).html](http://www.jusline.de/Bundeshaushaltsordnung_(BHO).html) [3.1.2008])
- Kägi, Wolfram/Siegrist, Stefan/Schäfli, Martin (2004): Einfluss des KVG auf die Verschiebung zwischen stationärer und ambulanter Medizin. Forschungsbericht 02/04. Bern: Bundesamt für Sozialversicherung.
- Kaluza, Gert/Lohaus, Arnold (2006): Psychologische Gesundheitsförderung im Kindes- und Jugendalter. Eine Sammlung empirisch evaluierter Interventionsprogramme. *Zeitschrift für Gesundheitspsychologie* 14(3): 119-134.
- Karger, Thomas/Berger, Christa/Uchtenhagen, Ambros (2000): Infektionsprophylaxe und Gesundheitsförderung im Strafvollzug. Forschungsberichte. Zürich: ISGF.
- Karlsson, Ove (1996): A critical dialogue in evaluation. How can the interaction between evaluation and politics be tackled? *Evaluation* 2(4): 405-416.
- Kaufmann, Philippe (2006): SEVAL Jahrestagung 2006. Wissensmanagement: Fehlendes Glied zwischen Politik und Evaluation? *Bulletin der Schweizerischen Evaluationsgesellschaft* 26: 2-4.
- Keller, Roger/Fabian, Carlo/Kaegi, Urs/Hornung, Rainer (2003): Wirksamkeit der Heimerziehung. Entwicklung eines Messinstrumentariums für geleistete Arbeit in Heimen. Basel: HPSABB (unveröffentlicht).
- Keller-Ebert, Cornelia (2006): Ziehen Auftraggebende Nutzen aus Evaluationen? *Sozialwissenschaften und Berufspraxis* 29(1): 101-107.
- Kempfert, Guy/Rolff, Hans-Günter (2005): Qualität und Evaluation ein Leitfaden für pädagogisches Qualitätsmanagement. 4. Auflage. Weinheim: Beltz.
- Kettiger, Daniel (2005): Gesetzesevaluation in der Schweiz: Stand – Einbettung in das politisch administrative System – Ausblick. In: Schäffer, Heinz (Hrsg.): Evaluierung der Gesetze / Gesetzesfolgenabschätzung in Österreich und im benachbarten Ausland. Wien: Manz, S. 47-73.
- KfW – Kreditanstalt für Wiederaufbau (2001): Neue Bundesländer – Einflussfaktoren für die Regionalentwicklung. Studie von IAB/DIW im Auftrag der KfW. Frankfurt: KfW.
- KfW – Kreditanstalt für Wiederaufbau (2006): Wege bereiten. Potenziale nutzen. Neunter Evaluierungsbericht (2004/2005) über die Projekte und Programme in Entwicklungsländern. Frankfurt am Main: KfW. (www.kfw-entwicklungsbank.de/DE_Home/Service/Online_Bibliothek/PDF-Dokumente_Auswertungsberichte/9_Evaluierungsbericht_d.pdf [10.3.2008])

- KGSt (1993): Das neue Steuerungsmodell. Begründungen. Konturen. Umsetzungen. KGSt-Bericht Nr. 5/1993. Köln: KGSt.
- KGSt (1994): Outputorientierte Steuerung in der Jugendhilfe. KGSt-Bericht 9/1994. Köln: KGSt.
- Khan, Jamil/Harmelink, Mirjam/Harmsen, Robert/Irrek, Wolfgang/Labance, Nicola (2006): Guidelines for the monitoring, evaluation and design of energy efficiency policies. How policy theory can guide monitoring and evaluation efforts and support the design of SMART Policies. Utrecht: Eco-fys.
- Kiendl-Wendner, Doris/Schwar, Beatrix (2007): Die Bedeutung des Vergaberecht bei Public Social Private Partnerships. Unveröffentlichtes Manuskript zum Endbericht PSPP. Graz: FH Joanneum.
- Kirkpatrick, Donald (1998): Evaluating training programs: The four levels. 2nd ed. San Francisco: Berrett-Koehler.
- Kirner, Leopold (2004): Ökonomische Auswirkungen der GAP-Reform 2003 auf Milchkuhbetriebe in Österreich. *Berichte über Landwirtschaft* 82(1): 58-80.
- Kirner, Leopold (2006): Reform der EU-Zuckermarktordnung. Ökonomische Auswirkungen auf den Rübenanbau in Österreich. Agrarpolitische Arbeitsbeihelfe 22. Wien: Bundesanstalt für Agrarwirtschaft.
- Kißler, Leo/Kersting, Norbert/Lange, Hans-Jürgen (Hrsg.) (2000): Politische Steuerung und Reform der Landesverwaltung. Baden-Baden: Nomos.
- Kissling-Näf, Ingrid/Knoepfel, Peter (1992): Politikflexibilität dank zentralstaatlichem Immobilismus? Handlungsspielräume kantonaler Vollzugspolitiken im schweizerischen politisch-administrativen System. In: Abromeit, Heidrun/Pommerehne, Werner (Hrsg.): Staatstätigkeit in der Schweiz. Bern: Haupt, S. 43-69.
- Kissling-Näf, Ingrid/Knoepfel, Peter (1994): Politikorientierter Lernprozess : Konzeptuelle Überlegungen. In: Bussmann, Werner (Hrsg.): Lernen in Verwaltungen und Policy-Netzwerken. Chur: Rüegg, S. 99-129.
- Kissling-Näf, Ingrid/Wälti, Sonja (2006): Der Vollzug öffentlicher Politiken. In: Klöti, Ulrich/Knoepfel, Peter/Kriesi, Hanspeter/Linder, Wolf/Papadopoulos, Yannis/Sciari, Pascal (Hrsg.): Handbuch der Schweizer Politik. 4. Auflage. Zürich: NZZ Verlag, S. 527-552.
- Kleinhanß, Werner/Bertelsmeier, Markus/Manegold, Dirk/Offermann, Frank/Osterburg, Bernhard/Salamon, Petra (2003): Folgenabschätzung der Legislativvorschläge zur Halbzeitbewertung der Agenda 2000. Braunschweig: FAL. (literatur.fal.de/fallitdok_extern/bitv/zi029815.pdf [3.4.2008])
- Klemisch, Herbert (2004): Umweltmanagement und ökologische Produktpolitik. München: Rainer Hampp.
- Klemisch, Herbert (Hrsg.) (1997): Öko-Audit und Partizipation. Die betriebliche Umsetzung von Umweltinformationssystemen in kleinen und mittelständischen Unternehmen. Köln: kni-Verlag.
- Klemisch, Herbert/Munier, Gerald/Pohl, Wolfgang/Scheffler, Monika/Schiller-Dickhut, Reiner (Hrsg.) (1994): Handbuch für Alternative Kommunalpolitik. Bielefeld: akp Verlag.
- Klemisch, Herbert/Potter, Philip (Hrsg.) (2006): Instrumente nachhaltigen Wirtschaftens in der Unternehmenspraxis. Düsseldorf: edition der Hans Böckler Stiftung.
- Klepp, Cornelia (2005): Qualitätssicherung durch Akkreditierung. Dargestellt am Beispiel der Fachhochschulen in Österreich, Bayern und der Schweiz. Klagenfurt (unveröffentlicht).
- Kletzan, Daniela (1999): Evaluierung der ökonomischen Effekte von Nationalparks. *Monatsberichte des Österreichischen Instituts für Wirtschaftsforschung* 72(11): 753-756.
- Klieme, Eckhard (2005): Zur Bedeutung von Evaluation für die Schulentwicklung. In: Maag Merki, Katharina/Sandmeier, Anita/Schuler, Patricia/Fend, Helmut (Hrsg.): Schule wohin? Schulentwicklung und Qualitätsmanagement im 21. Jahrhundert. Schriftenreihe zu "Bildungssystem und Humanentwicklung": Berichte aus dem Forschungsbereich Schulqualität & Schulentwicklung. Zürich: Pädagogisches Institut, Universität Zürich, S. 40-61.

- Klieme, Eckhard/Avenarius, Hermann/Blum, Werner/Döbrich, Peter/Gruber, Hans/Prenzel, Manfred/Reiss, Kristina/Riquarts, Kurt/Rost, Jürgen/Tenorth, Heinz-Elmar/Vollmer, Helmut J. (2003): Zur Entwicklung nationaler Bildungsstandards. Eine Expertise. Bonn: Bundesministerium für Bildung und Forschung.
- Klöti, Ulrich (1993): Verkehr, Energie und Umwelt. Die Infrastruktur und ihre Begrenzung. In: Schmid, Gerhard (Hrsg.): Handbuch Politisches System der Schweiz, Band 4: Politikbereiche. Bern: Haupt, S. 225-300.
- Klöti, Ulrich/Knoepfel, Peter/Kriesi, Hanspeter/Linder, Wolf/Papadopoulos, Yannis/Sciarini, Pascal (Hrsg.) (2006): Handbuch der Schweizer Politik. Zürich: Verlag NZZ.
- Klöti, Ulrich/Schneider, Gerald (1989): Die Informationsbeschaffung des Gesetzgebers: eine vergleichende Evaluationsstudie zur Bedeutung von Wirkungsanalysen im Schweizer Legiferierungsprozess. Grösch: Rüegger.
- Knapp, Gerald (Hrsg.) (2004): Soziale Arbeit und Gesellschaft. Entwicklungen und Perspektiven in Österreich. Klagenfurt: Verlag Hermagoras.
- Knoepfel, Peter/Bussmann, Werner (1997): Die öffentliche Politik als Evaluationsobjekt. In: Bussmann, Werner/Klöti, Ulrich/Knoepfel, Peter (Hrsg.): Einführung in die Politikevaluation. Basel: Helbing & Lichtenhahn, S. 58-77.
- Knoepfel, Peter/Nahrath, Stéphane (2006): Umwelt- und Raumordnungspolitik. In: Klöti, Ulrich/Knoepfel, Peter/Kriesi, Hanspeter/Linder, Wolf/Papadopoulos, Yannis/Sciarini, Pascal (Hrsg.): Handbuch der Schweizer Politik. 4. Auflage. Zürich: NZZ Verlag, S. 737-764.
- Knoepfel, Peter/Varone, Frédéric/Bussmann, Werner/Mader, Luzius (1997): Evaluationsgegenstände und Evaluationskriterien. In: Bussmann, Werner/Klöti, Ulrich/Knoepfel, Peter (Hrsg.): Einführung in die Politikevaluation. Basel: Helbing & Lichtenhahn, S. 78-118
- Kobi, Sylvie (2004): Evaluation individuelles Coaching in der Berufsbildung. Zürich: Fachhochschule Zürich.
- Koch, Benedikt/Rieder, Peter (2002): MAAP 2002 1. Teil: Getreidemarktanalyse (Auswirkungen der neuen Getreidemarktordnung unter besonderer Berücksichtigung verschiedener Zollsysteme), 2. Teil: Milchmarktanalyse (Auswirkungen staatlicher Massnahmen auf die Wettbewerbsfähigkeit der Milchwirtschaft), 3. Teil: Fleischmarktanalyse (Auswirkungen von staatlichen Massnahmen zur Preisstützung und Preisstabilisierung), 4. Teil: Eiermarktanalyse (Staatliche Marktinterventionen unter besonderer Berücksichtigung vertikaler Vertragsproduktion), Zürich: Bundesamt für Landwirtschaft.
- Koch, Peter/Haering, Barbara/Peters, Matthias/Seiler, Benno/Gsponer, Gabriele/Baur, Martin (2002): Evaluation der Verwaltungsreform wif! Schlussbericht. Zürich: Econcept.
- Kodama, Fumio (1995): Emerging patterns of innovation. Sources of Japan's technological edge. Boston: Harvard Business School Press.
- Kohler, Britta/Schrader, Friedrich-Wilhelm (Hrsg.) (2004): Ergebnisrückmeldung und Rezeption: Von der externen Evaluation zur Entwicklung von Schule und Unterricht. Themenheft. *Empirische Pädagogik* 18(1).
- Koller, Martin (2005): Wie erfolgreich sind Subventionen? – Investitionsförderung auf dem Prüfstand. IAB-Gutachten Nr. 1/2004. Nürnberg: Institut für Arbeitsmarkt- und Berufsforschung.
- Koller, Martin/Schwengler, Barbara/Zarth, Michael (2001): Zielerreichungsanalyse bei den Fördergebieten der Gemeinschaftsaufgabe "Verbesserung der regionalen Wirtschaftsstruktur". Beiträge zur Arbeitsmarkt und Berufsforschung Nr. 243. Nürnberg: Institut für Arbeitsmarkt- und Berufsforschung.
- König, Joachim (2000): Einführung in die Selbstevaluation. Ein Leitfaden zur Bewertung der Praxis sozialer Arbeit. Freiburg: Lambertus.

- Konle-Seidl, Regina (2005): Lessons learned. Internationale Evaluierungsergebnisse zu Wirkungen aktiver und aktivierender Arbeitsmarktpolitik. IAB Forschungsbericht 9. Nürnberg: Institut für Arbeitsmarkt- und Berufsforschung (IAB).
- Konsortium Bildungsberichterstattung (Hrsg.) (2006): Bildung in Deutschland. Ein indikatorengestützter Bericht mit einer Analyse zu Bildung und Migration. Bielefeld: W. Bertelsmann.
- Konzendorf, Götz (1998): Verwaltungsmodernisierung in den Ländern. Überblick und Einblicke. Speyerer Forschungsberichte 187. Speyer: Deutsche Hochschule für Verwaltungswissenschaften.
- Köppl, Franz (1984): Experimentelle arbeitsmarktpolitische Maßnahmen. In: Bundesministerium für Soziale Verwaltung (Hrsg.): Jugendliche auf dem österreichischen Arbeitsmarkt. Forschungsberichte aus Sozial- und Arbeitsmarktpolitik Nr. 12. Wien: Eigenverlag, S. 143-170.
- Kreß, Stefan/Maas, Stephanie/Wasem, Jürgen (2006): Effektivitäts-, Effizienz- und Qualitätsreserven im deutschen Gesundheitssystem. Expertise für die Hans-Böckler-Stiftung. Duisburg: Lehrstuhl für Medizinmanagement, Universität Duisburg-Essen. (www.boeckler.de/pdf_fof/S-2006-840-4-1 [10.5.2007])
- Kriz, Jürgen (2003): Gutachten über den Begriff der Wissenschaftlichkeit in der Psychotherapie. Zürich: Schweizer Berufsverband für Angewandte Psychologie. (www.sbp.ch/pdf/2003/GutachtenKriz2003.pdf [10.4.2008])
- Kromrey, Helmut (2001): Evaluation – ein vielschichtiges Konzept. Begriff und Methodik von Evaluierung und Evaluationsforschung. Empfehlungen für die Praxis. *Sozialwissenschaften und Berufspraxis* 24(2): 105-131. (www.hkromrey.de/SuB_2001-2-kromrey.pdf [29.5.2008])
- Kromrey, Helmut (2003): Evaluation in Wissenschaft und Gesellschaft. *Zeitschrift für Evaluation* 2003/1: 93-116.
- Kromrey, Helmut (2006a): Empirische Sozialforschung. 11. Auflage. Stuttgart: Lucius und Lucius.
- Kromrey, Helmut (2006b): Qualität und Evaluation im System Hochschule. In: Stockmann, Reinhard (Hrsg.): Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder. 3. Auflage. Münster: Waxmann, S. 234-259.
- Kübler, Daniel/Sager, Fritz/Schwab, Brigitte (2005): Governing without government: Metropolitan governance in Switzerland. In: Heinelt, Hubert/Kübler Daniel (eds): Metropolitan governance in the 21st century. London: Routledge, pp. 169-187.
- Kübler, Daniel/Schenkel, Walter/Leresche, Jean-Philippe (2003): Bright lights, big cities? Metropolization, intergovernmental relations and the new federal urban policy in Switzerland, *Swiss Political Science Review* 9(1): 35-60.
- Kuffner, Astrid (2000): Evaluation von Nachhaltigkeitsaspekten – Nachhaltige Evaluation? Eine vergleichende Analyse von deutschsprachigen Evaluationsberichten, die sich mit nachhaltiger Entwicklung beschäftigen. Wien: Wirtschaftsuniversität Wien.
- Kuhlmann, Sabine (2004): Evaluation lokaler Verwaltungspolitik: Umsetzung und Wirksamkeit des Neuen Steuerungsmodells in den deutschen Kommunen. *Politische Vierteljahresschrift* 45(3): 370-394.
- Kuhlmann, Sabine (2005): Selbstevaluation durch Leistungsvergleiche in deutschen Kommunen. *Zeitschrift für Evaluation* 1/2005: 7-28.
- Kuhlmann, Sabine (2008): Politik- und Verwaltungsreform in Kontinentaleuropa: Subnationaler Institutionenwandel im deutsch-französischen Vergleich. Habilitationsschrift an der Universität Potsdam. Baden-Baden: Nomos.
- Kuhlmann, Sabine/Wollmann, Hellmut (2006): Transaktionskosten von Verwaltungsreformen – ein „missing link“ der Evaluationsforschung. In: Jann, Werner/Röber, Manfred/Wollmann, Hellmut (Hrsg.): Public Management. Grundlagen, Wirkungen, Kritik. Wiesbaden: VS Verlag, S. 371-390.
- Kuhlmann, Stefan (1998a): Politikmoderation. Evaluationsverfahren in der Forschungs- und Technologiepolitik. Baden-Baden: Nomos.

- Kuhlmann, Stefan (1998b): Moderation of policy-making? Science and technology policy evaluation beyond impact measurement: The case of Germany. *Evaluation* 4(2): 130-148.
- Kuhlmann, Stefan (2001): Governance of innovation policy in Europe – three scenarios. *Research Policy* 30(6): 953-976.
- Kuhlmann, Stefan (2003a): Evaluation in der Forschungs- und Innovationspolitik. In: Stockmann, Reinhard (Hrsg.): *Evaluationsforschung*. 2. Auflage. Opladen: Leske+Budrich, S. 287-307.
- Kuhlmann, Stefan (2003b): Evaluation of research and innovation policies: A discussion of trends with examples from Germany. *International Journal of Technology Management* 26(2/3/4): 131-149.
- Kuhlmann, Stefan (2003c): Leistungsmessung oder Lernmedium? Evaluation in der Forschungs- und Innovationspolitik. *Technikfolgenabschätzung – Theorie und Praxis* 12(1): 11-19.
- Kuhlmann, Stefan/Edler, Jakob (2003): Changing governance in European research and technology policy. In: Edler, Jakob/Kuhlmann, Stefan/Behrens, Maria (eds.): *Changing governance of research and technology policy – the European Research Area*. Cheltenham: Edward Elgar, pp. 3-32.
- Kuhlmann, Stefan/Heinze, Thomas (2003): Informationen zur Forschungsevaluation in Deutschland – Erzeuger und Bedarf. Gutachten für die Geschäftsstelle der Deutschen Forschungsgemeinschaft. *Fraunhofer ISI Discussion Papers Innovation System and Policy Analysis*, 3/2003.
- Kuhlmann, Stefan/Heinze, Thomas (2004a): Evaluation von Forschungsleistungen in Deutschland: Erzeuger und Bedarf; Teil I: Konzeptionelle Grundlagen. *Wissenschaftsrecht* 37: 53-69.
- Kuhlmann, Stefan/Heinze, Thomas (2004b): Evaluation von Forschungsleistungen in Deutschland: Erzeuger und Bedarf; Teil II: Produktion und Verwendung evaluativer Information sowie Möglichkeiten ihrer künftigen Organisation. *Wissenschaftsrecht* 37: 125-149.
- Kuhlmann, Stefan/Holland, Doris (1995a): Evaluation von Technologiepolitik in Deutschland – Konzepte, Anwendung, Perspektiven. Heidelberg: Physica-Verlag.
- Kuhlmann, Stefan/Holland, Doris (1995b): Erfolgsfaktoren der wirtschaftsnahen Forschung. Heidelberg: Physica-Verlag.
- Kukartz, Udo/Dresing, Thorsten/Rädiker, Stefan/Stefer, Claus (2007): *Qualitative Evaluation. Der Einstieg in die Praxis*. Wiesbaden: VS Verlag.
- Kultusministerkonferenz Deutschland (2004): Stellungnahme der KMK zu den Ergebnissen von PISA 2003 (6. 12. 2004). Berlin: KMK. (www.kmk.org/aktuell/pm041206.htm [21. 2. 2008])
- Kultusministerkonferenz Deutschland (2006): Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring. München/Berlin: Luchterhand/KMK.
- Kultusministerkonferenz Deutschland (Hrsg.) (2003): *Bildungsbericht für Deutschland*. Opladen: Leske + Budrich.
- Kundert, Sonja/Sommerfeld, Peter (2001): Partizipation im Quartier. Das Beispiel „Quartierkommission Länggasse“ in Bern. Schlussbericht. Olten: Fachhochschule Solothurn Nordwestschweiz.
- Ladner, Andreas/Cemerin, Michael/Ritz, Adrian/Sager, Fritz (2007): *Evaluation der Parlaments- und Verwaltungsreform Neue Stadtverwaltung Bern NSB*. Bern: KPM-Verlag Bern.
- Lalive d'Epinay, Rafael/Zehnder, Tanya/Zweimüller, Josef (2006): Makroökonomische Evaluation der Aktiven Arbeitsmarktpolitik der Schweiz. Beiträge zur Arbeitsmarktpolitik Nr. 19. Bern: Seco.
- Lammers, Konrad/Niebuhr, Annekatrin (2002): Erfolgskontrolle in der deutschen Regionalpolitik: Überblick und Bewertung. HWWA-Report 214. Hamburg: Hamburgisches Welt-Wirtschafts-Archiv.
- Landert, Charles (1996): Externe und interne Evaluation – Schnittstellen und Übergänge. In: Heiner, Maja (Hrsg.): *Qualitätsentwicklung durch Evaluation*. Freiburg: Lambertus, S. 68-84.
- Landert, Charles (2002): *Klassencockpit: Evaluation im Auftrag der Projektleitung Klassencockpit*. Zürich: Landert Farago Davatz & Partner.
- Landert, Charles (2006): Schweizer Erfahrungen aus der Evaluation von Jugendprogrammen. In: Projekt eXe (Hrsg.): *Aushandlungsprozesse zur Steuerung externer Evaluationen. Dokumentation der Fachtagung München, 1.-3.12.2004*. München: Deutsches Jugendinstitut, S. 99-109. (www.dji.de/bibs/333_7045_eXe_Aushandlungsprozesse.pdf [05.03.2007])

- Landesinstitut für Schule und Weiterbildung (Hrsg.) (1996): Evaluation und Schulentwicklung. Ansätze, Beispiele und Perspektiven aus der Fortbildungsmaßnahme Schulentwicklung und Schulaufsicht. 2. Auflage. Soest: Verlag für Schule und Weiterbildung.
- Landwehr, Norbert/Steiner, Peter (2003): Q2E – Qualität durch Evaluation und Entwicklung. Konzepte, Verfahren und Instrumente zum Aufbau eines Qualitätsmanagements an Schulen - Schubert mit 5 Broschüren. Bern: h.e.p. Verlag.
- Landwehr, Norbert/Steiner, Peter (2007): Q2E – Qualität durch Evaluation und Entwicklung. Konzepte, Verfahren und Instrumente zum Aufbau eines Qualitätsmanagements an Schulen - Schubert mit 6 Broschüren. Bern: h.e.p. Verlag.
- Lang, Jochen (2003): Symbolische Implementation als Flexibilitätsreserve. Die Umsetzung europäischer regionaler Strukturpolitik in Deutschland, Irland und Schweden. Dissertation. FernUniversität Hagen.
- Larédo, Philippe/Mustar, Philippe (2001): Research and innovation policies in the new global economy. An international comparative analysis. Cheltenham: Edward Elgar.
- Laskowski, Wolfgang/Loidl-Keil, Rainer/Höhn, Martina (2006): Bedarfs- und Akzeptanzanalyse zum FH-Bakkalaureatsstudiengang "Sozialarbeit mit Ausbildungsschwerpunkt Sozialmanagement" und eines FH-Magisterstudienganges mit den Vertiefungsrichtungen "Sozialarbeit mit Erwachsenen und Älteren", "Sozialarbeitsforschung" und "Sozialmanagement". Bericht im Auftrag der FH Joanneum an den Fachhochschulrat Österreich. Graz (unveröffentlicht).
- Lassnigg, Lorenz/Leitner, Andrea/Wroblewski, Angela/Steiner, Mario/Steiner, Peter/Mayer, Kurt/Schmid, Günther/Schömann, Klaus (2000): Evaluationsschema für Maßnahmen der aktiven Arbeitsmarktpolitik in Wien. Studie im Auftrag von AMS Wien und WAFF. Wien: Institut für Höhere Studien (IHS).
- Lauber, Volkmar (1997): A latecomer who became pioneer. In: Andersen, Mikael Skou/Liefferink, Duncan (eds): European environmental policy: the pioneers. Manchester: Manchester University Press, pp. 81-117.
- Läubli Loud, Marlène M. (2004): Setting standards and providing guidelines. The means toward what end? *Evaluation* 10(2): 237-245.
- Lauterbach, Karl/Schrappe, Matthias (2004): Gesundheitsökonomie, Qualitätsmanagement und Evidence Based Medicine. 2. Auflage. Stuttgart: Schattauer.
- Lechner, Ferdinand/Reiter, Walter/Fehr-Duda, Helga (1988): Evaluation der Aktion 8000. BMAS (Hrsg.): Forschungsberichte aus Sozial- und Arbeitsmarktpolitik Nr. 23. Wien: Eigenverlag.
- Lechner, Michael (2006): Matching Estimating of Dynamic Treatment Models: Some Practical Issues, Universität St. Gallen, Department of Economics, Discussion Paper 3.
- Lechner, Michael/Frölich, Markus/Steiger, Heidi (2004): Mikroökonomische Evaluation aktiver Arbeitsmarktpolitik. Synthesis NFP 43. Bern: Schweizerischer Nationalfonds.
- Lechner, Michael/Miquel, Ruth/Werner, Stephan/Wiehler, Stephan (2007): Mikroökonomische Evaluation der Instrumente der Aktiven Arbeitsmarktpolitik in Österreich, Studie im Auftrag des BMWA, Projektbericht des Schweizerischen Instituts für Aussenwirtschaft und Angewandte Wirtschaftsforschung (SIAW). St. Gallen: Universität St. Gallen.
- Ledermann, Simone (2004): Wozu denn wissenschaftlich? Untersuchung zur Verwendung von Evaluationen in der schweizerischen Entwicklungszusammenarbeit (Lizentiatsarbeit). Bern: Universität, Institut für Politikwissenschaft.
- Ledermann, Simone/Hammer, Stephan/Sager, Fritz/Dubas, Daniel/Rüefli, Christian/Schmidt, Nicolas/Trageser, Judith/Vettori, Anna/ Zeyen Bernasconi, Petra (2006): Evaluation der Strategie Migration und Gesundheit 2002-2006. Bern: Büro Vatter.
- Lee, Barbara (2000): Theories of evaluation. In: Stockmann, Reinhard (Hrsg.): Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder. Opladen: Leske + Budrich, S. 127-164.

- Leeuw, Frans L. (2006a): Evaluation in Europe. In: Stockmann, Reinhard (Hrsg.): Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder. 3. Auflage. Münster: Waxmann, S. 64-84.
- Leeuw, Frans L. (2006b): Managing evaluations in the Netherlands and types of knowledge. In: Stame, Nicoletta/Rist, Ray C. (eds): From studies to streams. New Brunswick: Transaction, pp. 81-97.
- Leeuw, Frans L./Furubo, Jan-Eric (2008): Evaluation systems. What are they and why study them? *Evaluation* 14(2): 157-169.
- Leeuw, Frans L./Kerseboom, Jeroen S./Elte, Ronald (eds) (2007): Turven, tellen, toetsen. Over toezicht, handhaving en evaluatie en hun maatschappelijke betekenis in Nederland. Den Haag: Boom Juridische Uitgevers.
- Legewie, Heiner (2000): Psychotherapie unter Rationalisierungsdruck. *Psychomed* 2000/2: 45-50.
- Lehmann, Rainer H./Lenkeit, Jenny (2008): ELEMENT: Erhebung zum Lese- und Mathematikverständnis, Entwicklungen in den Jahrgangsstufen 4 bis 6 in Berlin - Abschlussbericht über die Untersuchungen 2003, 2004 und 2005 an Berliner Grundschulen und grundständigen Gymnasien. Berlin: Humboldt Universität zu Berlin.
- Lehmann, Rainer H./Peek, Rainer/Gänsfuß, Rüdiger (1997): Aspekte der Lernausgangslage von Schülerinnen und Schülern der fünften Klassen an Hamburger Schulen. Hamburg: Behörde für Schule, Jugend und Berufsbildung, Amt für Schule.
- Lehmann, Rainer H./Peek, Rainer/Gänsfuß, Rüdiger/Lutkat, Sabine/Mücke, Stephan/Barth, Ingola (2001): QuaSUM: Qualitätsuntersuchung an Schulen zum Unterricht in Mathematik. Brandenburg: Ministerium für Bildung, Jugend und Sport des Landes Brandenburg.
- Lehmbruch, Gerhard (1967): Proporzdemokratie: Politisches System und politische Kultur in der Schweiz und Österreich: Tübingen: Mohr.
- Leitner, Andrea/Wroblewski, Angela/Hofer, Helmut/Prenner, Peter/Mahringer, Helmut/Schuh, Andreas (2003): Aktive Arbeitsmarktpolitik im Brennpunkt IX: Arbeitsmarktpolitische Wirkungen des TBP Wien. Arbeitsmarktservice Österreich AMS Report 36. Wien: Verlag Hofstätter.
- Leitner, Karl-Heinz (2007): Intellectual Capital Reporting and Evaluation in Austrian Universities: Relationships and Complementarities. In: Platform Research and Technology Policy Evaluation, Austrian Council for Research and Technology Development (eds.): Evaluation of Austrian Research and Technology Policies. Vienna: Platform Research and Technology Policy Evaluation, pp. 97-105.
- Lemne, Marja/Sohlman, Åsa (2004): Governance, democracy and evaluation – From not evaluating at all to learning from ex post evaluations – the strenuous Swedish way towards accountability. Paper presented at the bi-annual conference of the European Evaluation Society, Berlin, September 29-October 2, 2004.
- Leschnik, Melanie (2007): Österreichische Evaluationen in den Bereichen Energie, Verkehr und Umwelt – Eine vergleichende Analyse ausgewählter Evaluationsprojekte. Wien: Wirtschaftsuniversität Wien.
- Leutenegger, Franziska (2006): Lärm und Raumluft in Bürogebäuden. Diplomarbeit unterstützt durch das Staatssekretariat für Wirtschaft Seco. Zürich: ETH Zürich.
- Leutner, Detlev (1999): Standards empirischer Forschung für Evaluation im Bildungsbereich. In: Thonhauser, Josef/Patry, Jean-Luc (Hrsg.): Evaluation im Bildungsbereich. Wissenschaft und Praxis im Dialog. Innsbruck: Studienverlag, S. 121-132.
- Lincoln, Yvonna S. (1990): Program review, accreditation processes, and outcomes assessment: Pressures on institutions of higher education. *Evaluation Practice* 11(1): 13-23.
- Linder, Wolf (1987): Politische Entscheidung und Gesetzesvollzug in der Schweiz. Bern: Haupt.
- Linder, Wolf (2005): Schweizerische Demokratie. 2. Auflage. Bern: Haupt.
- Loidl-Keil, Rainer (2002): Entwicklungsstand und Perspektiven des Sozialen Unternehmertums in Österreich. Eine Pilotstudie. Wien: Julius Raab Stiftung.

- Loidl-Keil, Rainer (2004): Plädoyer für erweiterte Konzepte für Evaluation im Gesundheitsbereich. In: Lauterbach, Karl/Schrappe, Matthias (Hrsg.): Gesundheitsökonomie, Qualitätsmanagement und Evidence Based Medicine. Eine systematische Einführung. 2. Auflage. Stuttgart: Schattauer, S. 15-24.
- Loidl-Keil, Rainer (2006): Dokumentieren, Kontrollieren, Sichern, Verbessern. Evaluationen in der Jugendarbeit – Zur Bewahrung von Komplexität. *Leitlinien der integrativen und interdisziplinären Jugendarbeit. Zeitschrift der pro mente Jugend Kärnten* 2: 11-12.
- Loidl-Keil, Rainer/Laskowski, Wolfgang (Hrsg.) (2006): Evaluation im Gesundheitswesen. Konzepte, Beispiele, Erfahrungen. München: Hampp.
- Loidl-Keil, Rainer/Posch, Klaus (2006): Über Kugeln und andere Wirklichkeiten. Begründung von Evaluation in einer lebensweltorientierten Sozialarbeit zwischen Anerkennung und Diskreditierung. In: Heinrich, Martin/Greiner, Ulrike (Hrsg.): Schauen, was rauskommt. Kompetenzförderung, Evaluation und Systemsteuerung im Bildungswesen. Wien: Lit-Verlag, S. 471-482.
- Loidl-Keil, Rainer/Tschemer, Marlies/Viechtbaur, Karin (2004): Evaluation Pflegeelternanstellung. Endbericht an das Land Oberösterreich, Abteilung Jugendwohlfahrt. Linz: Johannes Kepler Universität Linz, Institut für Soziologie.
- Loidl-Keil, Rainer/Viechtbaur, Karin (2003): Evaluation und Qualitätsmanagement als Rechtsauftrag? – Ein Vergleich im Deutschsprachigen Raum. In: Loidl-Keil, Rainer/Viechtbaur, Karin (Hrsg.): Evaluation in der außerschulischen Kinder- und Jugendarbeit in Österreich – Berichte aus der Praxis. Linz: Universitätsverlag Rudolf Trauner, S.17-38.
- Lucyshyn, Josef (2006): Implementation von Bildungsstandards in Österreich. Arbeitsbericht des Bundesinstituts für Bildungsforschung, Innovation und Entwicklung des Bildungswesens (BIFIE). Salzburg: BIFIE.
- Lüders, Christian/Haubrich, Karin (2006): Wirkungsevaluation in der Kinder- und Jugendhilfe: Über hohe Erwartungen, fachliche Erfordernisse und konzeptionelle Antworten. In: Projekt eXe (Hrsg.): Wirkungsevaluation in der Kinder- und Jugendhilfe – Einblicke in die Evaluationspraxis. München: Deutsches Jugendinstitut, S. 5-20.
- Luks, Fred (2002): Nachhaltigkeit. Hamburg: EVA.
- Lüthi, Jean Christophe/McLellan, William/Flanders, Dana/Pitts, Stephen/Burnand, Bernard (2002): Quality of health care surveillance systems: review and implementation in the Swiss setting. *Swiss Medical Weekly* 132: 461-469.
- Lutz, Hedwig/Mahringer, Helmut/Pöschl, Andrea/Huber, Peter/Novotny, Birgit/Schöberl, Marianne/Smeral, Kristin/Lassnigg, Lorenz/Leitner, Andrea/Scheibelhofer, Elisabeth/Steiner, Mario/Steiner, Peter/Wroblewski, Angela/Lechner, Ferdinand/Nemecek, Nicole/Riesenfelder, Andreas (2005a): Evaluierung europäischer Sozialfonds 2000-2006 Ziel 3 Österreich, Vorläufiger Endbericht, Studie im Auftrag des BMWA. Wien: WIFO.
- Lutz, Hedwig/Mahringer, Helmut/Pöschl, Andrea/Huber, Peter/Novotny, Birgit/Schöberl, Marianne/Smeral, Kristin/Lassnigg, Lorenz/Leitner, Andrea/Scheibelhofer, Elisabeth/Steiner, Mario/Steiner, Peter/Wroblewski, Angela/Lechner, Ferdinand/Nemecek, Nicole/Riesenfelder, Andreas (2005b): Evaluierung der österreichischen Arbeitsmarktförderung 2000-2003, Vorläufiger Endbericht, Studie im Auftrag des BMWA. Wien: WIFO.
- Lutz, Wolfgang/Grawe, Klaus (2001): Was ist „Evidenz“ in einer evidence based psychotherapy? *Integrative Therapie* 2: 11-28.
- Lutz, Wolfgang/Grawe, Klaus (2005): Psychotherapieforschung: Grundlagen, Konzepte und neue Trends. In: Strauss, Bernhard/Caspar, Franz/Hohagen, Fritz (Hrsg.): Lehrbuch der Psychotherapie. Göttingen: Hogrefe, S. 727-768.
- Lutz, Wolfgang/Tholen, Sven/Kosfelder, Joachim (2004): Modelle und Konzepte zur Entwicklung von Regeln zur empirischen Unterstützung und Evaluation des Psychotherapiebedarfs und der Therapieverlängerung am individuellen Patienten. *Verhaltenstherapie und Verhaltensmedizin* 25: 503-527.

- Mader, Luzius (2003): Zum aktuellen Stand der Gesetzesfolgenabschätzung in der Schweiz. In: Karpen, Ulrich/Hof, Hagen (Hrsg.): Wirkungsforschung zum Recht IV – Möglichkeiten einer Institutionalisierung der Wirkungskontrolle von Gesetzen. Baden-Baden: Nomos, S. 96-110.
- Mader, Luzius (2005): Artikel 170 der Bundesverfassung: Was wurde erreicht, was ist noch zu tun? *LeGes Gesetzgebung & Evaluation* 16(1): 29-37.
- Mader, Luzius (2006): Législation et légisprudence. In: Jean-Loup Chappelet (éd.): Contributions à l'action publique – Beiträge zum öffentlichen Handeln. Lausanne: Haupt, p. 193-204.
- Mader, Luzius (2007): Zwischen Anspruch und Wirklichkeit – Die Evaluation staatlicher Massnahmen in der Schweiz. *Neue Zürcher Zeitung* 67 (21.3.2007): 19.
- Mäding, Heinrich (1995): Erfolgskontrolle. In: Akademie für Raumforschung und Landesplanung (Hrsg.): Handwörterbuch der Raumordnung. Hannover: Akademie für Raumforschung und Landesplanung, S. 226-227.
- Maffli, Etienne/Wacker, Hans Rudolf/Mathey Marie-Claire (1995): 7-Jahres-Katamnese von stationär behandelten Alkoholabhängigen in der deutschen Schweiz. Forschungsbericht Schweizerische Fachstelle für Alkoholprobleme SFA. Lausanne: SFA.
- Magnin, Chantal (2005): Beratung und Kontrolle. Widersprüche in der staatlichen Bearbeitung von Arbeitslosigkeit. Zürich: Seismo.
- Mahringer, Helmut (2004): Data Warehouse (DWH) Monitoring in the Public Employment Service, Discussion paper prepared by the host country independent Expert, June 2004, Vienna. (peerreview.almp.org/pdf/austria04/dissAUS04.pdf [6.6.2007])
- Mair, Peter (2004): The Europeanization dimension. *Journal of European Public Policy* 11(2): 337-348.
- Mann, Stefan/Mack, Gabriele (2004): Wirkungsanalyse der Allgemeinen Direktzahlungen, FAT-Schriftenreihe Nr. 64. Ettenhausen: Forschungsanstalt für Agrarwirtschaft und Landtechnik Tänikon.
- Maradan, Olivier/Mangold, Max (2005): Bildungsstandards in der Schweiz. Das Projekt HarmoS. *ph akzente* 2005(2): 3-7.
- March, James G./Olsen, Johan P. (1989): Rediscovering institutions: The organizational basis of politics. New York: Free Press.
- Margraf, Jürgen (2007): Kosten und Nutzen der Psychotherapie. Eine aktuelle Literaturübersicht. Bern: Huber.
- Mark, Melvin M./Henry, Gary T. (2004): The mechanisms and outcomes of evaluation influence. *Evaluation* 10(1): 35-57.
- Marti, Michael/Osterwald, Stephan/Müller, André (2003): Prekäre Arbeitsverhältnisse in der Schweiz. Beiträge zur Arbeitsmarktpolitik Nr. 9. Bern: Seco.
- Martinovits-Wiesendanger, Alex/Ganzaroli, Dennis (2000): Wirkungsanalyse der Gastgewerbe-/Winwordkurse und Einsatzprogramme. Beiträge zur Arbeitsmarktpolitik Nr. 5. Bern: Seco.
- Martinuzzi, André (2004): Sustainable Development Evaluations in Europe – The key findings of EASY-ECO - Evaluation of Sustainability European Conferences. *Journal of Environmental Assessment Policy and Management* 3(4): 1-32.
- Martinuzzi, André (2006): Beratungsprogramme für Nachhaltiges Wirtschaften – Institutionelle Innovationen an der Schnittstelle zwischen Umweltpolitik und Umweltmanagement. In: Pfriem, Reinhard/Antes, Ralf/Fichter, Klaus/Müller, Martin/Paech, Niko/Seuring, Stefan/Siebenhüner, Bernd (Hrsg.): Innovationen für Nachhaltige Entwicklung, Wiesbaden: Deutscher Universitätsverlag, S. 171-200.
- Martinuzzi, André/Meyer, Wolfgang/Kopp, Ursula (2007): Evaluation zwischen Leistungsmessung und Lernschleife. *Ökologisches Wirtschaften* 2/2007: 43-46.
- Martinuzzi, André/Steurer, Reinhard (2003): The Austrian Strategy for Sustainable Development. Process Review and Policy Analysis. *European Environment* 13(5): 269-287.

- Martinuzzi, André/Steurer, Reinhard (2005): Die Umsetzung der Österreichischen Nachhaltigkeitsstrategie – Erfahrungen und Einsichten von Beteiligten. Schriftenreihe des Forschungsschwerpunkts Nachhaltigkeit und Umweltmanagement, Diskussionspapier 02/2005. Wien: Wirtschaftsuniversität Wien.
- Mastronardi, Philippe (2002): Kommentar zu Artikel 170. In: Ehrenzeller, Bernhard/Mastronardi, Philippe/Schweizer, Rainer J./Vallender, Klaus A. (Hrsg.): Die schweizerische Bundesverfassung. Zürich: Schulthess, S. 1677–1682.
- Mathiasen, David (1993): The separation of powers and political choice: Budgeting, Auditing, and Evaluation in the United States. In: Gray, Andrew/Jenkins, Bill/Segsworth, Bob (eds): Budgeting, auditing, and evaluation. Functions and integration in seven governments. New Brunswick: Transaction, pp. 19-39.
- Matschek, Markus (2004): Verwaltungsreform in Kärntens Gemeinden. In: Kärntner Verwaltungsakademie (Hrsg.): Bildungsprotokolle 9: New Public Management. Klagenfurt: Kärntner Verwaltungsakademie, S. 135-235.
- Mattes, Christoph/Fabian, Carlo (2007): Evaluation des Aktionspakets oberes Kleinbasel. Basel: FHNW (unveröffentlicht).
- Matzdorf, Bettina/Becker, Nadine/Reutter, Michaela/Tiemann, Silja (2005): Aktualisierung der Halbwertsbewertung des Plans zur Entwicklung des ländlichen Raums gemäß VO (EG) Nr. 1257/99 des Landes Brandenburg. Braunschweig: Selbstverlag.
- Mauch, Corine/Balthasar, Andreas (2005): Machbarkeitsstudie „Evaluation der bisherigen Umweltpolitik“ – Schlussbericht. Umwelt-Materialien Nr. 202. Bern: Bundesamt für Umwelt, Wald und Landschaft.
- Mauch, Corine/Haefeli, Ueli/Kanaan, Sami/Zellweger, Erich (2005): Evaluation des Ressourcen- und Umweltmanagements in der Bundesverwaltung (RUMBA). Bern: Parlamentsdienste.
- Mauch, Corine/Häfeli, Ueli/Landis, Flurina/Matti, Daniel/Rüefli, Christian/Willisegger, Jonas (2007): Evaluation der Umweltaußenpolitik der Schweiz. Schlussbericht zuhanden des Bundesamtes für Umwelt. Luzern/Bern: Interface/Büro Vatter.
- Mauch, Corine/Rieder, Stefan/Landis, Flurina/Frey, Rahel (2006): Verwesentlichung und Vereinfachung der Vollzugshilfen des BAFU. Luzern: Interface.
- Mayne, John/Divorski, Stan/Lemaire, Donald (1999): Locating evaluation: Anchoring evaluation in the executive or the legislature, or both or elsewhere? In: Boyle, Richard/Lemaire, Donald (eds): Building effective evaluation capacity. Lessons from practice. New Brunswick: Transaction, pp. 23-52.
- Mayntz, Renate/Scharpf, Fritz W. (1973a): Kriterien, Voraussetzung und Einschränkungen aktiver Politik. In: Mayntz, Renate/Scharpf, Fritz W. (Hrsg.): Planungsorganisation. Die Diskussion um die Reform von Regierung und Verwaltung des Bundes. München: Piper, S. 115-145.
- Mayntz, Renate/Scharpf, Fritz W. (Hrsg.) (1973b): Planungsorganisation. Die Diskussion um die Reform von Regierung und Verwaltung des Bundes. München: Piper.
- Mayring, Philipp (2002): Einführung in die Qualitative Sozialforschung. Eine Anleitung zu qualitativem Denken. 5. Auflage. Weinheim: Beltz-UTB.
- Mayring, Philipp/Huber, Günther L./Gürtler, Leo/Kiegelmann, Mechthild (eds) (2007): Mixed methodology in psychological research. Rotterdam: Sense Publishers.
- Mehl, Peter (2003): Gesetz der Bundesregierung: Entwurf eines Haushaltsbegleitgesetzes 2004 (Haushaltsbegleitgesetz 2004-HBeglG 2004), Stellungnahme zu Artikel 13, Änderung des Zweiten Gesetzes über die Krankenversicherung der Landwirte aus Anlass der öffentlichen Anhörung des Haushaltsausschusses des Deutschen Bundestages, Berlin, 8. Oktober 2003. (www.fal.de/cdn_045/nn_791232/SharedDocs/10_LR/DE/Publikationen/Sonstige/download_hbglG_10_2003_de,templateId=raw,property=publicationFile.pdf/download_hbglG_10_2003_de.pdf. [9.5.2007])

- Mehl, Peter (2006): Finanzielle Folgen einer Kapitalisierung neuer Dauerrenten in der landwirtschaftlichen Unfallversicherung. Braunschweig: Bundesforschungsanstalt für Landwirtschaft. (literatur.fal.de/fallitdok_extern/bitv/zi041605.pdf [9.5.2007])
- Mehl, Peter/Plankl, Reiner (2002): Regionale Agrarpolitik in Deutschland - Handlungsspielräume in einem verflochtenen Politikfeld. In: Europäisches Zentrum für Förderalismus-Forschung (Hrsg.): Jahrbuch des Förderalismus 2002. Förderalismus, Subsidiarität und Regionen in Europa. Baden-Baden: Nomos, S. 199-210.
- Merton, Robert K. (1936): The unanticipated consequences of purposive social action. *American Sociological Review* 1(6): 894-904.
- Merz, Annemarie/Walser, Fredi (1996): Soliwork – Beschäftigungsprogramm für ausgesteuerte Langzeitarbeitslose. Zürich: Sozialdepartement der Stadt Zürich.
- Merz, Annemarie/Walser, Fredi (1997): Soliwork - soziale Integration und Sozialhilfeablösung durch Erwerbstätigkeit: Wirkungsanalyse des Stadtzürcher Beschäftigungsprogramms für Langzeitarbeitslose. Zürich: Edition Sozialstatistik.
- Meusburger, Martina/Paulschin, Herbert (2007): Die Qualitätsdebatte in der Sozialarbeit. Die Entwicklung der Qualitätsdebatte auf Basis des Fachmagazins "Sozialarbeit in Österreich". In: EQUAL EntwicklungspartnerInnenschaft Donau – Quality in Inclusion (Hrsg.): Sozialer Sektor im Wandel. Zur Qualitätsdebatte und Beauftragung von Sozialer Arbeit. Linz: edition pro mente, S. 169-184.
- Meyer, Wolfgang (2001): Umweltevaluation. In: Schulz, Werner/Burschel, Carlo/Weigert, Martin/ Liedtke, Christa/Bohnet-Joschko, Sabine/Kreeb, Martin/Losen, Dirk/Geßner, Christian/Diffenhard, Volker/Maniura, Anja (Hrsg.): Lexikon Nachhaltiges Wirtschaften. München: Oldenbourg Verlag, S. 433-440.
- Meyer, Wolfgang (2007): Nachhaltige Diskussion. Geschichte und Vielfalt eines Begriffs. *Querbrief* 2(2007): 4-7.
- Meyer-Krahmer, Frieder (1999): Was bedeutet Globalisierung für Aufgaben und Handlungsspielräume nationaler Innovationspolitiken? In: Grimmer, Klaus/Kuhlmann, Stefan/Meyer-Krahmer, Frieder (Hrsg.): Innovationspolitik in globalisierten Arenen. Opladen: Leske+Budrich, S. 35-65.
- Meyer-Krahmer, Frieder/Schmoch, Ulrich (1998): Science-based technologies: university-industry interactions in four fields. *Research Policy* 27, 835-851.
- Meylan, Jean-Pierre (1983): Evaluation von Innovationen im Bereich der Primarschule. Bern: Haupt.
- Michaelowa, Katharina/Borrmann, Axel (2005a): Wer evaluiert was, wie und warum? Eine politökonomische Analyse am Beispiel der deutschen Entwicklungszusammenarbeit. In: Ahrens, Heinz (Hrsg.): Zur Bewertung der Entwicklungszusammenarbeit. Berlin: Duncker & Humblot, S. 57-86.
- Michaelowa, Katharina/Borrmann, Axel (2005b): What determines evaluation outcomes? Evidence from bi- and multilateral development cooperation. HWWA Discussion Paper 310. Hamburg: HWWA.
- Moed, Henk F./Glänzel, Wolfgang/Schmoch, Ulrich (2004): Handbook of quantitative science and technology research: The use of publication and patent statistics in studies of S&T Systems. Dordrecht: Kluwer.
- Morand, Charles-Albert (2001): L'évaluation des effets des mesures étatiques. In: Thürer, Daniel/Aubert, Jean-François/Müller, Jörg Paul (éds.): Droit constitutionnel suisse. Zürich: Schulthess, p. 1119-1129.
- Moser, Heinz (1975): Aktionsforschung als kritische Theorie der Sozialwissenschaften. München: Kösel.
- Müller, Sabine (2002): Schulinterne Evaluation – Gelingensbedingungen und Wirkungen. Dortmund: IFS-Verlag.
- Müller, C. Wolfgang (Hrsg.) (1978): Begleitforschung in der Sozialpädagogik. Analysen und Berichte zur Evaluationsforschung in der Bundesrepublik. Weinheim: Beltz.
- Müller-Kohlenberg, Hildegard/Beywl, Wolfgang (2003): Standards der Selbstevaluation – Begründung und aktueller Diskussionsstand. *Zeitschrift für Evaluation* 2(1): 65-75.

- Müller-Kohlenberg, Hildegard/Münstermann, Klaus (Hrsg.) (2000): Qualität von Humandienstleistungen. Evaluation und Qualitätsmanagement in Sozialer Arbeit und Gesundheitswesen. Opladen: Leske + Budrich.
- Münch, Richard (2007): Die akademische Elite. Frankfurt: Suhrkamp.
- Nadai, Eva/Maeder, Christoph (2006): The promises and ravages of performance. Enforcing the entrepreneurial self in welfare and economy. Schlussbericht im Rahmen des Nationalen Forschungsprogramms (NFP) 51. Olten/Kreuzlingen: FH Nordwestschweiz/PH Thurgau.
- Naschold, Frieder (1995): Ergebnissteuerung, Wettbewerb, Qualitätspolitik. Entwicklungspfade des öffentlichen Sektors in Europa. 3. Auflage. Berlin: Edition Sigma.
- Neeser, Kurt/Szucs, Thomas/Bulliard, Jean-Luc/Bachmann, Gaudenz/Schramm, Wendelin (2007): Cost-effectiveness analysis of a quality-controlled mammography screening program from the Swiss statutory health-care perspective: Quantitative assessment of the most influential factors. *Value in Health* 10: 42-53.
- Neidhardt, Friedhelm (1988): Selbststeuerung in der Forschungsförderung. Das Gutachterwesen der DFG. Opladen: Westdeutscher Verlag.
- Nevo, David (1995): School-based evaluation: A dialogue for school improvement. Oxford: Pergamon.
- Newburn, Tim (2001): What do we mean by evaluation? *Children & Society* 15: 5-13.
- Nitsch, Heike/Osterburg, Bernhard (2005): Cross Compliance (CC) in der EU und Ökologischer Leistungsnachweis (ÖLN) in der Schweiz – eine vergleichende Analyse, FAL (Bundesforschungsanstalt für Landwirtschaft). Braunschweig: Bundesamt für Landwirtschaft.
- Novak, Rudolf (2005): The Austrian Science Fund FWF. *Plattform fteval Newsletter* 25: 37-38.
- Nuspliger, Kurt (2005): Der Stellenwert der Evaluation am Beispiel des Kantons Bern. *LeGes Gesetzgebung & Evaluation* 16(1): 51-61.
- o.A. (2006): Gemeinsam für Deutschland. Mit Mut und Menschlichkeit. Koalitionsvertrag von CDU, CSU und SPD. (www.cdusu.de/upload/koavertrag0509.pdf [31.05.2007])
- O'Neil, Tammie/Foresti, Marta/Hudson, Alan (2007): Evaluation of citizen's voice and accountability: Review of literature and donor approaches. London: Overseas Development Institute.
- ÖAR Regionalberatung (2005): Evaluierung der Regionalentwicklungsstruktur in Tirol. Wien: ÖAR Regionalberatung.
- Oberender, Peter O. (Hrsg.) (2005): Clinical Pathways, Facetten eines neuen Versorgungsmodells. Stuttgart: Kohlhammer.
- Obermair, Anna (1999): New Public Sector Management und die Verwaltungsreform in Österreich. *WIFO-Monatsberichte* 3: 213-223.
- Oberösterreichischer Landesrechnungshof (2006): Initiativeprüfung Landwirtschaftliche Förderungen. Linz: OÖRH. (www.lrh-o-oe.at/_files/downloads/aktuelles/meldungen/2006/IP_Landwirtschaftliche_Foerderung_Bericht.pdf [2.7.2007])
- OECD – Organisation for Economic Co-operation and Development (2001): Territorial outlook. Paris: OECD.
- OECD – Organisation for Economic Co-operation and Development (2002): Territorial reviews. Switzerland. Paris: OECD.
- OECD – Organisation for Economic Co-operation and Development (2003): OECD Environmental performance reviews: Austria. Paris: OECD.
- OECD – Organisation for Economic Co-operation and Development (2004): Prüfungen im Bereich Regulierungsreform: Deutschland – Konsolidierung der wirtschaftlichen und sozialen Erneuerung. Paris: OECD.
- OECD – Organisation for Economic Co-operation and Development (2005): OECD Employment. Outlook 2005. (www.oecd.org/dataoecd/36/30/35024561.pdf [17.10.2007])
- OECD – Organisation for Economic Co-operation and Development (2007): OECD factbook 2007: Economic, environmental and social statistics. Paris: OECD.

- OECD – Organisation for Economic Co-operation and Development (diverse Jahrgänge): Agricultural policies in OECD countries: Monitoring and evaluation. Paris: OECD.
- Oedl-Wieser, Theresia (2004): Chancengleichheit im Rahmen des Österreichischen Programms für die Entwicklung des ländlichen Raumes: Mid Term Evaluierung 2003. Facts & features 28. Wien: Bundesanstalt für Bergbauernfragen.
- Oevreivit, John (2002): Evaluation gesundheitsbezogener Interventionen. Bern: Huber.
- ÖROK – Österreichische Raumordnungskonferenz (2002): Österreichisches Raumentwicklungskonzept 2001. Schriftenreihe Nr. 163. Wien: ÖROK.
- ÖROK – Österreichische Raumordnungskonferenz (1992): Österreichisches Raumordnungskonzept 1991. Schriftenreihe Nr. 96. Wien: ÖROK.
- ÖROK – Österreichische Raumordnungskonferenz (1998): Raumordnung in Österreich. Schriftenreihe Nr. 137. Wien: ÖROK.
- ÖROK – Österreichische Raumordnungskonferenz (1999): 10. ÖROK-Enquete vom 20. Mai 1999 in Wien; das Österreichische Raumordnungskonzept 2001. Schriftenreihe Nr. 151. Wien: ÖROK.
- ÖROK – Österreichische Raumordnungskonferenz (2003): Raumordnung im Umbruch – Herausforderungen, Konflikte, Veränderungen, Festschrift für Eduard Kunze. Sonderserie Raum & Region Heft 1. Wien: ÖROK.
- ÖROK – Österreichische Raumordnungskonferenz (2007): EU-Strukturfonds 2007 – 2013: Neue Rahmenbedingungen und Chancen für Österreichs Regionen und Städte. Wien: ÖROK. (www.oerok.gv.at/EU_Regionalpolitik_in_Oesterreich/eu_2007_2013.htm [22.8.2007])
- Ortner, Karl Michael (1998): Schätzung der Wirkungen und Akzeptanz von Maßnahmen zur Reform des Milchmarktes der EU. *Die Bodenkultur / Austrian Journal of Agricultural Research* 49(4): 251-272.
- Ortner, Karl Michael (2004): Der Reformbedarf der GAP im Bereich der Marktordnungen. In: Schönböck, Wilfried (Hrsg.): Neuausrichtung der österreichischen und europäischen Agrarpolitik. Der öffentliche Sektor – Forschungsmemoranden 1-2/2004. Wien: Institut für Finanzwissenschaft und Infrastrukturpolitik der Technischen Universität Wien, S. 27-42.
- Ortner, Karl Michael (Hrsg.) (2005): Assessing rural development policies of the Common Agricultural Policy. Selection of papers from the 87th Seminar of the EAAE. Kiel: Wissenschaftsverlag Vauk.
- ÖSB Consulting (2004): Zwischenevaluierung des Regionalmanagements in Kärnten. Wien: ÖSB Consulting.
- Österreichische Qualitätssicherungsagentur (2007a): Erhebung und Evaluierung der Gleichstellung und Frauenförderung an österreichischen Universitäten. Wien. (www.aqa.ac.at/download.274.gesamt.bericht-glst-ff-endfassung-inklanhang.pdf [20.08.2007])
- Österreichische Qualitätssicherungsagentur (2007b): Das QM-Verfahren der AQA. Wien. (www.aqa.ac.at/ [14.09.2007])
- Österreichische Qualitätssicherungsagentur (2007c): Hochschulvergleich 2007 - Ingenieurwissenschaften, Geisteswissenschaften, Erziehungswissenschaft und Psychologie; Bericht zu Ergebnissen der österreichischen Universitäten, Verfahren und zur Methodik. Wien. (www.hochschulranking.ac.at/downloads/HSV_Ergebnisbericht_2007.pdf [20.08.2007])
- Österreichischer Agrarverlag (1997): Österreichs Landwirtschaft im EU-Agrarsystem. Klosterneuburg: Österreichischer Agrarverlag.
- Owen, John M./Rogers, Patricia J. (1999): Program evaluation: Forms and approaches. London: Sage.
- Palumbo, Dennis J. (1987): Politics and evaluation. In: Palumbo, Dennis J. (ed.): The politics of program evaluation. Newbury Park: Sage, pp. 12-46.
- Parlamentarische Verwaltungskontrolle (2000): Das Kartellrecht: Standortbestimmung. Schlussbericht zuhanden der Geschäftsprüfungskommission des Nationalrates. Bern: PVK.
- Parlamentsdienste (2003): Evaluation des Bundesinventars der Landschaften und Naturdenkmäler von nationaler Bedeutung (BLN). Bern: Parlamentsdienste.
- Patton, Michael Quinn (1997): Utilization-focused evaluation. 3rd ed. Thousand Oaks: Sage.

- Patton, Michael Quinn (1998): Die Entdeckung des Prozessnutzens. Erwünschtes und unerwünschtes Lernen durch Evaluation. In: Heiner, Maja (Hrsg.): Experimentierende Evaluation. Ansätze zur Entwicklung lernender Organisationen. Weinheim: Juventa, S. 55-66.
- Patton, Michael Quinn (2008): Utilization-focused evaluation. 4th ed. Thousand Oaks: Sage.
- Pechar, Hans (2006): Vom Vertrauensvorschuss zur Rechenschaftspflicht. Der Paradigmenwechsel in der britischen Hochschul- und Forschungspolitik seit 1980. *Österreichische Zeitschrift für Politikwissenschaft* 35(1): 57-73.
- Pelinka, Anton (1996): Die (veränderte) Kultur bildungspolitischer Entscheidungen. In: Specht, Werner/Thonhauser, Josef (Hrsg.): Schulqualität. Entwicklungen, Befunde, Perspektiven. Innsbruck: Studienverlag, S. 22-36.
- Pesendorfer, Dieter (2007): Paradigmenwechsel in der Umweltpolitik – Von den Anfängen einer Umwelt- zu einer Nachhaltigkeitspolitik – Modellfall Österreich? Wiesbaden: VS Verlag.
- Peter, Kathrin/Hefti, Helena (2004): Bericht zur Planungs- und Dispositionsphase für die Berichterstattung zur Agrarökologie 1994-2005. Bern: Bundesamt für Landwirtschaft.
- Peter, Kathrin/Sutter, Jan/Hirsig, Peter/Bütler, Mirjam (2001): Evaluation der Nitrat- und Phosphorprogramme des Bundes, Ergebnisse aus der Evaluation der Pilotphase der Nitrat- und Phosphorprogramme des Bundes nach Gewässerschutzgesetz Art. 62a. Bern: Bundesamt für Landwirtschaft.
- Peter, Martin/Sutter, Daniel/Maibach, Markus (2005): Volkswirtschaftliche Bedeutung des Flughafens Zürich. Auswirkungen verschiedener Entwicklungsszenarien. Schlussbericht. Zürich: Kanton Zürich, Amt für Verkehr.
- Peters, Matthias/Furrer, Cornelia/Bolliger-Salzmann, Heinz (2002): Das Alkoholprogramm 1999-2002 "Alles im Griff?" Schlussbericht der Evaluation. Hergiswil: Arbeitsgemeinschaft Globalevaluation Alkoholprogramm 1999-2002.
- Peters, Matthias/Knoepfel, Peter (1980): Standard Setting and Implementation in SO₂ Air Quality Control Policies. Fallstudie Schweiz. In: Knoepfel, Peter/Weidner, Helmut/Hanf, Kenneth (eds): International Comparative Analysis of Program Formulation and Implementation in SO₂-Air Pollution Control Policies in the EC Countries and Switzerland. Berlin: o.V., o.S.
- Peters, Matthias/Müller, Verena/Luthiger, Philip (2001): Auswirkungen des KVG auf die Versicherten. Bern: Bundesamt für Sozialversicherung.
- Peters, Matthias/Wapf, Bettina (2006): Zur Wirksamkeit des Gemeindeansatzes in der Prävention von Sucht- und Drogenproblemen. Internationale Befunde und Schlussfolgerungen für die Schweiz. Zürich: Econcept.
- Petri, Gottfried (1983): Evaluation der Schulversuche in der Oberstufe der allgemeinbildenden höheren Schulen: Lehrerbefragung 1982 zum Wahlpflichtunterricht. Graz: Zentrum für Schulentwicklung.
- Petri, Gottfried (1984a): Evaluation der Schulversuche in den Schulen der Zehn- bis Vierzehnjährigen. 2 Bände. Wien: Österreichischer Bundesverlag.
- Petri, Gottfried (1984b): Evaluation der Schulversuche in der Oberstufe der allgemeinbildenden höheren Schulen: Lehrerbefragung 1982 zum System der Leistungsgruppen. Graz: Zentrum für Schulentwicklung.
- Petri, Gottfried (1985): Evaluation der Schulversuche in der Oberstufe der allgemeinbildenden höheren Schulen: Schülerbefragung 1983 zum Leistungsgruppensystem und zur vorgezogenen Teilprüfung der Reifeprüfung. Graz: Zentrum für Schulentwicklung.
- Pfingstner, Hubert (2007a): Tätigkeitsbericht 2006. Wien: Bundesanstalt für Agrarwirtschaft. (www.awi.bmlfuw.gv.at/publikation/TB_2006.pdf [2.7.2007])
- Pfingstner, Hubert (2007b): Arbeitsprogramm 2007. Wien: Bundesanstalt für Agrarwirtschaft.
- Pfusterschmid, Sophie (2004): Evaluierung Berufsbildung: Halbzeitevaluierung der Maßnahme Berufsbildung des ländlichen Entwicklungsprogramms. Agrarpolitische Arbeitsbehelfe 18. Wien: Bundesanstalt für Agrarwirtschaft.

- Pichl, Claudia/Puwein, Wilfried/Obernberger, Ingwald/Steininger, Karl (1999): Erneuerbare Energieträger in Österreichs Wirtschaft: Volkswirtschaftliche Evaluierung am Beispiel der Biomasse. Wien: Österreichisches Institut für Wirtschaftsforschung.
- Pinter, Karin (1988): Schulungsmaßnahmen für arbeitslose Frauen, Jugendliche und Langzeitarbeitslose. Forschungsberichte aus Sozial- und Arbeitsmarktpolitik, Nr. 22. Wien: Bundesministerium für Arbeit und Soziales.
- PISA-Konsortium Deutschland (2007): PISA 2006: Die Ergebnisse der dritten internationalen Vergleichsstudie Münster: Waxmann.
- Plankl, Reiner/Brand-Sassen, Henning/Daub, Regina/Doll, Helmut/Pohl, Christian/Rudow, Katja (2006): Aktualisierung der Halbzeitbewertung der Ausgleichszulage in benachteiligten Gebieten 2002 bis 2004, länderübergreifender Bericht. Braunschweig: Bundesforschungsanstalt für Landwirtschaft. (literatur.fal.de/fallitdok_extern/bitv/zi041338.pdf [27.7.2007])
- Platform Research and Technology Policy Evaluation/Austrian Council for Research and Technology Development (eds) (2007): Evaluation of Austrian Research and Technology Policies. A summary of Austrian evaluation studies from 2003 to 2007. Wien: Plattform FTEval.
- Plattform Forschungs- und Technologieevaluierung (Hrsg.) (2003): Standards der Evaluierung in Österreichs Forschungs- und Technologiepolitik. Wien: fteval.
- Plattform Forschungs- und Technologieevaluierung (Hrsg.) (2005): Standards der Evaluierung in der Forschungs- und Technologiepolitik. Wien: Plattform FTEval.
- Platzer, Reinhard (2004): Prüfungswesen, Innovation und New Public Management in der Kärntner Landesregierung mit besonderer Berücksichtigung der Rolle des Rechnungshofes. In: Kärntner Verwaltungsakademie (Hrsg.): Bildungsprotokolle 9: New Public Management. Klagenfurt: Kärntner Verwaltungsakademie, S. 11-75.
- Pleschberger, Werner (2004): Kommunale Leistungsvergleiche in Österreich – Erkundungen zu einem neuen Instrument der lokalen Verwaltungsmodernisierung. In: Kuhlmann, Sabine/Bogumil, Jörg/Wollmann, Hellmut (Hrsg.): Leistungsmessung und -vergleich in Politik und Verwaltung. Wiesbaden: VS Verlag, S. 302-322.
- Pluto, Liane/Gragert, Nicola/van Santen, Eric/Seckinger, Mike (2007): Kinder- und Jugendhilfe im Wandel. Eine empirische Strukturanalyse. München: Deutsches Jugendinstitut.
- Pointner, Wolfgang/Polz, Wolfgang (2005): Diffusionsorientierte Technologiepolitik. Eine vergleichende Wirkungsanalyse für Österreich, die Schweiz, Deutschland und die USA. Joanneum Research Int-TeReg Bd. 5. Wien: Leykam Verlag.
- Pollitt, Christopher (2006): Performance information for democracy. *Evaluation* 12(1): 38-55.
- Polz, Wolfgang/Rojo, Jaime/Zinöcker, Klaus/Fahrenkrog, Gustavo/Tübke, Alexander (eds.) (2002): RTD Evaluation Tool Box. Assessing the Socio-Economic Impact of RTD-Policies Brussels/Luxembourg: European Commission, Joint Research Centre, Institute for Prospective Technological Studies/CSC-EEC-EA.
- Polz, Wolfgang/Vonortas, Nicholas (2006): IST R&D Evaluation and Monitoring, Evaluation Plan, Blueprint of the Evaluation and Monitoring of IST R&D in 6th and 7th Framework Programmes. December 2006. Brussels: EC, DG Information Society.
- Polutta, Andreas (2006): Der Blick auf pädagogische Effekte und wissenschaftliche Evidenz. Wirkungsorientierung in Jugendhilfepraxis und Jugendhilfeforschung. *Forum Erziehungshilfen* 12(5): 267-272.
- Popp, Reinhold (2003): Soziale Arbeit auf dem Weg zur Profession? SiÖ Sozialarbeit in Österreich, Zeitschrift für soziale Arbeit, Bildung und Politik 2/03: 14-16.
- Posch, Peter/Altrichter, Herbert (1993): Schulautonomie in Österreich. 2. Auflage. Bildungsforschung des Bundesministeriums für Unterricht und Kunst, Bd. 1. Wien: BMUK.
- Posch, Peter/Altrichter, Herbert (1997): Möglichkeiten und Grenzen der Qualitätssicherung und Qualitätsentwicklung im Schulwesen. Bildungsforschung des Bundesministeriums für Unterricht und kulturelle Angelegenheiten, Bd. 12. Innsbruck: Studienverlag.

- Power, Michael (1997): The audit society. Rituals of verification. Oxford: Oxford University Press.
- Power, Michael (2008): Research evaluation in the audit society. In: Matthies, Hildegard/Simon, Dagmar (Hrsg.): Wissenschaft unter Beobachtung. *Leviathan Sonderheft* 24: 15-24.
- Preskill, Hallie/Caracelli, Valerie (1997): Current and developing conceptions of use: Evaluation use TIG survey results. *Evaluation Practice* 18(3): 209-225.
- Preußler, Annabell (2008): Wir evaluieren uns zu Tode – Möglichkeiten und Grenzen der Bewertung von Online-Lernen. Eine Meta-Evaluation. Dissertation. Hagen: Fernuniversität.
- Prittitz, Volker von (1988): Das Katastrophenparadox. Opladen: Leske + Budrich.
- Proeller, Isabella (2006): Forum „Lessons Learned“: Verwaltungsreform in Deutschland, Österreich und der Schweiz – Erfahrungen und Lehren unter Nachbarn. In: Hofmeister, Alfred (Hrsg.): Verwaltung wohin? Der öffentliche Sektor zwischen Stabilität und Veränderung. SGVW-Schriftenreihe 46. Bern: SGVW, S. 33-42.
- Projekt eXe (Hrsg.) (2006): Aushandlungsprozesse zur Steuerung externer Evaluationen. Dokumentation der Fachtagung München, 1.-3.12.2004. München: Deutsches Jugendinstitut. (www.dji.de/bibs/333_7045_eXe_Aushandlungsprozesse.pdf [05.03.2007])
- Promberger, Kurt/Bernhard, Josef/Niederkofler, Carmen (2006): Grundlagen zur Evaluation von Verwaltungsreformen. Wien: Linde.
- Promberger, Kurt/Greil, Leopold/Simon, Martha-Maria (2005): Pilotprojekt Flexibilisierungsklausel. Ein österreichischer Weg zum New Public Management. Wien: Neuer Wissenschaftlicher Verlag.
- Promberger, Kurt/Ruskala, Iris/Cecon, Franziska (2004): Public Management in Austria. Zentrum für Verwaltungsmanagement, Working Paper 15/2004. Innsbruck: Universität Innsbruck.
- Promberger, Kurt/Zimmermann, Carmen (2007): Evaluation des Wirkungsanalyse-Programms (WaP). Innsbruck: Universität Innsbruck.
- Radnitzky, Edwin/Schatz, Michael (1999): Der Blick in den Spiegel. Texte zur Praxis von Selbstevaluation und Schulentwicklung. Innsbruck: Studienverlag.
- Rahmstorf, Stefan/Schellnhuber, Hans-Joachim (2007): Der Klimawandel. Diagnose, Prognose, Therapie. München: Beck.
- Raich, Traudlinde (2005): Gesetzesfolgenabschätzung (GFA) in Vorarlberg. In: Schäffer, Heinz (Hrsg.): Evaluierung der Gesetze/Gesetzesfolgenabschätzung in Österreich und im benachbarten Ausland. Wien: Manz, S. 91-100.
- Ramseier, Erich/Allraum, Jürgen/Stalder, Ursula/Grin, François/Alliata, Roberta/Müller, Stephan/Willmann, Ivo/Dozio, Edo/Labudde, Peter/Maag Merki, Katharina/Pagnossin, Elisabetta/Stocker, Eugen (2005): Evaluation der Maturitätsreform 1995 (EVAMAR). Neue Fächerstruktur, pädagogische Ziele, Schulentwicklung. Schlussbericht zu Phase 1. Bern: Staatssekretariat für Bildung und Forschung.
- Raschert, Jürgen (1974): Gesamtschule: ein gesellschaftliches Experiment. Stuttgart: Klett-Cotta.
- Rat für Forschung und Technologieentwicklung (2005): Recommendation regarding the evaluation and monitoring of RTI Programmes. In: Platform Research and Technology Policy Evaluation, Austrian Council for Research and Technology Development (eds): Evaluation of Austrian research and technology policies. Vienna: Platform Research and Technology Policy Evaluation, pp. 60-67.
- Rathkolb, Oliver (2005): Die paradoxe Republik. Österreich 1945 bis 2005. Wien: Zsolnay.
- Rechnungshof (2002): Wahrnehmungsbericht des Rechnungshofes: New Public Management in den Ländern. Wien: Rechnungshof.
- Rechnungshof (2005): Wahrnehmungsbericht des Rechnungshofes 4. Wien: Rechnungshof.
- Rechnungshof Baden-Württemberg (2007): „Jahrhundertreform“ durch Neue Steuerungsinstrumente (NSI) entfaltet bisher wenig Wirkung. Pressemitteilung. Karlsruhe: Rechnungshof Baden-Württemberg.

- Rechnungshof von Berlin (2006): Bericht gemäß § 99 LHO über den Stand der Umsetzung des Verwaltungsreform-Grundsatzgesetzes (VGG) in den Senatsverwaltungen. Berlin: Rechnungshof von Berlin.
- Reger, Guido/Kuhlmann, Stefan (1995): European technology policy in Germany. The impact of European Community policies upon science and technology in Germany. Heidelberg: Physica.
- Rehm, Jürgen/Uchtenhagen, Ambros (2001): Metaanalyse schweizerischer Arbeiten zur Verwendung von Methadon als Substitutionsmittel bei der Behandlung von Opioidabhängigkeit. Forschungsbericht. Zürich: ISGF.
- Reichard, Christoph (2006a): 10 Jahre NPM in Deutschland, Österreich und die Schweiz im Vergleich. In: Hofmeister, Alfred (Hrsg.): Verwaltung wohin? Der öffentliche Sektor zwischen Stabilität und Veränderung. SGVW-Schriftenreihe 46. Bern: SGVW, S. 43-56.
- Reichard, Christoph (2006b): Verwaltungsmodernisierung in den Bundesländern. In: Jann, Werner/Bogumil, Jörg/Bouckaert, Geert/Budäus, Dietrich/Holtkamp, Lars/Kißler, Leo/Kuhlmann, Sabine/Mezger, Erika/Reichard, Christoph/Wollmann, Hellmut: Statusreport Verwaltungsreform – eine Zwischenbilanz nach zehn Jahren. 2. Auflage. Berlin: Edition Sigma, S. 87-99.
- Reiche, Danyel (Hrsg.) (2005): Grundlagen der Energiepolitik. Frankfurt: Lang.
- Reiter, Karin/Roggendorf, Wolfgang (2007): Nutzbarkeit vorhandener Datenbestände für Monitoring und Evaluierung – am Beispiel InVeKoS. In: BLE, Bundesanstalt für Landwirtschaft und Ernährung (Hrsg.): Monitoring und Indikatoren der Agrobiodiversität, Tagungsband eines Symposiums am 7. und 8. November 2006 in Königswinter. Agrobiodiversität, Schriftenreihe des Informations- und Koordinationszentrums für Biologische Vielfalt, H. 27. Bonn: BLE, S. 274-287.
- Rentsch, Hans (2003): Umweltschutz auf Abwegen. Zürich: Orell Füssli.
- Key, Alfred (2000): Föderalismusreform – Neuer Finanzausgleich. In: Knoepfel, Peter/Linder, Wolf (Hrsg.): Verwaltung, Regierung und Verfassung im Wandel. Gedächtnisschrift für Raimund E. Germann. Basel: Helbing & Lichtenhahn, S. 45-60.
- Rhomberg, Wolfram/Steindl, Claudia/Weber, Matthias (2006): Neue Entwicklungen im Bereich der Wirkungsanalyse und -abschätzung FTI-politischer Maßnahmen. Wien: ARC systems research.
- Rhyn, Heinz (2007): Heterogenität, Gerechtigkeit und Exzellenz in der Wissensgesellschaft. Länderbericht Schweiz. In: Rhyn, Heinz (Hrsg.): Heterogenität, Gerechtigkeit und Exzellenz in der Wissensgesellschaft. Innsbruck: Studienverlag, S. 11-66.
- Rhyn, Heinz/Widmer, Thomas/Roos, Markus/Nideröst, Bruno (2002): Zuständigkeiten und Ressourcen in Zürcher Volksschulen mit und ohne Teilautonomie (TaV), Evaluationsbericht. Zürich: Kompetenzzentrum für Bildungsevaluation und Leistungsmessung / Institut für Politikwissenschaft, Universität Zürich. (www.ipz.uzh.ch/institut/mitarbeitende/staff/widmer/dokumente/evaluation_TaV_sb.pdf [1.6.2008])
- Rieder, Stefan (2005): Leistungs- und Wirkungsmessung in NPM-Projekten. In: Lienhard, Andreas/Ritz, Adrian/Steiner, Reto/Ladner, Andreas (Hrsg.): 10 Jahre New Public Management in der Schweiz: Bilanz, Irrtümer und Erfolgsfaktoren. Bern: Haupt, S. 149-159.
- Rieder, Stefan/Bischof, Ueli/Maugué, Marc/Götz, Konrad (2003): Evaluation der Netzwerke von EnergieSchweiz. Bern: Bundesamt für Energie.
- Rieder, Stefan/Furrer, Cornelia (2000): Evaluation des Pilotprojektes Leistungsauftrag mit Globalbudget im Kanton Solothurn. Beurteilung der Umsetzung und Wirkungen bei Parlament, Regierung und Verwaltung. Schlussbericht zu Handen der WOV-Kommission. Luzern: Interface.
- Rieder, Stefan/Widmer, Thomas (2007): Kantone im Wandel. Bern: Haupt.
- Riedli, Anna Maria (2003): „Auch im Dschungel wachsen Kinder auf...“ Programmentwicklung im Bereich der Mütter- und Väterberatung. Züricher Beiträge zur Theorie und Praxis. Zürich: Hochschule für Soziale Arbeit.
- Riesenfelder, Andreas/Sorger, Claudia/Wetzel, Petra/Willsberger, Barbara (2006): Evaluierung der Einführung des Kinderbetreuungsgeldes. Studie im Auftrag der Bundeskammer für Arbeiter und

- Angestellte. Wien: L&R Sozialforschung. (www.lrsocialresearch.at/files/Endbericht_Eval_KBG_L&R_Sozialforschung.pdf [6.6.2007]).
- Riess-Passer, Susanne (2002): Die Aufgaben der Verwaltungsreform: ein großer Wurf der Bundesregierung. In: Khol, Andreas/Ofner, Günther/Burkert-Dottolo, Günther/Karner, Stefan (Hrsg.): Österreichisches Jahrbuch für Politik 2001. Wien: Verlag für Geschichte und Politik, S. 419-432.
- Riffert, Franz/Paschon, Andreas (2005): Selbstevaluation von Schulentwicklungsprojekten. Der Modulansatz MSS. Ein Praxisbuch für Schulpartner. Münster: LIT-Verlag.
- Rigby, John (2002): Expert panels and peer review. In: Fahrenkrog, Gustavo/Polt, Wolfgang/Rojo, Jaime/Tübke, Alexander/Zinöcker, Klaus (eds.): RTD evaluation toolbox. Assessing the socio-economic impact of RTD-policies. IPTS Technical Reports Series. Sevilla/Vienna: IPTS, pp. 163-172.
- Rigby, John (2004): Making decisions about science and technology – Between the devil and the deep blue sea? *Plattform fteval Newsletter* 21: 2-11.
- Rille-Pfeiffer, Christine/Kapella, Olaf (Hrsg.) (2007): Kinderbetreuungsgeld. Evaluierung einer familienpolitischen Maßnahme. Innsbruck: Studienverlag.
- Rist, Ray C. (1992): Program evaluation in the United States General Accounting Office: Reflections on question formulation and utilization. In: Mayne, John/Bemelmans-Videc, Marie-Louise/Hudson, Joe/Conner, Ross (eds): Advancing public policy evaluation. Learning from international experiences. Amsterdam: North-Holland, pp. 149-156.
- Rist, Ray C. (1994): The preconditions for learning: Lessons from the public sector. In: Leeuw, Frans L./Rist, Ray C./Sonnichsen, Richard C. (eds): Can governments learn? Comparative perspectives on evaluation & organizational learning. New Brunswick: Transaction, pp. 189-205.
- Rist, Ray C./Paliokas, Kathleen L. (2002): The rise and fall (and rise again?) of the evaluation function in the U.S. government. In: Furubo, Jan-Eric/Rist, Ray C./Sandahl, Rolf (eds): International atlas of evaluation. New Brunswick: Transaction, pp. 225-245.
- Ritz, Adrian (2003): Evaluation von New Public Management. Bern: Haupt.
- Ritz, Adrian (2005): 10 Jahre Verwaltungsreform in den Schweizer Kantonen. In: Lienhard, Andreas/Ritz, Adrian/Steiner, Reto/Ladner, Andreas (Hrsg.): 10 Jahre New Public Management in der Schweiz: Bilanz, Irrtümer und Erfolgsfaktoren. Bern: Haupt, S. 47-67
- Ritz, Adrian (2007): Performance management and output-based budgeting in Switzerland. Output and outcome data – An end in itself or relevant for steering? In: Hill, Hermann (ed.) Modernizing government in Europe. Baden-Baden: Nomos, pp. 149-171.
- Ritz, Adrian/Sager, Fritz (2007): Outcome-based public management and the balance of powers in the direct democratic context of a Swiss city. Berne (unpublished manuscript).
- Rolff, Hans-Günter (1991): Schulentwicklung als Entwicklung von Einzelschulen? Theorien und Indikatoren von Entwicklungsprozessen. *Zeitschrift für Pädagogik* 37(6): 865-886.
- Rolff, Hans-Günter (1993): Wandel durch Selbstorganisation. Theoretische Grundlagen und praktische Hinweise für eine bessere Schule. Weinheim: Juventa.
- Rolff, Hans-Günter (1996): Evaluation – ein Ansatz zur Qualitätsentwicklung von Schulen? In: Landesinstitut für Schule und Weiterbildung (Hrsg.): Evaluation und Schulentwicklung. Ansätze, Beispiele und Perspektiven aus der Fortbildungsmaßnahme Schulentwicklung und Schulaufsicht. 2. Aufl. Soest: Verlag für Schule und Weiterbildung, S. 293-310.
- Rossi, Peter H./Freeman, Howard E./Lipsey, Mark W. (1999): Evaluation – a systematic approach. 6th ed. Thousand Oaks: Sage.
- Rossi, Peter H./Lipsey, Mark W./Freeman, Howard E. (2004): Evaluation – a systematic approach. 7th ed. Thousand Oaks: Sage.
- Rossier-Affolter, Roger (2000): Cost analysis of the medical prescription of narcotics. In: Gutzwiller, Felix/Steffen, Thomas (eds): Cost-benefit analysis of heroin maintenance treatment. Basel: Karger, pp. 11-36.
- Rossmann, Bruno (2002): Schlanker Staat. *Wirtschaftspolitische Blätter* 49(3): 293-303.

- Rudolf, Peter (2007): Außenpolitikevaluation. Konzeptionelle Überlegungen zu einem vernachlässigten Bereich der Außenpolitikanalyse. *Zeitschrift für Internationale Beziehungen* 14(2): 319-330.
- Russ-Eft, Darlene F./Bober, Marcie J./de la Teja, Ileana/Foxon, Marguerite/Koszalka, Tiffany A. (2008): Evaluator competencies: Standards for the practice of evaluation in organizations. San Francisco: Jossey Bass.
- Rüst, Thomas/Debrunner, Anneliese (2005): "Supported employment". Modelle unterstützter Beschäftigung bei psychischer Beeinträchtigung. Zürich: Rüegger.
- Sager, Fritz (2002): Vom Verwalten des urbanen Raums. Institutionelle Bedingungen von Politikkoordination am Beispiel der Raum- und Verkehrsplanung in städtischen Gebieten. Bern: Haupt
- Sager, Fritz (2003): Kompensationsmöglichkeiten föderaler Vollzugsdefizite. Das Beispiel der kantonalen Alkoholpräventionspolitiken. *Swiss Political Science Review* 9(1): 309-333.
- Sager, Fritz/Ritz, Adrian/Busmann, Kristina (2009): Utilization-focused performance reporting. A three-step model for the development of matching goals and indicators. *Public Money & Management* (forthcoming).
- Sager, Fritz/Schenkel, Walter (2004): Evaluation Umweltverträglichkeitsprüfung UVP. Umwelt-Materialien Nr. 175. Bern: Bundesamt für Umwelt, Wald und Landschaft.
- Salzburger Landesregierung (2004): Evaluierung von Förderungsmaßnahmen der Landwirtschaft im Bundesland Salzburg. Zusammenfassung und Anhang. Salzburg: Landesregierung. (www.salzburg.gv.at/bf/pdf_evaluierung_foe-massn_zusammenfassung.pdf und www.salzburg.gv.at/pdf_evaluierung_foe-massn_anhang.pdf [2.7.2007])
- Sawillion, Martin (2003): Das Förderprogramm „Klimaschutz-Plus“ des Ministeriums für Umwelt und Verkehr Baden-Württemberg. Ergebnisse, Erfahrungen, Ausblick. Karlsruhe: Klimaschutz- und Energieagentur Baden-Württemberg.
- Schaaf, Susanne (2003): Arbeitszufriedenheit und Burnout-Niveau von Mitarbeitenden in der stationären Suchthilfe gesamtschweizerisch. Forschungsbericht. Zürich: ISGF.
- Schäffer, Heinz (2005a): Vom Beruf der Politik zur Gesetzgebung in unserer Zeit. In: Schäffer, Heinz (Hrsg.): Evaluierung der Gesetze/Gesetzesfolgenabschätzung in Österreich und im benachbarten Ausland. Wien: Manz, S. 9-28.
- Schäffer, Heinz (Hrsg.) (2005b): Evaluierung der Gesetze/Gesetzesfolgenabschätzung in Österreich und im benachbarten Ausland. Wien: Manz.
- Schaffner, Dorothee (2002): [Lehr]platz für Migrantinnen. Schlussbericht zur Evaluation einiger Teilaspekte des Projekts. Basel: HPSABB (unveröffentlicht).
- Schaffner, Dorothee (2003): Schul- und familienergänzende Tagesstrukturen für Schülerinnen und Schüler im Kanton Basel-Stadt. Basel: HPSABB (unveröffentlicht).
- Schaffner, Dorothee (2004): Evaluation Netzwerk ‚Log in‘. Basel: HPSABB (unveröffentlicht).
- Schaffner, Dorothee/Fabian, Carlo/Drilling, Matthias (2002): Schulsozialarbeit im Kanton Aargau. Evaluationsergebnisse aus zwei Gemeinden. Empfehlungen für eine zukünftige Ausrichtung. Rahmenkonzept. Basel: HPSABB (unveröffentlicht).
- Schalk, Hans-Joachim/Untiedt, Gerhard (1996): Technologie im neoklassischen Wachstumsmodell: Effekte auf Wachstum und Konvergenz. Empirische Befunde für die Arbeitsmarktregionen der Bundesrepublik Deutschland 1978-1989. *Jahrbücher für Nationalökonomie und Statistik* 215(5): 562-585.
- Schalk, Hans-Joachim/Untiedt, Gerhard (2000): Regional investment incentives in Germany: Impacts on factor demand and growth. *Annals of Regional Science* 34: 173-195.
- Scharpf, Fritz W. (1985): Die „Politikverflechtungsfall“: Europäische Integration und deutscher Föderalismus im Vergleich. *Politische Vierteljahresschrift* 26(4): 323-356.
- Scharpf, Fritz W. (1993): Positive und negative Koordination in Verhandlungssystemen. *Politische Vierteljahresschrift, Sonderheft* 24: 57-83.

- Scharpf, Fritz W./Schnabel, Fritz/Reissert, Bernd (1976): Politikverflechtung. Theorie und Empirie des kooperativen Föderalismus in der Bundesrepublik. Kronberg/Ts.: Scriptor.
- Scharpf, Fritz W./Schnabel, Fritz/Reissert, Bernd (1980): Politische Bedingungen der Effektivität von raumordnerischen Steuerungsinstrumenten in der Deutschen Bundesrepublik. Stuttgart: Fraunhofer IRB-Verlag.
- Schedler, Kuno/Proeller, Isabella (2000): New Public Management. Bern: Haupt.
- Scherer, Roland/Löning, Kathrin/Schnell, Klaus-Dieter/Walser, Manfred (2001): Ex-Ante Evaluation des INTERREG IIIa-Programmes Alpenrhein-Bodensee-Hochrhein. St. Gallen: IDT-HSG.
- Schibany, Andreas/Jörg, Leonhard (2005): Instrumente der Technologieförderung und ihr Mix. Unter Mitarbeit von Brigitte Nones. Joanneum Research, InTeReg Research Report No. 37. Vienna: Joanneum Research.
- Schibany, Andreas/Nones, Brigitte/Schmidmayer, Julia/Jörg, Leonhard/Warta, Katharina/Sheikh, Sonja/Edler, Jakob (2005): Evaluierung der Christian Doppler Gesellschaft CDG. Joanneum Research, KMU Forschung Austria, Fraunhofer ISI. Wien: Eigenverlag.
- Schleicher-Tappeser, Ruggero/Höllinger, Andreas/Scherer, Roland/Schroeder, Marco/Denny, Evelyne/Thierstein, Alain/Schnell, Klaus-Dieter (1999): Evaluation intermédiaire du Programme INTERREG II – Rhin supérieur Centre-Sud. Freiburg (D)/St. Gallen/Strasbourg: EURES/SIASR-HSG/Cabinet Denny Consultant.
- Schmid, Günther (2003): Moderne Dienstleistungen am Arbeitsmarkt: Strategie und Vorschläge der Hartz-Kommission. *Aus Politik und Zeitgeschichte* B 06-07: 3-6.
- Schmidt, Boris (2008): Warum so wirksam? Und warum manchmal wirkungslos? – Subjektive Erklärungen zur Wirkung von Lehrveranstaltungsevaluation aus Sicht von Nutzern und Anbietern. *Zeitschrift für Evaluation* 7(1): 7-33.
- Schmidt, Uwe (2005a): Zwischen Messen und Verstehen. Anmerkungen zum Theoriedefizit in der deutschen Hochschulevaluation. *evaNet-Positionen* 06/2005 (www.hrk-qm.de/de/download/dateien/05-2005_-_Theoriedefizit_in_der_deutschen_Hochschulevaluation_-_Schmidt.pdf [21.5.2008])
- Schmidt, Uwe (2005b): Evaluation und Qualitätsmanagement an innovativen Hochschulen. In: Cremer-Renz, Chista/Donner, Hartwig (Hrsg.): Die innovative Hochschule. Aspekte und Standpunkte. Bielefeld: Universitätsverlag Webler, S. 141-160.
- Schmidt-Eichstaedt, Gerd (1999): Autonomie und Regelung von oben. In: Wollmann, Hellmut/Roth, Roland (Hrsg.): Kommunalpolitik. Politisches Handeln in den Gemeinden. Opladen: Leske & Budrich, S. 323-337.
- Schneider, Christian/Kessler, Stefan/Iten, Rolf (2006): Globalbeiträge an die Kantone nach Art. 15 EnG. Wirkungsanalyse kantonaler Förderprogramme. Ergebnisse der Erhebung 2005. Bern: Bundesamt für Energie.
- Schnell, Klaus-Dieter/Pfister Giauque, Barbara (2006): Bericht zur Schlussevaluation INTERREG III Schweiz. St. Gallen: Universität St. Gallen, Institut für öffentliche Dienstleistungen und Tourismus.
- Scholl, Bernd/van den Berg, Max/Pinson, Gaëlle/Schremmer, Christoph/Sieverts, Thomas/Zetter, John (2006): Raumplanung und Raumentwicklung in der Schweiz. Beobachtungen und Anregungen der Internationalen Expertengruppe. Zürich: ETH Zürich, Institut für Raum- und Landschaftsentwicklung.
- Schön, Donald/Rein, Martin (1994): Frame reflection. Toward the resolution of intractable policy controversies. New York: BasicBooks.
- Schötz, Dirk/Wortmann, Klaus/Krieg, Oliver (Hrsg.) (2003): Werbung für Energieeffizienz. Evaluierung einer Energieeffizienzkampagne am Beispiel der Stand-by-Kampagne „Aus. Wirklich aus?“. Berlin: Erich Schmidt.

- Schrader, Friedrich-Wilhelm/Helmke, Andreas (2003): Evaluation – was danach? Ergebnisse der Schulleiterbefragung im Rahmen der Rezeptionsstudie Walzer. *Schweizerische Zeitschrift für Bildungswissenschaften* 25(1): 79-107.
- Schrader, Friedrich-Wilhelm/Helmke, Andreas (2004): Von der Evaluation zur Innovation? Die Rezeptionsstudie WALZER: Ergebnisse der Lehrerbefragung. *Empirische Pädagogik* 18(1): 140-161.
- Schramek, Jörg/Bergs, Rolf/Bieling, Alfons/Stegmann, Susanne/Steimel, Kirsten/Welz, Dietmar (2005): Aktualisierte Halbzeitbewertung des rheinland-pfälzischen Entwicklungsplans "Zukunftsinitiative für den ländlichen Raum" (ZIL) im Zeitraum 2000-2006. Frankfurt: Institut für Ländliche Strukturforschung (IfLS) an der Johann Wolfgang Goethe-Universität Frankfurt. (www.mwvlw.rlp.de/internet/med/cdd/cdd16f46-2f39-9013-3e2d-ce3742f22936,11111111-1111-1111-1111-111111111111.isDownload.zip [1.6.2008])
- Schatz, Michael/Iby, Manfred/Radnitzky, Edwin (2000): Qualitätsentwicklung. Verfahren, Methoden, Instrumente. Weinheim: Beltz.
- Schröder, Jan/Kettiger, Daniel (2001): Wirkungsorientierte Steuerung in der Sozialen Arbeit. Ergebnisse einer internationalen Recherche in den USA, den Niederlanden und der Schweiz. Schriftenreihe des Bundesministeriums für Familie, Senioren, Frauen und Jugend, Band 229. Stuttgart: Kohlhammer.
- Shubert, Hans-Joachim (2005): Entwicklung eines Leitbildes: Chancen und Gefahren. In: Lauterbach, Karl/Schrapppe, Matthias (2004): Gesundheitsökonomie, Qualitätsmanagement und Evidence Based Medicine. 2. Auflage. Stuttgart: Schattauer, S. 371-374.
- Schug, Liselotte/Federspiel, Barbara/Brügger, Urs (2006): Synthesebericht Ökonomische Evaluation in den Kernthemen Gesundes Körpergewicht – Psychische Gesundheit/Stress – Gesundheitsförderung und Prävention stärken. Bern: Gesundheitsförderung Schweiz.
- Schwab, Oliver/Toepel, Kathleen (2000): Strukturfondsreform 2000 - Evaluation im Spannungsfeld zwischen Wissenschaft und Politik. Internetdokumentation der Jahrestagung der DeGEval in Köln 1998. (www.degeval.de/koeln/1998/toepel.htm [9.5.2007])
- Schwappach, David/Blaudszun, Annette/Conen, Dieter/Ebner, Heinz/Eichler, Klaus/Hochreutener, Marc Anton (2003): „Emerge“ – benchmarking of clinical performance and patient's experience with emergency care in Switzerland. *International Journal for Quality in Health Care* 15: 473-485.
- Schwappach, David/Blaudszun, Annette/Conen, Dieter/Eichler, Klaus/Hochreutener, Marc Anton/Koeck, Christian (2004): Women's experiences with low-risk singleton in-hospital delivery in Switzerland. *Swiss Medical Weekly* 134: 103-109.
- Schwarz, Christine (2006): Evaluation als modernes Ritual. Zur Ambivalenz gesellschaftlicher Rationalisierung am Beispiel virtueller Universitätsprojekte. Hamburg: Lit Verlag.
- Schweizerische Eidgenossenschaft (2003): Botschaft zur Staatsrechnung der Schweizerischen Eidgenossenschaft für das Jahr 2002. Bern.
- Schweizerische Koordinationsstelle für Bildungsforschung (SKBF) (2007): Bildungsbericht Schweiz 2006. 2. Auflage. Aarau: SKBF.
- Schweizerischer Bauernverband (diverse Jahrgänge): Statistische Erhebungen und Schätzungen über Landwirtschaft und Ernährung, Brugg: SBV.
- Schweizerischer Bundesrat (2003): Bundesbeschluss zur Neugestaltung des Finanzausgleichs und der Aufgabenteilung zwischen Bund und Kantonen (NFA) vom 3. Oktober 2003. Bern: BBL.
- Scriven, Michael (1973): Goal-free evaluation. In: House, Ernest R. (ed.): School evaluation: The politics and processes. Berkeley: McCutchan, pp. 319-328.
- Scriven, Michael (1974): The doctrine of cost free evaluation. In: Popham, James W. (ed.): Evaluation in education. Berkeley: McCutchan Publishing, pp. 85-93.
- Scriven, Michael (1983): Costs in evaluation. Concept and practice. In: Alkin, Marvin C./Solmon, Lewis C. (eds): The costs of evaluation. Beverly Hills: Sage, pp. 27-44.
- Scriven, Michael (1991): Evaluation thesaurus. 4th ed. Newbury Park: Sage.

- SECO – Staatssekretariat für Wirtschaft [2005]: Logframe, User Manual. Bern: SECO. (siehe Link auf www.seco-cooperation.admin.ch/themen/entwicklung/00513/index.html?lang=de [1.6.2008])
- Sedlacek, Peter (2004a): Evaluation in der Stadt- und Regionalentwicklung. Herausforderungen für Wissenschaft und Praxis. In: Sedlacek, Peter (Hrsg.): Evaluation in der Stadt- und Regionalentwicklung. Wiesbaden: VS Verlag, S. 11-26.
- Sedlacek, Peter (Hrsg.) (2004b): Evaluation in der Stadt- und Regionalpolitik. Wiesbaden: VS Verlag.
- Seel, Andrea/Altrichter, Herbert/Mayr, Johannes (2006): Innovation durch ein neues Lehrerdienstrecht? Eine Evaluationsstudie zur Implementierung des LDG 2001. In: Heinrich, Martin/Greiner, Ulrike (Hrsg.): Schauen, was 'rauskommt. Kompetenzförderung, Evaluation und Systemsteuerung im Bildungswesen. Münster: LIT-Verlag, S. 95-111.
- Seitz, Helmut (2005): Die ökonomischen und fiskalischen Effekte der Verwaltungsreform in Mecklenburg-Vorpommern. Gutachten im Auftrag des Innenministeriums Mecklenburg-Vorpommern. April 2005. Dresden: Innenministerium Mecklenburg-Vorpommern.
- Senatsverwaltung für Inneres (1999): Die Reform der öffentlichen Verwaltung in Berlin – eine Zwischenbilanz. Berlin: Senatsverwaltung für Inneres.
- Shadish, William R. Jr./Cook, Thomas D./Leviton, Laura C. (1991): Foundations of program evaluation. Theories of practice. Newbury Park: Sage.
- Shapira, Philip/Kuhlmann, Stefan (eds) (2003): Learning from science and technology policy evaluation: Experiences from the United States and Europe. Cheltenham: Edward Elgar.
- Sheikh, Sonja/Bornett, Walter (2005): FFG – Bereich Basisprogramme – Projektevaluierung 2005. Im Auftrag der Österreichischen Forschungsförderungsgesellschaft (FFG), Bereich Basisprogramme. Wien: KMU Forschung Austria.
- Sheikh, Sonja/Mandl, Irene/Radauer, Alfred/Edler, Jakob/Lo, Vivien/Hafner, Sabine (2004): Evaluierung der österreichischen Betreuungsstrukturen für das 6. EU-Rahmenprogramm für Forschung, technologische Entwicklung und Demonstration. Im Auftrag des Bundesministeriums für Wirtschaft und Arbeit. Wien/Karlsruhe: KMU Forschung Austria/Fraunhofer ISI.
- Sheldon, George (2000): Evaluationsprogramm Arbeitsmarktpolitik: Die Effizienz der öffentlichen Arbeitsvermittlung. Arbeitsmarktpolitik Nr. 4. Bern: Seco.
- Sheldon, George (2005): Entwicklung der Performance der öffentlichen Stellenvermittlung der Schweiz im Zeitraum 1998–2003. Arbeitsmarktpolitik Nr. 14. Bern: Seco.
- Shulha, Lyn M./Cousins, Bradley J. (1997): Evaluation use: Theory, research, and practice since 1986. *Evaluation Practice* 18(3): 195-208.
- Simma, Angelika (2007): Gesetzesfolgenabschätzung in Österreich. In: Schäffer, Heinz (Hrsg.): Evaluierung der Gesetze/Gesetzesfolgenabschätzung (II). Weitere Erfahrungsberichte aus dem In- und Ausland/Staatspraxis und Wirtschaft. Wien: Manz, S. 47-61.
- Simon, Christian/Burla, Sophie (2004): CEST-Evaluationen. Grundsätze und Erfahrungen in Wissenschaftsevaluationen. CEST 2004/6. Bern: Center for Science and Technology Studies.
- Simon, Dieter (2000): Aspekte der Qualität. *Gegenworte* 5: 15-17.
- Sinabell, Franz (2006): Durch welche Maßnahmen konnten die beabsichtigten Effekte des Programms der ländlichen Entwicklung maximiert werden? Wien: BMLFUW. (www.dafne.at/dafne_plus_homepage/index.php?section=dafneplus&content=result&project_id=296 [2.7.2007])
- Sinabell, Franz/Salhofer, Klaus/Karagiannis, Giannis (2006): Das Programm der Ländlichen Entwicklung 2000-2006 – ausgewählte ökonomische Konsequenzen. Forschungsbericht im Auftrag des BMLFUW. Wien: Österreichisches Institut für Wirtschaftsforschung.
- Sinz, Manfred (2006): Neue Leitbilder für die Raumentwicklung in Deutschland, Beitrag zur Regionalplanertagung Baden-Württemberg 2006 in Überlingen. Berlin: Bundesministerium für Verkehr, Bau und Stadtentwicklung.
- Skinner, Denise (2004): Evaluation and change management: Rhetoric and reality. *Human Resource Management Journal* 14(3): 5-19.

- Smeral, Egon/Keller, Peter (2000): InnoTour – Erfolg eines tourismuspolitischen Programms. In: Bieger, Thomas/Laesser, Christian (Hrsg.): Schweizer Tourismusjahrbuch 2000/2001. St. Gallen: Schweizerische Tourismuswirtschaft, S. 141-159.
- SNF – Schweizerischer Nationalfonds (2007): Grundsätze für die Evaluation von mit öffentlichen Geldern finanzierter Forschung durch den SNF. Bern: SNF. (www.snf.ch/SiteCollectionDocuments/Web-News/news_070302_evaluation_d.pdf [21.8.2007])
- Sommerfeld, Peter/Baumgartner, Edgar (2001): Abschlussbericht Evaluation SRK Therapiezentrum für Folteropfer. Olten: FHNW (unveröffentlicht).
- Sommerfeld, Peter/Baumgartner, Edgar/Baur, Roland/Klemenz, Regina/Müller, Silke/Piller, Edith Maud (2006): Evaluation „Midnight Basketball“ Projekte. Olten: FHNW (unveröffentlicht).
- Sommerfeld, Peter/Castelli, Francesco (2006): Siedlungen KraftWerk1 und Regina Kägi-Hof. Olten: FHNW.
- Soratori, Robert (2004): Controlling und Tagessatzpolitik für Einrichtungen im Sozialbereich. *SiÖ Sozialarbeit in Österreich, Zeitschrift für soziale Arbeit, Bildung und Politik* 3/04: 28-29.
- Soratori, Robert (2005): Externes Controlling für soziale Einrichtungen - branchenspezifisches Controlling Paket. *SiÖ Sozialarbeit in Österreich, Zeitschrift für soziale Arbeit, Bildung und Politik* 1/05: 33-34.
- Specht, Werner (1993): Evaluation der Schulversuche zum gemeinsamen Unterricht behinderter und nichtbehinderter Kinder. Ergebnisse einer bundesweiten Befragung von Lehrerinnen und Lehrern im Schulversuch (Endbericht). Graz: Zentrum für Schulversuche und Schulentwicklung.
- Specht, Werner (1997): Jedes Kind ist Mittelpunkt. Ergebnisse und Gedanken aus der Evaluation der Schulversuche zur Integration behinderter Schüler in der Sekundarstufe I. *Behinderte in Familie, Schule und Gesellschaft* 20(4): 9-30.
- Specht, Werner (2007): Die Qualität des Bildungssystems, Standards und Monitoring. In: Rhyn, Heinz (Hrsg.): Heterogenität, Gerechtigkeit und Exzellenz in der Wissensgesellschaft. Innsbruck: Studienverlag, S. 195-226.
- Specht, Werner/Grabensberger, Eva (2007): Datenorientierte Schulentwicklung. Erfahrungen aus dem Pilotprojekt „Schule BEWUSST“ in der Steiermark. *Unser Weg* 62(3): 83-94.
- Spiegel, Hiltrud von (1997): Perspektiven der Selbstevaluation. In: Müller-Kohlenberg, Hildgard/Autrata, Otger (Hrsg.): Evaluation der sozialpädagogischen Praxis. Materialien zur Qualitätssicherung in der Kinder und Jugendhilfe, Heft 11. Berlin: BMFSFJ, S. 32-49.
- Spiel, Christiane (2004): Workshop III: Ergebnisse und Qualitätsentwicklung. In: Hochschulrektorenkonferenz (in Kooperation mit dem Verbund Norddeutscher Universitäten) (Hrsg.): Evaluation – ein Bestandteil des Qualitätsmanagements an Hochschulen. Beiträge zur Hochschulpolitik. Bonn: Hochschulrektorenkonferenz, S. 233-242.
- Spinatsch, Markus (2002): Evaluation in Switzerland: Moving toward a decentralized system. In: Furubo, Jan-Eric/Rist, Ray C./Sandahl, Rolf (eds): International atlas of evaluation. New Brunswick: Transaction, pp. 375-391.
- Spinatsch, Markus (2008): Using evaluative knowledge for policy design: A case study of the Swiss addiction policy. In: Boyle, Richard/Breul, Jonathan D./Dahler-Larsen, Peter (eds): Open to the public. Evaluation in the public arena. New Brunswick: Transaction, pp. 121-133.
- Stamm, Margrit (2003): Evaluation und ihre Folgen für die Bildung – eine unterschätzte pädagogische Herausforderung. Münster: Waxmann.
- Statistik Austria (2006): Statistischer Jahresbericht 2005. Wien: Statistik Austria.
- Statistisches Bundesamt (2003): Umweltnutzung und Wirtschaft. Bericht zu den Umweltökonomischen Gesamtrechnungen 2003. Wiesbaden: Statistisches Bundesamt.
- Steiner, Olivier/Fabian, Carlo/Schnorr, Valentin (2006): Evaluation des Projektes Pinto Prävention, Intervention, Toleranz. Basel: FHNW (unveröffentlicht).

- Steiner, Roald/Fischl, Iris/Warta, Katharina/Jörg, Leonhard/Bacher, Joachim (2007): Begleitende Evaluierung der Kampagne www.innovatives-oesterreich.at. Wien/Wien/Hamburg: KMU Forschung Austria/Technopolis/TNS Infratest Communication Research Centre.
- Steiner, Wolfgang (2005): Evaluierung und Gesetzesfolgenabschätzung in der Landesgesetzgebung. Erfahrung und Leistung am Beispiel des Landes Oberösterreich. In: Schäffer, Heinz (Hrsg.): Evaluierung der Gesetze/Gesetzesfolgenabschätzung in Österreich und im benachbarten Ausland. Wien: Manz, S. 101-116.
- Stephan, Gesine/Rässler, Susanne/Schewe, Torben (2006): Das TrEffeR-Projekt der Bundesagentur für Arbeit: Die Wirkung von Maßnahmen aktiver Arbeitsmarktpolitik. *Zeitschrift für Arbeitsmarktforschung* 39(3/4): 447-465.
- Steurer, Reinhard (2007): From Government Strategies to Strategic Public Management: An Exploratory Outlook on the Pursuit of Cross-Sectoral Policy Integration. *European Environment* 17(3): 201-214.
- Steurer, Reinhard/Martinuzzi, André (2005): Toward a New Pattern of Governance: First Experiences with National Strategies for Sustainable Development in Europe. *Environment and Planning C: Government and Policy* 23(3): 455-472.
- Stevens, Alex/Berto, Daniele/Frick, Uli/Hunt, Neil/Kersch, Viktoria/McSweeney, Tim/Oeuvray, Kerrie/Puppo, Irene/Santa Maria, Alberto/Schaaf, Susanne/Trinkl, Barbara/Uchtenhagen, Ambros/Werdenich, Wolfgang (2006): The Relationship between Legal Status, Perceived Pressure and Motivation in Treatment for Drug Dependence: Results from a European Study on Quasi-Compulsory Treatment. *European Addiction Research* 12(4): 197-209.
- Stockbauer, Uta (2000a): Was macht Evaluationen nützlich? Grundlagen und empirische Untersuchungen zum Thema Verwertung und Verwertbarkeit von Evaluationen. Unveröffentlichte Dissertation: Universität Salzburg.
- Stockbauer, Uta (2000b): Was macht Evaluationen nützlich? Überblick zum Forschungsstand – Ergebnisse von Fallstudien. In: Müller-Kohlenberg, Hildegard/Münstermann, Klaus (Hrsg.): Qualität von Humandienstleistungen. Evaluation und Qualitätsmanagement in Sozialer Arbeit und Gesundheitswesen. Opladen: Leske + Budrich, S. 121-128.
- Stockmann, Reinhard (1992): Die Nachhaltigkeit von Entwicklungsprojekten – Eine Methode zur Evaluierung am Beispiel von Berufsbildungsprojekten. Opladen: Westdeutscher Verlag.
- Stockmann, Reinhard (1996): Die Wirksamkeit der Entwicklungshilfe – Eine Evaluation der Nachhaltigkeit von Programmen und Projekten. Opladen: Westdeutscher Verlag.
- Stockmann, Reinhard (2006a): Evaluation in Deutschland. In: Stockmann, Reinhard (Hrsg.): Evaluationsforschung. Grundlage und ausgewählte Forschungsfelder. Münster: Waxmann, S. 15-46.
- Stockmann, Reinhard (2006b): Evaluation und Qualitätsentwicklung – eine Grundlage für wirkungsorientiertes Qualitätsmanagement. Münster: Waxmann.
- Stockmann, Reinhard (Hrsg.) (2006c): Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder. 3. Auflage. Münster: Waxmann.
- Stockmann, Reinhard (Hrsg.) (2007): Handbuch zur Evaluation. Eine praktische Handlungsanleitung. Münster: Waxmann.
- Stockmann, Reinhard/Meyer, Wolfgang/Gaus, Hansjörg/Kohlmann, Uwe/Urbahn, Julia (2001): Nachhaltige Umweltberatung. Evaluation eines Förderprogramms der Deutschen Bundesstiftung Umwelt. Opladen: Leske + Budrich.
- Stockmann, Reinhard/Meyer, Wolfgang/Krapp, Stefanie/Koehne, Godehard (2000): Wirksamkeit deutscher Berufsbildungszusammenarbeit. Ein Vergleich zwischen staatlicher und nicht-staatlicher Entwicklungszusammenarbeit mit der VR China. Opladen: Westdeutscher Verlag.
- Straubhaar, Thomas/Werner, Heinz (2003): Arbeitsmarkt Schweiz – ein Erfolgsmodell? *Mitteilungen aus der Arbeitsmarkt- und Berufsforschung* 36(1): 60-76.

- Streicher, Gerhard/Schibany, Andreas/Dinges, Michael (2004): Evaluation FWF - Impact Analysis. Background Report 4.2. Joanneum Research InTeReg Research Report No. 23. Vienna: Joanneum Research.
- Striewe, Ludwig/Loy, Jens-Peter/Köster, Ulrich (1996): Analyse und Beurteilung der einzelbetrieblichen Investitionsförderung in Schleswig-Holstein. *Agrarwirtschaft* 45(12): 423-434.
- Strittmatter, Anton (1999): Qualitätsevaluation und Schulentwicklung. In: Thonhauser, J./Patry, Jean-Luc (Hrsg.): Evaluation im Bildungsbereich. Wissenschaft und Praxis im Dialog. Innsbruck: Studienverlag, S. 173-188.
- Stroebe, Wolfgang/Hewstone, Miles/Stephenson, Geoffrey M. (eds) (1996): Sozialpsychologie. Berlin: Springer.
- Strohmeier, Gerhard/Holzinger, Elisabeth (2004): KAP-EVA Koordinations- und Arbeitsplattform Evaluierung Bilanzbericht. Wien: Österreichische Raumordnungskonferenz ÖROK. (www.oerok.gv.at/EU_Regionalpolitik_in_Oesterreich/strukturfonds_2000_2006_i_D/Evaluierung/KAP-EVA_Bilanzbericht_IFF_OeLR.pdf [13.5.2008])
- Strohmeier, Gerhard/Holzinger, Elisabeth [2006]: Die Koordinations- und Arbeitsplattform KAP-EVA: Ein Lernprozess in der Evaluierung der Strukturfonds-Programme in Österreich. Wien: Bundeskanzleramt. (www.bka.gv.at/DocView.axd?CobId=15358 [13.5.2008])
- Strubelt, Wendelin (2004): Evaluation in der Stadt- und Regionalentwicklung. Grußwort. In: Sedlacek, Peter (Hrsg.): Evaluation in der Stadt- und Regionalentwicklung. Wiesbaden: VS Verlag, S. 27-38.
- Stuck, Andreas/Egger, Mathias/Hammer, Andreas/Minder, Christoph/Beck, John (2002): Home Visits to Prevent Nursing Home Admission and Functional Decline in Elderly People. Systematic Review and Meta-regression Analysis. *Journal of the American Medical Association JAMA* 287: 1022-1028.
- Studer, Thomas/Lenz, Peter/Mettler, Monika (2004): Entwicklung von Instrumenten für die Evaluation von Fremdsprachenkompetenzen (Französisch/Englisch): Ziele, Kontext, Gegenstände und methodologische Aspekte des IEF-Projektes. *Schweizerische Zeitschrift für Bildungswissenschaften* 26(3): 419-434.
- Stufflebeam, Daniel L. (2000): Foundational models for the 21st century program evaluation. In: Stufflebeam, Daniel L./Madaus, George F./Kellaghan, Thomas (eds): Evaluation models viewpoints on educational and human services evaluation. 2nd ed. Boston: Kluwer Academic Publishers, pp. 33-83.
- Stufflebeam, Daniel L. (2001): Evaluation models. *New Directions for Evaluation* 89.
- Stufflebeam, Daniel L./Madaus, George F./Kellaghan, Thomas (eds) (2000): Evaluation models. Viewpoints on educational and human services evaluation. 2nd ed. Boston: Kluwer.
- Stutzer, Alois/Frey, Bruno S. (2000): Stärkere Volksrechte – Zufriedene Bürger: eine mikroökonomische Untersuchung für die Schweiz. *Swiss Political Science Review* 6(3): 1-30.
- Suchman, Edward A. (1967): Evaluative research. Principles and practice in public service & social action programs. New York: Russell Sage Foundation.
- Summa, Hilka (2002): Definitions and frameworks. In: Pollitt, Christopher/Girre, Xavier/Lonsdale, Jeremy/Mul, Robert/Summa, Hilka/Waerness, Marit (eds): Performance or compliance? Performance audit and public management in five countries. Oxford: Oxford University Press, pp. 11-29.
- SWR – Schweizerischer Wissenschaftsrat (1993): Revitalizing Swiss social science. Evaluation report to the Swiss Science Council by the international panel of experts looking into the Situation of the social sciences in Switzerland. SWR-Schriftenreihe Research Policy FOP 13/1993. Bern: SWR.
- SWTR – Schweizerischer Wissenschafts- und Technologierat (2002): Evaluation des Schweizerischen Nationalfonds (SNF) und der Kommission für Technologie und Innovation(KTI). Bericht des SWTR an den Bundesrat, 30. Mai 2002. Bern: SWTR.
- Szucs, Thomas/Sandoz, Menga/Keusch, Gérald (2004): The cost-effectiveness of losartan in type 2 diabetics with nephropathy in Switzerland – an analysis of the RENAAL study. *Swiss Medical Weekly* 134: 440-447.

- TAK - Tripartite Agglomerationskonferenz (2004): Horizontale und vertikale Zusammenarbeit in der Agglomeration. Bern: Tripartite Technische Arbeitsgruppe.
- Tálos, Emmerich (2006): Sozialpartnerschaft. Austrokorporatismus am Ende? In: Dachs, Herbert/Gerlich, Peter/Gottweis, Herbert/Kramer, Helmut/Lauber, Volkmar/Müller, Wolfgang C./Tálos, Emmerich (Hrsg.): Politik in Österreich. Das Handbuch. Wien: Manz, S. 425-442.
- Tamme, Oliver (2004a): Auswirkungen des Österreichischen Programms für die Entwicklung des ländlichen Raums auf die Bevölkerungsstruktur: Mid Term Evaluierung 2003. Facts & features 31. Wien: Bundesanstalt für Bergbauernfragen.
- Tamme, Oliver (2004b): Beschäftigungswirkungen des Österreichischen Programms für die Entwicklung des ländlichen Raums: Mid Term Evaluierung 2003. Facts & Features 30. Wien: Bundesanstalt für Bergbauernfragen.
- Tauber, Magdalena (2006): Ergebnisse der Fragebogenerhebung: Auswirkungen des BVergG auf betroffene Organisationen und Wünsche Richtung BVergG von Seiten betroffener/nicht-betroffener Organisationen. Kurzfassung. Improve – Qualität im Wettbewerb um soziale und arbeitsmarktpolitische Dienstleistungen. Wien: BDV – Bundesdachverband für soziale Unternehmen.
- Taut, Sandy/Brauns, Dieter (2003): Resistance to evaluation. A psychological perspective. *Evaluation* 9(3): 247-264.
- Taylor, Sandra/Bachtlor, John/Polverari, Laura (2001): Structural Fund evaluation as a programme management tool: Comparative assessment and reflections on Germany. *Informationen zur Raumentwicklung* 2001(6/7): 341-358.
- Teagasc (2007): AGMEMOD (Agricultural Member States Modelling). The Irish Food Development Authority. (www.tnet.teagasc.ie/agmemod/ [2.7.2007])
- TEC – Tsunami Evaluation Coalition (2006): Joint evaluation of the international response to the Indian Ocean tsunami: Synthesis Report. London: Tsunami Evaluation Coalition.
- Technopolis Forschungs- und Beratungsgesellschaft mbH (2005): Zwischenevaluierung der RIF 2000 Regionale Impulsförderung. Wien: Technopolis.
- Tenroth, Heinz-Elmar (2008): Bildungsstandards außerhalb der "Kernfächer": Herausforderungen für den Unterricht und die fachdidaktische Forschung – Zur Einleitung in den Thementeil. *Zeitschrift für Pädagogik* 54(2): 159-162.
- Thierstein, Alain (2002): Von der Raumordnung zur Raumentwicklung. Modeerscheinung oder Ausdruck neuer Akteure und Trends? *DISP* 148: 10-18.
- Thierstein, Alain/Behrendt, Heiko (2001): Überprüfung der Zielerreichung der Schweizer Regionalpolitik. Zürich: ETHZ, ORL.
- Thierstein, Alain/Boulianne, Louis Marie/Gabi, Simone/Reinhard, Mark (2004): Die Modellvorhaben der Agglomerationspolitik: Auswirkungen auf die wirtschaftliche Entwicklung. Zürich/Lausanne: ETH Zürich/C.E.A.T.
- Thierstein, Alain/Kruse, Christian/Glanzmann, Lars/Gabi, Simone/Grillon, Nathalie (2006): Raumentwicklung im Verborgenen. Untersuchungen und Handlungsfelder für die Entwicklung der Metropolregion Nordschweiz. Zürich: NZZ Verlag.
- Thierstein, Alain/Rey, Michel/Walser, Manfred/Pfister Giauque, Barbara (1999): Evaluation des Bundesbeschlusses INTERREG II. Schlussbericht. St. Gallen/Lausanne: IDT-HSG/ C.E.A.T.
- Thierstein, Alain/Schedler, Kuno/Bieger, Thomas (Hrsg.) (2000): Die lernende Region. Regionale Entwicklung durch Bildung. Zürich: Rüegger.
- Thole, Werner (Hrsg.) (2002): Grundriss Soziale Arbeit. Ein einführendes Handbuch. Opladen: Leske + Budrich.
- Thonhauser, Josef/Patry, Jean-Luc (Hrsg.) (1999): Evaluation im Bildungsbereich. Wissenschaft und Praxis im Dialog. Innsbruck: Studienverlag.
- Toepel, Kathleen (2000): Evaluation in der Regionalpolitik. *Vierteljahreshefte zur Wirtschaftsforschung* 69(3): 395-405.

- Toepel, Kathleen/Schwab, Oliver (2005): Evaluation in der europäischen Strukturpolitik. *Zeitschrift für Evaluation* 2005(1): 63-74.
- Toepel, Kathleen/Tissen, Günter (2000): Stand und Perspektiven der Evaluation in Deutschland – Eine Einführung. *Vierteljahrshefte zur Wirtschaftsforschung* 69(3): 347-349.
- Tötschel, Anja (2004): Soziale Arbeit 2010. Welche Schlüsselkompetenzen benötigt die Profession "soziale Arbeit"? Fragen zu "sozialarbeiterischen Grundqualitäten" - eine Methodendiskussion. Zusammenfassung zum Arbeitskreis AK1a unter Leitung von Friedhelm Vahsen. *SiÖ Sozialarbeit in Österreich, Zeitschrift für soziale Arbeit, Bildung und Politik* 4/04: 23.
- Transversal/SEREC (2000): Evaluation ex ante Programme INTERREG IIIA: France (Rhône-Alpes) - Suisse. Diagnostic stratégique et opérationnel. Rapport finale 1ère phase. Annemasse/Vissoie/Lausanne: Transversal/SEREC.
- Tribl, Christoph (2007): Auswirkungen der GAP-Reform auf die österreichische Landwirtschaft. Wien: BMLFUW (www.dafne.at/dafne_plus_homepage/index.php?section=dafneplus&content=result&project_id=2629 [2.7.2007])
- TU Wien, Institut für Rechtswissenschaften (2004): Vergleich und Evaluierung der raumordnungsrechtlichen Regelungen für Einkaufszentren in Vorarlberg. Wien: TU Wien.
- Turowski, Gerd (2005): Raumplanung. In: Akademie für Raumforschung und Landesplanung (Hrsg.): Handwörterbuch der Raumordnung. 4. Auflage. Hannover: Akademie für Raumforschung und Landesplanung, S. 893-898.
- Uchtenhagen, Ambros/Dobler-Mikola, Anja/Steffen, Thomas/Gutzwiller, Felix/Blättler, Richard/Pfeifer, Sylvie (1999a): Prescription of narcotics for heroin addicts. Main results of the Swiss national cohort study. Basel: Karger.
- Uchtenhagen, Ambros/Dombrowski, Daniela (2001): Effektevaluation des Projekts „Gsundi Schuel“. Forschungsbericht. Zürich : ISGF.
- Uchtenhagen, Ambros/Schaaf, Susanne/Berger, Christa (1999b): EUREHA. Vocational rehabilitation of drug users and drug dependent persons. Strasbourg: Groupe Pompidou.
- Umbach, Eberhard (2002): Ist Evaluitis heilbar? *Physik Journal* 1(2): 3.
- Umweltbundesamt (Hrsg.) (2002): Kampagnen für erneuerbare Energien. Die Evaluation von „Solar – na klar!“ und Empfehlungen für neue Kampagnen. Berlin: Umweltbundesamt.
- Untersuchungsausschuss „Allgemeine Verwaltungsorganisation“ des Arbeitskreises VI der Innenministerkonferenz (Hrsg.) (2002): Aktivitäten zur Staats- und Verwaltungsmodernisierung in den Ländern und beim Bund. Stand Dezember 2002. Magdeburg: Ministerium des Innern Sachsen-Anhalt. (www.inneres.sachsen-anhalt.de/min/r14/download/brosch_verwref.pdf [19.3.2008])
- Vader, Jean-Paul/Hämmig, Robert/Besson, Jacques/Eastus, Christopher/Eggenberger, Christina/Burnand, Bernard (2003): Appropriateness of methadone maintenance treatment for opiate addiction: evaluation by an expert panel. *Sozial- und Präventivmedizin/Social and Preventive Medicine* 48(Suppl. 1): S1-S14.
- Valovirta, Ville (2002): Evaluation utilization as argumentation. *Evaluation* 8(1): 60-80.
- Van Ackeren, Isabell/Bellenberg, Gabriele (2004): Parallelarbeiten, Vergleichsarbeiten und Zentrale Abschlussprüfungen. In: Holtappels, Heinz Günter/Klemm, Klaus/Pfeiffer, Hermann/Rolff, Hans-Günter/Schulz-Zander, Renate (Hrsg.): Jahrbuch der Schulentwicklung. Band 13. Daten, Beispiele und Perspektiven. Weinheim: Juventa, S. 125-159.
- Van der Knaap, Peter (1995): Policy evaluation and learning: Feedback, enlightenment or argumentation. *Evaluation* 1(2): 189-216.
- Vaney, Claude/Heinzel-Gutenbrunner, Monica/Schnelle, Martin/Reif, Marcus (2006): Hanf lindert Spasmen und macht mobiler – Ergebnisse einer wissenschaftlichen Studie mit einem bei MS-Betroffenen eingesetzten Cannabisextrakt. In: Bundesamt für Gesundheit (Hrsg.). Suchtforschung des BAG 1999-2001. Bern: BAG, S. 41-50.

- Vaterlaus, Stefan/Telser, Harry/Zweifel, Peter/Eugster, Patrick (2004): Was leistet unser Gesundheitswesen? Studie im Auftrag des Bundesamts für Sozialversicherung, von Galenica, Interpharma, Merian-Iselin-Spital, MSD, santésuisse, Staatssekretariat für Wirtschaft, Vereinigung der Importeure pharmazeutischer Spezialitäten, Verband Schweiz. Assistenz- und Oberärzte. Regensdorf: Plaut Economics.
- Vatter, Adrian (2002): Kantonale Demokratien im Vergleich. Entstehungsgründe, Interaktionen und Wirkungen politischer Institutionen in den Schweizer Kantonen. Opladen: Leske + Budrich.
- Vatter, Adrian (2006a): Die Kantone. In: Klöti, Ulrich/Knoepfel, Peter/Kriesi, Hanspeter/Linder, Wolf/Papadopoulos, Yannis/Sciarini, Pascal (Hrsg.): Handbuch der Schweizer Politik. 4. Auflage. Zürich: NZZ Verlag, S. 203-231.
- Vatter, Adrian (2006b): Einleitung. In: Vatter, Adrian (Hrsg.): Föderalismusreform. Wirkungsweise und Reformansätze föderativer Institutionen in der Schweiz. Zürich: Verlag NZZ, S. 13-19.
- Vatter, Adrian/Sager, Fritz (1996): Föderalismusreform am Beispiel des Ständemehrs. *Swiss Political Science Review* 2(2): 165-200.
- Vatter, Adrian/Sager, Fritz (2006): Das Ständemehr: Wirkungsweise und Reformen, in: Vatter, Adrian (Hrsg.). Föderalismusreform. Wirkungsweise und Reformmodelle föderativer Institutionen in der Schweiz. Zürich: NZZ Verlag, S. 73-98.
- Vedung, Evert (1997): Public policy and program evaluation. New Brunswick: Transaction.
- Vedung, Evert (1999): Evaluation im öffentlichen Sektor. Wien: Böhlau.
- Verhoest, Koen/Verschuere, Bram/Peters, Guy B./Bouckaert, Geert (2007): Pressure, legitimacy and innovative behavior by public organizations. *Governance* 20(3): 469-497.
- VN-Millenniums-Projekt (Hrsg.) (2005): In die Entwicklung investieren: Ein praktischer Plan zur Erreichung der Millenniums-Entwicklungsziele. Überblick. New York: Vereinte Nationen. (www.unmillenniumproject.org/documents/germanoverview.pdf [4.4.2008])
- Vodoz, Luc/Pfister Giauque, Barbara/Reinhard, Mark (2006): Evaluation eRegio. Evaluation du Centre de compétences pour les échanges d'informations et d'expériences des régions en Suisse. Lausanne: C.E.A.T.
- Voelzkow, Helmut/Hoppe, Andrea (1996): „Druck von oben“ und „von unten“. Zu den Reformansätzen der deutschen Regionalpolitik als Reaktion auf Implementationsprobleme des Europäischen Regionalfonds in den neuen Bundesländern. In: Heinelt, Hubert (Hrsg.): Politiknetzwerke und Europäische Strukturfondsförderung. Ein Vergleich zwischen EU-Mitgliedstaaten. Opladen: Leske + Budrich, S. 108-130.
- von Fischer, Christoph/Jouval, Norbert/Boillat, Patrick/Christ, Pascal/Guillaume-Gentil, Sylvain/Voillat, Françoise (2006): Räumliche Auswirkungen des Vue-des-Alpes-Tunnels - eine ex-post Analyse. Zusammenfassung. Bern: ARE.
- von Stokar, Thomas/Menegale, Sarah/Steinemann, Myriam/Hammer, Stephan (2003): Zwischenevaluation Interreg III. Schlussbericht. Zürich: seco/ARE.
- von Stokar, Thomas/Vettori, Anna/Steinemann, Myriam/Schmidt, Nicolas/Schoenenberger, Alain/Bohr, Nils (2004): Evaluation des Bundesbeschlusses zugunsten wirtschaftlicher Erneuerungsgebiete. Schlussbericht. Bern: INFRAS/Eco'Diagnostic.
- Wackernagel, Mathis/Rees, William (1997): Unser ökologischer Fußabdruck. Wie der Mensch Einfluss auf die Umwelt nimmt. Basel: Birkhäuser Verlag.
- Waldhauser Franz/Jürgenssen Olaf/Püspök Rudolf/Tatzer Ernst (Hrsg.) (2003): Weggelegt. Kinder ohne Medizin? Wien: Czernin Verlag.
- Watts, Ronald (1999): Comparing federal systems. Montreal: McGill.
- Weiss, Carol H. (1977): Research for policy's sake: The enlightenment function of social research. *Policy Analysis* 3: 531-545.
- Weiss, Carol H. (1987): Where politics and evaluation research meet. In: Palumbo, Dennis J. (ed.): The politics of program evaluation. Newbury Park: Sage, pp. 47-70.

- Weiss, Carol H. (1998a): Evaluation. Methods for studying programs and policies. 2nd ed. Upper Saddle River: Prentice Hall.
- Weiss, Carol H. (1998b): Have we learned anything new about the use of evaluation? *American Journal of Evaluation* 19(1): 21-33.
- Weiss, Carol H./Murphy-Graham, Erin/Birkeland, Sarah (2005): An alternate route to policy influence: How evaluations affect D.A.R.E. *American Journal of Evaluation* 26(1): 12-30.
- Weiß, Gerhard (1997): Evaluierung der Politik zum Schutz vor Naturgefahren: Grundlagen der Evaluation und Forschungsansatz. *Centralblatt für das gesamte Forstwesen* 114 (4): 189-211.
- Wendt, Heinz/Efken, Josef/Klepper, Reiner/Krah, Volker/Nölle, Joachim/Schäfer, Martin/Trefflich, Anette/Utrecht, Inge (2006): Aktualisierung der Zwischenbewertung der Förderung zur Marktstrukturverbesserung für den Förderzeitraum 2000-2006, länderübergreifender Bericht im Rahmen der Gemeinschaftsaufgabe Verbesserung der Agrarstruktur und des Küstenschutzes (GAK). Braunschweig: Selbstverlag.
- Wettach, Ralph (2005): Therapiezufriedenheit von Drogenabhängigen: Stellenwert von Persönlichkeitsaspekten, Behandlungsbeurteilung und Institutionsmerkmalen im stationären Setting. Frankfurt: Lang.
- Wicke, Lutz (1993): Umweltökonomie. München: Vahlen.
- Widmer, Thomas (1991): Evaluation von Massnahmen zur Luftreinhaltepolitik der Schweiz. Chur: Rüegger.
- Widmer, Thomas (1996): Meta-Evaluation. Kriterien zur Bewertung von Evaluationen. Bern: Haupt. (www.ipz.uzh.ch/institut/mitarbeitende/staff/widmer/dokumente/Meta-Evaluation.pdf [11.12.2007])
- Widmer, Thomas (2000): Evaluating evaluations: Does the Swiss practice live up to the 'Program Evaluation Standards'? In: Russon, Craig (ed.): Program evaluation standards in international settings. Occasional Papers Series, Vol. 17. Kalamazoo: Western Michigan University, The Evaluation Center, pp. 67-80.
- Widmer, Thomas (2002): Staatsreformen und Evaluation: Konzeptionelle Grundlagen und Praxis bei den Schweizer Kantonen. *Zeitschrift für Evaluation* 2002/1: 101-114. (www.zfev.de/fruehereAusgabe/ausgabe2002-1/download/widmer.pdf [12.3.2008])
- Widmer, Thomas (2004): The development and status of evaluation standards in Western Europe. *New Directions for Evaluation* 104: 31-42.
- Widmer, Thomas (2005): Instruments and procedures for assuring evaluation quality. A Swiss perspective. In: Schwartz, Robert/Mayne, John (eds): Quality matters. Seeking confidence in evaluating, auditing, and performance reporting. New Brunswick: Transaction, pp. 41-68.
- Widmer, Thomas (2006a): Evaluationsstandards in Europa. In: Joint Committee on Standards for Educational Evaluation (Hrsg.): Handbuch der Evaluationsstandards. 3. Auflage. Wiesbaden: VS Verlag, S. 301-312.
- Widmer, Thomas (2006b): Qualität der Evaluation – Wenn Wissenschaft zur praktischen Kunst wird. In: Stockmann, Reinhard (Hrsg.): Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder. 3. Auflage. Münster: Waxmann, S. 85-112.
- Widmer, Thomas (2007): Rahmenbedingungen und Praxis der Evaluation im schweizerischen Bundesstaat. *ÖHW – Das öffentliche Haushaltswesen in Österreich* 48(1-2): 69-93.
- Widmer, Thomas (2008a): Evaluationsansätze und ihre Effekte: Erfahrungen aus verschiedenen Politikfeldern. In: Matthies, Hildegard/Simon, Dagmar (Hrsg.): Wissenschaft unter Beobachtung. Effekte und Defekte von Evaluationen. *Leviathan Sonderheft* 24: 267-287.
- Widmer, Thomas (2008b): Evaluation in der Außenpolitik – Gründe für eine Evaluationslücke. *Zeitschrift für Internationale Beziehungen* 15(1): 125-137.
- Widmer, Thomas/Bisang, Kurt/Moser, Christian (2004): TARMED. Vorbereitung einer Wirkungsanalyse über die Einführung sowie die Wirkungen von TARMED. Bern: Bundesamt für Gesundheit.

- (www.ipz.uzh.ch/institut/mitarbeitende/staff/widmer/dokumente/Tarmed_2005_d.pdf [11.12.2007])
- Widmer, Thomas/Blaser, Cornelia/Falk, Chantal (2007a): Evaluating Measures Taken Against Right-Wing Extremism. *Evaluation* 13(2): 221-239.
- Widmer, Thomas/Frey, Kathrin (2006): Evaluation von Mehrebenen-Netzwerkstrategien. *Zeitschrift für Evaluation* 2006/2: 287-316.
- Widmer, Thomas/Landert, Charles/Bachmann, Nicole (2000): Evaluations-Standards der Schweizerischen Evaluationsgesellschaft (SEVAL-Standards). Bern/Genève: SEVAL. (www.seval.ch/de/documents/seval_Standards_2001_dt.pdf [17.9.2007])
- Widmer, Thomas/Neuenschwander, Peter (2004): Embedding evaluation in the Swiss federal administration. Purpose, institutional design and utilization. *Evaluation* 10(4): 388-409.
- Widmer, Thomas/Rieder, Stefan (2003): Schweizer Kantone im institutionellen Wandel. Ein Beitrag zur Beschreibung und Erklärung institutioneller Reformen. *Schweizerische Zeitschrift für Politikwissenschaft* 9(1): 201-232.
- Widmer, Thomas/Rothmayr, Christine/Serdült, Uwe (1996): Kurz und gut? Qualität und Effizienz von Kurzevaluationen. Chur: Rüegger.
- Widmer, Thomas/Rüegg, Erwin/Neuenschwander, Peter (2001): Stand und Aussichten der Evaluation beim Bund. Bern: Schweizerische Bundeskanzlei. (www.ipz.uzh.ch/institut/mitarbeitende/staff/widmer/dokumente/EvalBund_def_03072001.pdf [29.4.2008])
- Widmer, Thomas/Uffer, Sabina/Frey, Kathrin/Widmer, Alex/Widmer, Céline (2007b): Wirkungsanalyse Umweltbildung. Kurzbericht/Gesamtbericht. Bern: Bundesamt für Umwelt. (www.umwelt-schweiz.ch/uw-0708-d [19.3.2008])
- Wiesner, Reinhard (Hrsg.) (2006): SGB VIII Kinder und Jugendhilfe. Kommentar. 3. Auflage. München: Beck.
- Willi, Bruno/Neuhäusler, Irène/Gabriel-Schärer, Pia (2005): Evaluationsbericht des SIP-Pilotprojektes Luzern. Luzern: Fachhochschule Zentralschweiz (unveröffentlicht).
- Williams, Kevin/de Laat, Bastiaan/Stern, Elliot (2002): The use of evaluation in the Commission Services: Final report. Paris: Technopolis France. (www.forskningsradet.no/CSStorage/Flex_attachment/EvalCommissionServices.pdf [28.5.2008])
- Wissenschaftsrat (Hrsg.) (1994): Stellungnahme zur Umweltforschung in Deutschland, Bd.1 und 2. Köln: Wissenschaftsrat.
- Wissenschaftsrat (Hrsg.) (1996): Stellungnahme zur außeruniversitären Materialwissenschaft. Köln: Wissenschaftsrat.
- Wissenschaftsrat (Hrsg.) (1999): Stellungnahme zur Energieforschung. Köln: Wissenschaftsrat.
- Wissenschaftsrat (Hrsg.) (2000a): Systemevaluation der Blauen Liste – Stellungnahme des Wissenschaftsrates zum Abschluss der Bewertung der Einrichtungen der Blauen Liste. Köln: Wissenschaftsrat.
- Wissenschaftsrat (Hrsg.) (2000b): Drittmittel und Grundmittel der Hochschulen 1993-1998. Köln: Wissenschaftsrat
- Wissenschaftsrat (Hrsg.) (2001): Systemevaluation der HGF – Stellungnahme des Wissenschaftsrates zur Hermann von Helmholtz-Gemeinschaft Deutscher Forschungszentren. Köln: Wissenschaftsrat.
- Wollmann, Hellmut (2000): Evaluierung und Evaluierungsforschung von Verwaltungspolitik und -modernisierung – zwischen Analysepotenzial und -defizit. In: Stockmann, Reinhard (Hrsg.): Evaluationsforschung. Opladen: Leske & Budrich, S. 195-233.
- Wollmann, Hellmut (2001): Zwischen Management- und Politiksystem: Die Kommunalverwaltung der 90er Jahre auf einer Modernisierungswelle. In: Edeling, Thomas/Jann, Werner/Wagner, Dieter (Hrsg.): Reorganisationsstrategien in Wirtschaft und Verwaltung. Opladen: Leske & Budrich, S. 15-57.

- Wollmann, Hellmut (2002): Verwaltungspolitik und Evaluierung. Ansätze, Phasen und Beispiele im Ausland und in Deutschland. *Zeitschrift für Evaluation*, 1/2002: 75-99.
- Wollmann, Hellmut (2005a): Evaluation. In: Akademie für Raumforschung und Landesplanung (Hrsg.): Handwörterbuch der Raumordnung. 4. Auflage. Hannover: Akademie für Raumforschung und Landesplanung, S. 274-280.
- Wollmann, Hellmut (2005b): Evaluierung von Verwaltungsmodernisierung. In: Blanke, Bernhard/ Bandemer, Stephan von/Nullmeier, Frank/Wewer, Göttrik (Hrsg.): Handbuch zur Verwaltungsreform. Wiesbaden: VS Verlag, S. 502-510.
- Wottawa, Heinrich (2001): Qualitätsmanagement durch Zielvereinbarung. In: Spiel, Christiane (Hrsg.): Evaluation universitärer Lehre – zwischen Qualitätsmanagement und Selbstzweck. Münster: Waxmann, S. 151-163.
- Wottawa, Heinrich/Thierau, Heike (2003): Lehrbuch Evaluation. 3. Auflage. Bern: Huber.
- Wroblewski, Angela (2004): More Activation - More chances for the unemployed? Changes in Austria's active labour market policy after accessing the European Union. *European Journal of Social Security* 6(1): 60-82.
- Wroblewski, Angela/Gindl, Michaela/Leinter, Andrea/Pellert, Ada/Woitech, Birgit (2006): Wirkungsanalyse frauenfördernder Maßnahmen des bm:bwk, Materialien zur Förderung von Frauen in der Wissenschaft, Band 21. Wien: Verlag Österreich.
- Zimmer, Jürgen/Preissing, Christa/Thiel, Thomas/Heck, Anne/Krappmann, Lothar (1997): Kindergärten auf dem Prüfstand. Dem Situationsansatz auf der Spur. Seelze-Velber: Kallmeyer'sche Verlagsbuchhandlung.
- Zimmermann, Willi (1990): Un inventaire des évaluations au niveau fédéral. In: Horber-Papazian, K. (éd.): Evaluation des politiques publiques en Suisse: pourquoi? pour qui? comment? Lausanne: Presse polytechnique et universitaire romandes, p. 79-88.
- Zimmermann, Willi/Kolly, Doris/Mahon, Pascal (1988): Die wissenschaftliche Evaluation von Bundespolitik. Eine bibliographische und methodische Exploration. Bern: Forschungsstelle Wissenschaft und Politik.
- Zinöcker, Klaus/Dinges, Michael (2004): Evaluation of the Austrian Industrial Research Promotion Fund (FFF) and the Austrian Science Fund (FWF). Positionierung des FFF und FWF vis-a-vis anderen Finanzierungsinstrumenten. Modul 2. Joanneum Research InTeReg Research Report No. 23. Vienna: Joanneum Research.
- Zinöcker, Klaus/Neurath, Wolfgang/Topolnik, Michaela/Schmidmayer, Julia (eds) (2007): Evaluation of Austrian Research and Technology Policies. A Summary of Austrian Evaluation Studies from 2003 to 2007. Vienna: Platform Research and Technology Policy Evaluation.
- Zinöcker, Klaus/Schindler, Julia/Gude, Stefanie/Schmidmayer, Julia/Stampfer, Michael (2005a): Interimsevaluierung FIT – IT. Konzepte, Rahmenbedingungen, Design, Prozesse. Snapshots auf Wirkung und Additionalität. Studie im Auftrag des bmvit. Wien: InTeReg
- Zinöcker, Klaus/Sheikh, Sonja/Jörg, Leonhard/Polt, Wolfgang/Stampfer, Michael (2004): Ausbildungsworkshops Evaluierung – eine Dokumentation. Wien: Plattform fteval.
- Zinöcker, Klaus/Tempelmaier, Birgit/Radauer, Alfred/Fischl, Iris/Steiner, Roald/Ruegg, Rosalie (2005b): Austrian Genome Research Programme GEN-AU: Mid term programme management evaluation, Final report. Vienna: Joanneum Research.
- Zintl, Michaela (2006): Von Wirkungen und Nebenwirkungen. Evaluierung nach der Millenniumserklärung und der Pariser Erklärung. *EINS – Entwicklungspolitik Information Nord-Süd* 9/2006: 30-33.
- Zucker, Judith (2006): Analyse der Leistungsfähigkeit und des Nutzens von Evaluationen der Politik zur Entwicklung ländlicher Räume in Deutschland und Großbritannien am Beispiel der einzelbetrieblichen Investitionsförderung. Landbauforschung Völknerode, Sonderheft 293. Braunschweig: FAL.

Zürcher, Christoph/Koehler, Jan/Böhnke, Jan (2007): Assessing the Impact of Development Cooperation in North East Afghanistan. Interim Report, Evaluation Reports Nr. 28, Bonn: Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung.

Zürcher, Markus (1999): Wider ein betriebswirtschaftliche verkürztes New Public Management. Über Produkte, Indikatoren und die Wirksamkeit. *LeGes – Gesetzgebung heute* 10(2): 105-134.