

Schriften der Deutschen Sektion
des Internationalen Instituts
für Verwaltungswissenschaften
Band 38

Bewerten und Bewertet-Werden

**Wirkungskontrolle und Leistungssicherung
in der öffentlichen Verwaltung**

Herausgegeben von Jan Ziekow



Nomos

Schriftenreihe der Deutschen Sektion
des Internationalen Instituts
für Verwaltungswissenschaften
Band 38

Prof. Dr. Jan Ziekow (Hrsg.)

Bewerten und Bewertet-Werden

Wirkungskontrolle und Leistungssicherung
in der öffentlichen Verwaltung



Nomos

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

ISBN 978-3-8487-1059-1

1. Auflage 2014

© Nomos Verlagsgesellschaft, Baden-Baden 2014. Printed in Germany. Alle Rechte, auch die des Nachdrucks von Auszügen, der fotomechanischen Wiedergabe und der Übersetzung, vorbehalten. Gedruckt auf alterungsbeständigem Papier.

Vorwort

Werten und bewertet werden gehört in vielfältiger Hinsicht zu unserem Alltag und auch in den öffentlichen Sektor hat der Gedanke des Wettbewerbs längst Einzug gehalten – augenfällig versinnbildlicht nicht zuletzt durch die Aufnahme der Leistungsvergleiche nach Art. 91d in das Grundgesetz.

Wettbewerbe und Leistungsvergleiche bergen viel positives Potenzial zur Effizienz- und hoffentlich auch Effektivitätssteigerung im öffentlichen Bereich. Gleichwohl sind die bestehenden Nachteile bekannt, die man auch mithilfe der ausgefeiltesten Methodik bislang (noch) nicht hinreichend in den Griff bekommen hat. Und das angesichts der Tatsache, dass das Ergebnis derartiger Vergleiche gleichsam über „Leben und Tod“ einer Körperschaft oder einer Behörde entscheiden kann. Die Noten der internationalen Ratingagenturen im Finanzbereich sind ganz entscheidend für die Frage, ob und zu welchen Konditionen ein Staat Kredit bekommt – was sich wiederum darauf auswirkt, wie viel Geld er überhaupt noch ausgeben kann und ob er sozusagen wieder „auf die Beine kommt“.

Natürlich kann man – gerade aus wissenschaftlicher Sicht – nur immer wieder vor einem magischen Glauben an die Zahl warnen. Ob nun AAA oder ein anderer Buchstabe des Alphabets hinzukommt, dürfe nicht entscheidend sein für das Wohl und Wehe ganzer Staaten, klagt die Politik immer wieder. Allerdings hat es die Politik sonst sehr gerne, wenn sich komplizierteste Bewertungen auf einer Seite oder möglichst sogar in einer Zahl zusammenfassen lassen. Generell gilt deshalb: Man sollte immer schauen, was steht hinter den Zahlen und den knappen Bewertungen.

Die Messungen der sogenannten „Erfüllungskosten“ durch den Nationalen Normenkontrollrat sind ein gutes Beispiel dafür, wie Mess-Verfahren so eingesetzt werden können, dass positive Effekte für die Volkswirtschaft erzielt werden. Nach den Zahlen, die der Nationale Normenkontrollrat auf seiner *homepage* veröffentlicht hat, seien vom Kontrollrat seit dem 1. Juli 2011 ca. 350 Regelungsvorhaben abschließend geprüft worden. Diese Prüfung habe im Saldo zu einer Reduzierung des jährlichen Erfüllungsaufwands von rund 1,5 Mrd. Euro geführt, was allerdings im Wesentlichen auf eine einzige Maßnahme zurückgegangen sei, nämlich die Reduzierung der Aufbewahrungsfristen nach dem Steuer- und Handelsrecht. Aber auch ohne diese Maßnahme wäre ein Anstieg des Erfüllungsaufwands seit Juli 2011 von knapp 1 Mrd. Euro zu verzeichnen gewesen. Diese Zahlen sind sehr beeindruckend, zumal auch nicht unmittelbar zu erkennen ist, dass mit dieser Fristverkürzung gemeinwohlbezogene Nachteile zu entstehen drohen.

So positiv diese und andere Beispiele sein mögen, es gibt auch Risiken. Zum ersten ist auf das berühmte Bonmot zu verweisen, dass öffentliches und privates Management grundsätzlich gleich seien – aber nur in allen unwesentlichen Aspekten. Das heißt, dass die öffentliche Verwaltung zwar Steuerungsinstrumente aus der Privatwirtschaft übernehmen kann, aber sehr darauf achten muss, dass hier wirklich Äpfel mit Äpfeln verglichen werden und nicht Großkonzerne mit Jobcentern. Zum anderen ist der Umstand zu nennen, dass die Erhebung von Zahlen keineswegs als ein rein „unpolitischer“ Vorgang anzusehen ist. Vielmehr wissen wir alle, dass die Entscheidung darüber, welche Indikatoren gewählt werden, in mancher Hinsicht schon das Ergebnis vorwegnehmen kann. Wir wissen auch, dass es oftmals schwierig ist, die passenden Indikatoren zu finden.

Gefährlich wird es zum einen, wenn an unzutreffend ermittelte Indikatoren Ressourcenentscheidungen geknüpft werden. Aber nicht ungefährlich sind auch die von einer *bevorstehenden* Bewertung ausgehenden Anreize, wenn man die Evaluationsindikatoren schon kennt. Denn es kann sein, dass sich eine Organisation nur noch nach diesen ausrichtet. Sogenanntes „falsches“ Lernen ist die Folge, da die zu evaluierenden Aspekte zu Lasten anderer (etwa gemeinwohlbezogener) Themen bevorzugt werden. Im Bereich von PISA spricht man von einem „*teaching to the test*“, da vorwiegend dasjenige gelernt wird, das in den entsprechenden Tests auch abgefragt wird. Andere, ebenso wichtige Fächer oder Fähigkeiten können dabei an den Rand gedrängt werden.

Auch die Ergebnisse einer bereits durchgeführten Evaluation oder eines Leistungsvergleichs können sich schädigend auf die Motivation von Organisationen und ihrer Beschäftigten auswirken. So können diejenigen, die besonders *gut* abgeschnitten haben, veranlasst sein, zu glauben, sie bräuchten sich nicht mehr weiterzuentwickeln. Umgekehrt können diejenigen, deren Abteilung oder Behörde *besonders schlecht* abgeschnitten hat, oftmals frustriert werden und die Motivation verlieren, sich überhaupt derartigen Mess- und/oder Bewertungsprozessen zu unterziehen.

Es gab daher Grund genug, dass sich die Deutsche Sektion des Internationalen Instituts für Verwaltungswissenschaften und das Deutsche Forschungsinstitut für öffentliche Verwaltung bei Ihrer gemeinsamen Tagung des Themas „Bewerten und Bewertet-Werden“ dieser und anderer Aspekte der Wirkungskontrolle und Leistungssicherung in der öffentlichen Verwaltung angenommen haben. Die Tagung, die gleichzeitig die Jahrestagung 2012 der Deutschen Sektion darstellte, fand vom 18. bis 19. Oktober 2012 in Speyer statt. Die im Rahmen dieser Tagung präsentierten Beiträge sind in diesem Band dokumentiert.

Ohne die Unterstützung einer Vielzahl von Personen, denen an dieser Stelle herzlich gedankt werden soll, hätte die Veranstaltung nicht durchgeführt werden und dieser Band nicht erscheinen können. Das aus Mitgliedern des Vorstands der Deutschen Sektion und Mitgliedern des Forschungsinstituts gebildete Organisa-

tionskomitee hat die Tagung inhaltlich intensiv vorbereitet. Die Last der organisatorischen Vorbereitung lag beim Generalsekretär der Deutschen Sektion, Herrn Dr. *Christoph Hauschild* sowie den Mitarbeiterinnen und Mitarbeitern des Forschungsinstituts und des Tagungsbüros der Deutschen Universität für Verwaltungswissenschaften. Die sachkundige Formatierung des Tagungsbandes hat Frau *Elisabeth Gerhards*, Deutsches Forschungsinstitut für öffentliche Verwaltung, übernommen.

Speyer, im August 2013

Jan Ziekow

Inhaltsverzeichnis

Markus Müller

Anreize, Bewertungsverfahren und Informationssysteme in der obersten Verwaltungsbehörde. Kritische Anmerkungen zu Möglichkeiten und Grenzen 15

Literatur 22

Antoinette Weibel

Variabler Leistungslohn in der Verwaltung – Allheilmittel oder Medikament mit starken Nebenwirkungen? 23

1. Variabler Leistungslohn in der Theorie und in der Praxis 23

2. Arten der Motivation 25

3. Management intrinsischer Motivation: Verdrängungs- und Verstärkungseffekt 26

3.1 Der Verdrängungseffekt 26

3.2 Zwischenfazit 29

3.3 Der Verstärkungseffekt 29

3.3.1 Verstärkung durch Partizipation 30

3.3.2 Verstärkung durch prozedurale Fairness 30

3.3.3 Verstärkung durch normative geprägte Entscheidungskontexte 31

4. Fazit 32

Literatur 33

Sylvia Veit

Evidenzbasierte Politik durch Gesetzesfolgenabschätzungen? Erfahrungen aus Deutschland und Schweden 37

1. Einleitung: Evidenzbasierte Politik und Gesetzesfolgenabschätzungen. Konzeptionelle Vorüberlegungen 37

2. Institutionalisierung von Gesetzesfolgenabschätzungen in Deutschland 43

3. Institutionalisierung der GFA in Schweden 46

4. Empirische Daten zur Umsetzung der GFA in Deutschland und Schweden 49

5. Fazit: Evidenzbasierte Politik durch Gesetzesfolgenabschätzung? 54

Norbert Heller

Erfahrungen mit Bewertungen und Evaluation bei den EU-Strukturfonds 57

1.	Sagenhaft	57
2.	EU-Strukturfonds	58
3.	Bewertungskriterien	59
3.1	Makroebene und makroökonomische Modelle	59
3.2	Projektebene	60
3.2.1	Projektanzahl und Arten der Projekte	61
3.2.2	Input-Indikator	61
3.2.3	Output-Indikator	61
3.2.4	Ergebnisindikator	62
3.2.5	Zielsetzung und Zielerreichung auf Projektebene	63
3.3	Programmebene	64
3.3.1	Inputindikator und Ausgabekategorien	64
3.3.2	Output und Ergebnis	65
3.3.3	Zielsetzung und Zielüberprüfung	66
3.3.4	Umverteilungsentscheidungen	67
3.3.5	Gleiche Bedingungen	67
3.3.6	Theorie und Praxis	68
4.	Resümee	69

Werner Keggenhoff

Qualitätsmanagement in der Sozialverwaltung 71

1.	Das Landesamt	71
2.	Haushalt	71
3.	Unsere Aufgaben	72
4.	Kernaufgabe „Menschen mit Behinderungen“	72
5.	Aktuell	72
6.	Verwaltungsmodernisierung	73
7.	Betriebswirtschaftliche Elemente	73
8.	Auswertungsbogen	73
9.	Qualitätssicherung	74
10.	Aufbau unseres Qualitätsmanagements	74
11.	Ziele und Kennzahlen	74
12.	Indikatoren mit Gewichtung	75
13.	Weitere Indikatoren	75
14.	Auswertung	76
15.	Probleme	76
16.	Qualitätssicherung im Vergleich	76
17.	Weitere Produkte im Qualitätsmanagement	77

18. Qualitätsmanagement	77
19. Fazit	77

Jörg Bogumil

Möglichkeiten von Leistungsmessung und Leistungsvergleichen in öffentlichen Verwaltungen – ein zusammenfassender Kommentar	79
--	----

Literaturhinweise	82
-------------------	----

Margrit Seckelmann

Reflexive Reformprozesse und politische Selbstkorrekturfähigkeit: Erfolgsbedingungen von Evaluationen am Beispiel der „Hartz-Gesetze“	83
---	----

1. Ausgangsfrage: Erfolgsbedingungen von Evaluationen	83
2. Die Evaluation der „Hartz-Gesetze“	87
2.1 Erwartungen an Evaluationen	87
2.2 Ausgangssituation	89
2.3 Ziele der „Hartz“-Gesetze	91
2.4 Staatsleitbild des „aktivierenden Staats“	91
2.5 Die „Hartz-Gesetze“ im Einzelnen	93
2.5.1 Die Aushandlungsprozesse vor ihrem Inkrafttreten	93
2.5.2 Die Hauptinhalte der einzelnen Gesetze	94
2.6 Evaluationen der „Hartz“-Gesetze	95
2.6.1 Evaluationsarten	95
2.6.2 Akteure und Maßnahmen	97
2.6.2.1 Evaluation Hartz I – III: Fünf Arbeitspakete	100
2.6.2.1.1 Arbeitspaket 1: Wirksamkeit der Instrumente	100
2.6.2.1.2 Arbeitspaket 2: Organisatorischer Umbau der Bundesagentur für Arbeit	103
2.6.2.1.3 Arbeitspaket 3: Akzeptanz der Bundesagentur für Arbeit	103
2.6.2.1.4 Arbeitspaket 4: Administrative Unterstützung bei der Evaluation	104
2.6.2.1.5 Arbeitspaket 5: Aufbereitung evaluationsgeeigneter Daten	104
2.6.2.2 Gesamtzusammenfassung durch das BMAS	104
2.6.2.3 Fortlaufende Querschnittsevaluationen	105
2.6.2.4 Evaluation der kommunalen Option (§ 6c SGB II)	105
2.7 Evaluationsergebnisse	107
2.7.1 Personal-Service-Agenturen	108
2.7.2 Arbeitsbeschaffungsmaßnahmen	109
2.7.3 Ergebnisse der fortlaufenden Querschnittsevaluation	111
2.7.4 Evaluation der kommunalen Option (§ 6c SGB II)	112

2.8	Umsetzung der Evaluationsempfehlungen	112
2.8.1	Kommunale Option (ARGEn)	112
2.8.2	Personal-Service-Agenturen	113
2.8.3	Arbeitsbeschaffungsmaßnahmen	114
3.	Fazit: Erfolgsbedingungen von Evaluationen und von „Politiklernen“?	114
3.1	Die Interpretation der Evaluationsergebnisse	114
3.2	Umsetzung durch Politik und Verwaltung?	116
3.3	Die „Hartz-Reformen“: Erfolgreicher oder erfolgloser Versuch der Implementation eines „lernenden arbeitsmarktpolitischen Systems“?	117

Gottfried Konzendorf

Zum Einfluss von Evaluationen auf die politische Entscheidungsfindung	121
---	-----

1.	Einleitung und Zielsetzung	121
2.	Wissenschaftliche Grundlagen	122
3.	Erste Erwartungen (Hypothesen) und Befunde	124
4.	Direkte Nutzung von Evaluation: Einflussfaktoren	125
5.	Faktoren im Überblick und Ausblick	130
	Literatur	131

Alexander Kroll/Anne-Kathrin Meier

Erfahrungen mit Performance-Informationen in der Kommunalverwaltung: Datennutzung, Routinen und Soziale Kognition	133
---	-----

1.	Einleitung	133
2.	Performance Management in der Kommunalverwaltung	135
3.	Methodik der Umfrage	136
4.	Ergebnisse der Umfrage	138
4.1	Nutzung unterschiedlicher Kennzahlentypen	139
4.2	Einstellung gegenüber Kennzahlen	141
4.3	Erfahrungen im Umgang mit Kennzahlen	143
5.	Die Verwendung von Performance-Daten	144
5.1	Theoretische Erklärungen	144
5.2	Erste empirische Befunde	145
6.	Fazit	146
	Literatur	147

Rainer Christian Beutel

Indikatoren für eine wirkungsorientierte, erfolgreiche Steuerung von Städten und Landkreisen – ein Praxisbericht 149

1. Strategie 149
2. Wirkungsorientierung 150
3. Kennzahlen 152
4. Monitoringsysteme 152
- 4.1 Praxisbeispiele strategischer Steuerung 153

Benedikt Speer/Marika Gruber

Verwaltungswettbewerbe und -preise als Trojanisches Pferd der Leistungsmessung? 155

1. Einleitung 155
2. Ausgangslage in Österreich 155
3. (Qualitäts-)Wettbewerbe in der öffentlichen Verwaltung 156
- 3.1 Hintergrund und Zweck von Qualitätswettbewerben 157
- 3.2 Das Faible österreichischer Verwaltungen für Qualitätswettbewerbe 158
- 3.3 Zwischenfazit 161
4. Der „Speedy Offiziales“: ein „deviant case“ aus Kärnten? 162
- 4.1 Vom Verwaltungspreis für Schnelligkeit 162
- 4.2 ...zum Verwaltungspreis für Verfahrensqualität 163
5. Resümee 167
- Quellennachweis 168

Heinz Pansi

Die Auswirkungen von Bewertungen auf die öffentliche Verwaltung 171

1. Allgemeines 171
2. Die Bezirkshauptmannschaft 172
3. Die Bewertungen in der Praxis 173
- 3.1 Die Kunden/befragung/bewertung 174
- 3.1.1 Auswirkungen 174
4. Die (nicht selbstbestimmte) Qualitätsbewertung am Beispiel des „Speedy Offiziales“ 175
5. Benchlearning der Bezirkshauptmannschaften 175
- 5.1 Beispiel Führung 176
- 5.2 Beispiel „Fremdenwesen“ 176
- 5.3 Beispiel „Ethik in der Verwaltung“ 177
- 5.4 Beispiel „Sicherheitsagenden“ 177
- 5.5 Beispiel „Großschadensereignis“ 178

6.	Die Auswirkung von Bewertungen in der Praxis, unter dem Aspekt des „Benchlearning-Projektes“ der BH	179
6.1	Die Auswirkungen von Bewertungen auf personeller Ebene	179
6.2	Auswirkungen von Bewertungen auf die organisatorische Ebene	179
6.3	Die Außenwirkungen des Verhaltens der Organisation	179
7.	Conclusio	180
	Podiumsdiskussion	181
	<i>Christoph Emminghaus</i>	181
1.	Entwicklung von Evaluation und Wirkungsmessung	181
2.	Umgang mit Evaluation und Leistungskontrolle	182
3.	Aktuelle Trends	183
	<i>Beate Lohmann</i>	185
	<i>Veith Mehde</i>	187
	<i>Jann Meyer-Abich</i>	189
	Autorenverzeichnis	195

Anreize, Bewertungsverfahren und Informationssysteme in der obersten Verwaltungsbehörde. Kritische Anmerkungen zu Möglichkeiten und Grenzen

Markus Müller¹

Bevor wir uns angesichts meiner in diesem Punkte sicher nicht verallgemeinerbaren und in mancher Hinsicht vielleicht auch untypischen Praxis einzelnen Beispielen von Anreizen und Bewertungssystemen zur Diskussion ihrer Sinnhaftigkeit, ihrer Möglichkeiten und Grenzen, zuwenden, lassen Sie mich mit einer Beobachtung zweiter Ordnung im Sinne von Luhmann beginnen (zum letzteren siehe Schaal/Heidenreich 2011: 254 f.).

Ich finde es instruktiv, sich mit der Art und Weise zu beschäftigen, wie wir uns in der öffentlichen Verwaltung, namentlich den obersten Bundes- und Landesbehörden als Zentren der politischen Führung, mit der *Organisation von Wissen* befassen. Dabei will ich hier „Wissen“ durchaus in dem umfassenden Sinne verstehen, wie Sie den Begriff in der Postmodernen Theorie, namentlich bei Foucault vorfinden, insofern also auch einschließlich der *Sprache*, mit der wir über die Sachverhalte unserer Verwaltungsarbeit sprechen, und einschließlich der *Diskurse* und *Argumentationsmodi*, die wir zur Entscheidung über diese Sachverhalte abhalten oder zumindest für legitim und zielführend, für „rite et recte“, befinden.

In unserem Zusammenhang, also der Möglichkeit der Steuerung öffentlicher Verwaltung durch Anreize und Bewertungsverfahren, sind meines Erachtens mindestens zwei Aspekte dieser Betrachtungsebene interessant. Erstens: für die öffentliche Verwaltung in Deutschland jedenfalls gilt traditionell, dass nicht nur Entscheidungsstränge, sondern auch die Informationsverarbeitung, und damit das *Wissen*, *hierarchisch* geprägt sind. Wir behandeln daher Informationen und allgemein jedes Wissen streng unterschiedlich danach, woher es stammt, wer es paraphiert bzw. mit- oder freigezeichnet hat – und wofür –, und ob es nicht nur nach seiner inhaltlichen Qualität, sondern auch nach dieser, wenn man so will: formalen, Charakteristik einwandfrei, nützlich, einsetzbar, generell: zielführend ist.

1 Der Autor ist Honorarprofessor und gehört der Ministerialverwaltung Baden-Württembergs an. Er vertritt hier ausschließlich seine private Auffassung.

In der öffentlichen Verwaltung kennen wir – zumindest *offiziell* – keinen „Markt“ für Wissen, ebenso wenig wie für Zuständigkeiten, Zielsysteme, Instrumentenkästen oder Ressourcen sonstiger Art. Die Betonung liegt hier zugegebenermaßen auf „offiziell“. Nichtsdestoweniger entspricht die „Freiheit vom Markt“ unserem Selbstverständnis mindestens insoweit, als das uns umgebende Recht, und traditionell alle Regelmäßigkeit der öffentlichen Verwaltung, ganz wesentlich um diese Charakteristik herum konzipiert sind.

Zweitens: In der öffentlichen Verwaltung in Deutschland pflegen wir zumindest in ihrem überwiegenden Teil einen bestimmten Diskurs, zu dem eine eigene Sprache und ein System der Argumentation gehören, die sich der Dominanz der juristisch ausgebildeten Angehörigen des höheren Dienstes verdanken. Was überzeugt, was als plausibel und allgemein einsichtig verstanden wird, bemisst sich weit überwiegend nach einem Maßstab, den der juristisch gebildete Angehörige der öffentlichen Verwaltung durchschnittlich anlegt. Damit ist freilich weder gesagt, dass es sich hierbei um den allgemeinen Standard juristischen Argumentierens handelt – freiberufliche Rechtsanwälte sind hier nicht in gleicher Weise zu subsumieren – noch dass ausschließlich „juristische Argumentationsfiguren“ Geltung beanspruchen können. Es bedeutet hingegen durchaus, dass etwa sozial- oder wirtschaftswissenschaftliche Argumente ihre Durchsetzungskraft dramatisch steigern können, wenn sie sich mit juristischen *topoi* und Formulierungsstandards verbinden. Insgesamt jedenfalls ergibt sich so eine gewisse *Deutungsmacht* des juristischen Diskurses für die Bewältigung der sachlichen Aufgaben, freilich auch für die Bewertung der Aufgabenerfüllung und der hierzu zum Einsatz kommenden Mittel und Wege.

Nun, was heißt das für unser Thema? Eine öffentliche Verwaltung mit den genannten Charakteristika riskiert gewissermaßen einen „Strukturbruch“, um Lehmbruchs Konzept zu adaptieren, wenn sie zwei Logiken von Wissenssystemen und damit verbundenen Diskursen verknüpft. Das ist keine neue Erkenntnis, wie uns so ziemlich jedes Lehrbuch über „New Public Management“ zeigt. Und die Positionen hierzu sind durchaus umstritten. Jedoch ist etwa das Zitat eines ehemaligen dänischen Finanzministers vor zehn Jahren, wenn nicht in der Diktion, so doch dem Ziele nach, prototypisch: „The public sector must learn to think, act and be managed on the same terms as the private sector. The old bureaucrats must be smoked out!“ (zitiert nach Pollitt/Bouckaert 2004: 35).

Ein Ansatz, der systemimmanente Abläufe im weiten Sinne nach einer *marktimitierenden Logik* optimieren will, die zunächst einmal, oder sagen wir vorsichtiger: jedenfalls gleichzeitig, einer *marktaversen Logik* klassischer Administration folgen und weiter folgen müssen (!), stößt kraft Natur der Sache auf Probleme. Hinter den vielzitierten Bildern des New Public Management, etwa Osborne und Gaebler's „Steuern statt Rudern“, steht eben mehr als nur die ver-

meintliche Gegensätzlichkeit von ineffizienter Verwaltung im Bürokratismus versus effizientem Management mit Dienstleistungscharakter.

Die Fixierung von Zielen, etwa im Rahmen von Zielvereinbarungen, die Festlegung von Meilensteinen und Ressourcenaufwendungen, die Messung von Outputs und Wirkungen, all dieses liegt zunächst gewissermaßen quer zur klassischen Logik der Administration, die ja, und das ist hier aus meiner Sicht entscheidend, *weiterhin normativ Geltung beansprucht* und Beachtung finden muss.

Nun ist eine auf Zielen, Anreizen und Bewertungsverfahren basierende Steuerung großer operativ ausgerichteter Verwaltungen, mit entsprechend standardisierten „Produkten“, sicher nicht nur möglich, sondern auch dem Grunde nach sinnvoll. Auch die öffentliche Verwaltung, namentlich in nachgeordneten Behörden und vor allem auch im kommunalen Bereich, kennt so etwas wie ein „Massengeschäft“, das ebenfalls kraft Natur der Sache eine für (betriebs-)wirtschaftliche Effizienz- und Effektivitätsüberlegungen empfängliche Seite aufweist. Und selbst Kritiker der Übertragung betriebswirtschaftlicher Instrumente der freien Wirtschaft auf die Verwaltung, wie etwa der Rechnungshof Baden-Württemberg in einem Prüfbericht zum dortigen „NSI“ (Neue Steuerungsinstrumente) im Jahr 2006, halten grundsätzlich daran fest, dass für geeignete Verwaltungszüge die Sinnhaftigkeit gegeben sein kann. Dabei hängen Steuerungsverfahren und Organisationsaufbau naturgemäß eng miteinander zusammen (Stuttgarter Zeitung v. 17.10.2006, „Viel Beifall für Kritik des Rechnungshofs an Ämterreform“).

Das Problem beginnt dort, wo eine *Imitation* von solchen Verfahren und Schemata in Verwaltungsbereichen von Statten geht, die die Form über die Wirkung stellt, weil organisatorische Gegebenheiten unberücksichtigt bleiben, rechtliche Bedingungen der Bearbeitung eines Politikfelds, Akteurskonstellationen oder Modi der staatlich-gesellschaftlichen Interferenz Charakteristika aufweisen, die die „Steuerungsfähigkeit“ durch den Staat (bzw. die Verwaltung) jedenfalls mit den Mitteln betriebswirtschaftlicher Konzepte in Frage stellen.

Es entsteht so das Risiko, dass nicht nur informationelle „Parallelwelten“ entstehen, getrennte Wissenssysteme, die doppelten Aufwand und somit indirekt selbst Bürokratismus verursachen, sondern darüber hinaus auch eben jener Strukturbruch, der zu inkompatiblen, jedenfalls konfligierenden Zielen und Maßgaben für deren Umsetzung führt. Es gehört sicher nicht viel Mut dazu als These zu formulieren: Je politiknäher ein Verwaltungsbereich, desto schwieriger die Übernahme der Anreiz- und Bewertungslogik. Politiknahe Verwaltung lebt naturgemäß mit einem permanenten Primat der Politik. Seinem Wesen nach erzeugt dieses Primat der Politik flexible Zielsysteme. Genau betrachtet entsteht also auf diese Weise eine Trias von politischer, administrativer und betriebswirtschaftlicher Logiken, die, wie gesagt: in obersten Bundes- und Landesbehörden, zur Gleichzeitigkeit konfligierender Erfordernisse führen können.

Lassen Sie es mich an einem Beispiel erläutern, das frei erfunden ist. Nehmen wir an, Abteilung X aus dem Landwirtschaftsministerium eines ostdeutschen Bundeslandes plant für 2013 die Förderaktivitäten im sozialen Bereich für den ländlichen Raum, wir nennen es kurz „Sozialprogramm“. Förderempfänger sind unter anderem die Organisationen der „Landfrauen“ sowie der diversen „Landjugenden“. Fördergegenstand sind unter anderem die Bezuschussung der hauptamtlichen Strukturen dieser Organisationen von Landfrauen und Landjugend, ihre gemeinschaftlichen, ökologischen und fachlichen Aktivitäten sowie eine Sympathiekampagne für diese Organisationen.

Abteilung X dieses Landwirtschaftsministeriums schließt nun mit der Leitung des Hauses eine Zielvereinbarung zur Umsetzung dieser Förderaktivitäten „Sozialprogramm“. Deadlines der Umsetzung, Outputindikatoren etwa zu den erreichten Mitgliedern sowie – und das ist schon eher anspruchsvoll – Wirkungsindikatoren etwa zur Medienresonanz auf die Sympathiekampagne werden festgelegt. Hinter diesem soweit recht einfachen Zielvereinbarungsvorhaben stehen selbstverständlich, je nach Sachverhalt und bisheriger Praxis, ggf. komplexe Aufgaben, von der Vorbereitung der Bewilligung von Zuschüssen bis hin zu möglichen vergaberechtlichen Prüfungen.

Im Zeitablauf des betreffenden Haushaltsjahres 2013 wird durch eine Landtagsanfrage die vergaberechtliche Praxis der Regierung im Allgemeinen thematisiert, die Politik ist sensibilisiert und verlangt ab sofort eine besonders intensive Prüfung aller vergaberechtlich relevanten Vorgänge. Um Deadlines etc. weiterhin halten zu können, bedarf es, aufgrund des erhöhten Prüfaufwandes, wie wir einmal annehmen, eines erhöhten Personaleinsatzes. Hier wäre es nun dementsprechend an sich erforderlich, eine dezentrale Ressourcenplanung installiert vorzufinden. Abhängig von der operativen Umsetzung des „Sozialprogramms“, im Ministerium, durch einen beauftragten Dienstleister oder in einer nachgeordneten Behörde, innerhalb des Geschäftsbereichs des Ministeriums oder mit nur fachlicher Aufsicht, und so weiter, entscheidet sich, ob diese Flexibilität zur Steigerung der Vergabeprüfung möglich ist oder nicht.

Zielvereinbarungen definieren häufig nicht Prioritäten zwischen Zielen, so dass etwa eine Ressourcenverschiebung zur Verbesserung der Vergabequalität, etwa über Abteilungsgrenzen hinweg, kein praktikabler Weg sein dürfte. Und selbst innerhalb von Organisationseinheiten ist nicht unbedingt die freie Mobilität von Mitarbeitern gegeben.

Nehmen wir nun des Weiteren an, dass es aus politischen Gründen zu einer Verwerfung mit der evangelischen Landjugend kommt, in deren Folge die notwendigen Schritte auf Seiten des Förderempfängers zunächst unterbleiben, die für die Gewährung des Zuschusses erforderlich sind.

Das Zuwendungsrecht ist nun keineswegs ohne Beurteilungsspielräume, und gerade Förderpolitiken, die ministeriumsunmittelbar durchgeführt oder letztent-

schieden werden, kennen typischerweise hier eher Spielräume als in den nachgeordneten Bereich ausgelagerte Breitenprogramme. So kann es kommen, dass die Verwaltung in einen schwierigen Abwägungsprozess gerät, in dem sie die Zuschussgewährung rechtlich einwandfrei sicherzustellen hat, dabei Beurteilungsspielräume aber sowohl unter dem Gesichtspunkt eines letztlich oft nicht trennscharf formulierten politischen Primats als auch der geschlossenen Zielvereinbarung, die man ihrerseits freilich auch als politische Willensäußerung interpretieren kann, auszuüben hat.

Operative Einzelentscheidungen und konzeptionelle Zielsetzungen sind insofern gerade in der politiknahen Verwaltung nicht notwendig in Gleichklang zu bringen, denn sie laufen gewissermaßen in zwei Geschwindigkeiten. Ein angemessenes Steuerungsverfahren wird insofern etwa bei Zielvereinbarungen mit dem notwendigen Maß an Flexibilität und politischem Common Sense herangehen. Ein Grundkonflikt ist angesichts der verschiedenen Logiken von Politik, klassischem Verwaltungshandeln und betriebswirtschaftlichem Controlling aber latent vorhanden.

Ich will eine weitere These in den Raum stellen: Die Politikverflechtung, die sich aus der Europäisierung eines föderalen Rechtsstaats ergibt, erhöht *innerhalb des Controlling selbst* die Wahrscheinlichkeit *nicht synchroner Anreizsysteme und Bewertungsverfahren*. Lassen Sie mich in aller Abstraktion ein Beispiel benennen, die Durchführung der OPs der Europäischen Strukturfonds. Mit den entsprechenden Vereinbarungen zwischen der EU und den Mitgliedstaaten bzw. der N-2-Ebene wird neben dem einschlägigen europäischen Regelwerk auch das jeweils einschlägige nationale bzw. subnationale Recht, also etwa die Haushaltsordnungen oder das Vergaberecht, als für die Umsetzung verbindlich festgelegt. Doch sind die entsprechenden europäischen und nationalen Regelungen keineswegs durchgängig harmonisiert. So folgen etwa die Strukturfonds einem sieben Jahres-Zyklus und folgerichtig sind die Anreize und Bewertungsverfahren, also Kennziffern zu Outputs und Outcomes, auf die Gesamtperiode ausgerichtet. Gleiches gilt für den Haushaltsvollzug, einschließlich etwa einer n+2-Regel, die einen massiven Anreiz setzt, Programme und Projekte mit hohem, schnellem Mittelabfluss zu priorisieren.

Demgegenüber sind nationale und subnationale Haushalte auf ein oder zwei Jahre ausgerichtet und folgerichtig sind auch Kennzahlen des betriebswirtschaftlichen Haushalts-Controllings auf Jährlichkeit fixiert. Ungeachtet dessen, dass sich nationale Parlamente nicht von der EU die Kennzahlen vorschreiben lassen wollen, die sie für die Bewertung von Haushaltstiteln und Dotierungen zu interessieren haben, ergeben sich hier Friktionen. Zählt etwa ein Controlling-Haushalt für das Parlament die Bewilligungsfälle pro Jahr, so vermittelt die Bewilligung eines Drei-Jahres-Projekts aus einem Europäischen Strukturfonds in diesem Produkthaushalt das Bild einer zweijährigen Untätigkeit (nämlich in den

Jahren zwei und drei). Demgegenüber gibt es im Controllingsystem für die EU kein Problem, da hier andere Messgrößen angelegt werden. Eine weitere Erscheinungsform mangelnder Harmonisierung zwischen den Systemen der Ebenen ist die Unterschiedlichkeit in den Steuerungssensibilitäten. Die haushaltsmäßigen Regeln und betriebswirtschaftlichen Controllingverfahren kennen je ungleiche Schwerpunkte, Flexibilisierungen und Verbindlichkeiten. Sie mögen je in sich stimmig bzw. ausgewogen sein (z.B. hohe Wertgrenzen bei Vergaben, dafür strenge Maßstäbe der Einzelfallkontrolle), in der Kombination wirken sie allerdings additiv. Im Ergebnis wird die Controlling- und Kontrollumgebung von Ressourcen so zu einem besonders maßgeblichen Gesichtspunkt des Programmdesigns, der relative Aufwand der Verwaltung für die Verwaltung steigt.

Ich behaupte an dieser Stelle nicht, die Probleme seien systematisch nicht lösbar. Probleme dieser Art sind allerdings praktische Probleme und die Verknüpfung verschiedener Ebenen im europäisierten Bundesstaat bei der Administration von Fördermitteln setzt bei gegebener Souveränität nationaler (und im Föderalismus auch subnationaler) Parlamente ihrer Behebung Grenzen. Zumindest erzeugt die Friktion von Steuerungssystemen Rechtfertigungsdruck. Im Ergebnis stellt sich mitunter die Frage, ob der Aufwand noch in einem angemessenen Verhältnis zu dem Steuerungserfolg steht.

Lassen Sie mich eine dritte abschließende These formulieren. Vor kurzem habe ich eine Diskussionsrunde mit so genannten „Compliance-Beauftragten“ größerer mittelständischer Unternehmen besucht und deren Erfahrungsaustausch unmittelbar aufgenommen.

Paradoxerweise erlebt die Wirtschaft offenbar zur Zeit eine ähnliche Entwicklung, wie sie die öffentliche Verwaltung mit New Public Management schon seit längerem erfährt, nur in gewisser Weise mit umgekehrten Vorzeichen. Unter dem „Compliance“-Management verbirgt sich eine offene Reihe von Zielen und Anliegen, die sich eben gerade *nicht* aus einer effizienten und effektiven Betriebsführung ergeben, sondern gewissermaßen exogen in die klassischen Zielkoordinaten der freien Wirtschaft eingepflanzt werden müssen. Von ethischen, sozialen oder ökologischen Anliegen, wie der weltweiten Bekämpfung von Kinderarbeit oder von umweltschädlichen Stoffen etwa in Textilien, bis hin zu eher diffus gehaltenen Zielen „guter Unternehmensführung“, von „affirmative action“ und der Bekämpfung von Mobbing bis hin zur Korruptionsprävention, reichen die *topoi*, die von kaufmännisch ausgerichteten Managern sicher als ebenso fremdartig angesehen werden, wie die Sprache des betriebswirtschaftlichen Controllings von einem gediegenen Beamten (allgemein hierzu vergleiche auch Moosmayer 2012).

Aus dieser Diskussionsrunde habe ich drei Lehren mitgenommen, die – wie gesagt: mit umgekehrtem Vorzeichen – auch für die Verwaltung gelten können. Erstens, jede neue Werte und Normen in ein gegebenes System einführen will

(im Falle der Verwaltung: Effizienz, Marktorientierung, verbesserte Zielerfassung), hat es mit einem Kulturwandel zu tun. Einfach ein neues Wissenssystem darüber zu setzen, ist zu wenig. Es droht zur Entstehung einer Parallelwelt, schlimmstenfalls zur Paralyse zu führen, wenn eine hinreichende Abstimmung mit dem weiterhin gültigen Rahmen nicht erfolgt ist. Zweitens, neue Steuerungssysteme und organisatorische Vorkehrungen (wie etwa *Chief Compliance Officers*) können als bloße Exkulpierungsmechanismen für Eigentümer, Öffentlichkeit oder Aufsicht und Justiz zweckentfremdet werden. Ohne den erwähnten Kulturwandel bleiben alle institutionellen Lösungen gewissermaßen hohl. Und drittens, Compliance-Management ist dann sinnvoll, wenn es den Verantwortlichen durch Transparenz hilft, gute Entscheidungen auf ausreichender, breiter, aber eben auch verarbeitbarer Grundlage zu treffen. Die Verantwortung für „gute Unternehmensführung“ kann das Compliance-Management ebenso wenig übernehmen, wie betriebswirtschaftliche Steuerungsverfahren in der öffentlichen Verwaltung die Politik ersetzen können.

Ich komme deshalb zurück zum Anfang: Die Einrichtung einer neuen Sprache, eines neuen Informations- und letztlich Wissenssystems in der öffentlichen Verwaltung kann ohne Rücksicht auf seine Kompatibilität mit dem alten und weiterhin gültigen Wissenssystem, und damit der klassischen Ausrichtung der Verwaltung, einschließlich ihrer Regeln und Normen, einerseits, sowie dem Primat der Politik in der politiknahen Verwaltung andererseits nur schwerlich zu einem allgemeinen Verhaltenswandel in der gewünschten Richtung führen. Sie muss den Kontext von Steuerungsbehörde und Umsetzungsträgern, ihre jeweiligen Autonomieräume, ihre jeweiligen Handlungsrestriktionen durch das Recht und andere Anreiz- und Bewertungssysteme, gerade auch solche der weiteren staatlichen oder quasi-staatlichen Ebenen, bedenken. Und sie sollte sich davor hüten, Prozesse symbolischen Lernens zu vollziehen, bei denen formale Übertragungen anstelle materieller Anwendungen stattfinden. Das ist eine anspruchsvolle Aufgabe, deren Investitionskosten sicher besonders zu Beginn hoch sind.

Die Trias von *politischer Logik*, mit ihrem notwendigen Anspruch auf Flexibilität und Freiheit von Regelgebundenheit, von *klassischer Verwaltungslogik*, mit ihrem notwendigen Anspruch auf Gleichförmigkeit, Abstraktion und einer Ausrichtung an ein komplexes Konzept von Rechtsstaatlichkeit, und schließlich von *betriebswirtschaftlicher Controlling-Logik*, mit ihrem notwendigen Anspruch auf Zielfixierung sowie effektiver und effizienter Mittelwahl, konstituiert eine dreistellige Spannungsbeziehung. Sie schlägt sich nieder in unterschiedlichen Sprach- und Informationssystemen, die gleichzeitig nebeneinander bestehen. Der tatsächliche Umgang mit dieser Gleichzeitigkeit in den Behörden, die es betrifft, namentlich obersten Bundes- und Landesbehörden, erscheint mir weiterer Forschung wert. Dies zumal der Ruf nach besseren und stärkeren Formen der Partizipation auch das Verhältnis von Politik und Verwaltung, insbesondere im Hin-

blick auf Verfahren des Verwaltungshandelns (vergleiche hierzu Stender-Vorwachs 2012), per se einer Neubestimmung unterwerfen muss.

Literatur

Moosmayer, Klaus: Modethema oder Pflichtprogramm guter Unternehmensführung? Zehn Thesen zu Compliance, in: NJW 41/2012, S. 3013-3017.

Pollitt, Christopher/ Bouckaert, Geert: Public Management Reform. A Comparative Analysis. Oxford: Oxford UP 2004

Schaal, Gary S./ Heidenreich, Felix: Einführung in die politischen Theorien der Moderne. Opladen: Budrich 2011

Stender-Vorwachs, Jutta: Neue Formen der Bürgerbeteiligung?, in: NVwZ 17/2012, S. 1061-1066.

Variabler Leistungslohn in der Verwaltung – Allheilmittel oder Medikament mit starken Nebenwirkungen? *

Antoinette Weibel

Im Zuge des Anspruchs, öffentliche Verwaltungen nach betriebswirtschaftlichen Kriterien zu führen, wird in regelmäßigen Abständen die Forderung nach leistungsorientierter Belohnung laut. Meist wird mit der Modernisierung der Lohnsysteme die Hoffnung verbunden, die Produktivität der Verwaltungsmitarbeiter durch variablen Leistungslohn zu steigern und gleichsam die Kosten eines traditionell nach dem Senioritätsprinzip ausgerichteten Lohnsystems zu senken. Allerdings waren die bisherigen Reformbemühungen von wenig Erfolg gekrönt – die Möglichkeit zur leistungsorientierten Entlohnung wird nur zögerlich genutzt und in vielen Fällen werden solche Neuerungen schnell wieder eingestellt. Dieser Beitrag zeigt auf, warum leistungsvariable Entlohnung für die Verwaltung nur beschränkt geeignet ist, welche kostspieligen Nebenwirkungen entstehen, wenn Leistungslohn unreflektiert zum Einsatz kommt und welche Alternativen zur Erhöhung der Leistungsmotivation von Mitarbeitern bestehen.

1. Variabler Leistungslohn in der Theorie und in der Praxis

Als Inbegriff moderner Managementmethoden geltend hat sich das Prinzip der leistungsabhängigen Vergütung („pay for performance“) in vielen privatwirtschaftlichen Unternehmen durchgesetzt. Zur variablen Leistungsentlohnung zählen alle Lohnbestandteile, die an die Erreichung vorab definierter Ziele und Standards geknüpft sind und die im Zeitverlauf je nach Leistung variieren. Klassische Instrumente dieses Lohnsystems sind Akkordlöhne, Leistungsprämien und Bonussysteme. Variabler Leistungslohn setzt auf den „relativen Preiseffekt“. Er geht davon aus, dass ein höherer Preis ceteris paribus immer ein erhöhtes Angebot hervorruft, d.h., dass das Versprechen eines höheren Lohns zwangsläufig zu einer höheren Leistung führt. Dieser Effekt hat sich bei einfachen Tätigkeiten in

* Dieser Artikel beruht auf dem Beitrag von Margit Osterloh und Antoinette Weibel. "Wissenschaftliche Beiträge – Managing Motivation – Verdrängung und Verstärkung der intrinsischen Motivation aus Sicht der psychologischen Ökonomik", in: *WiSt Wirtschaftswissenschaftliches Studium* 37.8 (2008): 406. Aus Gründen der Lesbarkeit wurde darauf verzichtet jede wortwörtliche übernommene Passage zu kennzeichnen.

Form des Akkordlohns durchaus bestätigt. Ein viel zitiertes Beispiel ist das US-amerikanische Unternehmen Safelite Glass (Lazear, 2000): Nach der Umstellung von fixen Stundenlöhnen auf Stücklöhne stieg die Produktivität um erstaunliche 36% (Anreizeffekt: 20% und Selektionseffekt: 16%), während die Lohnkosten nur um 9% stiegen. Zudem wird erwartet, dass leistungsvariabler Lohn die Aufmerksamkeit der Mitarbeiter steuert: Mitarbeiter sollen genau die Tätigkeiten ausführen, die das Unternehmen von ihnen erwartet. Werden eben diese Tätigkeiten leistungsabhängig bezahlt, macht Leistungslohn das Mitarbeiterverhalten vorhersehbarer.

In jüngerer Zeit wurde das Prinzip des Akkordlohnes zunehmend auf alle Beschäftigungsverhältnisse übertragen, z.B. auf das mittlere und obere Management in Unternehmen (z.B. Bebchuk & Grinstein, 2005), in Behörden (z.B. Bertelli, forthcoming; Schneider, 2007) und im Rahmen der neuen W-Besoldung sogar auf Universitäten (Osterloh & Frey, 2007). Die Einführung leistungsvariabler Entlohnung in der Verwaltung wird jedoch heftig diskutiert. Die Reformen der Vergütungssysteme der öffentlichen Verwaltung sollen die Effizienz des öffentlichen Dienstes erhöhen, die Leistungsbereitschaft und Motivation der Angestellten positiv beeinflussen und eine stärkere Leistungsorientierung in der Verwaltung verankern. So empfiehlt eine Studie der Bertelsmann-Stiftung die Einführung leistungsvariabler Anreize in Deutschland, weil monetäre Anreize ein wichtiges Instrument seien, die Ernsthaftigkeit neuer leistungsorientierter Steuerungssysteme der Verwaltung zu unterstreichen (Bertelsmann Bertelsmann Stiftung, 2006, S. 33). Allerdings zeigt ein OECD-Bericht aus dem Jahr 2005, dass, obschon 2/3 der OECD-Mitgliedstaaten leistungsvariable Lohnkomponenten einsetzen, der Erfolg dieser Systeme auszubleiben scheint und folgert "there is no conclusive empirical evidence that such an approach has effectively helped to improve motivation and performance within the public service". Es stellt sich also insgesamt die Frage, ob die positiven Effekte leistungsvariabler Vergütung auch auf Tätigkeiten übertragbar sind, die sich fundamental von der typischen Akkordarbeit unterscheiden. Tätigkeiten im Management privater Unternehmen oder im gehobenen und höheren öffentlichen Dienst sind typischerweise durch einen großen Handlungsspielraum gekennzeichnet, schwer mess- und kontrollierbar und setzen deshalb ein hohes Ausmaß an Eigeninitiative und "Public Service Motivation" voraus (Osterloh, 2006; Osterloh & Frey, 2000).

Die universelle Gültigkeit des relativen Preiseffekts wird zudem schon seit längerem von Befunden der Sozialpsychologie relativiert, was einmal mehr die unkritische Rezeption privatwirtschaftlicher Instrumente wie leistungsvariable Löhne in Frage stellt. Studien zeigen, dass eine variable, leistungsabhängige Entlohnung die Leistung unter bestimmten Umständen mindert (Deci, 1985). Dies wird als „die verborgenen Kosten der Entlohnung“ (Lepper & Greene, 1978), oder als Korruptionseffekt (Kruglanski, 1975) bezeichnet. In der Ökono-

mik wurde dieser Effekt von Bruno Frey (1992; 1997) als sog. „Verdrängungseffekt“ eingeführt (zum Verdrängungs- und Verstärkungseffekt siehe auch Osterloh & Weibel, 2008).

Im Folgenden wird erläutert, warum leistungsvariable Löhne einen Verdrängungseffekt auslösen können. Dazu werden zuerst verschiedene Formen der Leistungsmotivation unterschieden. In einem zweiten Schritt wird die gegenteilige Wirkung von Preis- und Verdrängungseffekt aufgezeigt sowie erklärt, unter welchen Bedingungen leistungsvariable Entlohnung einen negativen Effekt auf die Mitarbeiterleistung ausübt. Schließlich werden Alternativen zum leistungsvariablen Lohn diskutiert.

2. *Arten der Motivation*

Motivation (lateinisch *movere* = bewegen) bezeichnet die Beweggründe des Handelns. Motivation verleiht dem Handeln Richtung, Stärke und Ausdauer im Hinblick auf die Befriedigung von Bedürfnissen (Heckhausen, 2006). Im Hinblick auf die Beeinflussbarkeit der Motivation ist es sinnvoll zwischen extrinsischer und intrinsischer Motivation zu unterscheiden. Extrinsische Motivation ist auf von außerhalb der Person kommende Anreize gerichtet (Belohnung oder Bestrafung). Diese ermöglichen eine mittelbare Bedürfnisbefriedigung, vor allem durch Geld. Intrinsische Motivation ist auf die unmittelbare Bedürfnisbefriedigung gerichtet, d.h. auf Aktivitäten, die um ihrer selbst willen ausgeführt werden. Intrinsische Motivation kann einerseits hedonisch, d.h. auf das Erleben von Freude und Spaß gerichtet sein. In diesem Fall vermittelt die Tätigkeit als solche ein sog. „Flow-Erlebnis“ (Csikszentmihalyi, 1991). Intrinsische Motivation kann andererseits auch prosozial sein, also auf das Wohlbefinden anderer Menschen abzielen. Zahlreiche empirische Befunde belegen, dass ein solches intrinsisch-prosoziales Handeln in hohem Ausmaß stattfindet (Meier, 2006). Dabei können wiederum zwei verschiedene Formen unterschieden werden: Altruismus bzw. unkonditional-prosoziale Motivation und Reziprozität bzw. konditional-prosoziale Motivation. Dabei ist Reziprozität nicht mit „Tausch“ gleichzusetzen. Es geht dabei nicht um den eigenen Nutzen und eine sofort erwartete Gegenleistung, sondern um eine freiwillige Gabe als Ausdruck einer sozialen Norm, die zwischen Altruismus und Anpassung angesiedelt ist (Göbel, Ortman, & Weber, 2006). Der Gebende geht davon aus, dass seine Handlung aufgrund sozialer Normen zum einen wertgeschätzt und zum andern durch normgerechtes Verhalten anderer bei Zeiten zurückerstattet wird. In Laborexperimenten zu prosozial motiviertem Verhalten zeigen sich unterschiedliche Anteile der vorherrschenden Motivation, jedoch lassen sich etwa 50 Prozent der Teilnehmenden als „Reziprokatoren“, 20 Prozent als Altruisten und 30 Prozent als Egoisten charakterisieren

(vgl. z.B. Andreoni & Miller, 2002; Fischbacher, Fehr, & Gächter, 2001). Feldexperimente zeigen einen geringeren Anteil an Altruisten (Frey & Meier, 2004).

Die Verwaltungswissenschaft geht davon aus, dass Bedienstete im öffentlichen Sektor im Schnitt hoch prosozial motiviert sind. Die verwaltungstypische Arbeitsmotivation wird auch als "Public Service Motivation" bezeichnet. Öffentliche Angestellte sehen in ihrer Tätigkeit häufig die Chance, ihr soziales Mitgefühl, die Orientierung am Gemeinwohl und öffentlichen Nutzen sowie ein genuines Interesse an Politik und Politikberatung auszuleben (Perry & Wise, 1990). Eine österreichische Untersuchung zeigt, dass Public Service Motivation auch in legalistischen Verwaltungsstrukturen wie in Deutschland und Österreich eine große Rolle spielt (Hammerschmid, Meyer, & Egger-Peitler, 2009). Die hedonisch-intrinsische Motivation ist ebenfalls stark ausgeprägt: 95% aller Befragten halten Selbständigkeit bei der Erledigung der Arbeit für sehr oder weitgehend wichtig und fast ebenso viele (94,7%) empfinden das Ausüben einer interessanten Tätigkeit als motivierend (Hammerschmid, et al., 2009). Zusammenfassend spielen beide Formen der intrinsischen Motivation für Verwaltungsangestellte eine maßgebliche Rolle.

3. Management intrinsischer Motivation: Verdrängungs- und Verstärkungseffekt

Extrinsische und intrinsische Motivation sind nicht additiv. Vielmehr wirken die beiden Motivationsarten in vorhersagbarer Weise aufeinander ein (Frey, 1997). Der Verdrängungseffekt besagt, dass die intrinsische Motivation für eine Tätigkeit durch extrinsische Belohnung (resp. Bestrafung) verdrängt wird. Spricht man vom sog. „Verstärkungseffekt“, ist die Verstärkung der intrinsischen Motivation für eine Tätigkeit durch bestimmte (extrinsische und intrinsische) Anreize gemeint.

3.1 Der Verdrängungseffekt

Variable, leistungsbezogene Entlohnung führt unter drei Voraussetzungen zu einer Reduktion von Arbeitsanstrengungen: (1) Die Tätigkeit war ursprünglich intrinsisch motivierend. (2) Die Entlohnung wird als kontrollierend empfunden (Deci, Koestner, & Ryan, 1999; Frey & Jegen, 2001). (3) Die durch Entlohnung erzeugte extrinsische Motivation gleicht den Verlust an intrinsischer Motivation nicht aus (Weibel, Rost, & Osterloh, im Druck). Mittlerweile konnten durch eine Vielzahl an Laborexperimenten genauere Einflussfaktoren des Verdrängungseffektes ermittelt werden: Der Verdrängungseffekt ist bei erwarteten Belohnun-

gen stärker als bei unerwarteten sowie bei monetären Entlohnungen stärker als bei symbolischen (Heckhausen, 2006). Er ist bei komplexen Tätigkeiten stärker als bei Routineaufgaben (Weibel, Rost, & Osterloh, 2007). Weiterhin wirkt er sich bei Beiträgen zu Kollektivgütern aus, die in der Öffentlichkeit erbracht werden, nicht aber immer bei privat erbrachten Kollektivgütern (Ariely, Bracha, & Meier, 2007). Neben den Laborexperimenten belegen und konkretisieren auch Feldexperimente den Verdrängungseffekt (vgl. z.B. Frey, Eichenberger & Oberholzer-Gee, 1996; Frey & Götte, 1999). Es konnte gezeigt werden, dass der Verdrängungseffekt dazu beiträgt, dass „pay for performance“ im Top-Management im Durchschnitt nicht zu einer Erhöhung sondern zu einer Verminderung der Unternehmensleistung führt (Rost & Osterloh, 2007). Der Verdrängungseffekt wurde mittlerweile auch durch Meta-Analysen bestätigt (vgl. vor allem Deci, et al., 1999) – allerdings stammen die Ergebnisse der verwendeten Primärstudien nur selten direkt aus dem Arbeitskontext.

Diese Lücke schließend zeigen Weibel, Rost und Osterloh (2007) in einer Meta-Analyse verschiedener Labor- und Feldexperimente in berufsnahen Situationen, dass der Verdrängungseffekt auch dort nachzuweisen ist. Konkret konnten sie folgende Zusammenhänge für den privatwirtschaftlichen Arbeitskontext identifizieren:

- Leistungsvariabler Lohn wirkt sich positiv auf die Mitarbeiterleistung aus, wenn langweilige Tätigkeiten im Fokus stehen.
- Leistungsvariabler Lohn wirkt sich negativ auf die Mitarbeiterleistung aus, wenn spannende oder am Gemeinwohl orientierte Tätigkeiten durch den zusätzlichen Anreiz angeregt werden sollen.

Weiterhin zeigen Studien, dass Preis- und Verdrängungseffekt gegenläufig wirken. Das Feldexperiment von Gneezy und Rustichini (2000) ist hierfür ein viel zitiertes Beispiel: Die Autoren untersuchen den Einfluss monetärer Anreize auf das ehrenamtliche Spendensammelverhalten von 180 Schülerinnen und Schülern. Diese wurden in drei Gruppen eingeteilt: Die erste Gruppe erhielt keine Erfolgsprämie, die zweite Gruppe erhielt 1% und die dritte Gruppe 10% des Sammelergebnisses. Die mit 1% beteiligte Gruppe sammelte deutlich weniger als jene, die nichts erhielt. Die dritte Gruppe sammelte mehr als die zweite Gruppe, blieb allerdings hinter dem Sammelergebnis der ersten Gruppe zurück. Anhand dieses Feldexperiments lässt sich die gegenläufige Wirkung von Preis- und Verdrängungseffekt aufzeigen (Frey & Osterloh, 2002): Gemäß dem Preiseffekt steigert ein Bonus die Arbeitsanstrengung (vgl. Abb. 1). Ohne Bonus – d.h. intrinsisch motiviert – sind die Kinder zu einem Arbeitsengagement A1 bereit. Solange kein Verdrängungseffekt wirkt, d.h. die intrinsische Motivation konstant bleibt, steigert ein Bonus B die Arbeitsanstrengung von A1 auf A2.

Abb. 1: Der Preiseffekt

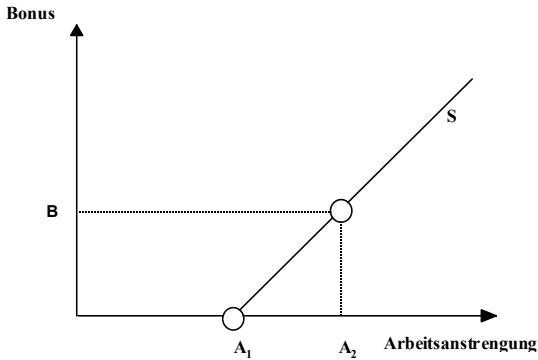
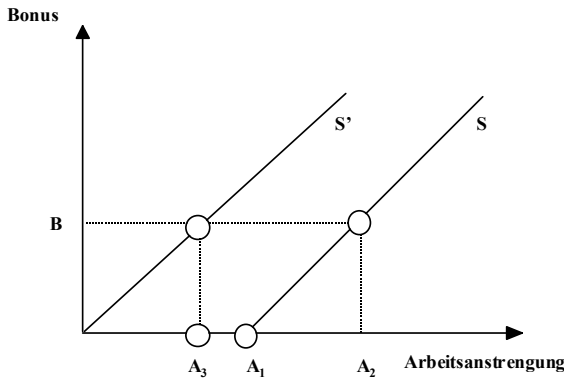


Abb. 2: Nettoeffekt zwischen Preis- und Verdrängungseffekt



Gemäß dem Verdrängungseffekt senkt ein Bonus die Arbeitsanstrengung sobald er als kontrollierend empfunden wird. Die Angebotskurve verschiebt sich von S nach S' (Abb. 2). Die intrinsische Motivation sinkt; im Beispiel in Abbildung 2 sinkt sie von A_1 auf 0. Im Ergebnis verringert sich die Arbeitsanstrengung auf A_3 . Ist der Verdrängungseffekt weniger stark, d.h., wird die intrinsische Motivation nicht auf null gesenkt, kann sich – je nach Stärke des Verdrängungseffekts – ein geringerer Abfall der Arbeitsanstrengung oder sogar eine leichte – aber teuer bezahlte – Erhöhung ergeben. Die Autoren Gneezy und Rustichini (2000) folgern daraus: „Pay enough or don’t pay at all“.

3.2 Zwischenfazit

Vor dem Hintergrund dieser Erkenntnisse lässt sich vermuten, dass Leistungslohn gerade im öffentlichen Sektor schlechte Erfolgsaussichten hat. Auf diesem Hintergrund kann man erklären, weshalb die im Rahmen des Neuen Steuerungsmodells eingeführte Leistungsorientierung in der Personalpolitik (konkretisiert durch an Zielvereinbarungen geknüpfte Boni) bislang nicht die gewünschte Wirkung erzielt: Zum einen scheinen viele Verwaltungsangestellte über eine hohe intrinsisch-prosoziale Motivation zu verfügen – es steht also viel auf dem Spiel. Zum anderen können die durch leistungsvariable Vergütung gesetzten Anreize den Verlust an intrinsischer Motivation häufig nicht ausgleichen, weil der öffentlichen Verwaltung enge finanzielle Grenzen gesetzt sind. Entsprechend ist mit einem starken Verdrängungs- und einem schwachen Preiseffekt zu rechnen.

Auch ist fragwürdig, ob die gesetzten Anreize die zentralen Arbeitsaufgaben erfassen. Tätigkeiten im gehobenen und höheren Dienst sind häufig durch Komplexität gekennzeichnet. Vielfach sind wichtige Merkmale der Arbeitserfüllung nicht messbar. So kann man beispielsweise die Summe eingeworbener Forschungsgelder von Professoren erfassen, während die gesellschaftliche Relevanz der Forschung kaum messbar ist. Belohnt man die einfacher messbaren Aufgaben, stellt sich eine Aktivitätenverlagerung ein: Den messbaren Aufgaben wird viel mehr Aufmerksamkeit geschenkt als den schlecht messbaren. In der Folge kann die Situation entstehen, die Stephen Kerr (1975) als "folly of rewarding A while hoping for B" bezeichnet: durchaus wichtige Aufgaben werden vernachlässigt, weil die Messbarkeit fehlt.

Aus diesen Gründen tut die öffentliche Verwaltung gut daran, die intrinsische Motivation der Mitarbeiter zu stärken und mit leistungsvariablen Löhnen behutsam umzugehen. Die Literatur zum Verstärkungseffekt zeigt auf, wie eine solche Stärkung der intrinsischen Motivation erreicht werden kann.

3.3 Der Verstärkungseffekt

Der Verstärkungseffekt ist weit weniger erforscht als der Verdrängungseffekt. Jedoch zeigen zahlreiche Befunde, dass verschiedene institutionelle Maßnahmen einen positiven Einfluss auf die intrinsische Motivation und damit auf die Arbeitsleistung haben.

3.3.1 Verstärkung durch Partizipation

Partizipation, d.h. Mitsprache in der Arbeit erhöht das intrinsisch motivierte Arbeitsengagement (Frey & Osterloh, 2002, S. 219) und neutralisiert zudem den Verdrängungseffekt variablen, leistungsabhängigen Einkommens (Benz & Frey, 2007). So haben Selbstständige etwa bei gleicher Einkommenshöhe einen höheren intrinsischen Nutzen aus ihrer Arbeit als Nichtselbstständige. Sie messen ihrer Selbstständigkeit einen Prozessnutzen bei, der abhängig vom Ausmaß der Selbstbestimmung ist, mit dem dieses Einkommen erzielt wurde (Benz & Frey, 2007). Ein zusätzlicher Beleg für die Wirkung von Partizipation auf prosozial-intrinsische Motivation stammt aus dem politischen Kontext: Teilhabe an der politischen Willensbildung verringert nachweislich das Ausmaß an Steuerhinterziehung (Feld & Frey, 2002; Frey & Torgler, 2007).

Reformbestrebungen der öffentlichen Verwaltung eröffnen die Möglichkeit, Angestellte vermehrt in Entscheidungsverfahren einzubeziehen. Gerade im Rahmen von Zielvereinbarungsprozessen – deren Einführung ein Kernelement vieler New Public Management ist – können Mitarbeiter in die Entscheidungsfindung eingebunden werden. Verwaltungswissenschaftliche Forschung zeigt, dass sich Motivation und Arbeitszufriedenheit der Angestellten erhöht, wenn Vorgesetzte einerseits einen partizipativen Führungsstil pflegen und andererseits Mitarbeiter in Zielvereinbarungsprozesse einbeziehen (Kim, 2005). Zudem kann die Motivation der Mitarbeiter gestärkt werden, indem sie an der Neuorganisation von Aufgaben und Prozessen mitwirken. Darüber hinaus kann eine partizipative Organisationsentwicklung in der Verwaltung die Wahrnehmung von Bürokratie nachhaltig ändern: anstatt sie als einengendes, stahlhartes Gehäuse zu begreifen, können Strukturen und Regeln als ermöglichend und Chancen öffnend gesehen werden (Adler & Borys, 1996). Eine solche positive Bürokratiewahrnehmung erhöht ebenfalls die intrinsische Motivation von Mitarbeitern im öffentlichen Dienst – ihre Public Service Motivation (Moynihan & Pandey, 2007; Sauer, 2013).

3.3.2 Verstärkung durch prozedurale Fairness

Die Fairnessforschung hat gezeigt, dass durch das Sicherstellen von Fairness mit relativ geringem Aufwand eine große Wirkung auf die Mitarbeitermotivation erzielt werden kann. Lind und Tyler, zwei der führenden Forscher auf diesem Gebiet, formulieren es so: *„[...] die Investition in gerechte Verfahren gehört wohl zu den am wenigsten teuren Methoden zur Verbesserung der Verhaltensweisen, des Zusammenhaltes und der Pflichterfüllung in einer Organisation. Dies spricht sehr dafür, prozedurale Fairness zu realisieren“* (Lind & Tyler,

1988, S. 201). Umfangreiche empirische Untersuchungen belegen, dass ein Verfahren dann als fair betrachtet wird, wenn es folgende Merkmale aufweist (Greenberg, 1996; Leventhal, 1976):

1. Konsistenz: Es wird für alle Personen und in allen vergleichbaren Situationen dasselbe Verfahren angewandt.
2. Unparteilichkeit: Die Stelle, die die Entscheidung über eine Verteilung fällt, ist unvoreingenommen.
3. Korrigierbarkeit: Es kann Einspruch erhoben werden, wenn gute Gründe dafür vorliegen.
4. Genauigkeit: Die Entscheidung wird auf Grundlage aller relevanten Informationen gefällt.
5. Partizipation: Die vom Entscheid betroffenen Mitarbeiter haben eine Mitsprachemöglichkeit.

Empirische Befunde zeigen, dass prozedurale Fairness dazu führt, dass zum einen die Ziele des Unternehmens und zum anderen die Vorgesetzten und Kollegen mehr unterstützt werden (Tyler & Blader, 2003). Somit stärken faire Verfahren die prosoziale Motivation der Mitarbeitenden. Zwei aktuelle Studien – ein Laborexperiment und eine Befragung von Mitarbeitern verschiedener Branchen – demonstrieren, dass sich prozedurale Fairness positiv auf die intrinsische Motivation auswirkt und dadurch auch die Arbeitsleistung der Mitarbeiter verbessert (Zapata-Phelan, Colquitt, Scott, & Livingston, 2009).

Weiterhin belegen zahlreiche wissenschaftliche Untersuchungen, dass prozedurale Fairness auch das Wohlbefinden und die Leistungswilligkeit von Verwaltungsmitarbeitern positiv beeinflusst (für einen Überblick vgl. Rubin, 2009). Wie sich die Verwaltungsmodernisierung auf die wahrgenommene prozedurale Fairness auswirkt ist noch unklar, scheint jedoch abhängig von der spezifischen Umsetzung zu sein. Eine Lichtung des "Regeldickichts" etwa kann nur dann zu erhöhter Mitarbeitermotivation führen, wenn Änderungen unter Einbezug der betroffenen Mitarbeiter stattfinden und sich Mitarbeiter weiterhin fair und gleichwertig behandelt fühlen. Allerdings stehen Studien zu diesem Zusammenhang noch aus.

3.3.3 Verstärkung durch normative geprägte Entscheidungskontexte

Individuen tragen mehr zu Gemeinschaftsgütern bei, wenn der Entscheidungskontext anzeigt, dass prosoziales Verhalten erwartet wird (Lindenberg, 2006). So ergab in einem Laborexperiment dieselbe Versuchsanordnung unterschiedliche Ergebnisse, je nachdem, ob sie als „Gemeinschafts-Spiel“ oder als „Wall-Street-Spiel“ gekennzeichnet wurde. Im ersten Fall trugen ca. 70 Prozent zu einem Gemeinschaftsgut bei, im zweiten Fall nur etwa 30 Prozent. Selbst subtile

Hinweise auf die soziale Eingebundenheit der Akteure genügen, um prosoziales Verhalten zu aktivieren (Haley & Fessler, 2005). Untersuchungen in Unternehmen zeigen, dass die bloße Erwartung einer ethischen Grundhaltung zu vermehrten prosozialen Verhaltensweisen führen. So verhalten sich Mitarbeiter nachweislich verantwortungsvoller, wenn Unternehmen einen schriftlichen Moralkodex verfassen (Adams, Tashchian, & Shore, 2001). Dieser Effekt tritt sogar dann ein, wenn die Mitarbeitenden nicht wissen, was genau in diesem Kodex verankert ist.

Wright (2007) untersucht den Zusammenhang zwischen Betriebskodex im Sinne einer klaren Vision und Motivation in der Verwaltung und kann nachweisen, dass sich eine solche Vision positiv auf die Public Service Motivation der Staatsangestellten auswirkt. In einer Folgestudie zeigen Wright und Ko-Autoren auf, wie Vision und Unternehmensziele noch stärker verankert werden können: Vorgesetzte, die inspirierend und zielgerichtet führen, stärken die Bedeutung der organisationalen Ziele für Mitarbeiter und beeinflussen deren Public Service Motivation positiv (Wright, Moynihan, & Pandey, 2012).

4. Fazit

Traditionell beruht der Arbeitsethos von Mitgliedern der öffentlichen Verwaltung auf intrinsischer Motivation und dem Bedürfnis, dem öffentlichen Wohl zu dienen. Reformen im Zuge der Verwaltungsmodernisierung können diese Form der Motivation schwächen oder stärken. Besonders problematisch, gerade für komplexe und interessante Tätigkeiten, ist dabei die Einführung leistungsvariabler Löhne. Diese können intrinsische und gemeinwohlorientierte Motivation verdrängen. Hingegen stellen sich positive Effekte der Reformen ein, wenn die Mitarbeiter der Verwaltung mehr Mitwirkungsmöglichkeiten erhalten, Regelsysteme gelichtet und, durch Anpassung an die Bedürfnisse der modernen Verwaltung, als fair wahrgenommen werden sowie wenn Führung in der Verwaltung partizipativ, ziel- und wertbezogen erfolgt.

Literatur

- Adams, J. S., Tashchian, A., & Shore, T. H. (2001). Codes of ethics as signals for ethical behavior. *Journal of Business Ethics*, 29(3), 199-211.
- Adler, P. S., & Borys, B. (1996). Two Types of Bureaucracy: Enabling and Coercive. *Administrative Science Quarterly*, 41(1), 61-89.
- Andreoni, J., & Miller, J. (2002). Giving according to garp: An experimental test of the consistency of preferences for altruism. *Econometrica*, 70(2), 737-753.
- Ariely, D., Bracha, A., & Meier, S. (2007). Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially. Paper presented at the SSRN eLibrary.
- Bebchuk, L., & Grinstein, Y. (2005). The growth of executive pay. *Oxford Review of Economic Policy*, 21(2), 283-303.
- Benz, M., & Frey, B. S. (2007). Being Independent is a Great Thing: Subjective Evaluations of Self-Employment and Hierarchy. *Economica*, online early access.
- Bertelli, A. M. (forthcoming). Motivation Crowding and the Federal Civil Servant: Evidence from the US Internal Revenue Service. *International Public Management Journal*.
- Bertelsmann Stiftung. (2006). *EU-Review: Vergütung im öffentlichen Dienst*. Bertelsmann Stiftung.
- Csikszentmihalyi, M. (1991). *Flow : The Psychology of Optimal Experience*. New York: Harper Perennial.
- Deci, E. L. (1985). *Intrinsic Motivation and Self-Determination in Human Behavior*. New York: Plenum Press.
- Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, 125, 627-668.
- Feld, L., & Frey, B. S. (2002). Trust breeds trust. *Economics and Governance*, 3(2), 87-99.
- Fischbacher, U., Fehr, E., & Gächter, S. (2001). Are People Conditionally Cooperative? Evidence from Public Good Experiments. *Economic Letters*, 71(3), 397-404.
- Frey, B. S. (1992). Tertium Datur – Pricing, Regulating and Intrinsic Motivation. *Kyklos*, 45(2), 161-184.
- Frey, B. S. (1997). *Not Just for the Money: An Economic Theory of Personal Motivation*. Cheltenham, UK/Brookfield, USA: Edward Elgar.
- Frey, B. S., Eichenberger, R., & Oberholzer-Gee, F. (1996). The Old Lady Visits Your Backyard: A Tale of Morals and Markets. *Journal of Political Economy*, 104, 193-209.
- Frey, B. S., & Götte, L. (1999). *Does Pay Motivate Volunteers?* University of Zurich, Institute for Empirical Research in Economics, (7). Zurich.
- Frey, B. S., & Jegen, R. (2001). Motivation crowding theory: A survey of empirical evidence. *Journal of Economic Surveys*, 15(5), 589-611.
- Frey, B. S., & Meier, S. (2004). Social comparisons and pro-social behavior: Testing "conditional cooperation" in a field experiment. *American Economic Review*, 94(5), 1717-1722.
- Frey, B. S., & Osterloh, M. (Eds.). (2002). *Managing Motivation. Wie Sie die neue Motivationsforschung für Ihr Unternehmen nutzen können* (2. Auflage ed.). Wiesbaden: Gabler.
- Frey, B. S., & Torgler, B. (2007). Tax morale and conditional cooperation. *Journal of Comparative Economics*, 35(1), 136-159.

- Gneezy, U., & Rustichini, A. (2000). Pay enough or don't pay at all. *Quarterly Journal of Economics*, 115(3), 791-810.
- Göbel, M., Ortmann, G., & Weber, C. (2006). *Geben und Nehmen – Reziprozität – State of the Art*. Arbeitspapier der Universität Bundeswehr Hamburg.
- Greenberg, J. (1996). *The Quest for Justice on the Job*. Thousand Oaks: Sage.
- Haley, K. J., & Fessler, D. M. T. (2005). Nobody's watching? Subtle cues affect generosity in an anonymous economic game. *Evolution and Human Behavior*, 26(3), 245-256.
- Hammerschmid, G., Meyer, R. E., & Egger-Peitler, I. (2009). Das Konzept der Public Service Motivation, Status Quo der internationalen Diskussion und erste empirische Evidenzen für den deutschsprachigen Raum. *der moderne staat, Zeitschrift für Public Policy, Recht und Management*, 2(1).
- Heckhausen, J. (2006). *Motivation und Handeln* (3., überarb. und aktual. Aufl. ed.). Berlin: Springer.
- Kerr, S. (1975). On the Folly of Rewarding A, while Hoping for B. *Academy of Management Journal*, 18, 769-783.
- Kim, S. (2005). The role of trust in the modern administrative state: an integrative model. *Administration & Society*, 37(5).
- Kruglanski, A. W. (1975). Endogenous Exogenous Partition in Attribution Theory. *Psychological Review*, 82(6), 387-406.
- Lazear, E. P. (2000). Performance pay and productivity. *American Economic Review*, 90(5), 1346-1361.
- Lepper, M. R., & Greene, D. (1978). The hidden costs of reward: new perspectives on the psychology of human motivation. Hillsdale, NJ: Erlbaum.
- Leventhal, G. S. (1976). The Distribution of Rewards and Resources in Groups and Organizations. *Advances in Experimental Social Psychology*, 9, 91-131.
- Lind, E. A., & Tyler, T. R. (1988). *The Social Psychology of Procedural Justice*. New York: Plenum.
- Lindenberg, S. (2006). Prosocial Behavior, Solidarity and Framing Processes. In D. Fetschenhauer, A. Flache, B. Buunk & S. Lindenberg (Eds.), *Solidarity and Prosocial Behavior: An Integration of Sociological and Psychological Perspectives* (pp. 23-44). New York: Springer.
- Meier, S. (2006). The economics of non-selfish behaviour : decisions to contribute money to public goods. Cheltenham, UK ; Northampton, Mass.: Edward Elgar Pub.
- Moynihan, D. P., & Pandey, S. K. (2007). The role of organizations in fostering public service motivation. *Public Administration Review*, 67(1), 40-53.
- Osterloh, M. (2006). Human Resources Management and Knowledge Creation. In I. Nonaka & I. Kazuo (Eds.), *Handbook of Knowledge Creation* (pp. 158-175). Oxford: Oxford University Press.
- Osterloh, M., & Frey, B. (2007, 21. Juli 2007). Die Krankheit der Wissenschaft, *Frankfurter Allgemeine Zeitung*, p. 13.
- Osterloh, M., & Frey, B. S. (2000). Motivation, Knowledge Transfer, and Organizational Forms. *Organization Science*, 11(5), 538-550.
- Osterloh, M., & Weibel, A. (2008). Managing Motivation – Verdrängung und Verstärkung der intrinsischen Motivation aus Sicht der psychologischen Ökonomik. *WIST*, 37(8), 406-411.
- Perry, J., & Wise, L. (1990). The motivational bases of public service. *Public Administration Review*, 50(3), 367-373.

- Rost, K., & Osterloh, M. (2007). *Management Fashion Pay-for-Performance*. Paper presented at the SSRN eLibrary.
- Rubin, E. V. (2009). The role of procedural justice in public personnel management: Empirical results from the Department of Defense. *Journal of Public Administration Research and Theory*, 19(1), 125-143.
- Sauer, H. (2013). *Formalization and Happiness at Work*. Diss., Univ. Liechtenstein, Vaduz.
- Schneider, M. (2007). Zielvorgaben und Organisationskultur. Eine Fallstudie. *Die Betriebswirtschaft* (6), 619-637.
- Tyler, T. R., & Blader, S. L. (2003). The group engagement model: Procedural justice, social identity, and cooperative behavior. *Personality and Social Psychology Review*, 7(4), 349-361.
- Weibel, A., Rost, K., & Osterloh, M. (2007). Crowding-Out of Intrinsic Motivation–Opening the Black Box.
- Weibel, A., Rost, K., & Osterloh, M. (im Druck). Disziplinierung der Agenten oder Crowding-out? – Gewollte und ungewollte Anreizwirkungen von variablen Löhnen. *Zeitschrift für Betriebswirtschaftliche Forschung*.
- Wright, B. E. (2007). Public service and motivation: does mission matter? *Public Administration Review*, 67(1), 54-64.
- Wright, B. E., Moynihan, D. P., & Pandey, S. K. (2012). Pulling the Levers: Transformational Leadership, Public Service Motivation, and Mission Valence. *Public Administration Review*, 72(2), 206-215.
- Zapata-Phelan, C. P., Colquitt, J. A., Scott, B. A., & Livingston, B. (2009). Procedural justice, interactional justice, and task performance: The mediating role of intrinsic motivation. *Organizational Behavior and Human Decision Processes*, 108(1), 93-105.

Evidenzbasierte Politik durch Gesetzesfolgenabschätzungen? Erfahrungen aus Deutschland und Schweden

Sylvia Veit

1. *Einleitung: Evidenzbasierte Politik und Gesetzesfolgenabschätzungen: Konzeptionelle Vorüberlegungen*

In modernen repräsentativen Demokratien lassen sich idealtypisch zwei Legitimitätsressourcen unterscheiden: (1) Die *Input-Legitimation* politischer Entscheidungen beruht auf der Delegation politischer Macht an gewählte Politikerinnen und Politiker sowie auf spezifischen formalen Verfahren und Abstimmungsmodi in Entscheidungsprozessen. Die Beteiligung der Ministerialbürokratie und nachgeordneter Verwaltungsbehörden an der Gestaltung von Politikinhalten ist ebenfalls über Institutionen legitimiert, in der BRD über das Verfassungsprinzip der Ministerverantwortlichkeit, über einen hierarchischen Verwaltungsaufbau und über damit verbundene Zugriffsrechte in Form der Fach- und Dienstaufsicht. (2) Dessen ungeachtet kann es zu Legitimitätsproblemen des politisch-administrativen Systems kommen, wenn politische Entscheidungen Prinzipien der Effektivität und Effizienz nicht hinreichend berücksichtigen. Diese *Output-Legitimation* politischer Entscheidungen bezieht sich auf deren Rationalität „durch den Bezug auf gesichertes und in der Wissenschaft konsentiertes Wissen“¹. Ausgehend von diesen Vorüberlegungen kann die Programmatik einer „evidenzbasierten Politik“ (EBP) – und damit die Popularität von Gesetzesfolgenabschätzungen (GFA) als Kernelement der Reformagenda zur besseren Rechtsetzung in den EU-Mitgliedsstaaten seit nunmehr einem reichlichen Jahrzehnt – als Reaktion von Regierungen auf wachsende Anforderungen an die *Output-Legitimation* politischen Handelns interpretiert werden. Um diesen Gedanken näher zu erläutern und die Möglichkeiten und Grenzen der mit Hilfe einer evidenzbasierten Politik angestrebten „Rationalisierung“ politischer Entscheidungsprozesse kritisch zu diskutieren, wird in einem ersten Schritt geklärt, was unter dem Konzept der „evidenzbasierten Politik“ zu verstehen ist und wie die GFA in dieses Konzept einzuordnen ist.

- 1 *Weingart, Peter*: Erst denken, dann handeln? Wissenschaftliche Politikberatung aus der Perspektive der Wissens(chaft)soziologie, in: Svenja Falk/Dieter Rehfeld/Andrea Römmele/Martin Thunert, Handbuch Politikberatung, Wiesbaden 2006, S. 36.

Das Konzept der evidenzbasierten Politik bzw. des *evidence-based policy-making*² geht auf die 1997 in die Regierungsverantwortung zurückgekehrte britische *Labour Party* zurück. *New Labour* propagierte die Einführung eines neuen, reformorientierten Regierungsstils, bei dem die politischen Entscheidungsträger bei der Gestaltung von Politikinhalt möglichst viel unabhängige (wissenschaftliche) Expertise berücksichtigen. Ziel dieses Ansatzes ist das Treffen von Entscheidungen, welche die Lebensbedingungen der Bevölkerung nachhaltig verbessern.³ Hinter dem Konzept des EBP steht die Kernannahme, dass die Relevanz sachlicher (rationaler) Argumente gegenüber Fragen des politischen Kalküls, parteipolitischen Interessen und Ideologien durch die systematische Nutzbarmachung vorhandenen Wissens für den politischen Entscheidungsprozess erhöht werden könne. Unter Evidenz wird in diesem Zusammenhang empirisch abgesichertes, in der Wissenschaft konsentiertes, aggregiertes Wissen über gesellschaftliche Entwicklungen und deren Ursachen verstanden. Das Konzept der evidenzbasierten Politik verbreitete sich von Großbritannien ausgehend seit Ende der 1990er Jahre in westlichen Demokratien. Die Parallelen von EBP zur Planungseuphorie und den Ideen eines „Aktiven Staates“ in den 1960er und 1970er Jahren sind offensichtlich, wobei die zum Teil sehr ernüchternden Erfahrungen dieser Jahrzehnte in der EBP-Debatte nur ungenügend reflektiert werden.

Versteht man evidenzbasierte Politik in einem weiteren Sinne als die Gesamtheit institutionenpolitischer Entscheidungen, die auf eine Erhöhung der *Output-Legitimation* politischer Entscheidungen durch die Bezugnahme auf empirisch abgesichertes Wissen abzielen, lassen sich vielfältige praktische Erscheinungsformen von EBP identifizieren. Dazu gehört die Intensivierung wissenschaftlicher Politikberatung, beispielsweise in Form einer Erhöhung der Ressourcen für Auftragsforschung oder in einer verstärkten Einbindung von Expertenkommissionen in den Prozess der Politikentwicklung. In Großbritannien förderte *New Labour* vor allem die Zusammenarbeit mit externen *Think Tanks* sowie mit Universitätsinstituten und Forschungsnetzwerken.⁴ Ein anderer Ansatzpunkt für EBP ist die Stärkung der ressortübergreifenden Zusammenarbeit und der „positi-

- 2 Begrifflich lehnt sich EBP an das seit Anfang der 1990er Jahre an Bedeutung gewinnende Konzept einer „evidenzbasierten Medizin“ an. Dieses Konzept betont, dass medizinische Behandlungsentscheidungen möglichst auf Basis einer empirisch nachgewiesenen Wirksamkeit getroffen werden sollten. Der Transfer aktueller wissenschaftlicher Erkenntnisse in die individuelle Behandlungspraxis rückt damit ins Zentrum der Aufmerksamkeit. Für den Arzt als Entscheider besteht eine wesentliche Kompetenz darin, sich wissenschaftliche Erkenntnisse anzueignen, diese für die Praxis zu interpretieren und in Behandlungsentscheidungen umzusetzen.
- 3 Uwe Jun/Karsten Grabow: Mehr Expertise in der deutschen Politik? Zur Übertragbarkeit des „Evidence-based policy approach“, *Zukunft Regieren. Beiträge für eine gestaltungsfähige Politik* 1/2008, Gütersloh, 2008, S. 22.
- 4 Jun/Grabow (Fn. 3), S. 25.

ven Koordination“⁵ in der Exekutive, beispielsweise durch die Schaffung entsprechender Koordinations-, Strategie- oder Planungseinheiten in der Regierungszentrale oder durch die Etablierung inter-ministerieller Arbeitsgruppen. Ein dritter Ansatz für EBP zielt auf die Förderung organisationalen und individuellen Lernens ab. Praktische Erscheinungsformen hierfür sind alle Arten von administrativen Lernnetzwerken (z.B. im Rahmen von *Benchmarking*-Prozessen) sowie Strategien der Personalentwicklung für die Zentralverwaltung. Schließlich legt ein viertes Bündel von Reformmaßnahmen den Fokus auf Verfahrensregeln zur Berücksichtigung von Evidenz im Policy-Prozess. Derartige prozedurale Ansätze können unter dem Stichwort der „Meta-Regulierung“ zusammengefasst werden.

Die im Mittelpunkt dieses Beitrags stehende Gesetzesfolgenabschätzung (GFA) gehört zur letztgenannten Kategorie von Ansätzen einer evidenzbasierten Politik. Mit Hilfe von Gesetzesfolgenabschätzungen werden die Notwendigkeit einer Regelung sowie deren intendierte und nicht intendierte Folgen mit wissenschaftlichen Methoden evaluiert. Der Terminus „Gesetzesfolgenabschätzung“ hat sich vor allem für Verfahren der ex ante Folgenabschätzung etabliert (prospektive und begleitende GFA)⁶, während retrospektive Folgenabschätzungen in der Regel unter dem Stichwort „Evaluation“ diskutiert werden. Im Folgenden geht es ausschließlich um ex ante Folgenabschätzungen. Deren Intention besteht idealtypisch darin, die politischen Entscheidungsträger dabei zu unterstützen, aus einer Vielzahl möglicher Steuerungsvarianten diejenige Regelungsalternative zu selektieren, mit welcher vorgegebene politische Ziele am effektivsten und effizientesten erreicht werden können.

Dieses Ansinnen ist, wie bereits eingangs erwähnt, keineswegs neu. Die Idee, politische Steuerungsentscheidungen durch eine Erhöhung der Komplexität und Informationsverarbeitungskapazitäten des „Systems“ (d.h. der Strukturen und Prozesse politisch-administrativer Entscheidungsfindung) zu verbessern, war Ausgangspunkt für die Entwicklung der Policy-Analyse in den USA in der Mitte des letzten Jahrhunderts. Die Pioniere der Disziplin Daniel Lerner und Harold Lasswell verbreiteten die Idee einer „interdisziplinären Problemlösungswissenschaft für die Gesellschaft“ (Policy Science), die sich (vermeintlich) exakter, wissenschaftlicher Methoden bedient und normativ auf eine rationale Planung und Steuerung gesellschaftlicher Entwicklungen mit Hilfe einer ganzheitlichen Betrachtung von *Policies* abzielt. Theoretisch wurden diese Vorstellungen durch die rationalistische Entscheidungstheorie untermauert, welche ein ideales, lineares Modell politischer Prozesse bestehend aus den Teilschritten Problemanalyse,

5 Fritz W. Scharpf: Games Real Actors Could Play. Positive and Negative Coordination in Embedded Negotiations, in: Journal of Theoretical Politics 6 (1994), S. 27-53.

6 Böhret, Carl/Götz Konzendorf: Handbuch Gesetzesfolgenabschätzung (GFA). Gesetze, Verordnungen, Verwaltungsvorschriften, Baden-Baden 2001.

Zieldefinition, Identifizierung von Handlungsalternativen, Bewertung dieser Handlungsalternativen im Hinblick auf Effektivität und Effizienz, Auswahl der optimalen Handlungsalternative, Implementation und Evaluation zeichnet.⁷

In der Realität zeigten sich jedoch bald die praktischen Grenzen der Vorstellung einer umfassenden und rationalen Planung durch das politisch-administrative System. Die ernüchternden Erfahrungen in den USA der 1960er Jahre mit großangelegten sozialpolitischen Reformprogrammen und das Scheitern vieler Initiativen der „Aktiven Politik“ und Planung in der BRD der späten 1960er sowie frühen 1970er Jahre sind bekannte Beispiele hierfür. In der *Policy*-Forschung spiegelte sich dies in einer zunehmenden Konzentration auf empirische Implementations- und Evaluationsstudien wider. Theoretisch-konzeptionell setzte sich die Erkenntnis durch, dass der rational-analytischen Erfassung komplexer Systeme und insbesondere der Prognostizierung gesellschaftlicher Entwicklungen erhebliche Grenzen gesetzt sind. Akteure seien aufgrund ihrer begrenzten kognitiven und zeitlichen Ressourcen, wenn überhaupt, nur zu „begrenzt rationalen Entscheidungen“ in der Lage. Herbert Simon, Ökonom und geistiger Vater dieser Theorierichtung, konstatierte, dass die Vorstellung einer umfassenden Suche nach Alternativen sowie deren ausführlicher Analyse und Bewertung in Entscheidungssituationen unrealistisch sei. In der Praxis wird die Suche nach einer Handlungsalternative bereits dann eingestellt, wenn eine Lösung gefunden wurde, die zwar möglicherweise nicht optimal, aber für den oder die an der Entscheidung beteiligten Akteure zufriedenstellend (*satisficing*) ist.⁸

In der Politikwissenschaft entwickelte sich ausgehend von der Pluralismustheorie und auf Basis empirischer Untersuchungen eine Forschungsrichtung, welche den Vorstellungen der rationalistischen Entscheidungstheorie und der synoptischen Richtung der *Policy*-Analyse skeptisch gegenüber stand. Die empirische Realität der Politikentwicklung wurde von Vertretern dieser Forschungsrichtung als inkrementalistisch und fragmentiert beschrieben (Politik der kleinen Schritte nach dem Prinzip des *trial and error*).⁹ Politikwissenschaftler wie Robert A. Dahl und Charles E. Lindblom vertraten nicht nur die Auffassung, dass die inkrementalistische Entscheidungstheorie eher der empirischen Realität ent-

7 Dieses idealtypische Verständnis von Politikentwicklung bildete die Grundlage für das Modell des Policy Cycle. Der Policy Cycle wird heute allerdings üblicherweise nicht mehr als normatives Modell im Sinne der oben genannten Schritte verstanden, sondern als neutraler analytischer Referenzrahmen (Problemdefinition, Agenda-Setting, Politikformulierung, Entscheidung, Implementation, Evaluation) für die Analyse policy-bezogener Fragestellungen.

8 Herbert Simon: Theories of Decision Making in Economics and Behavioural Science, in: American Economic Review, 49 (1959), S. 253-283.

9 Charles E. Lindblom: The Science Of Muddling Through, in: Public Administration Review, 19 (1959), S. 79-88; Lindblom, Charles E.: Still Muddling, Not yet through, in: Public Administration Review, 39 (1979), S. 517-526.

sprache, sondern sie behaupteten darüber hinaus, dass die Dezentralität und die Akteursvielfalt in inkrementellen Politikprozessen auch normativ wünschenswert sei. Politikentwicklung in demokratischen Systemen ist auf die Integration einer Vielzahl von gesellschaftlichen Teil-Interessen angelegt. Die hierfür notwendigen Verhandlungsprozesse und argumentativen Diskurse erfordern ein politisch-administratives System, welches die Differenziertheit und Komplexität der Gesellschaft „spiegelt“.¹⁰ Deshalb wird die Vorstellung einer per se übergeordneten objektiven Gemeinwohlorientierung wissensbasierter politischer Steuerung durch die Inkrementalisten abgelehnt. Fritz Scharpf zeigte darüber hinaus, dass im föderalen System der BRD die Tendenz zur inkrementellen Politikentwicklung durch ein komplexes System der horizontalen und vertikalen Politikverflechtung manifestiert und verstärkt wird.¹¹

Ähnliche Erkenntnisse setzten sich auch in der Verwaltungswissenschaft und in der sozialwissenschaftlichen Forschung zur wissenschaftlichen Politikberatung durch. Die ursprünglich dominante Idealvorstellung von Politik und Verwaltung sowie von Politik und Wissenschaft als zwei getrennten Welten¹², die jeweils bestimmte Funktionen im politischen Prozess erfüllen, hat in den letzten Jahren im Zusammenhang mit den Erkenntnissen der Bürokratieforschung¹³, der Usability-Forschung¹⁴, der „Entdeckung“ von Policy-Netzwerken¹⁵, der Diskus-

10 Robert A. Dahl: Who governs? Democracy and Power in an American City. New Haven 1961.

11 Fritz W. Scharpf: Föderalismusreform. Kein Ausweg aus der Politikverflechtungsfalle? Frankfurt 2009.

12 Sowohl das auf Max Weber zurückgehende dezisionistische Modell der Politikberatung als auch das von Schelsky entwickelte technokratische Modell gehen von der Prämisse eines linearen Prozesses aus, in welchem Politiker und Wissenschaftler jeweils bestimmte, klar voneinander abgrenzbare Funktionen erfüllen. Während nach Weber die politische Entscheidung allein vom Politiker nach dessen irrationalen Kriterien gefällt wird und es somit mehr oder weniger dem Zufall überlassen bleibt, ob irrationale Werte und Orientierungen des Politikers mit den rationalen Erkenntnissen des Wissenschaftlers zusammenfallen, sind es nach Schelsky's Vorstellungen die Sachzwänge der modernen Welt, die dem Politiker einen „one best way“ vorgeben und diesen damit zum Erfüllungsgehilfen wissenschaftlicher Sachrationalität machen (Jürgen Habermas: Technik und Wissenschaft als Ideologie, Frankfurt am Main 1969).

13 Siehe z.B. Kai-Uwe Schnapp: Ministerialbürokratien in westlichen Demokratien. Wiesbaden 2004.

14 Empirische Studien haben gezeigt, dass sozialwissenschaftliche Erkenntnisse nur selten direkt bzw. instrumentell genutzt werden, sondern der Wissenstransfer eher als ein diffuser Prozess der Veränderung von Wahrnehmungsmustern und Paradigmen beschrieben werden kann (siehe z.B. Carol Weiss (ed.): Using Social Science in Public Policy Making, Lexington 1977).

15 Siehe z.B. Renate Mayntz: Policy-Netzwerke und die Logik von Verhandlungssystemen, in: Adrienne Héritier (Hrsg.): Policy-Analyse. Kritik und Neuorientierung, Politische Vierteljahresschrift 34/Sonderheft 24 (1993), Opladen, 39-56.

sion über *Governance*¹⁶ und den differenzierten Beobachtungen zu den vielfältigen Erscheinungsformen sektoraler Politikentwicklung zunehmend an Bedeutung verloren. Thesen über die Politisierung der Bürokratie und die Bürokratisierung der Politik¹⁷, über die Politisierung der Wissenschaft und die Verwissenschaftlichung der Politik¹⁸ bringen zum Ausdruck, dass sich die Funktionen und Rollenzuschreibungen von Spitzenbürokraten und Politikern ebenso wie die Handlungsorientierungen und Diskursrahmen politischer und wissenschaftlicher Akteure in der modernen Gesellschaft immer mehr vermischen. Wenn heute von einer Tendenz zur Verwissenschaftlichung der Politik gesprochen wird, dann ist damit weniger Schelsky's Idee eines technisch vorgegebenen *one best way* (technokratisches Modell wissenschaftlicher Politikberatung) gemeint¹⁹, sondern eher die Tatsache, dass aufgrund eines exorbitanten Bedeutungszuwachses der Wissenschaft als funktionales Teilsystem der Gesellschaft immer mehr *Policy*-Akteure mit wissenschaftlichen Standards, Methoden und Argumentationsmustern vertraut sind und deshalb die Bedeutung wissenschaftlicher Argumente im Diskurs über die Gestaltung von Politikinhalten immer wichtiger wird.²⁰

Seit der Jahrtausendwende wurden in vielen europäischen Ländern und auf EU-Ebene systematische Ansätze zur Integration der GFA in den Politikformulierungsprozess entwickelt, indem verpflichtende Vorschriften zur Durchführung einer GFA für neue Politikentwürfe eingeführt oder bestehende Vorschriften erweitert wurden. Zur Umsetzung dieser GFA-Vorschriften wurden länderspezifisch unterschiedliche Zuständigkeiten und Verfahren definiert. Diese Renaissance rationalistischer Vorstellungen im Gewand der GFA kann dabei durchaus als Gegenreaktion auf eine zunehmende funktionale Differenzierung der Gesellschaft, auf hochgradig ausdifferenzierte Organisationsstrukturen im politisch-administrativen System, auf Phänomene einer vertikalen Versäulung in Folge NPM-inspirierter Verwaltungsreformen sowie auf einen wachsenden Zeitdruck in politischen Entscheidungsprozessen verstanden werden.

Im Folgenden werden am Beispiel Deutschlands und Schwedens zwei unterschiedliche Ansätze der Institutionalisierung von Gesetzesfolgenabschätzungen

- 16 Für einen Überblick: Arthur Benz/Susanne Lütz/Uwe Schimank/Georg Simonis: Handbuch Governance. Theoretische Grundlagen und empirische Anwendungsfelder, Wiesbaden 2007.
- 17 Carl Böhrer: Politik und Verwaltung, Beiträge zur Verwaltungspolitologie, Opladen 1983; Werner Jann/Sylvia Veit: Politicisation of administration or bureaucratisation of politics? The case of Germany. Potsdamer Diskussionspapiere zur Verwaltungswissenschaft 6, Potsdam 2010.
- 18 Weingart (Fn. 1).
- 19 Helmut Schelsky: Der Mensch in der wissenschaftlichen Zivilisation, in: Helmut Schelsky (Hrsg.): Auf der Suche nach Wirklichkeit, Düsseldorf 1965, S. 439-480.
- 20 Siehe z.B. Claudio M. Radaelli: The role of knowledge in the policy process, in: Journal of European Public Policy, 2 (1995), S. 159-183.

aufgezeigt. Anschließend werden einige empirische Daten zur Umsetzung der GFA in den beiden Ländern vorgestellt. Im letzten Teil des Beitrags werden Erklärungsansätze für die zu beobachtenden Unterschiede in der Implementation der GFA in den beiden untersuchten Ländern diskutiert und es wird thesenhaft reflektiert, inwiefern Gesetzesfolgenabschätzungen die „Evidenzbasiertheit“ politischer Entscheidungen unterstützen und damit deren *Output*-Legitimation erhöhen.

2. *Institutionalisierung von Gesetzesfolgenabschätzungen in Deutschland*

Ein zentraler Reforminhalt der Politik zur besseren Rechtsetzung in Deutschland besteht seit mehr als dreißig Jahren darin, Vorschriften zur standardmäßigen Prüfung von Gesetzesfolgen im Rechtsetzungsprozess zu verankern. So wurden im Zuge einer Novellierung der Gemeinsamen Geschäftsordnung der Bundesministerien (GGO) bereits Ende 1976 die Vorschriften zur Darstellung der Auswirkungen von Gesetzentwürfen der Bundesregierung in der Gesetzesbegründung stark erweitert. Besondere Schwerpunkte bildeten dabei die Folgen für die öffentlichen Haushalte, Preise und Verbraucher sowie für die Umwelt. Die Novellierung der GGO umfasste außerdem die Einführung eines Gesetzesvorblattes, das auf einer Druckseite Auskunft über Zielsetzung, Lösungsvorschlag, Alternativen und Kosten eines Gesetzentwurfes gibt. Gesetzesbegründung und Gesetzesvorblatt weisen noch heute eine ähnliche Struktur auf, wobei die Vorgaben zur Darstellung von Gesetzesfolgen im Laufe der Jahre sukzessive erweitert wurden. Einen besonders umfassenden Versuch zur Institutionalisierung eines Prüfverfahrens im Rechtsetzungsprozess stellten die 1984 vom Bundeskabinett verabschiedeten „Blauen Prüffragen“ dar, welche trotz (oder gerade wegen einer geringen praktischen Relevanz) Mitte der 1990er Jahre als Anhang in die GGO aufgenommen wurden.

Im Jahr 2000 wurde die GGO umfassend novelliert. Die „Blauen Prüffragen“ wurden als Prüfkatalog abgeschafft und deren Inhalte in kondensierter und modifizierter Form direkt in die GGO aufgenommen. Der Terminus „Gesetzesfolgen“ erhielt eine zentrale Stellung (§§ 43, 44 GGO). Damit verbunden war der Anspruch, nicht nur die Berücksichtigung bestimmter Folgenaspekte wie der Kosten für die öffentlichen Haushalte im Rechtsetzungsprozess sicherzustellen, sondern alle wesentlichen Auswirkungen eines Regelungsentwurfes standardmäßig zu evaluieren. Ergänzend zur Novellierung der GGO ließ das Bundesministerium des Innern als eines der Leitprojekte des Programms „Moderner Staat – Moderne

Verwaltung“ ein Handbuch zur Gesetzesfolgenabschätzung²¹ sowie einen dazugehöriger Leitfaden erarbeiten. Beide Handreichungen wurden in einer Pilotphase anhand von acht ausgewählten Gesetzesvorhaben erprobt. Einige Ressorts entwickelten zudem eigene Unterstützungsmaterialien zu den Folgenabschätzungsanforderungen der novellierten GGO.²²

Mit der Übernahme der Regierungsverantwortung durch die „Große Koalition“ unter Bundeskanzlerin Angela Merkel im Herbst 2005 verengte sich der Hauptfokus der Reformen zur „besseren Rechtsetzung“ von der Idee einer umfassenden Gesetzesfolgenabschätzung auf die durch Regulierungen hervorgerufenen bürokratischen Lasten. CDU und SPD verankerten die Einführung des Standardkosten-Modells verbunden mit der Einrichtung eines unabhängigen Nationalen Normenkontrollrates (NKR) im Koalitionsvertrag. Bürokratieabbau wurde zur „Chefsache“ erklärt und die zentrale Zuständigkeit für das Thema vom Bundesministerium des Innern ins Bundeskanzleramt verlagert. Im April 2006 konkretisierte die Regierung ihre politischen Vorhaben im Rahmen des „Programms für Bürokratieabbau und bessere Rechtsetzung“, wenige Monate später wurde auf Basis eines von den Fraktionen der CDU/CSU und der SPD gemeinsam in den Bundestag eingebrachten und beschlossenen Gesetzes (Gesetz zur Einrichtung eines Nationalen Normenkontrollrates – NKR) ein unabhängiger Nationaler Normenkontrollrat mit Sekretariat im Bundeskanzleramt berufen.

Kernelement der neuen Strategie war das aus den Niederlanden stammende und sich im gesamten europäischen Raum rasch verbreitende Standardkosten-Modell (SKM) zur Messung von Bürokratiekosten. Bürokratiekosten werden im SKM sehr eng definiert als diejenigen Kosten, welche Unternehmen, Bürgern und der öffentlichen Verwaltung durch auf staatliche Regulierungen zurückzuführende Informationspflichten (z.B. das Ausfüllen von Anträgen oder das Führen von Statistiken) entstehen. Um diese Kosten zu errechnen, befragt man Vertreter der jeweiligen Zielgruppe (z.B. einzelne Unternehmen) zu dem Zeit- und Kostenaufwand, welcher für sie mit der Erfüllung staatlicher Informationsanforderungen verbundenen ist. Die standardisierten und in Zusammenarbeit mit Fachexperten validierten Daten werden anschließend segmentiert und mit dem Tarif, der Häufigkeit der Erfüllung und der Anzahl der Betroffenen multipliziert. Am Ende steht eine monetäre Angabe zu den mit einer Rechtsvorschrift, einem Rechtsbereich oder einem bestimmten Ereignis (z.B. Unternehmensgründung) verbundenen Bürokratiekosten. Während das SKM ursprünglich vorrangig zur

21 Böhret/Konzendorf (Fn. 6).

22 Bundesministeriums für Familie, Senioren, Frauen und Jugend: Geschlechterdifferenzierete Gesetzesfolgenabschätzung. Gender Mainstreaming bei der Vorbereitung von Rechtsvorschriften; Bundesministerium für Wirtschaft und Technologie: Merkblatt zur Ermittlung der Kostenfolgen und Preiswirkungen von Gesetzesvorlagen, Vorlagen von Rechtsverordnungen und von Verwaltungsvorschriften.

Evaluation des bestehenden Rechtsbestandes („Nullmessung“) entwickelt wurde, wird es heute in vielen europäischen Staaten auch eingesetzt, um die mit einer neuen gesetzlichen Regelung verbundenen Bürokratiekosten ex ante abzuschätzen.

In Deutschland wurde die Bürokratiekostenabschätzung mit dem SKM als ein Prüfstein bzw. ein Modul der GFA konzeptionalisiert: Im Dezember 2006 wurde die GGO um die Verpflichtung erweitert, auf dem Gesetzesvorblatt und im Rahmen der Gesetzesbegründung Transparenz über Informationspflichten und Bürokratiekosten für Unternehmen, Bürger und die öffentliche Verwaltung herzustellen. Der NKR prüft die Umsetzung dieser Vorschriften im Zuge der Ressortabstimmung und gibt eine Stellungnahme ab, welche obligatorischer Bestandteil der Kabinettsvorlage ist. Zur praktischen Realisierung der ex ante Abschätzung von Bürokratiekosten im Rechtsetzungsprozess entwickelten der NKR und die Geschäftsstelle Bürokratieabbau im Bundeskanzleramt in Abstimmung mit den Bundesministerien einen „Leitfaden für die ex ante Abschätzung der Bürokratiekosten nach dem Standardkosten-Modell“. Bei der Durchführung von Bürokratiekostenabschätzungen zu „ihren“ Referentenentwürfen bietet das Sekretariat des NKR den Ministerialbeamten Unterstützung in methodischen Fragen an.

Nachdem die Abschätzung der Bürokratiekosten für Unternehmen rasch implementiert worden war, entwickelte sich ab 2008 eine neue Diskussion um die GFA, die aus verschiedenen Richtungen Impulse erhielt. Zwei sollen hier erwähnt werden: Erstens hatte die Bundesregierung inspiriert durch Entwicklungen auf EU-Ebene bereits im Fortschrittsbericht zur Nationalen Nachhaltigkeitsstrategie vom Oktober 2008 angekündigt, zukünftig im Rahmen der GFA auch Auswirkungen auf eine nachhaltige Entwicklung zu untersuchen. Im Mai 2009 folgte eine entsprechende GGO-Änderung. In § 44 GGO wurde die Vorgabe neu eingefügt, dass im Rahmen der Gesetzesbegründung darzustellen sei, „ob die Wirkungen des Vorhabens einer nachhaltigen Entwicklung entsprechen, insbesondere welche langfristigen Wirkungen das Vorhaben hat“. Zweitens wurde beklagt, dass die Tätigkeit des NKR aufgrund der engen Definition von Bürokratiekosten für die einzelnen Unternehmen kaum zu spürbaren Entlastungen führe. Nach intensiven Debatten um eine mögliche Ausweitung des NKR-Mandates wurde schließlich im März 2011 das NKR-Gesetz geändert und sein Prüfauftrag auf die „Darstellung des Erfüllungsaufwandes neuer Regelungen für Bürgerinnen und Bürger, Wirtschaft und öffentliche Verwaltung auf ihre Nachvollziehbarkeit und Methodengerechtigkeit sowie die Darstellung der sonstigen Kosten der Wirtschaft, insbesondere für die mittelständischen Unternehmen“ (§ 1 (3) NKR-G) erweitert. Darüber hinaus kann sich der NKR auch zu anderen Fragen wie zur Verständlichkeit der Regelungsziele, zu alternativen Lösungsmöglichkeiten sowie

zur Möglichkeit einer Befristung und der Frage, ob eine Evaluierung des Gesetzes geplant ist, äußern.

Zusammenfassend sind folgende Merkmale der Institutionalisierung der GFA in Deutschland zu betonen: (1) Formale Vorgaben zur GFA sind auf Geschäftsordnungsebene geregelt. (2) Die Zuständigkeit zur Durchführung von Gesetzesfolgenabschätzungen liegt beim federführenden Ministerium. (3) Gesetzesfolgenabschätzungen werden in der Regel nicht an Externe vergeben. (4) Die Ergebnisse der GFA werden auf dem Gesetzesvorblatt und in der Gesetzesbegründung dokumentiert. (5) Die Plausibilität der Angaben zu Gesetzesfolgen wird im Falle der Bürokratiekosten und seit 2011 für Erfüllungsaufwand und sonstige Kosten vom NKR geprüft, alle anderen Folgenaspekte sind Gegenstand einer dezentralen, kaum formalisierten Prüfung durch andere Ressorts (z.B. prüft das Finanzministerium die Plausibilität der angegebenen Haushaltsauswirkungen).

3. Institutionalisierung der GFA in Schweden

Bevor die Institutionalisierung der GFA in Schweden erläutert werden kann, sollen drei Charakteristika des politisch-administrativen Systems hervorgehoben werden: der Verwaltungsdualismus, das Kommissionswesen und das Öffentlichkeitsprinzip.

Unter Verwaltungsdualismus wird die spezifische Aufgabenteilung zwischen Ministerien und nachgeordneten Behörden in Schweden verstanden. Die schwedischen Ministerien sind organisatorisch in der Kanzlei der Ministerien zusammengefasst. In der gesamten Kanzlei der Ministerien arbeiten derzeit 4.345 Mitarbeiter/innen²³, die schwedischen Ministerien sind also vergleichsweise kleine Organisationen. Die Ministerien sind primär mit Aufgaben der Politikformulierung und Vorbereitung von Entscheidungen befasst, während die Umsetzung von Gesetzen und Verordnungen (ebenso wie die Konkretisierung des Vollzugs durch die Ausfertigung von Ausführungsvorschriften) nachgeordneten, vergleichsweise großen, zentralen Verwaltungsbehörden überlassen ist. Ein einzelner Minister kann in Schweden prinzipiell nicht steuernd in die Tätigkeiten der zentralen Verwaltungsbehörden eingreifen, da es keine individuelle Ministerverantwortlichkeit wie in Deutschland gibt, sondern alle Beschlüsse (formal) von der Regierung als Kollektiv getroffen werden. Die zentralen Verwaltungsbehörden in Schweden sind in der Ausführung der Gesetze relativ frei. Die Regierung kann zwar Ziele setzen und bestimmte Rahmenbedingungen festsetzen, die Wahl der Mittel bleibt aber den Fachbehörden überlassen. Ein wichtiges Instrument für

23 Regeringens årsbok 2011.

Regierung und Ministerialverwaltung, um die Verwaltungsbehörden zu steuern, ist deshalb (neben anderen wie Personal und Budget) die Festlegung von Prozeduren über Verfahren zur Entscheidungsfindung. Zu dieser Art von „Institutionenpolitik“ gehört auch die im Zentrum dieses Beitrags stehende Regulierung von Verfahren zur Folgenabschätzung. Die GFA ist in Schweden demnach nicht nur ein Ansatz zur Qualitätsverbesserung von Gesetzen und anderen Rechtsvorschriften (mit der Ministerialverwaltung als Adressat), sondern auch ein Steuerungsinstrument zur Kontrolle der relativ unabhängigen zentralen Behörden.

Ein zweites wichtiges Charakteristikum des politisch-administrativen Systems Schwedens ist das Kommissionswesen, dessen historische Wurzeln bis ins 16. Jahrhundert zurückreichen. Während in Deutschland die meisten Gesetze von Ministerialbeamten ausgearbeitet werden (Referentenentwurf), wird ein erheblicher Teil der Vorbereitungsarbeit in Schweden an unabhängige Untersuchungskommissionen („*utredningar*“) delegiert. Die Steuerung der Kommission findet über Direktiven der Regierung statt (Festlegung des Budgets, der Mitglieder und der Untersuchungsschwerpunkte und -ziele). In der Ausführung ihrer Arbeit und in ihrer Entscheidungsfindung sind die Kommissionen formal unabhängig und können von der Regierung keine Weisungen erteilt bekommen. Fast alle wichtigen politischen Entscheidungen werden in Schweden durch Untersuchungskommissionen vorbereitet. Werden neue Regelungen in Schweden nicht durch eine Kommission vorbereitet (was auf circa die Hälfte der Gesetzesvorhaben zutrifft), dann sind es häufig zentrale Verwaltungsbehörden, welche der Regierung zuarbeiten und vorbereitende Studien durchführen. Die Delegation von Vorbereitungsarbeiten an obere Bundesbehörden oder Kommissionen und Beiräte ist zwar auch in Deutschland übliche Praxis, wird aber nicht so umfassend und weitreichend praktiziert wie in Schweden. Die Ministerialverwaltung in Schweden muss sich der beschriebenen Aufgabenteilung entsprechend kaum mit Aufgaben der Wissensgenerierung und Informationsbeschaffung auseinandersetzen, sondern sie erhält meist sehr detailliert ausformulierte Vorschläge von den Kommissionen, die bereits ein öffentliches Konsultationsverfahren durchlaufen haben. Aufgabe der Ministerialverwaltung ist die finale politische Koordination und Abstimmung der Entwürfe.

Ein weiteres für das Verständnis der folgenden Ausführungen wesentliches Merkmal des schwedischen Systems ist das in der Verfassung verankerte Öffentlichkeitsprinzip. Zu neuen Politikentwürfen wird in Schweden ein offenes Konsultationsverfahren („*remiss*“) durchgeführt. Die im Rahmen dieses Verfahrens abgegebenen Stellungnahmen werden veröffentlicht. Im Begleitmaterial zu Gesetzentwürfen, das nach Kabinettsbeschluss zusammen mit dem Entwurf dem Parlament zugeleitet wird, werden die Ergebnisse des *remiss*-Verfahrens zusammengefasst und es wird dargestellt, welche Positionen, aus welchen Gründen und von welchen Akteuren übernommen wurden. In Deutschland hingegen wer-

den zwar auch externe Akteure im exekutiven Verfahren der Gesetzesentwicklung konsultiert, deren Positionen und der Einfluss dieser auf eventuelle Änderungen im Entwurf werden jedoch nicht systematisch öffentlich zugänglich gemacht.

Folgenabschätzungen in Schweden sind aufgrund der genannten Systemmerkmale in Schweden anders institutionalisiert als in Deutschland: Sie spielen nicht nur auf Ebene der Ministerialbürokratie eine Rolle, sondern sind auch für die Rechtsetzungstätigkeit der nachgeordneten Behörden von Bedeutung. Reformanstrengungen in diesem Bereich gehen bis in die 1970er Jahre zurück, wobei der Fokus anfangs primär auf Kostenfolgen lag. Später erfolgte eine sukzessive Ausweitung in Richtung einer umfassenden GFA.²⁴

Vorschriften zur Folgenabschätzung für Gesetzesentwürfe der schwedischen Regierung, die im Mittelpunkt der folgenden Ausführungen stehen werden, existieren in Schweden (wie auch in Deutschland) bereits seit mehreren Jahrzehnten. Ende der 1970er Jahre erließ die schwedische Regierung verschiedene Direktiven für das Kommissionswesen, die auf die Darstellung der Kostenfolgen der von Kommissionen ausgearbeiteten Gesetzesvorschläge abzielten. Im Verlauf der 1980er Jahre wurden die Vorschriften zur Beteiligung von Kommunen und Verbänden verschärft und die Vorschriften zur Folgenabschätzung für Kommissionsentwürfe schrittweise ausgeweitet. Für die Ministerien gab es Checklisten zur Folgenabschätzung, verbindlichere Vorgaben wurden jedoch erstmals Ende der 1990er Jahre mit den Regelungen zur sog. Simplex-Folgenabschätzung für kleine und mittlere Unternehmen verabschiedet. In den 2000er Jahren konzentrierten sich die Reformaktivitäten vor allem auf die Reduzierung administrativer Lasten für Unternehmen, wobei in den ersten Jahren vor allem die Identifikation von Entlastungen im bestehenden Normenbestand im Fokus stand. Eine ex ante Bürokratiekostenabschätzung für neue Rechtsvorschriften wurde erst Ende 2007 im Rahmen der neuen Folgenabschätzungsverordnung formal vorgeschrieben.

Seit 2008 gibt es in Schweden zudem ein per Kabinettsbeschluss eingerichtetes unabhängiges Beratungsgremium („*regelråd*“), dessen Aufgabe die Prüfung der Qualität der durchgeführten Folgenabschätzungen ist. Wenn zutreffend, nimmt der *regelråd* außerdem Stellung zu den administrativen Lasten für Unternehmen. Obgleich der *regelråd* formal für den gesamten Bereich der GFA zuständig ist, liegt sein Fokus auf Folgen im ökonomischen Bereich (inklusive Bürokratiekosten), was sich auch in der organisatorischen Anbindung an das Wirtschaftsministerium widerspiegelt. Ähnlich wie der NKR in Deutschland nimmt der *regelråd* Stellung zu Gesetzesvorschlägen der Regierung, prüft aber

24 Sylvia Veit: Bessere Gesetze durch Folgenabschätzung? Deutschland und Schweden im Vergleich, Wiesbaden 2010, S. 119-137.

auch die GFA für Kommissionsvorschläge. Die Ergebnisse dieser Prüfungen werden in schriftlichen Stellungnahmen fixiert. Darüber hinaus veröffentlicht der *regelråd* einen jährlichen Tätigkeitsbericht.

Zusammenfassend sind folgende Merkmale der Institutionalisierung der GFA in Schweden zu betonen: (1) Formale Vorgaben zur GFA existieren in Form einer Folgenabschätzungs-Rechtsverordnung und in diversen Kommissionsdirektiven. (2) Die Zuständigkeit zur Durchführung von Gesetzesfolgenabschätzungen sind auf drei Akteursgruppen verteilt: auf die zur Vorbereitung von Politikentwürfen eingesetzten staatlichen Untersuchungskommissionen, das federführende Ministerium und die zentralen Verwaltungsbehörden. (3) Die Ergebnisse der GFA werden in Kommissionsberichten, in Gesetzesbegründungen und im Begleitmaterial zu behördlichen Ausführungsvorschriften dokumentiert. (4) Die Qualität der GFA wird von einem unabhängigen Beratungsgremium, dem *regelråd*, geprüft.

4. Empirische Daten zur Umsetzung der GFA in Deutschland und Schweden

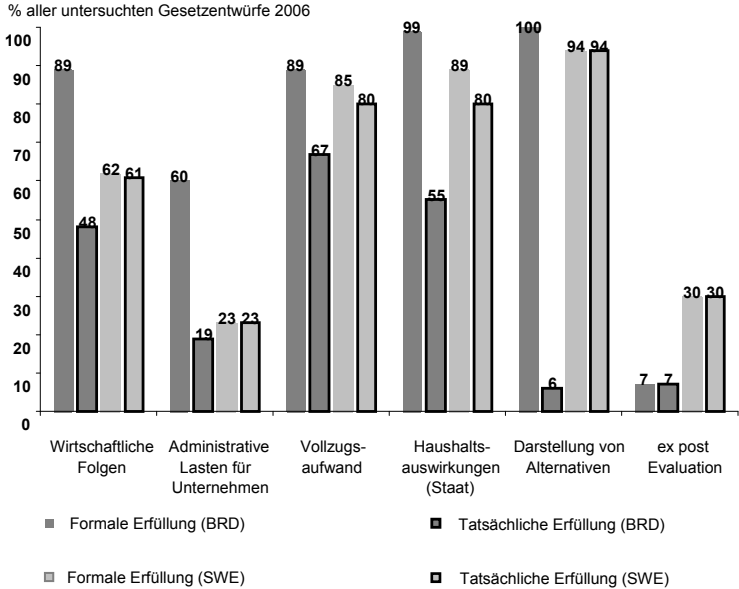
Nachfolgend werden exemplarisch einige Ergebnisse einer von mir durchgeführten empirischen Studie zur Umsetzung der Vorschriften zur Gesetzesfolgenabschätzung in Deutschland und Schweden vorgestellt werden. Für den schwedischen Fall beziehe ich mich ausschließlich auf die in den Gesetzentwürfen der Regierung dokumentierten Angaben zu Gesetzesfolgen. Diese Angaben gehen zum Teil auf die von Kommissionen durchgeführten GFAs zurück, zum Teil stammen sie aus der Feder von Ministerialbeamten. GFAs für die Ausführungsvorschriften nachgeordneter Behörden in Schweden werden in der folgenden Darstellung nicht betrachtet.²⁵ Datenbasis sind alle im Jahr 2006 von der deutschen (n=152) und von der schwedischen Regierung (n=132) beschlossenen Gesetzentwürfe sowie eine Stichprobe von Entwürfen aus dem Jahr 1977 (n=30 für BRD, n=27 für SWE). Für Deutschland wird darüber hinaus eine Stichprobe von Gesetzentwürfen aus dem Jahr 2010 (n=15) mit in die Betrachtung einbezogen.

Bei der vergleichenden Analyse dieser Materialien fallen zunächst einige grundlegende Unterschiede ins Auge. Die schwedischen Gesetzesbegründungen sind im Durchschnitt wesentlich länger als die deutschen (Schweden: 54 Seiten plus 45 Seiten Anhang; Deutschland: 13 Seiten). Die durchschnittlich 13-seitige Gesetzesbegründung der Regierungsentwürfe in Deutschland ist meist in einen „Allgemeinen Teil“ und einen „Besonderen Teil“ gegliedert. Der „Allgemeine

25 Ausführlich hierzu siehe Veit (Fn. 24), S. 198-205.

Teil“ enthält Aussagen zur Problemstellung, zu den politischen Zielen, zu Alternativen, zu Gesetzesfolgen sowie zu einigen juristischen Fragen. Trotz dieses weiten inhaltlichen Spektrums des „Allgemeinen Teils“ der Gesetzesbegründungen, ist dieser Teil in fast einem Viertel der Fälle nur eine Seite lang. Einen mehr als sieben Seiten langen „Allgemeinen Teil“ weisen weniger als 6% aller Entwürfe auf. Schwedische Gesetzesbegründungen sind, auch das sicher zum Teil bedingt durch die unterschiedlichen Konventionen bezüglich des Umfangs der Gesetzesbegründungen, in geringerem Maße als die deutschen nach einem einheitlichen formalen Schema strukturiert. Formal-juristische Aspekte spielen kaum eine Rolle, stattdessen wird in den Gesetzesbegründungen ausführlich dokumentiert, welche Akteure in den Prozess der Entwurfserstellung involviert waren, welche unterschiedlichen Positionen es gab und welche Änderungen im Zuge des Konsultationsverfahrens am Entwurf vorgenommen wurden. Informationen über Beteiligungen und Konflikte im Prozess der Entwurfserstellung sind den deutschen Gesetzesbegründungen im Regelfall nicht zu entnehmen. Die Dokumentation der Gesetzesfolgen ist eher knapp gehalten und folgt häufig dem gleichen formalen Schema. Bestimmte Textbausteine sind in fast allen deutschen Gesetzesbegründungen zu finden.

Abbildung 1: Formale und tatsächliche Erfüllung von Folgenabschätzungsanforderungen in Deutschland und Schweden



In Abbildung 1 wird zwischen „formaler Erfüllung“ und „tatsächlicher Erfüllung“ von Anforderungen zur Gesetzesfolgenabschätzung unterschieden. Unter „tatsächlicher Erfüllung“ wird eine fundierte inhaltliche Prüfung bestimmter Kriterien der besseren Rechtsetzung verstanden, die sich in der Regierungsvorlage in Form von ausführlichen und mit einer Begründung versehenen Aussagen niederschlägt. Eine „formale Erfüllung“ hingegen liegt schon dann vor, wenn im Gesetzesvorblatt und/oder in der Gesetzesbegründung eine Aussage zur entsprechenden Variable zu finden ist, selbst wenn das Statement nur sehr kurz ist und nicht begründet wird. Die „tatsächliche Erfüllung“ wird also als Teilmenge der „formalen Erfüllung“ definiert. Die Differenz zwischen „formaler Erfüllung“ und „tatsächlicher Erfüllung“ wird im Folgenden als „formale Anpassung“ bezeichnet. „Formale Anpassung“ ist dadurch gekennzeichnet, dass ein Folgenabschätzungskriterium durch ein floskelhaftes, in vielen Gesetzentwürfen so oder in ähnlicher Art und Weise zu findendes Statement „abgehakt“ wird, ohne dass substantielle Begründungen geliefert werden.

Wie Abbildung 1 zeigt, existieren sowohl in Deutschland als auch in Schweden Implementationsdefizite der GFA, die je nach Prüfkriterium unterschiedlich stark ausgeprägt sind. Im deutsch-schwedischen Vergleich fallen jedoch auch einige Unterschiede auf: So ist die „formale Erfüllung“ der Anforderungen zur Folgenabschätzung in Deutschland bei fünf von sechs Variablen höher als in Schweden, während die „tatsächliche Erfüllung“ der in Abb. 1 aufgeführten Prüfkriterien bei allen sechs Vergleichsvariablen in Schweden höher ist als in Deutschland. Am deutlichsten tritt diese Differenz in der „tatsächlichen Erfüllung“ bei der Darstellung der Alternativen zu Tage. Während Alternativen in fast allen schwedischen Gesetzentwürfen diskutiert werden, enthalten die Begründungen der deutschen Gesetzentwürfe nur sehr selten Aussagen zu Alternativen. Allerdings ist auch in den schwedischen Gesetzesvorlagen keine umfassende Analyse von Alternativen im Geiste rationalistischer Vorstellungen zu finden (z.B. in Form einer Kosten-Nutzen-Analyse verschiedener Regelungsoptionen), sondern es werden Alternativen vor allem in Bezug auf Regelungsdetails diskutiert. So enthalten die schwedischen Entwürfe beispielsweise häufig Vorschläge von Interessengruppen, die eine abweichende Gestaltung bestimmter Einzelvorschriften beinhalten. Diese Alternativvorschläge werden meist nicht nur kurz erwähnt, sondern es wird auch diskutiert und begründet, weshalb der Vorschlag nicht übernommen wurde. In Deutschland hingegen werden derartige Vorschläge nur selten in den Gesetzesbegründungen dokumentiert. Auf dem Vorblatt findet sich unter Gliederungspunkt C („Alternativen: ...“) in den meisten Fällen (94% der Gesetzentwürfe) nur die knappe Aussage, dass „keine Alternativen“ bestünden.

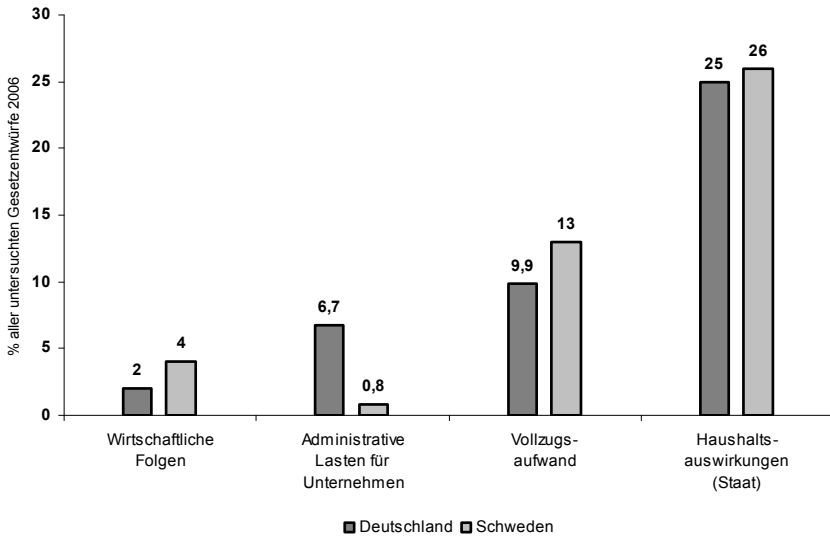
Bezüglich der Quantifizierung von Gesetzesfolgen sind die Ergebnisse für Deutschland und Schweden ähnlich (Abb. 2): Quantifizierungen sind selten und

am häufigsten für die zu erwartenden Haushaltsauswirkungen zu finden, die in beiden Staaten bereits seit den 1970er Jahren verpflichtend darzustellen sind. Kosten für die Wirtschaft (ohne Bürokratiekosten), das mag vor dem Hintergrund der umfassenden Reformbemühungen in diesem Bereich überraschen, werden nur in 2% (Deutschland) bzw. 4% (Schweden) der Fälle quantifiziert. Die Betrachtung einer Stichprobe von deutschen Regierungsentwürfen aus dem Jahr 2010 (n=15) ergibt, dass dieser Wert 2010 etwas höher ist als 2006 (7% Quantifizierungen im Jahr 2010), sich aber immer noch auf einem sehr niedrigen Niveau befindet. Ähnliches gilt für den Vollzugaufwand: Hier finden wir 2006 unter 10% Quantifizierungen in Deutschland, in der Stichprobe 2010 13% Quantifizierungen. Zu vermuten sind also leichte „Ausstrahleffekte“ der intensiven politischen Bemühungen um eine Bürokratiekostenabschätzung auf andere Bereiche der Kostenfolgenabschätzung im Rahmen der GFA.

Angaben zu den Bürokratiekosten für Unternehmen enthielten 2006 knapp 7% der Regierungsentwürfe in Deutschland (Abb. 2). Da der NKR in Deutschland erst im Spätsommer 2006 konstituiert wurde und es in Schweden 2006 weder die formale Verpflichtung zur ex ante Abschätzung der Bürokratiekosten für Unternehmen noch den *regelråd* gab, ist davon auszugehen, dass diese Zahlen heute deutlich höher liegen. Eine Analyse von 429 Regierungsentwürfen, die zwischen dem 1.3.2007 und dem 8.5.2009 in den Bundestag eingebracht wurden, bestätigt dies: Für diejenigen Entwürfe, die Informationspflichten für Unternehmen verändern, neu einführen oder abschaffen, werden die Bürokratiekosten mehrheitlich quantifiziert. Bezüglich der Qualität dieser Quantifizierungen und der Nachvollziehbarkeit der Berechnung sind jedoch Differenzierungen notwendig: So wird beispielsweise die Anzahl der betroffenen Unternehmen in 42% dieser Entwürfe nicht benannt. Angaben zu den Datenquellen enthalten sogar nur 18% der mit veränderten Informationspflichten für Unternehmen verbundenen Entwürfe, und nur in 10% der Fälle werden Datenquellen und Erhebungsmethoden im Detail transparent gemacht.²⁶

26 Veit (Fn. 24), S. 177-180.

Abbildung 2: Quantifizierung von Gesetzesfolgen in Deutschland und Schweden



Interessant ist die Frage, inwiefern sich die politischen Maßnahmen der letzten Jahrzehnte zur verbesserten GFA von Regierungsentwürfen substantiell in der Dokumentation von Gesetzesfolgen niederschlagen. Aufschluss hierüber gibt ein Vergleich mit Gesetzentwürfen aus dem Jahr 1977.

Für den schwedischen Fall sind kaum Veränderungen bezüglich der Darstellung von Alternativen, der Darstellung des Vollzugsaufwandes sowie von Auswirkungen auf Kommunen und die Wirtschaft zu erkennen. Differenzen sind für Schweden im Vergleich zwischen 1977 und 2006 im Hinblick auf drei Aspekte festzustellen: (1) Ex post-Evaluationen werden 2006 in 30% der Fälle angekündigt, 1977 in 19% der Fälle. (2) Auswirkungen auf KMU werden 1977 in 10% der Entwürfe thematisiert, 2006 in 29% der Entwürfe. Hier scheinen sich die Maßnahmen zur Simplex-Folgenabschätzungen deutlich niederschlagen. (3) Angaben zu Haushaltsauswirkungen werden 2006 in 89% der Fälle getätigt, 1977 hingegen nur in 52% der Fälle. Quantifizierungen von Haushaltsauswirkungen werden jedoch 1977 häufiger vorgenommen als 2006.

Dieses Phänomen zeigt sich auch im deutschen Fall: 2006 werden Haushaltsauswirkungen seltener quantifiziert als 1977, dafür wird der Vollzugsaufwand besser dargestellt. Auch bei den Ankündigungen einer ex post-Evaluation zeigt sich ein ähnliches Ergebnis wie in Schweden, allerdings auf viel niedrigerem Niveau: 2006 wird für 8% der deutschen Regierungsentwürfe eine ex post-Evaluation angekündigt, 1977 für keinen einzigen Entwurf. Für alle anderen

Prüfkriterien zeigt der Vergleich zwischen 1977 und 2006 nur ein Ergebnis: Substantiell hat sich wenig verändert, aber formale Anpassungen sind 2006 viel häufiger als 1977. Zynisch formuliert könnte man also die Zunahme von Formalismen als das augenfälligste Resultat jahrzehntelanger Bemühungen um eine bessere Rechtsetzung durch Folgenabschätzungen bezeichnen. Besonders deutlich sind die formalen Anpassungen bei allen Querschnittsthemen, die innerhalb der Exekutive nicht ressortmäßig organisiert sind: Nachhaltige Entwicklung, Gleichstellung der Geschlechter, Alternativen, Auswirkungen auf Verbraucher.

Welche Schlussfolgerungen lassen sich aus diesen empirischen Daten in Bezug auf den Nutzen von Folgenabschätzungen als Instrument einer evidenzbasierten Politik ziehen? Nachfolgend soll nicht dafür plädiert werden, Ansätze einer Versachlichung politischer Entscheidungen mit Hilfe von Folgenabschätzungen ganz aufzugeben. Aber es wird argumentiert, dass es an der Zeit ist, sich von unrealistischen Vorstellungen bezüglich der Effekte von GFAs zu verabschieden.

5. *Fazit: Evidenzbasierte Politik durch Gesetzesfolgenabschätzungen?*

Forschungsarbeiten zur Umsetzung der GFA in Deutschland zeigen sehr deutlich, dass eine integrierte GFA – also eine Abschätzung aller intendierten und nicht-intendierten Gesetzesfolgen – durch das federführende Ministerium nicht passfähig zum existierenden Prozess der Gesetzbereitung durch die Ministerialbürokratie ist. Die Ursachen hierfür sind vielfältig²⁷, besonders relevant ist jedoch die Tatsache, dass für das federführende Ressort keine positiven Anreize existieren, Transparenz über Gesetzesfolgen herzustellen. Im Gegenteil: Eine ausführliche Dokumentation der Gesetzesfolgen und Alternativen besitzt aus Sicht des Ressorts das Potential, die regierungsinterne und später die parlamentarische Durchsetzungsfähigkeit eines Gesetzentwurfes zu gefährden. Die ausführlichere Dokumentation möglicher Gesetzesfolgen und Diskussion von Alternativen in Schweden ist primär durch die Institution des Kommissionswesens und die lange Tradition des Öffentlichkeitsprinzips zu erklären. Da in Deutschland keine vergleichbare institutionalisierte Vorstrukturierung von Politikgehalten außerhalb der Ministerialbürokratie stattfindet, sind die Widersprüche zwischen einer integrierten GFA und den dominierenden Handlungsrationaltäten in der Ministerialverwaltung (z.B. Herstellung von Durchsetzungsfähigkeit für Gesetzentwürfe, informationelle Entlastung der Leitungsebene) besonders virulent.

27 Siehe hierzu Veit (Fn. 24), S. 241-311.

Diese Widersprüche sind ursächlich für die in Deutschland stark ausgeprägte Tendenz zur „formalen Anpassung“ an GFA-Vorschriften.

Diese Diagnose des strukturell bedingten Scheiterns gilt für die Durchführung integrierter GFAs durch Fachministerien, nicht jedoch für extern erstellte Folgenabschätzungen (Beispiel: Kommissionswesen in Schweden) und auch nicht für Versuche einer systematischen Abschätzung einzelner Folgenaspekte. Wie das Beispiel der Bürokratiekostenabschätzung zeigt, ist eine verstärkte Berücksichtigung einzelner Folgenaspekte (z.B. der durch Gesetze verursachten administrativen Lasten für Unternehmen) implementierbar, wenn organisatorische Zuständigkeiten definiert werden und eine verfahrensmäßige Integration in den Rechtsetzungsprozess stattfindet. Vorschriften zur Meta-Regulierung sind also durchaus dafür geeignet, die Argumentationsfähigkeit bestimmter Akteure im *Policy*-Prozess zu erhöhen²⁸ und damit die systematische Berücksichtigung bestimmter Folgenaspekte im regierungsinternen Verfahren der Gesetzesvorbereitung zu erhöhen. Organisationsentscheidungen spielen hierfür eine erhebliche Rolle: Folgenaspekte, für deren Prüfung niemand zuständig ist, werden weder formal noch tatsächlich erfüllt. Diese Tatsache belegen die empirischen Daten zur Umsetzung der Vorschriften zur ex post-Evaluation und zur Alternativendarstellung sehr deutlich. Gleichzeitig ist darauf hinzuweisen, dass verbindliche Verfahrensvorgaben und methodische Vorgaben zur Durchführung von Folgenabschätzungen zwar die Implementation unterstützen, aber auch erheblichen Aufwand verursachen können.

Was bedeutet dies für die GFA als einen Ansatz zur Umsetzung des Konzepts der evidenzbasierten Politik? Gesetzesfolgenabschätzungen sind aus dieser Perspektive nicht primär dazu geeignet, die Wissensbasis politischer Entscheider im Hinblick auf *alle* relevanten, intendierten und nicht-intendierten Gesetzesfolgen zu verbessern und eine Alternativenauswahl anzuleiten. Ihr wesentlicher Zweck besteht auch nicht darin, die Kommunikation zwischen Regierung und Parlament zu verbessern, wie in der Literatur verschiedentlich diskutiert. Stattdessen sind zwei andere Aspekte hervorzuheben: Erstens können GFA-Vorschriften die Entscheidung darüber, *zu welchen Folgenaspekten* in welchem Umfang Wissen systematisch nutzbar gemacht und innerhalb des PAS archiviert wird, beeinflussen. Und zweitens können Gesetzesfolgenabschätzungen – und das ist ein Aspekt, der in der bisherigen akademischen Debatte zur GFA kaum eine Rolle spielt und zukünftig besser untersucht werden sollte – verwaltungsinterne Prozesse des organisationalen Lernens strukturieren.

28 Werner Jann/Kai Wegrich: Wie bürokratisch ist Deutschland? Und warum? Generalisten und Spezialisten im Entbürokratisierungsspiel, in: der moderne staat (dms), 1 (2008), S. 49-72.

Erfahrungen mit Bewertungen und Evaluation bei den EU-Strukturfonds

Norbert Heller

1. Sagenhaft

Wie sich die 13. Fee an Dornröschen, deren Eltern und deren ganzem Gefolge rächte, das wissen sie. Viel raffinierter noch war aber Eri, die griechische Göttin des Streites und der Zwietracht. Sie befand sich in einer ganz ähnlichen Lage wie die 13. Fee. Hatte doch Peleus sie anders als die anderen Göttinnen leichtsinniger Weise nicht zu seiner Hochzeit eingeladen. Peleus wurde viel später der Vater des berühmten Achilles, der mit der Achillesferse. Aber das ist eine andere Geschichte!

Die nicht zur Feier geladene Göttin der Zwietracht, machte ihrem Namen alle Ehre. Sie warf einen goldenen Apfel unter die Gäste. Dieser trug die Aufschrift "Für die Schönste!" Die Göttinnen Hera, Athene und Aphrodite beanspruchten den Titel für sich. Das rief nun förmlich nach Bewertung und einem Evaluator. Und schon damals zeigte sich, dass das Bewerten eine heikle Sache sein kann. Zeus, um seine Beurteilung gebeten, lehnte dankend ab. Er, der Göttervater, fürchtete den Zorn der unterlegenen Göttinnen. So bestimmte er den trojanischen Königssohn Paris zum Schiedsrichter. Jede der Göttinnen versuchte, diesen für sich zu gewinnen. Hera versprach ihm politische Macht und die Vormachtstellung in Asien, Athene Weisheit und Kriegsglück, Aphrodite aber versprach ihm Helena. Diese war zwar schon verheiratet, aber die schönste Frau der Welt. Wie die Evaluation ausging, wissen Sie: Aphrodite gewann das Ranking und Paris die schöne Helena. Das Ergebnis war der trojanische Krieg und der Untergang Trojas!

Bewerten – und Bewertet-Werden! Beides war also schon immer keine leichte Sache! Bei allem Fortschritt, zumindest bei den EU-Strukturfonds, hat sich dies bis heute nicht groß geändert.

2. EU-Strukturfonds

Die EU-Strukturfonds als Gegenstand der Bewertung. EU-Strukturfonds, was ist das?

Fonds, fondis – also Quelle und in unserem Zusammenhang eine Geldquelle! Geld der Europäischen Union, das diese von den Mitgliedsstaaten erhält. Genau genommen handelt es sich um ca. 308 Mrd. € in Preisen von 2004, also knapp 350 Mrd. € in jeweiligen Preisen über die wir hier für den Zeitraum von 2007 – 2013 sprechen. So viel Geld! Das schreit geradezu nach Bewertung! Wie aber schon das Urteil des Paris beweist, ist das keine leichte Sache.

Erst einmal ist die Summe von 350 Mrd. € richtig zu erfassen. Sie erscheint schon überschaubarer, wenn Sie sie durch 9 teilen. 9, weil die Mittel innerhalb von 9 Jahren auszugeben sind. Das macht pro Jahr dann durchschnittlich knapp 40 Mrd. €. Wenn Sie diesen Betrag nun auf nahezu 300 europäische Regionen verteilen relativiert er sich noch weiter. Vergleichen Sie die Summen mit den Beträgen der Bankenrettung nach 2008 oder der Euro-Rettungsschirme. Alles relativ! Was sind jährlich 40 Mrd. € EU-Strukturfondsmittel im Vergleich zu einem Gesamtvolumen der Haushalte aller EU-Mitgliedsstaaten von mehr als 6.300 Mrd. € oder einer EU-Wirtschaftsleistung von mehr als 14.000 Mrd. €?

Nichtsdestoweniger sind 350 Mrd. € eine große Summe, von denen auch Deutschland immerhin 25 Mrd. € empfängt, 3,4 Mrd. davon erhält das Land Sachsen-Anhalt, das wie die anderen ostdeutschen Länder derzeit noch zu den Konvergenzregionen zählt. Konvergenzregionen erhalten die höchste Förderung aus Brüssel. Dies wird sich für die ostdeutschen Länder für die Periode 2014 – 2020 ändern.

Interessanter aber für die Frage der Bewertung ist die Frage nach dem Verwendungszweck dieser Mittel. Den Hinweis gibt das Wort "Struktur" im Begriff EU-Strukturfonds. In allen Europäischen Mitgliedsstaaten sollen die Lebensverhältnisse und Strukturen vergleichbar sein. Städtische Gebiete sollen zu städtischen, ländliche Gebiete zu ländlichen Gebieten annähernd gleich sein. Ziel ist die Kohäsion, also der wirtschaftliche und soziale Zusammenhalt der Mitgliedsstaaten der EU. Mit anderen Worten, ein Auseinandertriften oder gar Auseinanderbrechen der Union soll verhindert werden. Die EU-Strukturfondsmittel sollen nachhaltig wettbewerbsfähige Strukturen fördern.

Erreichen sollen dies der Europäische Fonds für Regionale Entwicklung (EF-RE). Der Europäische Sozialfonds (ESF) und der Kohäsionsfonds (KF). Daneben gibt es noch den Europäischen Fonds für die Entwicklung im Ländlichen Raum (ELER) und den Europäischen Fischereifonds. Beide wollen nicht mit den EU-Strukturfonds zusammen gebracht werden, sondern lieber ihre eigenen Wege gehen.

Die EU-Strukturfondspolitik hat allerdings grundsätzlich eine integrierende Funktion. Die EU-Strukturfonds liegen horizontal über den vertikalen Politikbereichen der fachpolitischen Sektoren und verbinden diese mit integrierten Ansätzen.

3. Bewertungskriterien

Wirkung und Ergebnis der eingesetzten Mittel sehen die Befürworter des Mitteleinsatzes als positiv an. Gegner der Kohäsionspolitik bewerten diese naturgemäß kritisch und beklagen die Verschwendung von Steuermitteln. Wer hat Recht?

Eine einfache Methode, den Mitteleinsatz zu bewerten, wäre das Bauchgefühl. Diese erscheint dann allerdings doch als zu „gefühlsmäßig“ und zudem auch noch sehr subjektiv. Es reicht schließlich nicht nur eine Entscheidung zu treffen, sondern diese allen oder wenigsten einer Mehrheit an Hand von nachvollziehbaren Kriterien zu begründen. Zu unterscheiden sind die verschiedenen Ebenen der Bewertung: Die Makroebene, die Projektebene und die Programmebene.

3.1 Makroebene und makroökonomische Modelle

Ziel der Finanztransfers sind der wirtschaftliche und soziale Zusammenhalt, die Kohäsion. Es bietet sich also an, bei Fragen nach der Zielerreichung bei Kriterien anzusetzen, die Auskunft über den Grad an Zusammenhalt geben. Die Verteilung der Mittel erfolgt nach dem Entwicklungsstand der europäischen Regionen. Die Kriterien dafür sind das Bruttoinlandsprodukt pro Kopf und die Beschäftigtenquote. Es bietet sich also an, das Ergebnis und die Wirkung des Fördereinsatzes mit Hilfe dieser Indikatoren zu erfassen.

Die Entwicklung des Bruttoinlandsproduktes pro Kopf und die Beschäftigung werden regelmäßig gemessen. Es stellt sich nun aber die Frage, ob diese Entwicklung vom Einsatz der EU-Strukturfondsmittel beeinflusst worden ist und, wenn ja, in wie weit.

Makroökonomische Modelle versuchen, die Wirkung der Förderung und deren Nachhaltigkeit zu erfassen. Die zu untersuchende Volkswirtschaft wird dazu in dem Modell abgebildet. Mit Hilfe des Modells kann man dann darstellen, wie sich eine Volkswirtschaft mit und ohne Förderung entwickelt hat und entwickelt hätte. Um die Nachhaltigkeit darzustellen werden diese Rechnungen als Prognosen in die Zukunft fortgeschrieben.

Im Auftrag der Bundesregierung untersuchte beispielsweise das GEFRA-Institut im Jahr 2010 die Bedeutung der EU-Strukturfonds für die Förderperiode von 2014 bis 2020 für Ostdeutschland. Prämissen waren, dass das bisherige

Förderszenario bezüglich Höhe, Einsatzkategorien und Verteilung im Zeitverlauf beibehalten wird.

Es kommt zu einem jährlichen Impuls von 0,6 % des ostdeutschen BIP. Das BIP und die Erwerbstätigkeit liegen im Implementierungszeitraum (2000 – 2020) mit 1,50 % deutlich über dem Basisansatz und langfristig bis 2030 ergibt sich beim BIP noch ein Plus von 0,7 %, bei der Erwerbstätigkeit immerhin noch von 0,4 %.

Gegenüber diesem Förderszenario im Zeitraum 2000 – 2020 reduzieren sich BIP und Erwerbstätigkeit ohne diese Förderung im Durchführungszeitraum um bis zu 0,9 % jährlich. Im Zeitraum danach bis 2030 immerhin das BIP noch um 0,3 % und die Beschäftigung um 0,15 %.

Insgesamt zeigt dieses Szenario, dass die Niveauverluste im BIP und in der Erwerbstätigkeit insbesondere in der Implementationsphase anfallen, während sie in der Post-Implementationsphase moderater sind. Bei der Bewertung dieser Ergebnisse ist zu beachten, dass wie erwähnt der jährliche Impuls in der Implementationsphase auch „nur“ 0,6 % des ostdeutschen BIP beträgt.

Mit dem Modell lässt sich also belegen, dass mehr dabei herauskommt, wenn man mehr hineinsteckt – ja, sogar mehr als man hineingesteckt hat! Auch ein nachhaltiger Effekt auf der Makroebene lässt sich belegen. Es lohnt sich danach, die Mittel einzusetzen.

Die ostdeutschen Länder und das Bundeswirtschaftsministerium sowie diejenigen Generaldirektionen der EU-Kommission, die für diese Förderung verantwortlich sind, benutzten solche Ergebnisse in der Debatte um die Fortführung der EU-Strukturfondsförderung überhaupt.

So ermittelte Ergebnisse sind immer nur so gut wie die Modellierung. Jeder Modellierung liegen Annahmen zu Grunde. Diese Annahmen könnten auch falsch sein. Auch erzeugen makroökonomische Modelle umso verlässlichere Ergebnisse, umso größer die betrachtete Volkswirtschaft ist. Die EU mit Ihren gut 500 Mio. Einwohnern und ihrem Bruttoinlandsprodukt von mehr als 14.000 Mrd. EUR lässt sich deutlich besser modellieren als Ostdeutschland oder gar das Land Sachsen-Anhalt. Je kleiner eine Volkswirtschaft ist, desto größer ist die Wirkung exogener Einflüsse.

3.2 Projektebene

Wenn die Bewertung der Makroeffekte das eine Ende des Untersuchungsgegenstandes darstellt, dann ist die Ebene des Projektes das andere Ende. Wenden wir uns nun dem anderen Ende zu.

3.2.1 Projektanzahl und Arten der Projekte

Förderprojekte oder auch Projekttypen gibt es viele. So förderte beispielsweise das Land Sachsen-Anhalt in der Förderperiode 2000 – 2008 mit einem EU-Strukturfondsbeitrag von 3,5 Mrd. € ca. 60.000 Förderprojekte mit einem Gesamtinvestitionsvolumen von 12 Mrd. €. Die Palette der Förderung reicht dabei von Großprojekten großer Unternehmen über die Förderung kleiner und mittlerer Unternehmen, über wirtschaftsnahe Infrastrukturen hin zu ergänzenden Infrastrukturen im städtischen Bereich, über den Umwelt- oder Bildungsbereich weiter zur Landwirtschaftsförderung und die Förderung des ländlichen Raums im allgemeinen bis hin zu Qualifizierungs- und ähnlichen Projekten.

Neben der Bewertung eines einzelnen Projektes lassen sich alle ähnlichen Projekte bewertend miteinander vergleichen. Dies kann dann Aufschluss geben, welche Eigenschaften, Voraussetzungen oder Verfahrensweisen über den Erfolg eines Projektes entscheiden können.

3.2.2 Input-Indikator

Ein einzelnes Projekt heraus zu greifen und dieses zu bewerten, scheint einfacher zu sein. Doch auch dies ist nicht ohne Tücken. Einfach ermitteln – und vor allem einheitlich – lässt sich immer die Fördersumme, d.h. der Input.

3.2.3 Output-Indikator

Was dann mit dem Geld gemacht wird, der Output also, ist schon nicht mehr so klar, aber auch noch vergleichsweise einfach zu beschreiben. Es werden Produktionsanlagen oder Infrastrukturen gebaut oder Schulungsmaßnahmen durchgeführt. Die Indikatoren, die zur Messung herangezogen werden, sind nun aber nicht einheitlich, sondern sehr heterogen und lassen sich nicht vergleichen.

- Arbeitsplatzkapazität einer Produktionsanlage
- Anzahl der gebauten Straßenkilometer
- Länge eines Abwassernetzes
- Kapazität einer Tiermastanlage
- Länge der gebauten Hochwasserschutzanlage
- Teilnehmerzahl einer Eingliederungs- und Schulungsmaßnahme für Schulabbrecher
- Teilnehmerzahl einer Qualifikationsmaßnahme für Postgraduierte

Das sind nur einige wenige Beispiele. Da es nun nicht darum geht, jedes Projekt für sich zu bewerten, sondern alle Projekte einer Förderaktion, empfiehlt es sich

dringend, sich bereits im Vorfeld über die Projekttypen klar zu werden und sich mit den Verantwortlichen auf eine klare Beschreibung und Definition der Indikatoren zu verständigen. Dabei muss es das Ziel sein, die Zahl der insgesamt verwendeten Output-Indikatoren zu optimieren. Lässt man den Verantwortlichen auf den Umsetzungsebenen freie Hand bei der Benennung und Definition der zu verwendenden Indikatoren landet man bei schnell bei 1000 oder mehr Indikatoren. Dies zumindest bei einem regionalen deutschen OP, das alle Fonds umfasst!

3.2.4 Ergebnisindikator

Doch was nützt es, die Anzahl von Teilnehmern, die Kapazitäten von Produktionsanlagen oder die Länge von Infrastrukturen zu erfassen? Neben Werten für den Inputindikator, können wir z.B. die gebauten Straßenkilometer als Output angeben. Aber: Auch für eine Straße, die niemand braucht, auf der niemand fährt, gibt es einen In- und Output-Wert! Wir wissen wie viel Geld wir eingesetzt haben. Wir wissen, dass da ein Unternehmen für dieses Geld so und so viele Kilometer einer Straße gebaut hat. Was aber ist das Ergebnis der Förderung? Es muss also neben Output-Indikatoren auch Ergebnisindikatoren geben. Im Straßenbeispiel also z.B. die Fahrzeugfrequenz auf der Straße und die durchschnittlich Fahrzeit. Bei der Deichanlage, dass sie auch den Jahrhunderthochwasser standhält.

Im Deichbau ist das einfach: Wir warten ab. Im Straßenbau erscheint das schon schwieriger, aber noch lösbar. Woran aber messen wir das Ergebnis einer Qualifizierungsmaßnahme, des Baues einer Produktionsanlage den einer Messförderung oder einer Unternehmensberatung?

Theoretisch fällt uns dazu bestimmt eine Menge ein. Es liegt mir aber daran, die Aufmerksamkeit auf ein damit verbundenes Problem zu lenken – gerade auch bei den Ergebnisindikatoren. Das Problem der Erhebbarkeit. Die Erhebung muss praktisch zu leisten und natürlich auch finanzierbar sein.

Das Ergebnis der Messförderung müsste sich an Hand der Geschäftsabschlüsse und der Steigerung des Umsatzes zeigen, das Ergebnis der Unternehmensberatung ebenfalls am Umsatz und am Gewinn. Die Schulungsmaßnahme müsste in eine Beschäftigung münden oder sogar in eine bessere Beschäftigung als dies ohne die Förderung möglich gewesen wäre. Um dieses Ergebnis zu messen, müssten die Projekte über einen längeren Zeitraum betrachtet werden. Die Projektträger müssten zur Auskunft z.B. über ihre Umsatz- und Gewinnentwicklung oder ihre Berufslaufbahn bereit sein. Zumindest nicht alle sind auskunftswillig, insbesondere nicht die Unternehmen und schon gar nicht über einen län-

geren Zeitraum. Erstens handelt es sich um betriebsinterne Daten und Zeit ist schließlich Geld.

Dann ist dann da noch das Problem der Masse. Von den 60.000 Projekten des Zeitraums 2000 – 2006 waren 15.000 ESF-Projekte. Mit diesen wurden 300.000 Projektteilnehmer gefördert. Natürlich ist Masse für sich genommen kein Problem. Mit einem guten Datenbanksystem lässt sich das alles bewältigen. Bleibt nur das Problem, dass irgendjemand die Daten des Teilnehmers über einen definierten längeren Zeitraum erheben und eintragen muss. Wenn der Teilnehmer aber seinen Qualifizierungslehrgang beendet, löst sich der Kontakt zwischen Bildungsträger und Teilnehmer mehr oder weniger schnell. Ob der Teilnehmer dann z.B. eine dauerhafte Stelle gefunden hat oder nach kurzer Zeit wieder auf der Straße steht, lässt sich nur schwer oder gar nicht nachverfolgen.

Die EU-Kommission sieht die Probleme bei den Ergebnisindikatoren sehr wohl. Bei den Output-Indikatoren gibt es immerhin eine Liste von Kernindikatoren. Bei den Ergebnisindikatoren hält sich die EU-Kommission von Beispielen abgesehen zurück. Sie legt die Verantwortung für diese Übung in die Hände der Mitgliedsstaaten. Dies ist gleichzeitig ein Beleg dafür, dass das Subsidiaritätsprinzip auch für die EU vorteilhaft sein kann.

3.2.5 Zielsetzung und Zielerreichung auf Projektebene

Will man ein Ergebnis bewerten, dann ist die Zielsetzung, die man verfolgt, wichtig. Wenn ich weiß, was ich will, ergeben sich daraus die Bewertungskriterien. Bei der Antragstellung sollte man wissen, welche Kriterien für die Förderfähigkeit, die Projektauswahl und für die Projektbewertung am Ende wichtig sind.

Es stellt sich aber ebenso wie auf der Makroebene die Frage nach dem Kausalzusammenhang. Ist es die Förderung, die das Ergebnis bewirkt hat oder war etwas anderes ursächlich? Um das herauszufinden empfiehlt die EU den Einsatz von kontrafaktischen Vergleichsgruppen. Idealtypischer Weise benötigt man zwei identisch zusammengesetzte Gruppen, von denen die eine gefördert, die andere nicht gefördert wird. Man verfolgt die Entwicklung beider Gruppen und erhält somit Hinweise, ob die Förderung für eine bestimmte Entwicklung in den Gruppen ursächlich ist oder nicht. Der Gedanke Vergleichsgruppen einzusetzen ist gut. Der damit verbundene Aufwand ist jedoch nicht unerheblich. Auch eignen sich bestimmte Projekttypen besser, andere dagegen nicht für Vergleichsgruppen. Oft ist es auch nicht leicht eine entsprechende Vergleichsgruppe überhaupt zu finden und mit dieser parallel eine Bewertung durchzuführen.

Beispielhaft ist hier der Innovationsvoucher eines Mitgliedsstaats zu nennen. Von den 4.000 eingegangenen zulässigen Anträgen kleiner und mittelständischer

Unternehmen ermittelte die zuständige Stelle 2.000 Zuwendungsempfänger per Losverfahren. Auf diese Weise entstanden zwei vergleichbare Untergruppen der Grundgesamtheit. Nun ließen sich statistisch sehr elegant die geförderte und die nicht geförderte Gruppen im Zeitverlauf vergleichen und bewerten. Bei anderen Konstellationen ist der Einsatz dieser Methode dagegen viel aufwendiger oder unmöglich.

3.3 Programmebene

Richtig interessant wird es, wenn man sich zwischen die Makro- und die Projektebene begibt. Das heißt auf die Ebene der Operationellen Programme. Hier werden alle Projekte oder ein Teil eines Programmes auf den Prüfstand gestellt. Operationelle Programme sind dabei untergliedert in Prioritäten und Aktionen. Ziel dieser Übung auf Programmebene kann es sein, unterschiedliche Programme miteinander vergleichen zu können. Es ist aber auch erforderlich alle Programmteile einheitlich zu bewerten, wenn es zu Umschichtungen im Verlauf der Förderperiode kommen muss. Dies erfolgt etwa bei Halbzeitbewertungen oder aber, wenn sich die Notwendigkeit ergibt oder der Wunsch besteht, die Verteilung der Mittel zu ändern.

3.3.1 Inputindikator und Ausgabekategorien

Bei der Bewertung der Programmumsetzung bedient man sich immer auch des Indikators Input. Der bis zum Evaluierungszeitpunkt getätigte Input im Verhältnis zum verfügbaren Budget ist die Absorptionsrate.

Der Input-Indikator ist in allen Programmen, Prioritäten und Projekten einheitlich. Er ist einfach zu ermitteln. Allerdings sagt der Mittelabfluss alleine eben nicht sehr viel aus. Der Indikator besagt nur, dass Geld für die Prioritäten eines Programmes über die dazugehörigen Projekte geflossen ist.

Da die Prioritäten und Aktionen in den unterschiedlichen Programmen der EU-Regionen sehr unterschiedlich und daher nicht vergleichbar sind, hat die EU ein System von Ausgabekategorien vorgegeben. Sie will den Input zusätzlich zur Untergliederung nach Prioritäten nach den einheitlichen Ausgabekategorien ermitteln.

Die Ausgabekategorien kann man bisher in solche mit und solche ohne Lissabon-Relevanz – u.a. Forschung und Entwicklung oder Information und Kommunikation – unterscheiden. Konvergenzregionen müssen als Zielwert mindestens 60 % ihrer Ausgaben Lissabon-relevant tätigen, Regionen des Ziels Wettbewerbs und Beschäftigung müssen 75 % erreichen. Die Agenda 2020 setzt vergleichbare

Ziele und unterteilt nun die Kategorien in solche nach intelligentem Wachstum, nachhaltigem Wachstum und integrativem Wachstum.

Beispielsweise für den EFRE wird zurzeit eine Quotierung der Ausgaben diskutiert. In stärker entwickelten Regionen und in Übergangsregionen sollen mindestens 80 % der Mittel für Energieeffizienz und erneuerbare Energien, Forschung, Innovation und KMU-Unterstützung eingesetzt werden. 20 % davon sollen wiederum für Energieeffizienz und erneuerbare Energien verwendet werden. Da die Regionen, die nicht mehr unter das Ziel „Konvergenz“ fallen, nach wie vor Umstrukturierungsbedarf aufweisen, wird der Mindestsatz auf 60 % verringert. Für weniger entwickelte Regionen soll die Mindestquote für Energieeffizienz und erneuerbare Energien, Forschung, Innovation und KMU-Unterstützung 50 % der Mittel betragen; 6 % davon sollen für Energieeffizienz und erneuerbare Energien reserviert sein.

Die EU versucht somit, den Mittelabfluss über die Kategorisierung der Ausgaben in bestimmte „erwünschte“ Kanäle zu lenken.

Die Absorptionsrate für sich genommen könnte zu unzutreffenden Schlüssen führen. Mittel müssen noch nicht ausgegeben sein, sie können aber trotzdem bereits fest für Projekte gebunden sein. Insbesondere größere Projekte bedürfen einer längeren Umsetzungszeit. Es empfiehlt sich daher nicht nur auf die bisher getätigten Zahlungen, sondern auch auf die versprochenen Zahlungen, d.h. die gebundenen Mittel zu schauen.

Was in einem Operationellen Programm festgeschrieben ist oder für Projekte gebunden wird, muss aber auch kontrolliert werden. Bei auffälligen Abweichungen muss die Verwaltungsbehörde des Operationellen Programmes nachhaken. Auffällig könnte z.B. sein, dass zwar bereits alles Geld gebunden ist, aber kaum Auszahlungen erfolgt sind. Besteht die Auffälligkeit zu Recht oder zu Unrecht? Es könnte sein, dass das Geld nur „pro forma“ gebunden wurde, damit andere an die noch nicht verausgabten Mittel nicht so leicht herankommen. In solchen Fällen müsste eingeschritten werden. Dazu muss sich die Verwaltungsbehörde Informationen zu dieser Priorität oder Aktion beschaffen.

Dieses Beispiel macht auch deutlich, dass es nicht ausreicht, sich auf die Zahlen aus dem Monitoring einer Datenbank zu verlassen. Es werden immer auch weitere Informationen benötigt, die unter anderem auch durch vertiefende Evaluationen ermittelt werden können.

3.3.2 Output und Ergebnis

Interessant ist es neben dem Input auch Output und Ergebnis mehrerer Projekte zu vergleichen. Bei gleichen Projekten kann man die Frage stellen, ob und, falls ja, warum es bei gleichem Input zu unterschiedlichem Output und Ergebnis

kommt. Welche Lehren lassen sich aus der Betrachtung ziehen? Evaluation ist ja nicht Selbstzweck, sondern soll uns helfen, in Zukunft unsere Entscheidungen auf besserer Grundlage zu treffen.

Bei unterschiedlichen Projekttypen stellt sich die Frage, wie man diese miteinander vergleichen kann. Warum sollte man aber Äpfel und Birnen vergleichen? Nun, weil bei der Mittelverteilung und jeder Mittelumverteilung die Frage beantwortet werden muss, wer wie viel vom noch verfügbaren Budget zugeteilt bekommen soll. Vernünftiger Weise sollte man die Verteilung so vornehmen, dass sie das beste Gesamtergebnis verspricht.

Wenn sich die Projekttypen nicht miteinander vergleichen lassen, dann könnte eine Gruppierung der Projekttypen, z.B. entsprechend den Ausgabekategorien der EU, erfolgen. Projekttypen die auf die Verbesserung der Wettbewerbsfähigkeit ausgerichtet sind, müssten sich miteinander messen lassen; genauso wie alle Projekttypen, die dem Schutz und dem Erhalt der Umwelt dienen oder solche, die die Beschäftigung erhöhen wollen. Dazu wäre es nützlich, einen Ergebnisindikator oder wenige Ergebnisindikatoren pro Gruppe zu finden. Diese Indikatoren müssen von allen oder doch von vielen Projekttypen einer Gruppe bedient werden können.

3.3.3 Zielsetzung und Zielüberprüfung

Die Überlegungen zum angestrebten Ergebnis sind bereits bei der Aufstellung eines Operationellen Programmes zu berücksichtigen. Dies gilt für jede seiner Prioritätensetzungen und Aktionen, die es untergliedern. Prioritätensetzung ist sinnvoller Weise an Hand von Zielen vorzunehmen. Schwer tut sich dann der, der gar kein inhaltliches Ziel oder nur sehr globale Ziele hat. Wer kein Ziel hat, ist orientierungslos. Wer nicht weiß, wo er ankommen will, kann jeden beliebigen Weg einschlagen und auch überall landen. Evaluation wird dann zu einer lästigen und überflüssigen und durch aus nicht Kosten günstigen Übung. Wer kein Ziel vor Augen hat, kann auch nicht bewerten, ob er sein Ziel erreicht.

Die Tatsache fehlender oder sachfremder, unklarer, unbestimmter und nicht quantifizierter Ziele ist das Haupthindernis der Bewertung.

Hat man Ziele, muss man bei der Setzung von Prioritäten oder zumindest von Aktionen aufpassen, dass es nicht zu Zielkonflikten kommt. Es macht keinen Sinn, mit einer Aktion Projekte mit unterschiedlicher Zielrichtung fördern zu wollen. So schließen sich z.B. die Ziele der Erhöhung der Produktivität als auch das der Erhöhung der Zahl der Arbeitsplätze aus. Will man beide Ziele mit dem Operationellen Programm verfolgen, benötigt man unterschiedliche Aktionen. Mit anderen Worten es sind unterschiedliche Wege einzuschlagen, um unterschiedliche Ziele zu erreichen.

Sind klare Ziele definiert, kann man den Zielerreichungsgrad für die Indikatoren (Input, Output, Ergebnis) der Aktionen überprüfen. Man kann bei der Betrachtung auch die bereits erfolgte Absorption und die verbindliche Mittelbindung für weitere, bereits konkret geplante Ausgaben berücksichtigen. Haben die Prioritäten eines operationellen Programms unterschiedliches politisches Gewicht, kann man dieses auch noch über einen entsprechenden Faktor in die Bewertung einfließen lassen. So ermittelt sich für jede Aktion ein entsprechender Wert.

3.3.4 Umverteilungsentscheidungen

Die Voraussetzungen der Zielsetzung müssen gegeben sein. Es lassen sich über ein Regelwerk Umschichtungsvorschläge errechnen. Vorgeschlagen wird, dass diejenigen Aktionen mit schlechten Werten Mittel abgeben müssen, diejenigen mit guten Werten sollen dagegen zusätzliche Mittel erhalten. Im mittleren Bereich sollen die Ansätze konstant bleiben.

Ein solch rechnerischer Umverteilungsvorschlag unterliegt selbstverständlich vielfältiger Kritik. Trotzdem gibt dieser Ansatz für eine Verwaltungsbehörde von EU-Strukturfondsmitteln eine gute Grundlage für die schwierige Aufgabe einer Programm-Neujustierung. Viele der zu berücksichtigenden Fakten sind bereits in die Überlegungen eingeflossen. Die Beweislast, dass etwas doch anders zu bewerten sei, liegt nun bei den Verantwortlichen für eine Aktion. Diesen gefällt der Vorschlag ggf. ja nicht. Über Stellschrauben dieses Systems lassen sich Änderungen manuell einpflegen. Dabei werden sofort auch die Auswirkungen auf die anderen Ebenen sichtbar. Man muss sich nämlich auch klar darüber sein, dass die Umverteilungsmasse umso kleiner wird, wenn man weniger bei anderen wegnimmt.

3.3.5 Gleiche Bedingungen

Wer klare Ziele pro Aktion hat, kann dann auch unterschiedliche Aktionen über ihren Zielerreichungsgrad mit einander vergleichen. Diese Vorgehensweise setzt allerdings voraus, dass die Erfüllung der gesetzten Ziele tatsächlich gemessen werden kann und gemessen wird.

Notwendig setzt dies natürlich voraus, dass die Zielsetzung in den unterschiedlichen Aktionen einheitlich realistisch erfolgte. Dies erfolgt nicht immer. Wenn es darum geht, überhaupt an die Fördermittel zu gelangen, sind Aspiranten der Förderung bemüht die zu erwartenden Ergebnisse in möglichst rosigen Farben zu beschreiben. Sind mit der Nichterreichung der Ergebnisse allerdings Nachteile, mit einer Erreichung oder Übererreichung Vorteile zu erreichen, dann

versuchen die Akteure die Zielwerte deutlich niedriger anzusetzen. Dies macht eine Zielerreichung oder Übererfüllung und damit Boni doch viel wahrscheinlicher. So hätten die Landwirtschaftsreferate des Landes Sachsen-Anhalt bei oberflächlicher Betrachtung der Ergebnisse zur Halbzeit des Programmes 2000 – 2006 die damals zur Verteilung anstehende Effizienzreserve komplett erhalten müssen. Allerdings konnte im konkreten Fall die Verwaltungsbehörde dann darauf hinweisen, dass die Landwirtschaftsreferate bei der Zieldefinition mehr als vorsichtig verfahren waren.

Noch besser wäre es, wenn diese Messlatte direkt und konkret schon bei der Mittelverteilung abgefordert wird. Dann nämlich kann sich der politische Entscheider bereits bei der Verteilung an den gesetzten Zielen messen und nicht erst wenn es darum geht den Erfolg zu messen.

3.3.6 Theorie und Praxis

Soweit so gut! In der Praxis kommt es zu erheblichen Problemen. Das fängt bei der Auswahl und Beschreibung der Indikatoren an. Auch der Faktor Mensch bei der Erhebung, Eingabe und Pflege der Werte ist nicht zu unterschätzen. So werden z.B. selbst einfach scheinende Indikatoren oft einfach nicht richtig geführt.

Beispielsweise kam bei einer Wirtschaftlichkeitsprüfung seitens des Europäischen Rechnungshofes zu Tage, dass statt der tatsächlich geschaffenen Arbeitsplätze immer nur die im Bescheid angegebenen geführt wurden, wenn diese niedriger waren. Grund war, dass die Höhe der Förderung auch von der Zahl der geschaffenen Arbeitsplätze abhängig ist. Leistet der Zuwendungsempfänger mehr als er muss, ist das natürlich unschädlich. Würde aber der Indikatorwert nach oben gesetzt, würde dies bei einem späteren Absinken im elektronischen System der zwischengeschalteten Stelle einen Teilwiderruf auslösen. Diese Auswirkung oder eine zusätzliche Erfassung wollte die zuständige Stelle verhindern. Allerdings ohne dies zu kommunizieren! Sie entschloss sich, immer den im Bescheid genannten Indikatorwert zu führen und diesen bei Änderungen nach oben nicht anzupassen. Im Ergebnis wurde dieser Indikator für die Richtung nach oben, also im Falle der Zielübererfüllung, bereits auf Projektebene unbrauchbar. Eine Bewertung des Zielerreichungsgrades auf Aktionsebene war damit nicht möglich.

Auch die Wissenschaft und die Evaluationsinstitute spielen eine eigene Rolle. Es entsteht insgesamt der Eindruck, dass diese gerne die Systeme und Methoden immer weiter verfeinern möchten. So kommt man z.B. weg von statischer Betrachtungsweise und hin zu einer dauernden begleitenden Bewertung. Dies setzt natürlich ein sehr gutes Monitoringsystem und eine sehr gute Datenbanklösung voraus. Die Bewertung des Erfolgs eines Operationellen Programmes ist immer

auch eine Bewertung der politischen Beschlüsse, die zur Mittelverteilung in diesem Programm geführt haben. Immerhin besteht die Möglichkeit, ein gesetztes Ziel auch zu verfehlen. Wer aber möchte in einem solchen Fall seine Beschlüsse, seine Politik und damit sich selbst wirklich bewerten lassen?

4. Restimee

Wer würde es angesichts dieser vielen kleinen und großen methodischen oder praktischen Probleme nicht wie Göttervater Zeus machen wollen? Der konnte es sich leisten, sich zu drücken. Aber wie Paris sehen wir uns zur Entscheidung in der Bewertungsfrage gezwungen. Er wollte in den Genuss seines verlockenden Preises kommen, wir in den Genuss der EU-Strukturfonds! Es bleibt uns wie ihm also nichts anderes übrig als die Schönste der Schönen zu küren.

Nicht jeder aber findet das gut. Nicht jeder möchte sich in die Karten schauen lassen oder seine Ziele offen legen. Dies gilt für die Projektebene und für die Programmebene, das gilt auf der Arbeits- und auf der politischen Ebene.

Bewerten finden alle ja gut. Nicht wenige würden es aber gerne vermeiden, selbst, insbesondere negativ, bewertet zu werden. Es ist schon schwer genug, das Geld rechtmäßig und im Zeitrahmen auszugeben! Bewertung der Förderfälle bedeutet immer auch weitere Arbeit. Bewertung bedeutet darüber hinaus eine Bewertung nicht nur des Projektes, sondern eine Bewertung dessen, der die Förderung gewährt oder diese politisch verantwortet.

Zu guter Letzt muss sich aber selbst die Evaluation auch bewerten lassen. Bei aller Verfeinerung und Neuerfindung von Methoden befinden wir uns nicht am Ziel. Wir befinden uns allenfalls auf dem Weg!

Die Beobachtung der Methodik der EU-Strukturfondsförderung zeigt, dass die Bewertungskultur durch die Kohäsionspolitik über die EU-Strukturfondsverordnungen sehr gefördert wird. Bessere Bewertungsverfahren sind bestimmt nicht verkehrt und ganz im Sinne des Steuerzahlers, insbesondere des Steuerzahlers in den Nettozahler-Mitgliedsstaaten! Mögen unsere Kriterien zu besseren Ergebnissen führen als die des Paris aus Troja!

Qualitätsmanagement in der Sozialverwaltung

Werner Keggenhoff

1. Das Landesamt

Das Landesamt für Soziales, Jugend und Versorgung Rheinland-Pfalz ist an den Standorten Mainz, Koblenz, Landau und Trier vertreten. Die Zentrale ist Mainz. Als Fachverwaltung stützt sich das Landesamt auf entsprechendes Fachpersonal. Insgesamt sind dies 906 Fachleute mit unterschiedlicher Vorbildung:

Ärzte	34
Pharmazeuten	11
Juristen	28
Pädagogen, Psychologen	13
Sozialpädagogen, Sozialarbeiter	45
Verwaltungsmitarbeiter	703
Mitarbeiter EDV	10
Wirtschaftswissenschaftler	18
Auszubildende	44

Während die klassische Verwaltung stark von Juristen dominiert ist, gibt es in dieser Fachverwaltung mehr Sozialpädagogen (45) und auch mehr Ärzte (34) als Juristen (28). Das Fachwissen dieser Mitarbeiterinnen und Mitarbeiter aus den unterschiedlichen Bereichen ist die Basis für unsere Arbeit als Fachverwaltung.

2. Haushalt

Das Landesamt trug im Jahr 2011 Verantwortung für ein Haushaltsvolumen von 2.702.835.273 €. Diese Ausgaben beinhalten EU-, Bundes- und Landesmittel. Den größten Teil der Landesmittel bilden dabei die Ausgaben in der Sozialhilfe mit 894.388.297 € und die Ausgaben für Kindertagesstätten mit 434.563.915 €. Einen großen Anteil der Bundesmittel machen immer noch die Aufwendungen für die Kriegsopferversorgung/-fürsorge mit 98.283.540 € aus.

3. *Unsere Aufgaben*

Die Aufgaben des Landesamts sind sehr vielfältig. Die Zuständigkeit reicht von A wie Adoptionen, Apothekenwesen, Arbeitsmarktprojekte, Ausgleichsabgabe bis Z wie zahnärztliche Prüfungen und Zentrale Medizinische Untersuchungsstelle. Die Themen reichen von der Wiege bis zur Bahre und darüber hinaus. Das geht von der Förderung der Schwangerenberatung und der Kindertagesstättenberatung bis zur Aufsicht über die Altenheime und zur Versorgung von Kriegsbeschädigten und deren Witwen. Der Aufgabenbestand ist dynamisch. Dem Landesamt werden immer wieder neue Aufgaben wie das Landestariftreuegesetz, die Anlauf- und Beratungsstelle für ehemalige Heimkinder und die Projekte gegen Rechtsextremismus übertragen.

4. *Kernaufgabe „Menschen mit Behinderungen“*

Das Landesamt ist die Fachverwaltung des Landes für behinderte Menschen. In Rheinland-Pfalz sind rund 417.000 Menschen, also gut 10 % der Bevölkerung als schwerbehindert anerkannt.

Dem Landesamt obliegen die Feststellung der Schwerbehinderteneigenschaft, die Erteilung des Schwerbehindertenausweises, die Integration behinderter Menschen in Arbeit, aber auch die Finanzierung stationärer und teilstationärer Hilfen und die Aufsicht über die Heime.

5. *Aktuell*

Das Landesamt bleibt in Bewegung und stellt sich Veränderungsprozessen. Dies ist eine wichtige Grundhaltung, denn nur eine moderne, effizient arbeitende, kostengünstige und zugleich bürgerfreundliche Dienstleistungsverwaltung wird angesichts sich verschärfender Haushaltsengpässe die gesellschaftlichen und sozialen Herausforderungen der Zukunft bewältigen. Aktuell steht das Projekt „Zukunft Landessozialverwaltung“ zur Umsetzung an. Der Landtag Rheinland-Pfalz hat dazu das Landesgesetz zur Eingliederung der Ämter für soziale Angelegenheiten in das Landesamt für Soziales, Jugend und Versorgung beschlossen, das am 1.1.2013 in Kraft tritt. Zeitgleich wird ein neues EDV-Verfahren zur Durchführung der Schwerbehinderten-Feststellung eingeführt, das kürzere und klarere Bescheide ermöglicht. Schließlich tritt ein neuer Schwerbehindertenausweis im Scheckkartenformat an die Stelle der etwas unhandlichen Ausweise.

6. *Verwaltungsmodernisierung*

Verwaltungsmodernisierung ist eine Daueraufgabe, der sich das Landesamt seit seiner Gründung am 1.1.1996 verschrieben hat. Dahinter steht die Überzeugung, dass nur eine reformbereite Verwaltung in der Lage ist, sich wechselnden Herausforderungen zu stellen und bei knappen finanziellen Möglichkeiten bestmögliche Ergebnisse zu erzielen. Bei der Zusammenlegung des Landesversorgungsamtes und des Landesamtes für Jugend und Soziales zum Landesamt für Soziales, Jugend und Versorgung war die Vorgabe, dass 30 Prozent der Personalstellen binnen 10 Jahren abzubauen sind. Dieser Auftrag konnte erfüllt werden, indem Verwaltungsstrukturen vereinfacht und verbessert wurden.

Modernisierungsmaßnahmen erfolgten weitestgehend aus eigener Kraft und wurden unter Beteiligung der Mitarbeiterinnen und Mitarbeiter durchgeführt. Projektarbeit mit eigenen Mitteln erwies sich als besonders erfolgreich, weil der Sachverstand der Betroffenen eingebracht und genutzt werden konnte.

7. *Betriebswirtschaftliche Elemente*

Wichtiger Bestandteil der Veränderungsprozesse war der Einsatz betriebswirtschaftlicher Methoden. Im Jahr 1999 wurde die Kostenstellenrechnung eingeführt, die Budgetierung der Personal- und Sachkosten sowie Investitionen im Jahr 2000 und die Kosten- und Leistungsrechnung im Jahr 2006. Vor der Vergabe des Auftrags für eine geeignete Software für die Kosten- und Leistungsrechnung wurden die Anforderungen wiederum in Projektarbeit selbst entwickelt. Den Zuschlag erhielt die Firma Mach AG. Als später landesweit in der Verwaltung die Kosten- und Leistungsrechnung eingeführt wurde, kam wiederum diese Firma zum Zuge. Das Landesamt hatte also den Vorteil, eine KLR nach eigenen Vorstellungen zu erhalten, und musste nicht eine für andere Bereiche entwickelte Software übernehmen.

8. *Auswertungsbogen*

Mit Hilfe der Kosten- und Leistungsrechnung können die angefallenen Kosten im Landesamt verursachungs- und periodengerecht zugeordnet und ausgewertet werden. Die vorgegebenen Budgets werden in regelmäßigen Abständen kontrolliert und in Controllinggesprächen mit Führungskräften monatlich vorgestellt und erörtert. Bei Abweichungen wird steuernd eingegriffen.

Mit der Kosten- und Leistungsrechnung wird der Verbrauch an personellen und sächlichen Ressourcen erfasst. Die alleinige Betrachtung der Kosten einer

Behörde reicht aber nicht aus, um Erkenntnisse zu Effektivität und Effizienz der erstellten Verwaltungsprodukte zu erhalten. Um Aussagen darüber treffen zu können, müssen neben den Kosten auch die Quantität und Qualität einer Dienstleistung und deren Bestimmungsfaktoren betrachtet werden.

9. Qualitätssicherung

In einzelnen Fachbereichen des Landesamtes bestehen schon länger Verfahren zur Betrachtung der Qualität. Dies gilt z.B. für die Pharmazie, die an einem bundesweit gültigen Qualitätssicherungssystem teilnimmt, wobei Begutachtungen der Arzneimittelüberwachung – insbesondere der Durchführung von Inspektionen und der Umgang mit Arzneimittelrisiken – vorgenommen werden.

Ziel war nun, eine breit angelegte kontinuierliche Betrachtung der Qualität in den maßgeblichen Aufgabenbereichen des Landesamtes einzuführen.

10. Aufbau unseres Qualitätsmanagements

Es sollte ein auf Kennzahlen gestütztes Wertesystem entwickelt werden, mit dem der Gesamtprozess der Leistungserstellung des Landesamtes betrachtet werden kann. Damit verbunden war die Grundsatzentscheidung, die Verwaltung über strategische Ziele zu steuern, die mittels Kennzahlen kontrolliert werden. Wirkungsindikatoren helfen, die Behörde zu steuern. Daher werden Indikatoren benötigt, die neben den Kosten auch Quantität und Qualität beschreiben, damit diese gemessen werden können.

Die Frage „Wie messen wir Qualität?“ wurde im Vorfeld umfassend erörtert. Aus dem vorhandenen Produktkatalog des Landesamts wurden Modellprodukte, u. a. das größte Pilotprodukt „Teilhabe schwerbehinderter Menschen – Feststellungsverfahren nach § 69 SGB IX“ ausgewählt.

11. Ziele und Kennzahlen

In Arbeitsgruppen zu den jeweiligen Produkten wurden Ziele festgelegt, die die Qualität und Quantität der Produkte beschreiben. Diese Ziele wurden mit Hilfe von Kennzahlen in eine einheitliche Skalierung mit Werten von 1 bis 10 übersetzt. Der Wert 1 steht für ein sehr schlechtes Ergebnis, der Wert 10 für ein fehlerfreies Ergebnis. Hinter diesen Werten stehen je Produkt die jeweiligen Istwerte zu den zuvor gebildeten Zielen.

Die Anzahl der Ziele ist von Produkt zu Produkt unterschiedlich. Liegen mehrere Ziele vor, wird festgelegt, welche Wichtigkeit das jeweilige Ziel einnimmt, und hierfür eine Gewichtung hinterlegt, so dass am Ende eine Kennzahl die Qualität bestimmt. Die Kosten und die Quantität werden ebenfalls mit Hilfe einer Kennzahl bewertet, woraus sich dann gemäß einer entsprechenden Gewichtung mit 30 % Qualität, 20 % Quantität und 50 % Kosten eine Gesamtkennzahl ergibt.

12. Indikatoren mit Gewichtung

Die Qualität des Beispielprodukts „Teilhabe schwerbehinderter Menschen – Feststellungsverfahren nach § 69 SGB IX“ wird mit Hilfe von fünf qualitativen Indikatoren bestimmt:

1. Einhaltung der Personal- und Leistungskennzahlen : 15%
(Fallzahl pro Mitarbeiter)
2. Einhaltung des Beweiserhebungskostenbudgets : 15%
(Zahl der pro Verfahren angeforderten kostenpflichtigen Befundberichte)
3. Korrekte Rechtsanwendung : 30%
(Entspricht die Anerkennungsquote der Erstanträge den langjährigen Erfahrungswerten?)
4. Nachvollziehbarkeit der Entscheidungen: 20%
(Zahl der gegen die Entscheidung eingelegten Widersprüche)
5. Bearbeitungsdauer: 20%
(Ist der vorgegebene Zeitraum eingehalten, in dem mindestens 50% der eingegangenen Anträge erledigt werden sollen?)

Diese Kennzahlen und die dazugehörige Skalierung wurden aufgrund der Erfahrungen und bereits vorhandener Statistiken bestimmt.

13. Weitere Indikatoren

Die Quantitätskennzahl wird durch die Anzahl der Anträge bestimmt. Die Kosten je Produkt werden aus der Kosten- und Leistungsrechnung ermittelt. Aufgrund der einheitlichen Kennzahlenvorgabe hinsichtlich Gewichtung der Ziele und Skalierung sind Vergleiche der Ergebnisse an den vier Standorten möglich. Um eine aussagekräftige Steuerung zu gewährleisten, ist eine Betrachtung aller drei Säulen, Kosten, Qualität und Quantität, notwendig.

Die vorhandene KLR-Software wurde dahingehend erweitert, dass zu den tatsächlichen Kosten je Produkt auch die Kennzahlen zu den Kosten, der Quantität und der Qualität abgebildet werden.

14. Auswertung

Die erfassten Informationen gehen in einen Auswertungsbogen ein, in dem die Kennzahlen auf einen Blick zu erkennen sind. Diese Kennzahlen werden von dem jeweiligen Produktverantwortlichen zu jedem Produkt quartalsweise gebucht. Dabei spielen auch Planwerte eine Rolle, welcher Wert bei den jeweiligen Zielen zu erreichen sein sollte.

Dieser Auswertungsbogen erfüllt den Zweck, dass die vorhandenen Kennzahlensysteme leicht ausgewertet werden können.

Mit Hilfe dieses Auswertungsbogens erhält die Dienststellenleitung schnelle Hinweise auf die wichtigsten Kennzahlen. Eine dezentrale Auswertung findet durch die Produktverantwortlichen und den Bereich des Controllings statt. Auch diese Auswertungen werden regelmäßig vorgestellt, um bei Abweichungen von vereinbarten Zielen steuern zu können und Planwerte einzuhalten. Dies setzt voraus, dass Vergleiche von Kennzahlenwerten vorgenommen werden. So bieten sich durch den Auswertungsbogen wertvolle Erkenntnisse zur Analyse der eigenen Kennzahlenwerte und der zu Grunde liegenden Ursachen. Aus diesen Analysen lässt sich Handlungsbedarf erkennen. In der Anfangsphase wurden auch Erkenntnisse darüber gewonnen, ob angestrebte Ziele realistisch sind oder gegebenenfalls Skalierungen angepasst werden müssen.

15. Probleme

Der Auswertungsbogen hat gerade bei diesem Produkt kritische Tendenzen aufgezeigt. Die Kennzahlen haben sich aufgrund steigender Antragszahlen bei knapperen Personalressourcen verschlechtert. Im Bereich des Feststellungsverfahrens gehen so viele Anträge ein wie nie zuvor. Dabei ist bei knappen Ressourcen an Geld und Personal abzuwägen, ob zugunsten gründlicher Bearbeitung eher längere Bearbeitungszeiten hingenommen werden sollen.

16. Qualitätssicherung im Vergleich

Eine der Erkenntnisse der Qualitätssicherung im Bereich des Feststellungsverfahrens ist, dass weitaus weniger Rechtsmittel eingelegt werden als im Bereich anderer Sozialleistungsträger. Außerdem ist festzustellen, dass in anderen Ländern, die diese Aufgabe kommunalisiert haben und vorrangig auf die Kosten achten, stark divergierende Entscheidungen zustandekommen.

17. Weitere Produkte im Qualitätsmanagement

Folgende 11 weitere Pilotprodukte werden im Landesamt bisher mit Kennzahlen zu Kosten, Qualität und Quantität bewertet:

- Kündigungsschutz/Teilhabe schwerbehinderter Menschen im Arbeitsleben
- Hilfen für Menschen in besonderen sozialen Schwierigkeiten
- Gutachten/Stellungnahmen im Schwerbehindertenrecht
- Betriebserlaubnisse für Kindertagesstätten
- Investitions- und Personalkostenzuschüsse des Landes für Kindertagesstätten
- Vereinbarung der Vergütungen (Behindertenhilfe)
- Personalplanung, -ab- und -zugang
- Landestariftreuegesetz
- Landesgesetz über Wohnformen und Teilhabe (früher Heimgesetz)
- Bewilligung und Begleitung von arbeitsmarktpolitischen Projekten
- ESF-Bescheinigungsbehörde

18. Qualitätsmanagement

Diese Vorgehensweise bei der Bewertung der Produkte mit Hilfe der Kennzahlen zu Kosten, Quantität und Qualität wurde durch die Deutsche Gesellschaft zur Zertifizierung von Managementsystemen nach der ISO 9001:2008 im Jahr 2011 zertifiziert. Dabei wurden nicht nur die Kennzahlen unserer Verwaltungsleistungen betrachtet. Vielmehr wurden die gesamten Prozesse des Landesamtes analysiert. Die Zertifizierung bewertet neben der Wirtschaftlichkeit die Produkt- und Prozessqualität sowie die Kunden- und Mitarbeiterzufriedenheit. So wurden auch die Bereiche der organisatorischen Gliederung, der personellen Ressourcen und die Arbeitsplatzumgebung im Zertifizierungsverfahren betrachtet. Die Wirksamkeit des Systems wurde Anfang 2012 erfolgreich ein zweites Mal begutachtet. Damit ist bescheinigt, dass das Landesamt mit seinem System auf dem richtigen Weg ist.

19. Fazit

Das Landesamt sieht sich als Fachverwaltung besonders gefordert, neben den Kosten auch der Qualität einen hohen Stellenwert einzuräumen. Dies ist für eine Fachverwaltung von existentieller Bedeutung.

Möglichkeiten von Leistungsmessung und Leistungsvergleichen in öffentlichen Verwaltungen – ein zusammenfassender Kommentar

Jörg Bogumil

Ich möchte im Folgenden einige grundsätzliche Anmerkungen zu dem Thema dieses Blockes machen. Das Gemeinsame, worüber wir reden, ist die Frage, inwieweit man mit Leistungsmessung und Leistungsvergleichen dazu beitragen kann, politische Entscheidungen und politische Steuerung insgesamt zu verbessern. Bringen uns Leistungsmessungen und Leistungsvergleiche neue, wichtige und bessere Informationen, die wir vorher nicht hatten, und helfen diese Informationen, politische Entscheidungen zu optimieren?

Die Untersuchung von Wirkung und Folgen von politischem Handeln ist in Deutschland keine Neuigkeit, wenngleich Evaluationen in der politischen Praxis eher selten sind. Seit Anfang der 90er Jahre hat sich die Diskussion im Zusammenhang mit der Einführung von Public Management- und Performance Management-Elementen intensiviert. Vor acht Jahren habe ich schon einmal eine Zwischenbilanz zu den Verfahren der Leistungsmessung und des Leistungsvergleiches in Deutschland gezogen. Ich hatte damals formuliert, dass diese Verfahren bis jetzt nur wenig Handlungskonsequenzen zeitigen, da es kaum Anreize für ein Lernen aus Vergleichsringen gibt angesichts fehlender verwaltungsinterner und verwaltungsexterner Öffentlichkeit, weil die deutsche Verwaltungskultur auf Fehlersuche und -vermeidung ausgerichtet ist (und nicht vergleichbar ist z.B. mit der skandinavischen Verwaltungskultur) und weil wir keinerlei Verknüpfungen mit Finanzverteilungsverfahren haben und insofern der Druck auf die Verwaltungen gering ist, zumal es in vielen Bereichen für die Abnehmer von Verwaltungsleistungen keine Exit-Option gibt (Bogumil 2004, S. 396).

Diese Einschätzung war also eher pessimistisch. Jetzt sind acht Jahre vergangen, wir haben viele neue Verfahren, den Normkontrollrat und Ähnliches. Insofern, glaube ich, können wir insgesamt ein bisschen positiver in die Welt schauen. Wichtig ist jedoch die Art und Weise, wie solche Verfahren institutionalisiert werden. Bevor ich hierauf zu sprechen komme, möchte ich zunächst noch einmal auf zwei Grundprobleme von Leistungsmessungen in öffentlichen Verwaltungen aufmerksam machen.

Das eine Grundproblem ist das *Messbarkeitsproblem*. Albert Einstein hat einmal gesagt: Nicht alles, was gezählt werden kann, zählt und nicht alles, was zählt, kann gezählt werden. Dem ist eigentlich nichts hinzuzufügen. Zwar ist es völlig unstrittig, dass wir durch viele quantitative Daten hilfreiche Informationen

haben, die möglicherweise Hinweise auf eine Verbesserung politisch-administrativen Handelns geben. Wir müssen aber feststellen, dass die Art und Weise der Leistungsmessung und Leistungsvergleiche extrem quantifiziert ist, weil wir nur auf das zurückgreifen, was wir haben, und das ist in der Regel quantifiziert. Ob man das so intensiv machen muss wie bei dem vorhin dargestellten Beispiel aus der EU (Beitrag Heller, in diesem Band), da habe ich so meine Zweifel, denn dann stellt sich die Frage der Transaktionskosten doch in erheblichem Umfang.

Das zweite Grundproblem ist noch wichtiger: Wir unterstellen, dass mit einer besseren Informationsversorgung quasi automatisch die politische Steuerung besser wird. Das kann so sein, aber das muss nicht so sein, weil eine Verbesserung der politischen Steuerung nicht nur von besseren Informationen abhängt, sondern immer auch *untrennbar mit Machtprozessen verbunden* ist. Ergebnisse von Leistungsvergleichen und Leistungswettbewerben werden in jeder Organisation mikropolitisch interpretiert und machtpolitisch genutzt. Wir haben es bei politischen Entscheidungsprozessen nicht mit einem herrschaftsfreien Raum zu tun. Aaron Wildavsky, hat mal gesagt, dass Benchmarks „weapons in the political wars“ sind und weiter „I started out thinking it was bad for organizations not to evaluate, and I ended up wondering why they ever do it“ (Aaron Wildavsky 1979, S. 212). Hier möchte ich ihm uneingeschränkt zustimmen.

Wenn man nun diese beiden Problemlagen berücksichtigt, das Problem der Messbarkeit und das Problem des Machtprozesses, und dennoch glaubt, Leistungsmessung und Leistungsvergleiche können sinnvoll sein – und ich glaube das –, dann muss man sich die Frage stellen, wie institutionalisiere ich diese Verfahren so, dass sie einen fruchtbaren Effekt haben können? Die *Institutionalisierung von Leistungsmessung und Leistungsvergleichen ist entscheidend* und Sylvia Veith hat auf diese Institutionalisierungsfragen z.B. beim Normenkontrollrat hingewiesen. Die Grundidee dort ist, dass man Fachpolitiken versucht zurückzudrängen, indem man ein Gremium schafft, welches eine zentrale Steuerungslogik schafft. Man könnte jetzt mal überprüfen, wie erfolgreich das ist, ob mit dieser Form der Institutionalisierung also die intendierten Effekte erreicht wurden.

Mit der institutionellen Ausgestaltung in verschiedensten Ländern hat sich im letzten Jahr Sabine Kuhlmann sehr schön auseinander gesetzt und gezeigt, dass neben der institutionellen Ausgestaltung auch Faktoren wie Verwaltungskultur oder das Ausmaß von Parteienwettbewerb wichtig sind (Kuhlmann 2011). So funktioniert vieles, was in der Schweiz als erfolgreich gilt, z.B. in Deutschland nicht, wo das Ausmaß an Parteienwettbewerb deutlich größer ist.

Bezüglich der Institutionalisierung wird in einer neueren Veröffentlichung ausgeführt (Jäkel/Kuhlmann 2013), dass man in Deutschland von freiwillig initiierten, mit geringer Transparenz und keiner Sanktion verbundenen Verfahren ausgehen kann und von einem mittleren Verbreitungsniveau. In Schweden und

England haben wir eine hohe Transparenz und eine Veröffentlichung solcher Vergleichsverfahren mit allerdings unterschiedlichen Konsequenzen. In Schweden gibt es freiwillige Verfahren mit informellen Sanktionen und in England verpflichtende Verfahren mit finanziellen Sanktionen. Wir wissen auch, dass dies in England nicht zu sinnhaften Prozessen führt, da der Anreiz bei direkten finanziellen Sanktionen hoch ist, dass die Verwaltung anfängt „zu lügen“ bzw. die Zahlen ein bisschen zu schönen.

Bezüglich dieser Institutionalisierungsformen möchte ich nun auf eine Form der Weiterentwicklung in Deutschland zu sprechen kommen. Prinzipiell kann eine Leistungsmessung *intern* oder *extern* durchgeführt werden, sie kann *freiwillig* oder *erzwungen* und *mit* oder *ohne Sanktionen* versehen sein. Zudem kann man die Ergebnisse *transparent* machen oder nicht. Es gibt in Deutschland ein relativ neues Modell, welches ich vielversprechend finde, das ist die Art der Institutionalisierung der *Gemeindeprüfungsanstalt in Nordrhein-Westfalen* (vgl. ausführlich Ebinger/Bogumil 2012, Bogumil/Ebinger 2013). Dies ist eine neue Institutionalisierungsform, *extern*, *verpflichtend* für die Kommunen und *transparent* (da die Ergebnisse dem Rechnungsprüfungsausschuss vorgestellt werden), aber *ohne Sanktionen*.

Sicherlich wird der Punkt „verpflichtend“ am meisten kritisch diskutiert. Ich glaube allerdings, dass Lernprozesse oder Weiterentwicklung von öffentlichen Verwaltungen in Deutschland ganz ohne einen externen Druck nicht vorankommen. Damit will ich Führungskultur und andere wichtige Personalentwicklungsinstrumente nicht herunterspielen, aber wir tun auch in Deutschland gut daran, ein bisschen mehr darüber nachzudenken, wie wir Verbindlichkeit herstellen. Der Trick bei diesem Institutionalisierungsmix ist, dass die Verpflichtung nicht dazu führt, dass man versucht die Ergebnisse zu schönen, denn die Ergebnisse werden zwar öffentlich gemacht, sind aber nicht mit formalen oder finanziellen Konsequenzen verbunden. Die Ergebnisse der Leistungsmessungen und Leistungsvergleiche der Gemeindeprüfungsanstalt erhält die Kommune mit ihrem Ranking, sie hat dann aber die Möglichkeit, gegenüber der Politik zu begründen, warum sie so ausgefallen sind. Sie kann sich kritisch damit auseinandersetzen und braucht keine direkten Konsequenzen zu fürchten. Dies sorgt dafür, dass ein produktiver Umgang möglich ist.

Wenn also Leistungsmessung und Leistungsvergleiche in öffentlichen Verwaltungen vorangebracht werden sollen, ist eine intelligente Institutionalisierung wichtig. Was man zudem aus den bisherigen Erfahrungen lernen kann ist, dass *ausufernde Berichte zu vermeiden* sind, da sie zu erheblichen Transaktionskosten führen. Ein gutes Beispiel hierfür sind die Universitäten. Was dort in den Fakultäten an Berichten erstellt werden muss, die danach kaum jemand liest oder kontrolliert, ist beachtlich. Wenn man wirklich mit Berichtswesen steuern will, dann sollte man auf ganz wenige Maßzahlen beschränken, die dann aber auch kontrol-

liert werden können (vgl. hierzu Bogumil u.a. 2013). Zudem muss ich, wenn ich vergleiche, „fair“ *vergleichen*. Die Bundesagentur für Arbeit macht das mittlerweile besser, indem sie jetzt nicht jede Arbeitsagentur miteinander vergleicht, sondern innerhalb bestimmter Cluster, die über ähnliche Rahmenbedingungen verfügen. Ähnlich kann man auch vorgehen, wenn Haushaltsergebnisse von Kommunen miteinander verglichen werden.

Zum Schluss soll noch ein letzter Punkt angesprochen werden, der wahrscheinlich umstritten ist. Man sollte vorsichtig mit dem Versprechen sein, dass es durch Performance-Management zu einer besseren politischen Steuerung kommt. So könnte man an Herrn Heller die Frage stellen, ob aus den ganzen Zahlen, die Sie uns präsentiert haben, irgendeine Konsequenz für die Veränderung der EU-Strukturmittelvergabe resultiert? Und was passiert eigentlich mit den ganzen Daten der Doppik in den Kommunen? Gibt es hier Hinweise, dass dadurch die politische Steuerung verbessert wird (kritisch dazu Bogumil/Holtkamp 2012)? Dies alles heißt nicht, dass man diese Instrumentarien nicht braucht, aber ein bisschen mehr Bescheidenheit und Realitätssinn im Umgang mit ihnen könnte nicht schaden.

Literaturhinweise

- Bogumil, Jörg, 2004: Probleme und Perspektiven der Leistungsmessung in Politik und Verwaltung, in: Kuhlmann, Sabine/Bogumil, Jörg/Wollmann, Hellmut (Hrsg.), Leistungsmessung und -vergleich in Politik und Verwaltung. Konzepte und Praxis, Wiesbaden: VS, S. 392-398.
- Bogumil, Jörg/Burgi, Martin/Heinze, Rolf G./Gerber, Sascha/Gräf, Ilse-Dore/Jochheim, Linda/Schickentanz, Maren/Wannöfel, Manfred, 2013: Modernisierung der Universitäten. Umsetzungsstand und Wirkungen neuer Steuerungsinstrumente, Berlin: edition sigma.
- Bogumil, Jörg/Ebinger, Falk, 2013: Leistungsvergleiche zwischen Lernen und Kontrolle. Die Institutionalisierung von Innovation durch die Gemeindeprüfungsanstalt NRW, erscheint in: „der moderne staat – Zeitschrift für Public Policy, Recht und Management“, Sonderheft "Wissen und Expertise in Politik und Verwaltung". Im Erscheinen.
- Bogumil, Jörg/Holtkamp, Lars, 2012: Doppik in der Praxis. Bisher vor allem intransparent und ineffizient, in: Verwaltung & Management, Heft 3, 18 Jg. S. 115-117.
- Ebinger, Falk/Bogumil, Jörg, 2012: Angeleitetes Lernen über Leistungsvergleiche? Zwischenbilanz nach acht Jahren Gemeindeprüfungsanstalt NRW, in: Die Verwaltung, 45(1), S. 123-140.
- Jäkel, Tim/Kuhlmann, Sabine, 2013: Lernen von den Besten? Steuerung und Nutzung von Leistungsvergleichen in europäischen Verwaltungen, in: „der moderne staat – Zeitschrift für Public Policy, Recht und Management“, Sonderheft "Wissen und Expertise in Politik und Verwaltung". Im Erscheinen.
- Kuhlmann, Sabine, 2011: Messung und Vergleich von Verwaltungsleistungen: Benchmarking-Regime in Westeuropa, in: Die Verwaltung, 2/2011, S. 155-178.
- Wildavsky, Aaron, 1979: Speaking Truth to Power. The Art and Craft of Policy Analysis, Boston/Toronto.

Reflexive Reformprozesse und politische Selbstkorrekturfähigkeit: Erfolgsbedingungen von Evaluationen am Beispiel der „Hartz- Gesetze“

Margrit Seckelmann¹

1. Ausgangsfrage: Erfolgsbedingungen von Evaluationen

Beschäftigt man sich mit der Wirkungsforschung von jüngeren Gesetzen in Deutschland, so kommt man an den arbeitsmarktpolitischen Reformen nicht vorbei, welche unter dem Obertitel „Hartz-Gesetze“ zusammengefasst werden. ‚Taufpate‘ für diese Gesetze war der Vorsitzende der ihre Grundlagen entwickelnden Regierungskommission „Moderne Dienstleistungen am Arbeitsmarkt“, der damalige VW-Personalvorstand Peter Hartz. Die Regierungskommission hatte vor der Bundestagswahl 2002 publikumswirksam ihre Vorschläge in 13 Modulen vorgelegt, die in der Folgezeit in vier Gesetze eingingen, die in diesem Beitrag näher vorgestellt werden sollen. Zwei dieser Gesetze („Hartz I“² und „Hartz II“³) traten zum 1. Januar 2003, ein weiteres zum 1. Januar 2004 („Hartz III“⁴) und ein letztes, das besonders umstrittene sog. „Hartz IV-Gesetz“⁵, zum 1. Juli 2004 in Kraft.

Die mithilfe dieser Gesetze vorgenommene Umstrukturierung des Systems der Arbeitsvermittlung wie der Hilfsleistungen kann man mit Recht als umfassendste Arbeitsmarktreform der Bundesrepublik Deutschland bezeichnen.⁶ Sie vollzog sich vor dem Hintergrund einer veränderten Staatsauffassung, des sog. „aktivierenden Staats“, auf die im Folgenden noch zu sprechen zu kommen sein wird.

- 1 Ich danke *Wolfram Lamping* für seine hilfreichen Anregungen und seine Mitarbeit bei der Vortragsversion sowie *Hans Peter Bull* für seine wertvollen Hinweise.
- 2 Erstes Gesetz für moderne Dienstleistungen am Arbeitsmarkt, BGBl. I/2002 Nr. 87, S. 4607.
- 3 Zweites Gesetz für moderne Dienstleistungen am Arbeitsmarkt, BGBl. I/2002 Nr. 87, S. 4621.
- 4 Drittes Gesetz für moderne Dienstleistungen am Arbeitsmarkt, BGBl. I/2003 Nr. 65, S. 2848.
- 5 Viertes Gesetz für moderne Dienstleistungen am Arbeitsmarkt, BGBl. I/2003 Nr. 66, S. 2954.
- 6 In diesem Sinne etwa *Klaus Brandner*: Evaluation Hartz I-III. Mehr Evaluation und mehr Qualität. Power-Point-Präsentation, Friedrich-Ebert-Stiftung, Berlin 15. März 2006.

Nachfolgend soll allerdings nur ein spezieller Aspekt der sog. „Hartz-Reformen“ näher untersucht werden. Und zwar enthielten die „Hartz-Gesetze“ eine explizite politische Selbstverpflichtung zur wissenschaftlichen Evaluation. Bereits im Entschließungsantrag der Koalition aus SPD und Bündnis 90/Die Grünen vom 14. November 2002 befand sich ein Passus zur umfassenden Evaluation bezüglich der „Neuorganisation der Bundesanstalt für Arbeit“, der „Weiterentwicklung der arbeitsmarktpolitischen Instrumente“, der „Änderungen des Arbeitnehmerüberlassungsgesetzes“, der „Auslagerungen von Aufgaben auf Dritte“ und der „Akzeptanz der Bundesanstalt für Arbeit bei den Kunden“ innerhalb von drei Jahren nach Inkrafttreten.⁷

Im Trend der Zeit zur Einführung sog. „sunset clauses“, also der Befristung von Gesetzen, wie sie auf Landesebene schon teilweise praktiziert wurde,⁸ lag der Griff zu einer zeitlichen Befristung zahlreicher Reform-Instrumente. Dieser wurde verbunden mit einem auf Bundesebene⁹ relativ neuartigen gesetzgeberischen Instrument, der frühzeitigen und gesetzlich verbindlichen Evaluation der Gesetze. Ein weiterer Anlass zur Aufnahme evaluativer Elemente in das Gesetz war der Umstand, dass sich Bund und Länder sowie Regierung und Opposition im Bund nicht auf ein bestimmtes Modell zur Organisation der Verwaltung der Leistungen nach dem Sozialgesetzbuch hatten einigen können. Denn der Deutsche Landkreistag hatte recht erfolgreich die Opposition sowie die CDU- bzw. CSU-geführten Länder dafür gewinnen können, dass nicht nur die Arbeitsagen-

- 7 *Deutscher Bundestag*: Entschließungsantrag der Fraktionen SPD und Bündnis 90/Die Grünen zu der dritten Beratung des Gesetzentwurfs der Fraktionen SPD und Bündnis 90/Die Grünen zum Entwurf eines Ersten Gesetzes für moderne Dienstleistungen am Arbeitsmarkt, Drs. 15/98 (14. 11. 2002): „Der Bundestag wolle beschließen: [...] II. [...] 5. Der Deutsche Bundestag fordert die Bundesregierung auf, die Umsetzung der Hartz-Vorschläge sowie des Zweistufenplans der Bundesregierung insgesamt zeitnah evaluieren zu lassen. Neben der Neuorganisation der Bundesanstalt für Arbeit müssen auch die Weiterentwicklung der arbeitsmarktpolitischen Instrumente, die Änderungen des Arbeitnehmerüberlassungsgesetzes, die Auslagerungen von Aufgaben auf Dritte und die Akzeptanz der Bundesanstalt für Arbeit bei den Kunden umfassend untersucht werden. Erste belastbare Ergebnisse der Untersuchungen sollen nach drei Jahren vorliegen“.
- 8 *Werner Jann/Kai Wegrich/Sylvia Veit*: Verfahren und Instrumente erfolgreicher (De-)Regulierung, in: Stefan Empter/Frank Frick/Robert B. Vehrkamp (Hrsg.), Auf dem Weg zu moderner Regulierung. Eine kritische Bestandsaufnahme, Bielefeld/Gütersloh 2005, S. 47-77; *Bastian Janz/Sylvia Veit*: Bessere Rechtsetzung durch Befristungs- und Evaluationsklauseln? Empfehlungen zur Befristung von Bundesgesetzen in Deutschland auf Basis einer empirischen Auswertung internationaler Erfahrungen, Gütersloh 2010.
- 9 Allerdings waren (und sind) seit dem Jahr 2000 nach den §§ 43, 44 der *Gemeinsamen Geschäftsordnung der Bundesministerien* (GGO) die Folgen eines Gesetzes *ex ante* zu prüfen und auf dem Gesetzesvorblatt darzustellen, vgl. dazu u. a. *Götz Konzendorf*: Institutionelle Einbettung der Evaluationsfunktion in Politik der Verwaltung in Deutschland, in: Thomas Widmer/Wolfgang Beywl/Carlo Fabian (Hrsg.), *Evaluation. Ein systematisches Handbuch*, Wiesbaden 2009, S. 27-39, 32 ff.

turen, sondern auch die Kommunen (und in der Praxis zumeist die Landkreise) zu tauglichen Trägern der SGB II-Verwaltung erklärt werden sollten. Man hatte sich – wie unter 2.6.2.4 im Einzelnen dargestellt – letztlich darauf verständigt, beide Optionen zuzulassen und bis zu 69 Landkreisen oder kreisfreien Städten (entsprechend der Zahl der Sitze im Bundesrat) auf Antrag die alleinige Trägerschaft der Grundsicherung für Arbeitssuchende zuzuerkennen. Dieser Kompromiss war mit einer „Experimentierklausel“ in § 6a SGB II ‚abgesichert‘ worden, die auf sechs Jahre befristet war (vom 1. Januar 2005 bis zum 31. Dezember 2010). Auf Basis einer gesetzlich vorgeschriebenen Evaluation gem. § 6c SGB II sollte das Bundesministerium für Wirtschaft und Arbeit (später: für Arbeit und Soziales) „die Wahrnehmung der Aufgaben durch die zugelassenen kommunalen Träger im Vergleich zur Aufgabenwahrnehmung durch die Agenturen für Arbeit“ überprüfen und „den gesetzgebenden Körperschaften des Bundes bis zum 31. Dezember 2008 über die Erfahrungen mit den Regelungen nach den §§ 6a und 6b“ SGB II berichten, wobei die Länder „bei der Entwicklung der Untersuchungsansätze und der Auswertung der Untersuchung“ beteiligt werden sollten. Der Bundesgesetzgeber sollte aufgrund dieses Berichts eine Entscheidung über die Fortführung des Optionsmodells fällen.

Durch die Festschreibung der *vergleichenden Begleitforschung* in § 6c SGB II a. F., der Evaluation in § 6c SGB II sowie die Verpflichtung zur Evaluation bestimmter arbeitsmarktpolitischer Instrumente im Wege einer *laufenden Wirkungsforschung* in § 55 SGB II und der *Arbeitsmarkt- und Berufsforschung* nach § 282 SGB III in das „Hartz IV“-Gesetzespaket sollte die Arbeitsmarktpolitik eine Vorreiterstellung im Bereich der „Evaluationskultur“ auf Seiten des Bundes einnehmen, dem sonst eher – auch gegenüber anderen europäischen Ländern¹⁰ – ein „Evaluationsdefizit“ nachgesagt wurde.¹¹

Warum man gerade die „Hartz-Gesetze“ zur Implementation von evaluativen Elementen in der Gesetzgebung des Bundes nutzte, hatte verschiedene Gründe. Einer davon ist sicherlich die fundamentale Bedeutung des Vorhabens für die Wirtschafts- und Sozialordnung der Bundesrepublik und seine Einbettung in das

10 Beispielsweise im Vergleich zu Schweden, vgl. *Sylvia Veit*, Versachlichung gesetzgeberischer Entscheidungen durch Folgenanalysen? Eine vergleichende Untersuchung zu Reforminhalten und -ergebnissen in Deutschland und Schweden, dms 1 (2008), S. 73-98; *dies.*: Versachlichung gesetzgeberischer Entscheidungen durch Folgenabschätzung? Deutschland und Schweden im Vergleich, Wiesbaden 2010; *dies.*: National Experiences with General Forms of Ex Ante Evaluation of Legislation: The Cases of Germany and Sweden, in: Jonathan Verschuren (Hrsg.), *The Impact of Legislation. A Critical Analysis of Ex Ante Evaluation*, Leiden/Boston 2009, S. 175-197.

11 *Sabine Kuhlmann*: Die Evaluation von Institutionenpolitik in Deutschland: Verwaltungsmodernisierung und Wirkungsanalyse im föderalen System, in: Thomas Widmer/Wolfgang Beywl/Carlo Fabian (Hrsg.), *Evaluation. Ein systematisches Handbuch*, Wiesbaden 2008, S. 371-380, 373.

Programm zur Reform desselben, die „Agenda 2010“, das die zweite Kanzlerschaft *Gerhard Schröders* prägte. Ein zweiter ist die Möglichkeit, über die zeitliche Befristung von Gesetzen und über Evaluationsklauseln Kompromissfindung zu ermöglichen.¹² Ein dritter ist die im Kodifikationszeitpunkt bestehende relativ hohe Unsicherheit über die Wirkungsweise der in Deutschland bislang nicht erprobten Instrumente. Ein vierter schließlich ist der Umstand, dass aktive Arbeitsmarktpolitik auch zuvor in der Tradition intensiver Eigen- und Auftragsforschung zahlreicher Forschungsinstitute (und nicht zuletzt auch der Ressortforschung¹³) stand. Und schließlich wurde aktive Arbeitsmarktpolitik, gerade von der rot-grünen Koalition, zunehmend als ein „lernendes System“ beschrieben, indem „neue Instrumente [...] gleichsam zur Erprobung zeitlich befristet eingeführt“ wurden, wobei in einem zweiten Schritt dann Erkenntnisse der Wirkungsforschung „auf ‚politische Bestellung‘ gezielt und zeitnah zur Überprüfung des Erreichten und bei Diagnose von Fehlsteuerungen zur Anpassung des Rechtsrahmens“ genutzt werden sollten.¹⁴

Im Folgenden soll am Beispiel der „Hartz-Gesetze“ der Frage nach den Erfolgsbedingungen von (*on going*-)Evaluationen¹⁵ nachgegangen werden. Definiert man den „Erfolg“ von Evaluationen als eine möglichst weitgehende Umsetzung der Evaluationsergebnisse durch Politik und Verwaltung, so lässt sich fragen: Welche Rolle haben Evaluationen nach der Umsetzung gespielt? Inwieweit zeigten sich Regierung und Verwaltung in diesem Zusammenhang bereit, aufgrund von Evaluationsergebnissen eine Selbstkorrektur vorzunehmen? Lässt sich also bezogen auf dieses Gesetzgebungswerk von (erfolgreich) implementierten reflexiven Reformprozessen reden? Kann man in diesem Zusammenhang gar von „politischem Lernen“ sprechen?

12 Vgl. dazu u. a. *Evert Vedung*: Evaluation im öffentlichen Sektor, Wien/Köln/Graz 1999.

13 Zur Ressortforschung vgl. u. a. *Eva Barlösius*: Ressortforschung, in: Dagmar Simon/Andreas Knie/Stefan Hornbostel (Hrsg.), *Handbuch Wissenschaftspolitik*, Wiesbaden 2010, S. 377-389; *dies.*: „Forschen mit Gespür für politische Umsetzung“ – Position, interne Strukturierung und Nomos der Ressortforschung, *dms* 2 (2009), S. 347-366; *Thomas Groß*, Ressortforschung, Agenturen und Beiräte – zur notwendigen Pluralität der staatlichen Wissensinfrastruktur, in: Hans Christian Röhl (Hrsg.), *Wissen – Zur kognitiven Dimension des Rechts*, DV Beiheft 9, Berlin 2010, S. 135-155.

14 *Gerd Heyer*, Zielsetzung und Struktur der „Hartz-Evaluation“, *Zeitschrift für Arbeitsmarktforschung* (heute: *Journal for Labour Market Research*) 2006, S. 467-476, 469.

15 Im Folgenden sollen, wenn nicht ausdrücklich anders vermerkt, unter „Evaluationen“ Analysen von Gesetzen und anderen Maßnahmen verstanden werden. *On going* meint, dass diese während der Laufzeit von Gesetzen bzw. Maßnahmen durchgeführt wurden.

2. Die Evaluation der „Hartz-Gesetze“

An den Beginn der Analyse sollen zwei Zitate gestellt werden, die unterschiedlicher nicht sein können. Das erste lautet:

„Die im Rahmen der Hartz-Reformen gesetzlich verordnete Wirkungsanalyse ist ein vorbildlicher, bisher nicht dagewesener Einstieg in eine ‚Evaluationskultur‘ im Hinblick auf staatliches Handeln“.¹⁶

Das zweite kommt – zumindest auf den ersten Blick – zu einem genau entgegengesetzten Ergebnis:

„Der Anspruch, im Zuge der Hartz-Gesetzgebung erstmalig so etwas wie ein ‚lernendes arbeitsmarktpolitisches System‘ zu installieren, scheint – von Ausnahmen abgesehen – nicht gelungen“.¹⁷

Welchem der Autoren, die jeweils aufgrund intensiver Studien zu ihren Aussagen gekommen sind, ist nun zuzustimmen? Um diese Frage beantworten zu können, sollen nachfolgend (1.) zunächst mögliche Erwartungen an Evaluationen skizziert werden, um sodann (2.) die Ausgangssituation und die Ziele (3.) der „Hartz-Gesetze“ näher vorzustellen. Letztere sollen eingebettet werden in das diesen Gesetzen zugrunde liegende Staatsleitbild des „aktivierenden Staats“ (4.). In einem nächsten Schritt sollen die „Hartz-Gesetze“ im Einzelnen kurz umrissen werden (5.), bevor dann (6.) die Evaluationen dieser Gesetze eingehend dargestellt werden. Zwei nächste Abschnitte widmen sich den Evaluationsergebnissen (7.), und deren politischer Umsetzung (8.). Im Anschluss daran soll (III.) ein Fazit gezogen werden, das sich mit den möglichen Erfolgsbedingungen von Evaluationen und von „Politiklernen“ am Beispiel der untersuchten Reformmaßnahmen beschäftigt.

2.1 Erwartungen an Evaluationen

Die vorangestellten Zitate können nur dann sinnvoll bewertet werden, wenn man sich zunächst über die Erwartungen an Evaluationen Klarheit verschafft. Evaluationen können – neben anderen – vor allem folgende Funktionen erfüllen: 1. eine Erkenntnisfunktion für Politik und Verwaltung (als diejenigen, in deren Auftrag die Evaluation durchgeführt wird), 2. eine Kontrollfunktion bezüglich der (Wirk-

16 Bernd Fitzenberger: Nach der Reform ist vor der Reform? Eine arbeitsökonomische Analyse ausgewählter Aspekte der Hartz-Reformen (ZEW Discussion Paper 08/011) Mannheim 2008, S. 4.

17 Volker Baethge-Kinsky/Peter Bartelheimer/Alexandra Wagner: Die „Hartz-Gesetze“, ihre wissenschaftliche Evaluation und deren Verarbeitung (Schlussbericht des Transferprojekts „Monitor Arbeitsmarktpolitik“), Göttingen/Berlin 2010, S. 97.

samkeit der) Umsetzung staatlicher Maßnahmen, 3. eine Dialogfunktion, die insbesondere bei sog. „formativen“ Evaluationen eintritt, bei denen durch die Evaluierten ihre Organisation selbst an die Organisationsergebnisse angepasst werden kann und 4. eine Legitimationsfunktion, die der Akzeptanzschaffung für die Umsetzung späterer Maßnahmen dient.¹⁸

Insgesamt wird für die Bundesrepublik eher ein ‚zu wenig‘¹⁹ an Evaluationen beklagt. Zu konstatieren ist jedenfalls, dass (in Deutschland) vergleichsweise wenige Empfehlungen von Evaluationen umgesetzt werden. Evaluationen werden häufig ignoriert oder politisiert.²⁰ Insofern könnte man darauf hoffen, dass mit einer neuen „Evaluationskultur“²¹ auch brachliegendes Potential für eine „evidenzbasierte Politikgestaltung“²² genutzt werden könne.²³ Allerdings könnten sich – auf der anderen Seite – übertriebene Erwartungen an die Effekte von Evaluationen kontraproduktiv auswirken, wenn sich herausstellen sollte, dass ihr tatsächlicher Effekt auf die Politik(re)formulierung, zumindest in der hier untersuchten Bundesrepublik Deutschland, eher gering ist.

Daher sollen nachfolgend – am Beispiel der „Hartz-Gesetze“ – Überlegungen dazu angestellt werden, wann Evaluationen überhaupt in Auftrag gegeben werden und wann deren Ergebnisse einen Einfluss auf Politik haben können. Die „Hartz-Reformen“ können, so die Vermutung, ein besonderes Gelegenheitsfenster für die Implementation evaluativer Elemente in die Gesetzgebung geboten haben. Denn sie vollzogen sich – wie oben bereits angedeutet – zu einem Zeitpunkt, in dem nicht nur die Wirkungsforschung von Gesetzen (wieder) vermehrt ‚in Mode kam‘,²⁴ sondern auch die Politik angesichts der Neuartigkeit der arbeitsmarktpolitischen Instrumente in Deutschland tatsächlich einen Bedarf an zeitnahen Informationen über deren Wirkungsweise verspürte. Des Weiteren ließ sich durch die Aufnahme von Befristungs- und Evaluationsklauseln die Konsens-

18 Vgl. im Einzelnen dazu *Vedung* (Fn. 12).

19 *Kuhlmann* (Fn. 11), S. 373.

20 *Michael Quinn Patton*: Utilization-Focused Evaluation, Thousand Oaks 1997, S. 7.

21 So z. B. *Fitzenberger* (Fn. 16), S. 4; zu ‚evaluationskulturellen Problemen‘ vgl. *Jörg Bogumil*: Probleme und Perspektiven der Leistungsmessung in Politik und Verwaltung, in: Sabine Kuhlmann/ders./Helmut Wollmann (Hrsg.), *Leistungsmessung und -vergleich in Politik und Verwaltung. Konzepte und Praxis*, Wiesbaden 2004, S. 392-398, 396.

22 *Ian Sanderson*: Making Sense of „What Works“: Evidence Based Policy Making as Instrumental Rationality?, *Public Policy and Administration* 17 (2002), S. 61-75.

23 In diese Richtung gehend u. a. die Beiträge von *Veit* (Fn. 10).

24 Vgl. u. a. *Hagen Hof/Getrude Lübke-Wolf* (Hrsg.): *Wirkungsforschung zum Recht I. Wirkungen und Erfolgsbedingungen von Gesetzen*, Baden-Baden 1999; *Hermann Hill/Hagen Hof* (Hrsg.): *Wirkungsforschung zum Recht II, Verwaltung als Adressat und Akteur*, Baden-Baden 2000; *Hagen Hof/Martin Schulte* (Hrsg.): *Wirkungsforschung zum Recht III, Folgen von Gerichtsentscheidungen*, Baden-Baden 2001; *Ulrich Karpen/Hagen Hof* (Hrsg.): *Wirkungsforschung zum Recht IV. Möglichkeiten einer Institutionalisierung der Wirkungskontrolle von Gesetzen*, Baden-Baden 2003.

findung bei der Gesetzgebung selbst verbessern. Und schließlich verfügte, wie erwähnt, das Politikfeld „Arbeitsmarktpolitik“ über ein hinreichendes Angebot an ausgewiesener Expertise, die mit der Evaluation der Gesetze betraut werden konnte.

Allerdings gab es auch, so die These, Aspekte, die eher als ‚Erfolgshindernisse‘ für Evaluationen anzusehen sind. Dazu gehört in erster Linie, dass Regierungen *ex post*-Evaluationen ihrer politischen Maßnahmen deswegen scheuen, da sie fürchten, dass etwaige Befunde über Verbesserungspotentiale, um ein geflügeltes Wort von Horst Ehmke zu benutzen, eine „Abhakliste für die Opposition“²⁵ darstellen können. Mit anderen Worten: Kritische Ergebnisse können mit der Rationalität politischer Akteure (etwa der Selbstdarstellung der Regierung, dem Verhalten der Opposition) konfliktieren. Und handelt es sich um ein hochpolitisches Politikfeld wie dasjenige der Arbeitsmarkt- bzw. der Sozialpolitik, wird Evaluation noch brisanter. Das gilt umso mehr dann, wenn die Regierung ihre eigene Meßlatte sehr hoch gelegt hat. Denn der damalige Bundeskanzler, Gerhard Schröder, hatte die Vorschläge der „Hartz-Kommission“ im Wahlkampf von 2002 als eine Art ‚Rettungsanker‘ genutzt, mit der er seine prognostische Einschätzung von Beginn der Legislaturperiode 1998

„[w]enn wir es nicht schaffen, die Arbeitslosigkeit signifikant zu senken, dann haben wir es weder verdient, wiedergewählt zu werden, noch werden wir wiedergewählt“²⁶

durch einen ‚Kunstgriff‘ retten wollte. Er hatte argumentiert, dass zwischen 1998 und 2002 der Bundesregierung noch das passende Instrumentarium gefehlt habe, welches erst bis zum Jahr 2002 von der „Hartz-Kommission“ entwickelt worden sei und welches dann in der kommenden Legislaturperiode umgesetzt werden solle.

2.2 Ausgangssituation

Die Klagen über den Zustand des Wirtschaftsstandort Deutschlands als „kranken Mann Europas“ hatten sich um die Jahrtausendwende gehäuft.²⁷ Nicht zuletzt

25 Zitiert nach *Martin Brüggemeier*: Von der Kunst erfolgreich zu scheitern – Wirkungsorientiertes Controlling in öffentlichen Verwaltungen, in: Sabine Kuhlmann/Jörg Bogumil/Helmut Wollmann (Hrsg.), *Leistungsmessung und -vergleich in Politik und Verwaltung. Konzepte und Praxis*, Wiesbaden 2004, S. 374-391, 383.

26 So der damalige Kanzlerkandidat *Gerhard Schröder* in einem Spiegel-Interview am 21.09.1998, eine Woche vor der Bundestagswahl am 27.09.1998, zitiert nach *Margarete Hasel/Herbert Hönigsberger*: *Schröder verstehen – Kanzlerstrategie und Kanzlerkommunikation*, in: Frans Becker/Karl Duffek/Tobias Mörschel (Hrsg.), *Sozialdemokratische Reformpolitik und Öffentlichkeit*, Wiesbaden 2007, S. 65-112, 92.

aufgrund der Terroranschläge auf das World Trade Center am 11. September 2001 hatten die Aufschwungjahre der späten Neunziger Jahre ein Ende gefunden. Infolgedessen waren die Arbeitslosenzahlen immer weiter angestiegen und waren seither weit von dem von Gerhard Schröder vollmundig erklärten Ziel von 1998 entfernt.

Den unmittelbaren Anlass für die Einsetzung der Kommission „Moderne Dienstleistungen am Arbeitsmarkt“ unter dem damaligen VW-Personalvorstand Peter Hartz, der unter anderem auch Vertreter von Gewerkschaften angehörten, hatte dann der Umstand gebildet, dass die Presse aufgedeckt hatte, dass Arbeitsmarktstatistiken durch die Bundesanstalt für Arbeit ‚geschönt‘ worden waren. Diese Behörde war ohnehin als schwerfällig und ineffizient empfunden worden und passte zudem nicht mehr in den Trend zur „*agencification*“.²⁷

Wie eingangs erwähnt, erarbeitete die sog. „Hartz-Kommission“ recht zügig Vorschläge zur Umgestaltung der Arbeitsvermittlung, aber auch zur Zusammenlegung von Arbeits- und Sozialhilfe, die sie in 13 Module aufteilte. Die im Jahr 2002 vorgeschlagenen Maßnahmen seien, so die Kommission, in der Lage, die Arbeitslosigkeit bis um 50% zurückgehen zu lassen.

Die Bundesregierung griff, wie bereits geschildert, diese Vorschläge sofort begierig auf und kündigte an, diese noch vor den Bundestagswahlen „eins zu eins“ umzusetzen. Dieser Ankündigung wurde indes nur beschränkt Folge geleistet. Zwar wurden die Vorschläge tatsächlich in sehr hohem Tempo in Gesetzesvorhaben überführt, die entsprechenden Gesetze traten aber weder vor der Bundestagswahl in Kraft, noch entsprachen die Gesetzesvorhaben (und noch viel weniger die letztlich verabschiedeten Gesetze) „eins zu eins“ demjenigen, was die Kommission vorgeschlagen hatte. Das führte dazu, dass sich einige der Kommissionsmitglieder, namentlich aus den Gewerkschaften, später von der Art und Weise der Umsetzung ihrer Vorschläge distanzieren.²⁹

27 Vgl. Klaus Peter Weinert: Der Erfinder der Agenda 2010. Vor zehn Jahren verkündete Bundeskanzler Gerhard Schröder seine umstrittenen Hartz-Reformen, Deutschlandradio Kultur, „Kalenderblatt“ vom 14.03.2013, <http://www.dradio.de/dlf/sendungen/kalenderblatt/2036581/> [letzter Zugriff am 26.06.2013].

28 Unter „*agencification*“ wird die Auslagerung von Aufgaben aus der Ministerialverwaltung an Einrichtungen verstanden, die mit einer gewissen Unabhängigkeit ausgestattet sind und ergebnisorientiert gesteuert werden, vgl. Marian Döhler: Vom Amt zur Agentur? Organisationsvielfalt, Anpassungsdruck und institutionelle Wandlungsprozesse im deutschen Verwaltungsmodell, in: Werner Jann/ders. (Hrsg.), *Agencies in Westeuropa*, Wiesbaden 2007, S. 12-47; Christopher Politt: *Decentralized management. Agencies and "arm's length" bodies*, in: Tony Bovaird/Elke Löffler (Hrsg.), *Public Management and Governance*. 2. Aufl., London 2009, S. 249-259.

29 So äußerte Isolde Kunkel-Weber, Mitglied des Vorstandes der Gewerkschaft ver.di und Mitglied der Hartz Kommission, später: „Was Bundeskanzler Gerhard Schröder, seine Zuträger im Kanzleramt, die Roland Kochs und Wolfgang Clements dieser Republik, aus den Hartz-Vorschlägen gemacht haben, war schlicht kontraproduktiv und unsozial. Sie

2.3 Ziele der „Hartz“-Gesetze

In seiner Regierungserklärung vom 14. März 2003 machte Gerhard Schröder, der mit einem nicht allzu großen Vorsprung zusammen mit der Koalition aus SPD und Bündnis 90/Die Grünen wiedergewählt worden war, die arbeitsmarkt- und sozialpolitischen Reformen dann zum zentralen Thema der neuen Legislaturperiode („Agenda 2010“). Er kündigte zum einen die Zusammenlegung von Arbeitslosen- und Sozialhilfe an:

„Wir brauchen [...] Zuständigkeiten und Leistungen aus einer Hand. Damit steigern wir die Chancen derer, die arbeiten können und arbeiten wollen. Das, meine Damen und Herren, ist der Grund, warum wir die Arbeitslosen- und Sozialhilfe zusammenlegen werden. Und zwar einheitlich auf eine Höhe, und auch das gilt es auszusprechen, die in der Regel dem Niveau der Sozialhilfe entsprechen wird.“³⁰

In seiner sog. „Agenda-Rede“ führte Schröder aus:

„Entweder wir modernisieren, und zwar als soziale Marktwirtschaft, oder wir werden modernisiert, und zwar von den ungebremsten Kräften des Marktes, die dann das Soziale beiseite drängen.“³¹

2.4 Staatsleitbild des „aktivierenden Staats“

Hinter der „Agenda 2010“ und den „Hartz-Reformen“ stand das – unter anderem im Regierungsprogramm „Moderner Staat – Moderne Verwaltung“ der rot-grünen Koalition vom 1. Dezember 1999³² skizzierte – Staatsleitbild³³ des „aktivierenden Staats“: Darunter wurde

tragen die Verantwortung für die sozialen Folgen, nicht die Hartz-Kommission und ihre Mitglieder“, vgl. *Weinert* (Fn. 27).

30 Zitiert nach *Weinert* (Fn. 27).

31 Zitiert nach *Weinert* (Fn. 27).

32 „Moderner Staat – Moderne Verwaltung“ – Leitbild und Programm der Bundesregierung, Kabinettsbeschluss vom 1. Dezember 1999, download unter http://epractice.eu/files/media/media_418.pdf [letzter Zugriff am 26.6.2013].

33 Zu Leitbildern vgl. u. a. *Werner Jann*: State, Administration and Governance in Germany: Competing Traditions and Dominant Narratives, *Public Administration* 81 (2003), 95-118; *ders.*: Governance als Reformstrategie – Vom Wandel und der Bedeutung verwaltungspolitischer Leitbilder, in: *Gunnar Folke Schuppert* (Hrsg.), *Governance-Forschung. Vergewisserung über Stand und Entwicklungslinien*, Baden-Baden 2005, S. 21-43; *ders.*: Wandel verwaltungspolitischer Leitbilder: Vom Management zur Governance, in: *Klaus König* (Hrsg.), *Deutsche Verwaltung an der Wende zum 21. Jahrhundert*, Baden-Baden 2002, 279-304; *Gunnar Folke Schuppert*: Der Gewährleistungsstaat – modisches Label oder Leitbild sich wandelnder Staatlichkeit?, in: *ders.* (Hrsg.), *Der Gewährleistungsstaat – Ein Leitbild auf dem Prüfstand*, Baden-Baden 2005, S. 11-53.

„ein Staat verstanden, der zwar an einer umfassenden öffentlichen Verantwortung für gesellschaftliche Aufgaben festhält, jedoch nicht alle Leistungen selbst erbringen muss. Seine Aufgabe ist vielmehr, die Gesellschaft einschließlich der Beschäftigten des öffentlichen Dienstes zu fordern und zu fördern, sich selbst als Problemlöser zu engagieren. Dieser Staat ist keineswegs ein Minimalstaat, der nur dort tätig wird, wo die Gesellschaft dies fordert, sondern im Gegenteil, der aktivierende Staat tritt der Gesellschaft und ihren Individuen fordernd und fördernd gegenüber.“³⁴

Als charakteristisch für den „aktivierenden Staat“ wurde seine Verortung zwischen der tendenziell „bürokratisierende[n] Wohlfahrtsproduktion“³⁵ des sog. „aktiven“ Staats und dem eher marktliberalen Zugang des sog. „schlanken Staats“ ausgegeben, die diesem Staatsleitbild zeitlich vorangingen.³⁶

Die unter dem Leitbild des „aktivierenden Staats“ entwickelte „Agenda 2010“ enthielt ein Reformprogramm für Strukturreformen für mehr Wachstum und Beschäftigung, das allerdings nicht darauf angelegt war, „den Sozialstaat zu gefährden, sondern ihn angesichts eines sich verstärkenden weltweiten Änderungsdrucks durch eine neue Balance zwischen sozialem Ausgleich und ökonomischer Effizienz zu erhalten. Zum Fördern sollte das Fordern kommen.“³⁷ Dementsprechend fallen jüngere Beurteilungen der „Agenda 2010“ und der in ihrem Zusammenhang durchgeführten „Hartz-Reformen“ durch führende deutsche Wirtschaftsforscher recht positiv aus,³⁸ während die sozialpolitischen Auswirkungen durch Sozialwissenschaftler oftmals eher negativ beurteilt werden.³⁹

- 34 Vgl. u. a. *Stephan von Bandemer/Josef Hilbert*: Vom expandierenden zum aktivierenden Staat, in: Bernhard Blanke/Stephan von Bandemer/Frank Nullmeier/Göttrik Wewer (Hrsg.), *Handbuch zur Verwaltungsreform*, 3. Aufl., Wiesbaden 2005, S. 26-35, 30.
- 35 *Wolfram Lamping/Henning Schridde/Stefan Plaf/Bernhard Blanke*: Der Aktivierende Staat. Positionen, Begriffe, Strategien. Studie für den Arbeitskreis Bürgergesellschaft und Aktivierender Staat der Friedrich-Ebert-Stiftung, Bonn 2002.
- 36 *Margrit Seckelmann*: Die effiziente Verwaltung – Zur Entwicklung der Verwaltungsleitbilder seit dem Ende der 1970er Jahre, in: Peter Collin/Klaus-Gert Lutterbeck (Hrsg.), *Eine intelligente Maschine? Handlungsorientierungen moderner Verwaltung* (19./20. Jhd.), Baden-Baden 2009, S. 245-261.
- 37 *Karl Brenke/Klaus F. Zimmermann*: Reformagenda 2010 – Strukturreformen für Wachstum und Beschäftigung, download unter http://www.diw.de/documents/publikationen/73/diw_01.c.79926.de/08-11-1.pdf, S. 1 [letzter Zugriff am 26.6.2013].
- 38 Ebd. sowie *Ulrich Blum*: Agenda 2010: Neues unter Deutschlands Himmel?, *Wirtschaftsdienst* 2008, S. 151-155; *Wolfgang Franz*: Die Agenda 2010: Symbol eines wirtschaftspolitischen Kurswechsels, ebd., S. 156-159; *Christoph M. Schmidt*: Agenda 2010 und Hartz-Reformen: Erfolg oder Reformirrtum?, ebd., S. 166-170; *Klaus F. Zimmermann*: Die Agenda 2010: Ein geschichtsträchtiger gesellschaftspolitischer Reformaufbruch, ebd., S. 170-174; differenzierend *Gustav A. Horn/Camille Logeay*: Gibt es eine Reformdividende? Eine vorläufige Bilanz der Arbeitsmarktreformen, ebd., S. 160-165.
- 39 Beispielsweise bei *Baethge-Kinsky/Bartelheimer/Wagner* (Fn. 17).

2.5 Die „Hartz-Gesetze“ im Einzelnen

2.5.1 Die Aushandlungsprozesse vor ihrem Inkrafttreten

Das von der „Hartz-Kommission“ vorgeschlagene Maßnahmenbündel fand (mit Modifikationen, s. o.) Eingang in vier Gesetze („Hartz I – IV“ genannt). Diese Gesetze bildeten aber – anders als von Bundeskanzler Schröder angekündigt – keinesfalls „eins zu eins“ die von der Kommission vorgeschlagenen Maßnahmen ab. Vielmehr wurde manches schon in der Konzeption der Gesetzentwürfe, vieles aber auch erst im Wege des politischen Kompromisses im Gesetzgebungsverfahren eingefügt.

In den parlamentarischen Beratungen erwiesen sich insbesondere die Fragen der gesellschaftspolitischen Ziele, der damit verbundenen ‚Förderphilosophie‘, des Instrumenteneinsatzes, der möglichen Wirkungen von Maßnahmen und der Kriterien für ihre Bewertung als kontrovers. So wollten insbesondere die CDU/CSU-Fraktion und die FDP-Fraktion stärker noch die Abkehr vom bisherigen System hin zu einem „Fördern *und Fordern*“ der Leistungsempfänger. Sie forderten auch erfolgreich eine Absenkung des vorgesehenen ALG II-Betrags, da – so ihre Grundüberzeugung – aufgrund der zu setzenden Anreize zwischen den ALG II-Leistungen und dem Erwerbseinkommen eine deutliche Spanne liegen solle. Sie setzten auch weitere Einschränkungen durch, denn, so der Antrag der CDU/CSU-Fraktion vom 11. September 2002, der ebenfalls den Vorschlag zu einer Wirkungsüberprüfung enthielt:⁴⁰

„Es muss der Grundsatz gelten: Wer arbeitet, muss mehr verdienen als wenn er nicht arbeitet. Sozialhilfe ist eine zweite Chance – sie darf kein Lebensstil sein! Arbeitslose müssen sich ernsthaft und nachprüfbar um Arbeit bemühen, sich weiterqualifizieren oder gemeinnützige Arbeit leisten. Wer dazu nicht bereit ist, erhält nur noch gekürzte Leistungen. Die Zumutbarkeitsregeln müssen verschärft und die Beweislast umgekehrt werden. Arbeitslose müssen gefördert aber auch gefordert werden.“⁴¹

Ein anderer, heftig umstrittener Punkt war – wie erwähnt – die Verwaltung der Leistungen nach dem SGB II. Diese waren insbesondere Teil der Aushandlungsprozesse mit dem Bundesrat.⁴² Nicht zuletzt deswegen wurde die Frage der Or-

40 *Deutscher Bundestag*: Antrag der Abgeordneten Horst Seehofer u. a. und der Fraktion der CDU/CSU „Zeit für Taten – Offensive für mehr Beschäftigung“ vom 11.9.2002, Drs. 14/9944, Punkt 6: „Initiative Flexibler Arbeitsmarkt“, S. 7: „Alle Instrumente der aktiven Arbeitsmarktpolitik müssen auf ihre Wirksamkeit hin überprüft werden. Das gilt auch für Arbeitsbeschaffungsmaßnahmen, die vor allem in Bereichen mit guter Arbeitsmarktlage zurückgefahren werden können“.

41 Ebd., S. 6.

42 Die „Hartz II–IV-Gesetze“ waren zustimmungsbedürftig, „Hartz I“ nicht.

ganisation der sog. „ARGEn“ (heute: „Jobcenter“) ausdrücklich in einer „Experimentierklausel“ (mit Berichtspflicht, vgl. § 6 c SGB II a. F.⁴³) geregelt.

2.5.2 Die Hauptinhalte der einzelnen Gesetze

Die „Hartz-Gesetze“, die von Januar 2003 bis Juli 2004 nach und nach in Kraft traten, hatten folgende Primärinhalte:

1. die Modernisierung der Vermittlung mit dem Ziel der Reduzierung der Arbeitslosigkeit bis 2005 um bis zu 50% („Hartz I und II“),
2. die Reform der Bundesanstalt für Arbeit (BA) zu einem modernen Dienstleister („Hartz III“) und
3. die Zusammenführung von Arbeitslosen- und Sozialhilfe zu „Arbeitslosengeld II“ („Hartz IV“).

Im Einzelnen betraf das zum 1. Januar 2003 in Kraft getretene *Erste Gesetz für moderne Dienstleistungen am Arbeitsmarkt* („Hartz I“) die beabsichtigte Erschließung von Beschäftigungsmöglichkeiten und die erhoffte Schaffung neuer Arbeitsplätze durch die Schaffung sog. „Personal-Service-Agenturen“ (PSAn). Damit verbunden war eine Liberalisierung der Arbeitnehmerüberlassung. Weitere Maßnahmen zielten auf die verbesserte Integration älterer Beschäftigungssuchender in den Arbeitsmarkt ab.

Das zeitgleich in Kraft getretene *Zweite Gesetz für moderne Dienstleistungen am Arbeitsmarkt* („Hartz II“) befasste sich mit der Regelung von Existenzgründungszuschüssen („Ich-AGs“) und „Mini-Jobs“ und von Voraussetzungen für die Errichtung von Jobcentern.

Ein Jahr später, am 1. Januar 2004, trat das *Dritte Gesetz für moderne Dienstleistungen am Arbeitsmarkt* („Hartz III“) in Kraft. Dieses regelte den Umbau der Bundesanstalt für Arbeit in eine Dienstleistungsagentur („Agentur für Arbeit“). Daneben hatte es die Vereinfachung des Leistungs- und Förderungsrechts der Arbeitslosenversicherung und von arbeitsmarktpolitischen Instrumenten wie Arbeitsbeschaffungsmaßnahmen (ABM) und Eingliederungszuschüssen (EGZ) zum Inhalt.

43 § 6c SGB II (Wirkungsforschung zur Experimentierklausel) lautete in der bis zum 11.08.2010 geltenden Fassung: „Das Bundesministerium für Wirtschaft und Arbeit [später: Arbeit und Soziales] untersucht die Wahrnehmung der Aufgaben durch die zugelassenen kommunalen Träger im Vergleich zur Aufgabenwahrnehmung durch die Agenturen für Arbeit und berichtet den gesetzgebenden Körperschaften des Bundes bis zum 31. Dezember 2008 über die Erfahrungen mit den Regelungen nach den §§ 6a und 6b. Die Länder sind bei der Entwicklung der Untersuchungsansätze und der Auswertung der Untersuchung zu beteiligen.“

Das damals wie heute am heftigsten umstrittene *Vierte Gesetz für moderne Dienstleistungen am Arbeitsmarkt* („Hartz IV“) trat zum 1. Juli 2004 in Kraft. Es betraf die Zusammenlegung der Arbeitslosenhilfe und der Sozialhilfe für erwerbsfähige Sozialhilfeempfänger zum sog. „Arbeitslosengeld II“ (ALG II) zum 1.1.2005. Daneben wurde die Bezugsdauer des prozentual an die Höhe des Einkommens der letzten Monate gekoppelten Arbeitslosengeldes (jetzt „ALG I“ genannt) radikal verkürzt: Nicht mehr die Zahl der Beitragsmonate sollte für die Dauer der Auszahlung des ALG I ausschlaggebend sein, vielmehr wurde die Bezugsdauer generell auf 12 Monate (ab dem 55. Lebensjahr: 18 Monate) verkürzt. Nach Ende der Zahlungen nach ALG I bestand nicht mehr das Instrument der Arbeitslosenhilfe, sondern – bei „Bedürftigkeit“ – konnte das ALG II beantragt werden. An die „Bedürftigkeit“ wurden relativ hohe Anforderungen gestellt, so durfte etwa der Partner oder die Partnerin, mit dem der Antragsteller oder die Antragstellerin in einer sog. „Bedarfsgemeinschaft“ lebt, nicht allzu viel verdienen. Auch durften nur noch geringe Sparvermögen in der „Bedarfsgemeinschaft“ (einschließlich der Kinder) vorhanden sein. Daneben wurden die Kriterien, nach welchen eine angebotene Arbeit als „zumutbar“ zu gelten hat, erheblich ausgedehnt.

Allerdings gab es auch eine Gruppe, für die das ALG II eine Verbesserung darstellte: Die *erwerbsfähigen* Sozialhilfeempfänger (z. B. arbeitslose Berufseinsteiger und Selbstständige, die keinen Versicherungsanspruch hatten) wurden denjenigen, die länger als ein Jahr (bzw. anderthalb Jahre, s. o.) auf Arbeitssuche waren, gleichgestellt und erhielten so nicht nur einen höheren Regelsatz, sondern auch Ansprüche auf Fördermaßnahmen der Arbeitsagenturen. Daneben wurde die Sozialhilfe für diejenigen Personen weitergeführt, die nicht erwerbsfähig waren, sei es, weil sie schon das Rentenalter erreicht hatten oder weil sie nicht länger als drei Stunden am Tag arbeiten konnten (Erwerbsunfähigkeit), und zwar in Form von Eingliederungshilfen für behinderte Menschen, einer Grundsicherung für Ältere, denen Rente nicht zum Leben reicht, einer Hilfe zur Pflege, einer Hilfe zum Lebensunterhalt und bestimmten Hilfen zur Gesundheit für Bedürftige.

2.6 Evaluationen der „Hartz“-Gesetze

2.6.1 Evaluationsarten

Die „Hartz-Gesetze“ enthielten eine explizite politische Selbstverpflichtung zur wissenschaftlichen Evaluation. Allerdings lassen sich verschiedene Evaluationsarten und -formen unterscheiden: Hinsichtlich der „kommunalen Option“ war

eine (neuartige) *vergleichende Begleitforschung* gem. § 6c SGB II (a.F.) vorgesehen.⁴⁴ Demgegenüber baute die *laufende Wirkungsforschung* nach § 282 SGB III⁴⁵ auf Vorläufern auf. Denn der gesetzliche Auftrag der (damaligen) Bundesanstalt für Arbeit zur Arbeitsmarkt- und Berufsforschung fand sich erstmals im Arbeitsförderungsgesetz von 1969. Mit der Überführung des Arbeitsförderungsgesetzes (AFG) in das SGB III 1998 wurde der gesetzliche Auftrag in dieses übernommen und schrittweise konkretisiert. Die letzte Konkretisierung geschah im Rahmen der „Hartz-Gesetze“.

Nach § 280 SGB III betreibt die Bundesagentur für Arbeit (zuvor: Bundesanstalt für Arbeit) Arbeitsmarkt- und Berufsforschung. Sie hat dieser Bestimmung zufolge 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, indem sie erstens Statistiken erstellt, zweitens Arbeitsmarkt- und Berufsforschung betreibt und drittens Bericht erstattet“. Eine Konkretisierung hiervon enthält § 282 SGB III, und zwar bezüglich dessen, wer bei der Festlegung der Forschungsinhalte mitwirkt, nämlich die Bundesagentur (zuvor: die Bundesanstalt), die bei der Festlegung von Inhalt, Art und Umfang der Arbeitsmarkt- und Berufsforschung ihren eigenen Informationsbedarf zu berücksichtigen hat sowie denjenigen des Bundesministeriums für Arbeit und Soziales (bis

44 Vgl. Fn. 43. Heute ist die regelmäßige vergleichende Evaluation in § 55 Abs. 2 SGB II geregelt.

45 Die Absätze 1-3 des § 282 SGB III lauteten in der damals gültigen Fassung: „(1) Die Bundesanstalt für Arbeit [später: Bundesagentur] hat bei der Festlegung von Inhalt, Art und Umfang der Arbeitsmarkt- und Berufsforschung ihren eigenen Informationsbedarf, den des Bundesministeriums für Wirtschaft und Arbeit sowie den des Bundesministeriums für Gesundheit und Soziale Sicherung zu berücksichtigen, soweit er sich auf die Berücksichtigung der beruflichen Teilhabe behinderter und schwerbehinderter Menschen bezieht. Die Bundesanstalt [später: Bundesagentur] für Arbeit hat den Forschungsbedarf mindestens in jährlichen Zeitabständen mit dem Bundesministerium für Wirtschaft und Arbeit [später: für Arbeit und Soziales] unter Beteiligung des Bundesministeriums für Gesundheit und Soziale Sicherung [später: für Gesundheit] abzustimmen. (2) Die Untersuchung der Wirkungen der Arbeitsförderung ist ein Schwerpunkt der Arbeitsmarktforschung. Sie soll zeitnah erfolgen und ist ständige Aufgabe des Instituts für Arbeitsmarkt- und Berufsforschung. (3) Die Wirkungsforschung soll unter Berücksichtigung der unterschiedlichen Zielsetzungen des Gesetzes insbesondere 1. die Untersuchung, in welchem Ausmaß die Teilnahme an einer Maßnahme die Vermittlungsaussichten der Teilnehmer verbessert und ihre Beschäftigungsfähigkeit erhöht, 2. die vergleichende Ermittlung der Kosten von Maßnahmen in Relation zu ihrem Nutzen, 3. die Messung von volkswirtschaftlichen Nettoeffekten beim Einsatz arbeitsmarktpolitischer Instrumente, 4. die Analyse von Auswirkungen auf Erwerbsverläufe unter Berücksichtigung der Gleichstellung von Frauen und Männern umfassen.“

2005: Wirtschaft und Arbeit), unter Beteiligung des Bundesministeriums für Gesundheit (bis 2005: Gesundheit und soziale Sicherung).“⁴⁶

Speziell für die Instrumente der „Hartz-Gesetze“ wurde der seinerzeit neugefasste § 55 SGB II formuliert: Danach waren die „Wirkungen der Leistungen zur Eingliederung und der Leistungen zur Sicherung des Lebensunterhalts [...] regelmäßig und zeitnah zu untersuchen und in die Arbeitsmarkt- und Berufsforschung nach § 282 des Dritten Buches einzubeziehen.“⁴⁷ Die Einzelheiten seien zwischen dem Bundesministerium und der Bundesanstalt/Bundesagentur festzulegen, wobei, soweit zweckmäßig, Dritte mit der Wirkungsforschung beauftragt werden konnten (§ 55 Abs. 1 SGB II).

Durch die Forschung im Rahmen des § 55 SGB II können – aufgrund des Verweises auf das SGB III – Synergien „bei den Wirkungsanalysen zu den arbeitsmarktpolitischen Instrumenten, bei der Forschung zu spezifischen Personengruppen – etwa Jugendlichen, Älteren oder Rehabilitanden – sowie bei den Analysen zu den regionalen und gesamtwirtschaftlichen Effekten“ erzielt werden.“⁴⁸ Allerdings umfasst § 55 SGB II ein deutlich größeres Themenspektrum als § 282 SGB III „da auch solche Wirkungen der Grundsicherung umfassend untersucht werden sollen, die sich nicht unmittelbar auf dem Arbeitsmarkt zeigen. Letztlich geht es also darum – viel stärker als im SGB III –, auch die Erfüllung der sozialpolitischen Ziele des Gesetzes zu untersuchen“.⁴⁹

Nach der Evaluation der „kommunalen Option“ wurde die vergleichende Begleitforschung und die laufende Wirkungsforschung im neugefassten § 55 SGB II verbunden, die regelmäßige vergleichende Evaluation ist nunmehr in § 55 Abs. 2 SGB II geregelt.

2.6.2 Akteure und Maßnahmen

Das damalige Bundesministerium für Wirtschaft und Arbeit (BMWA, seit der Bundestagswahl von 2005 war dann das wieder vom Bundeswirtschaftsministerium getrennte Bundesministerium für Arbeit und Soziales, BMAS, für die

46 Vgl. <http://www.iab.de/de/ueberblick/gesetzlicher-auftrag.aspx> [letzter Zugriff am 26.06.2013].

47 § 55 SGB II (Wirkungsforschung) lautete seinerzeit: „(1) 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 einzubeziehen. Das Bundesministerium für Wirtschaft und Arbeit [später: für Arbeit und Soziales] und die Bundesanstalt [später: Bundesagentur] für Arbeit können in Vereinbarungen Einzelheiten der Wirkungsforschung festlegen. Soweit zweckmäßig, können Dritte mit der Wirkungsforschung beauftragt werden.“

48 Vgl. Fn. 46.

49 Ebd.

„Hartz-Evaluationen“ zuständig) vergab zur Vorbereitung der Evaluation zwei konkurrierende Vorstudien⁵⁰ an zwei Projektkonsortien (eines bestehend aus dem Institut für Sozialforschung und Gesellschaftspolitik [ISG] und dem Rheinisch-Westfälischen Institut für Wirtschaftsforschung [RWI]⁵¹ und ein weiteres aus dem Zentrum für Europäische Wirtschaftsforschung [ZEW] und der Gesellschaft für Sozialforschung und Marktforschung mbH [sinus]⁵²). Dieser Verfahrensschritt hatte das Ziel, möglichst das gesamte wissenschaftliche Spektrum der Arbeitsmarktforschung zu erschließen. Der „Zweck dieser Studien war es, Vorschläge für die inhaltliche Ausgestaltung und Organisation der umfangreichen Forschungsarbeiten zu erarbeiten. Die Vorstudien hatten einen Überblick zu erstellen, welche – dem aktuellen wissenschaftlichen Standard entsprechende – Methoden zur Bearbeitung welcher Fragestellung geeignet sind und diese darstellen.“⁵³ Sofern mehrere geeignete methodische Ansätze in Betracht kamen, sollte ein Vorschlag unterbreitet werden, welcher Ansatz gewählt werden solle.⁵⁴

Die im November 2003 fertig gestellten Vorstudien wurden auf einem Workshop des Bundesministeriums für Wirtschaft und Arbeit am 4. und 5. Dezember 2003 in Berlin vorgestellt und mit Expertinnen und Experten aus der Arbeitsmarktforschung und -praxis diskutiert. Die Vorstudien enthielten eine Aufteilung der vorzunehmenden Evaluation in verschiedene Teilpakete und relativ strikte inhaltliche und methodische Angaben, die der Ausschreibung der einzelnen Evaluationsteile zugrunde gelegt wurden. Ein Vergabeverfahren war aufgrund des Einsatzes erheblicher finanzieller Mittel erforderlich. Es bewarben sich nicht nur wegen dieser umfangreichen Mittel zahlreiche potentielle Evaluatoren, sondern auch weil ihnen ein privilegierter Daten- und Feldzugang gewährt wurde.

50 Die Vorstudien und Berichte unterfallen oftmals in mehrere Bände (Bericht plus Anlagen), bzw. Zwischenbericht und Endbericht. Aus Platzgründen wird in den folgenden Fußnoten immer der aus Sicht der Verf. ‚anschaulichste‘ zitiert, die anderen lassen sich über die Homepage der mit der Koordination der Gesamtevaluation (= Arbeitspaket 4) beauftragten Politikberater Dr. *Bruno Kaltenborn* leicht erschließen, vgl. <http://www.wi-pol.de/hartz/evaluierung.htm> [letzter Zugriff am 26.06.2013].

51 Näheres bei der Projektbeschreibung unter http://www.isg-institut.de/index.php?b=single&id_B=1&id_UB=23&id_Nummer=46 [letzter Zugriff am 26.06.2013]] und in der aus dem Projekt hervorgegangenen Publikation *Michael Fertig/Jochen Kluge/Christoph M. Schmidt/Helmut Apel/Werner Friedrich/Helmut Haegle*, Die Hartz-Gesetze zur Arbeitsmarktpolitik – Ein umfassendes Evaluationskonzept (RWI-Schriften 74), Berlin 2004.

52 Näheres bei der Projektbeschreibung unter <http://www.zew.de/de/forschung/projekte.php3?action=detail&nr=356> [letzter Zugriff am 26.06.2013] und in der aus dem Projekt hervorgegangenen Publikation *Tobias Hagen/Alexander Spermann*, Hartz-Gesetze – Methodische Ansätze zu einer Evaluierung (ZEW Wirtschaftsanalysen 74), Baden-Baden 2004.

53 *Kaltenborn* (Fn. 50).

54 *Kaltenborn* (Fn. 50).

Forschungsorganisatorisch drängte das Bundesministerium auf die Bildung von Forschungskonsortien (zusammengesetzt aus Forschungsinstituten bzw. anderen Einrichtungen sowie Lehrstuhlinhabern und/oder privaten Politikberatungsunternehmen). Das war relativ neu im deutschen Forschungs- und Politikberatungssystem und brachte dort im Hinblick auf (künftige) Kooperationen einiges in Bewegung.

Ein anderer Aspekt waren die recht klaren Vorgaben, die den Evaluatoren gemacht wurden: Die relativ strikte inhaltliche wie methodische Festlegung der Herangehensweise hatte nicht nur vergaberechtliche Gründe, sondern zugleich den Vorteil, eine gewisse Vergleichbarkeit der Evaluationsstudien herzustellen. Allerdings zwang sie die Evaluatoren in ein relativ enges inhaltliches wie zeitliches Korsett, das auch wieder die Untersuchung bestimmter unerwünschter Nebeneffekte der Maßnahmen erschwerte.

Nach Durchführung eines Vergabeverfahrens wurde im Jahr 2004 nach und nach mit der Evaluation der „Hartz-Gesetze I-III“ (sowie der Organisation der „Job-Center“ aufgrund der Experimentierklausel durch mehr als 20 Forschungsinstitute⁵⁵) begonnen, die neben die Ressortforschung des BMAS durch das Institut für Arbeitsmarkt- und Berufsforschung (IAB), das Forschungsinstitut der Bundesanstalt (später: Bundesagentur) für Arbeit, und freiwillige Evaluationen anderer Akteure⁵⁶ traten.

55 Zum Beispiel: Deutsches Institut für Wirtschaftsforschung (DIW), Forschungsinstitut zur Zukunft der Arbeit (IZA), Gender-Institut Sachsen Anhalt (GISA), Institut Arbeit und Technik (IAT), Institut für Arbeitsmarkt- und Berufsforschung der Bundesagentur für Arbeit (IAB), Institut für angewandte Sozialwissenschaft (infas), Institut für Wirtschaftsforschung an der Universität München (ifo, heute Leibniz-Institut für...), Institut für Medienforschung und Urbanistik (IMU), Institut für Sozialforschung und Gesellschaftspolitik Köln (ISG), Institut für Sozialwirtschaft und Sozialforschung (iso), Institut für Sozialökonomische Strukturanalysen (SÖSTRA), Institut für Wirtschaftsforschung Halle (IWH), Progress-Institut für Wirtschaftsforschung (PIW), Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI Essen), Wissenschaftszentrum Berlin für Sozialforschung (WZB), Zentrum für Europäische Wirtschaftsforschung (ZEW) sowie die weiteren, im Folgenden genannten Einrichtungen und Einzelautoren.

56 Vgl. etwa die durch die Hans-Böckler-Stiftung geförderte Meta-Evaluation der „Hartz-Reformen“ durch *Baethge-Kinsky/Bartelheimer/Wagner* (Fn. 17) sowie die Untersuchung des Charakterwechsels der Vermittlungseinrichtungen im Zusammenhang mit den „Hartz-Reformen“ (als Teil einer längerfristig angelegten Studie, die schon vor 2003 beginnt) von *Ingo Bode*: Die Dynamik organisierter Beschäftigungsförderung. Eine qualitative Evaluation, Wiesbaden 2005, S. 32.

2.6.2.1 Evaluation Hartz I – III: Fünf Arbeitspakete

Im Einzelnen wurde die Evaluation der „Hartz I-III-Gesetze“ in fünf Arbeitspakete aufgeteilt, deren erstes wiederum in sechs Teilmodule untergliedert wurde. Das erste Arbeitspaket betraf die „Wirksamkeit der Instrumente“, das zweite den „Organisatorische[n] Umbau der Bundesagentur für Arbeit“, das dritte die „Akzeptanz der Bundesagentur für Arbeit“, das vierte Paket, das zugleich die Evaluation insgesamt koordinierte, die „Administrative Unterstützung bei der Evaluation“ und das letzte schließlich (ebenfalls von Bedeutung für die Gesamtkoordination) die „Datenbereitstellung“.

Im Folgenden sollen die einzelnen Arbeitspakete ganz kurz umrissen werden.

2.6.2.1.1 Arbeitspaket 1: Wirksamkeit der Instrumente

– Modul 1a „Neuausrichtung der Arbeitsvermittlung“

Das Modul 1a „Neuausrichtung der Arbeitsvermittlung“⁵⁷ sollte die Neuausrichtung der *Arbeitsvermittlung* untersuchen, wie sie durch die Vorschläge der „Hartz-Kommission“ konzipiert und in die Gesetzespakete „für moderne Dienstleistungen am Arbeitsmarkt“ („Hartz I-III“) eingegangen war. Forschungsgegenstand war vor allem die Frage, ob aufgrund der gesetzlich eingeführten Instrumente „eine Verbesserung der Qualität und Schnelligkeit der Arbeitsvermittlung nachgewiesen werden“ könne.⁵⁸

Evaluiert wurde insbesondere „der Vermittlungsprozess und seine Neuausrichtung einschließlich des verstärkten Wettbewerbs durch die seit 2002 erfolgten Reformen am Arbeitsmarkt (Umsetzung des Zweistufenplans der Bundesregierung und der sog. Hartz-Gesetze)“ und zwar auf die intendierten und nicht intendierten Wirkungen sowie die Effektivität und Effizienz.⁵⁹ Unter „Vermittlung“ wurden dabei „alle Aktivitäten und Prozesse der Bundesagentur für Arbeit

57 *Wissenschaftszentrum Berlin für Sozialforschung (WZB)/infas Institut für angewandte Sozialwissenschaft*, Modul 1a „Neuausrichtung der Arbeitsvermittlung“, Bericht, Berlin/Bonn 2005, download unter http://www.bmas.de/SharedDocs/Downloads/DE/PDF-Publikationen/fb-f348-1-forschungsbericht-1a-neuausrichtung-der-vermittlungsprozesse-bericht.pdf?__blob=publicationFile [letzter Zugriff am 26.06.2013].

58 *Forschungsinstitut zur Zukunft der Arbeit (IZA)/ Deutsches Institut für Wirtschaftsforschung (DIW)/Institut für angewandte Sozialwissenschaft (infas)*, Evaluation der Maßnahmen zur Umsetzung der Vorschläge der Hartz-Kommission, Modul 1b: Förderung beruflicher Weiterbildung und Transferleistungen, Bericht, Bonn/Berlin 2005, download unter http://www.bmas.de/SharedDocs/Downloads/DE/PDF-Publikationen/fb-f348-2-forschungsbericht-1b-foerderung-beruflicher-weiterbildung-haupttext.pdf?__blob=publicationFile [letzter Zugriff am 26.06.2013], S. 5.

59 *Kaltenborn* (Fn. 50).

(BA) und ihrer Partner/innen im Sinne des Sozialgesetzbuches zur Zusammenführung von Arbeitsuchenden und Arbeitgeber/innen“ verstanden.⁶⁰ Die Evaluation umfasste auch die Effektivität und Effizienz sowie die Wirkungen der Instrumente „Virtueller Arbeitsmarkt (VAM)“, „Vermittlungsgutschein“ zur Einschaltung Dritter im Zusammenhang mit der Vermittlung (§ 37a a.F. SGB III, jetzt § 37 und § 421i SGB III), der „Personal-Service-Agenturen (PSAs)“, des „Profiling“, der „Eingliederungsvereinbarungen“, der verpflichtenden frühzeitige Arbeitssuche (§ 37b SGB III) sowie der Neuregelung von Zumutbarkeit und Sperrzeiten.

– Modul 1b: „Förderung beruflicher Weiterbildung und Transferleistungen“
Modul 1b betraf die „Förderung beruflicher Weiterbildung und Transferleistungen“⁶¹. Ermittelt wurden die intendierten und nicht intendierten Wirkungen sowie die Effektivität und Effizienz der Förderung der beruflichen Weiterbildung (FbW) einschließlich des Einsatzes von Bildungsgutscheinen sowie der Änderungen der FbW durch die „Hartz-Gesetze“. Darüber hinaus wurden auch die Wirkungen der neu geschaffenen Transferleistungen (Transfermaßnahmen und das Transferkurzarbeitergeld) untersucht.⁶²

– Modul 1c: „Arbeitsbeschaffungsmaßnahmen“
Im Rahmen des Moduls 1c „Arbeitsbeschaffungsmaßnahmen“⁶³ wurde analysiert, inwieweit Arbeitsbeschaffungsmaßnahmen Wirkungen auf die Beschäftigungsfähigkeit der Geförderten und ihre Integration in den allgemeinen Arbeitsmarkt zeitigten. Dabei war auf regional unterschiedliche Arbeitsmarktlagen und -strategien Rücksicht zu nehmen. Besonderes Augenmerk war auf die (intendierten und unintendierten) Effekte der Zusammenführung von Arbeitsbeschaffungsmaßnahmen und Strukturanpassungsmaßnahmen zu legen.⁶⁴ Daneben wurde

60 Kaltenborn (Fn. 50).

61 Forschungsinstitut zur Zukunft der Arbeit/Deutsches Institut für Wirtschaftsforschung/Institut für angewandte Sozialwissenschaft (Fn. 58).

62 Kaltenborn (Fn. 50).

63 COMPASS Gesellschaft für Informationsmanagement und Projektentwicklung/IMU Institut für Medienforschung und Urbanistik/Institut für Sozialökonomische Strukturanalysen SÖSTRA/PIW Progress-Institut für Wirtschaftsforschung/Universität Hamburg (im Unterauftrag des PIW), Evaluation der Maßnahmen zur Umsetzung der Vorschläge der Hartz-Kommission – Arbeitspaket 1: Wirksamkeit der Instrumente, Modul 1c: Arbeitsbeschaffungsmaßnahmen – Zweiter Bericht, Berlin/Bremen/Hamburg 2005, download unter http://www.bmas.de/SharedDocs/Downloads/DE/PDF-Publikationen/fb-f348-3-1c-arbeitsbeschaffungsmaßnahmen.pdf?__blob=publicationFile [letzter Zugriff am 26.06.2013].

64 Kaltenborn (Fn. 50).

analysiert, welche Wirkungen der Wegfall der Versicherungspflicht zur Bundesagentur für Arbeit (BA) auf die Inanspruchnahme des Instrumentes hatte.⁶⁵

– Modul 1d: „Eingliederungszuschüsse und Entgeltsicherung“

Modul 1d befasste sich mit der Evaluation der Wirksamkeit der arbeitsmarktpolitischen Instrumente „Eingliederungszuschüsse“ (EGZ) und „Entgeltsicherung“.⁶⁶ Untersucht wurden die intendierten und nicht intendierten Wirkungen sowie die Effektivität und Effizienz von EGZn sowie diejenigen des Beitragsbonus für Arbeitgeberinnen und Arbeitgeber bei der Einstellung Älterer.⁶⁷ Hinsichtlich der gleichen Fragestellungen wurde auch das Instrument der „Entgeltsicherung“ analysiert.

– Modul 1e „Existenzgründungen“

Ziel der Untersuchungen in Modul 1e war es, sowohl qualitativ als auch quantitativ belastbare Aussagen über intendierte wie nicht intendierte Wirkungen der beiden Förderinstrumente „Überbrückungsgeld“ und „Existenzgründungsschuss bei Gründungen aus Arbeitslosigkeit“ zu bekommen und auf dieser Basis Effektivität und Effizienz der Programme und ihres Fördereinsatzes zu bewerten und Politikempfehlungen zu ihrer weiteren Ausgestaltung zu geben.⁶⁸

65 Kaltenborn (Fn. 50).

66 Zentrum für Europäische Wirtschaftsforschung (ZEW)/Institut für Arbeitsmarkt- und Berufsforschung (IAB)/Institut Arbeit und Technik (IAT), Evaluation der Maßnahmen zur Umsetzung der Vorschläge der Hartz-Kommission, Arbeitspaket 1: Wirksamkeit der Instrumente, Modul 1d: Eingliederungszuschüsse und Entgeltsicherung, Bericht, Nürnberg, Gelsenkirchen und Mannheim 2005, download unter http://www.bmas.de/SharedDocs/Downloads/DE/PDF-Publikationen/fb-f348-4-forschungsbericht-1d-eingliederungszuschuesse-und-entgeltsicherung.pdf?__blob=publicationFile [letzter Zugriff am 26.06.2013].

67 Kaltenborn (Fn. 50).

68 Institut für Arbeitsmarkt- und Berufsforschung der Bundesagentur für Arbeit (IAB)/Deutsches Institut für Wirtschaftsforschung (DIW)/sinus Gesellschaft für Sozialforschung und Marktforschung mbH, Alexander Kritikos (Gesellschaft für Arbeitsmarktaktivierung GfA)/infas Institut für angewandte Sozialwissenschaft: Evaluation der Maßnahmen zur Umsetzung der Vorschläge der Hartz-Kommission, Arbeitspaket 1: Wirksamkeit der Instrumente, Modul 1e: Existenzgründungen, Bericht, Nürnberg/Berlin/Bonn/München 2006, download unter http://www.bmas.de/SharedDocs/Downloads/DE/PDF-Publikationen/fb-f356e-evaluation-der-massnahmen-hartz-arbeitspaket-1e.pdf;jsessionid=51009FBD889B83ED973C6E8399DF8144?__blob=publicationFile [letzter Zugriff am 26.06.2013].

– Modul 1f: „Verbesserung der beschäftigungspolitischen Rahmenbedingungen“ und Makrowirkungen der aktiven Arbeitsmarktpolitik“

Modul 1f betraf die Frage nach der „Verbesserung der beschäftigungspolitischen Rahmenbedingungen“ durch die neu eingeführten Instrumente und die „Makrowirkungen der aktiven Arbeitsmarktpolitik“. ⁶⁹

2.6.2.1.2 Arbeitspaket 2: Organisatorischer Umbau der Bundesagentur für Arbeit

Im Gegensatz zum ersten Arbeitspaket, das primär die Wirkungen der arbeitsmarktpolitischen Instrumente betraf, hatte das zweite Arbeitspaket den organisatorischen Umbau der Bundesanstalt für Arbeit zum Gegenstand. Dabei untersuchte es zum einen die Entwicklung der Organisation der Arbeitsverwaltung hin zu einer „modernen Dienstleistungseinrichtung“ sowie zum anderen die Einzelaspekte der Implementierung des Neuen Steuerungsmodells sowie veränderter Arbeitsabläufe (von der Aufbau- zur Ablauforganisation) und Arbeitsmethoden. ⁷⁰

2.6.2.1.3 Arbeitspaket 3: Akzeptanz der Bundesagentur für Arbeit

Gegenstand des dritten Arbeitspakets ⁷¹ war die Untersuchung der Entwicklung der Akzeptanz (oder des Images) der Bundesagentur für Arbeit im zeitlichen Verlauf, insbesondere im Hinblick auf die Wahrnehmung der angebotenen

69 *Rheinisch-Westfälisches Institut für Wirtschaftsforschung/Institut für Sozialforschung und Gesellschaftspolitik*: Evaluation der Umsetzung der Vorschläge der Hartz-Kommission – Arbeitspaket 1, Modul 1f: Verbesserung der beschäftigungspolitischen Rahmenbedingungen und Makrowirkungen der aktiven Arbeitsmarktpolitik. Forschungsvorhaben im Auftrag des Bundesministeriums für Arbeit und Soziales, Endbericht, Juni 2006: Zusammenfassung und Schlussfolgerungen, download unter http://www.bmas.de/SharedDocs/Downloads/DE/PDF-Publikationen/fb-f356f-evaluation-der-massnahmen-hartz-arbeitspaket-1f.pdf?__blob=publicationFile [letzter Zugriff am 26.06.2013].

70 *Daniel Bieber/Volker Hielscher/Peter Ochs/Christine Schwarz/Simon Vaut* (= *Organisationsberatung Peter Ochs/Institut für Sozialwirtschaft und Sozialforschung e. V. [iso]*): Organisatorischer Umbau der Bundesagentur für Arbeit, Evaluation der Maßnahmen zur Umsetzung der Vorschläge der Hartz-Kommission, Arbeitspaket 2, Bericht, Berlin 2005, download unter http://www.bmas.de/SharedDocs/Downloads/DE/PDF-Publikationen/fb-f348-7-arbeitspaket-2.pdf?__blob=publicationFile [letzter Zugriff am 26.06.2013].

71 *infas Institut für angewandte Sozialwissenschaft GmbH*, Evaluation der Maßnahmen zur Umsetzung der Vorschläge der Hartz-Kommission, Arbeitspaket 3: Akzeptanz der Bundesagentur für Arbeit, Bericht, Bonn 2005, download unter http://www.bmas.de/SharedDocs/Downloads/DE/PDF-Publikationen/fb-f348-8-arbeitspaket-3.pdf?__blob=publicationFile [letzter Zugriff am 26.6.2013].

Dienstleistungen und ihrer Leistungsfähigkeit durch die Arbeitgeber und Arbeitnehmer und die Bevölkerung. Die Akzeptanz wurde aufgrund repräsentativer telefonischer Befragungen im Frühjahr der Jahre 2004, 2005 und 2006 ermittelt.

2.6.2.1.4 Arbeitspaket 4: Administrative Unterstützung bei der Evaluation

Die Koordination der Hartz-Evaluierung umfasste die Unterstützung des *Bundesministeriums für Arbeit und Soziales (BMAS)* bei der Konzeption der Evaluierung, die laufende Koordination der Evaluierung und die Auswertung und Aufbereitung der Evaluationsergebnisse. Hierzu gehörte auch die Erstellung von jährlichen Gesamtberichten auf Basis der Einzelberichte der Evaluatoren.⁷²

2.6.2.1.5 Arbeitspaket 5: Aufbereitung evaluationsgeeigneter Daten

Das Arbeitspaket 5 beschäftigte sich schließlich mit der Aufbereitung evaluationsgeeigneter Daten für den gesamten Evaluationsprozess. Hierzu gehörten auch Sonderauswertungen von Dateien der Bundesagentur für Arbeit, die Ziehung von Stichproben sowie die Verknüpfung von Daten der Bundesagentur für Arbeit mit Erhebungsdaten anderer Institute. Hiermit wurde das Institut für Arbeitsmarkt- und Berufsforschung (IAB), das Forschungsinstitut der Bundesagentur für Arbeit, beauftragt.

2.6.2 Gesamtzusammenfassung durch das BMAS

Die Untersuchungen der Forschungsinstitute und Forschungskonsortien wurden durch das BMAS bzw. durch das IAB als Grundlage einer eigenen Gesamtzusammenfassung und -bewertung genommen.⁷³

72 Näheres hierzu bei Heyer (Fn. 14), vgl. auch http://www.bmas.de/SharedDocs/Downloads/DE/PDF-Publikationen/fb-hartz-evaluierung-ausgangslage.pdf?__blob=publicationFile [letzter Zugriff am 26.06.2013].

73 Vgl. z. B. *Bundesministerium für Arbeit und Soziales: Die Wirksamkeit moderner Dienstleistungen am Arbeitsmarkt. Bericht 2006 des Bundesministeriums für Arbeit und Soziales zur Wirkung der Umsetzung der Vorschläge der Kommission Moderne Dienstleistungen am Arbeitsmarkt (ohne Grundsicherung für Arbeitsuchende)*, Langfassung, Berlin, 20. Dezember 2006, download unter http://www.bmas.de/SharedDocs/Downloads/DE/PDF-Publikationen/f356-hartz-bericht-langfassung.pdf?__blob=publicationFile [letzter Zugriff am 26.06.2013].

2.6.2.3 Fortlaufende Querschnittsevaluationen

Zusätzlich zu den wissenschaftlichen Untersuchungen zu den Wirkungen der Grundsicherung für Arbeitsuchende, die im Rahmen des gesetzlichen Evaluationsauftrags nach § 55 des Zweiten Buches Sozialgesetzbuch (SGB II) bzw. der allgemeinen Arbeitsmarkt- und Berufsforschung nach § 282 des Dritten Buches Sozialgesetzbuch (SGB III) durchgeführt wurden, gab das Bundesministerium für Arbeit und Soziales (BMAS) sog. „fortlaufende Querschnittsevaluationen“ zu sog. „Querschnittsthemen“ in Auftrag und begleitete diese. Diese hatten z. B. die familien- oder gleichstellungspolitischen Auswirkungen der „Hartz-Reformen“⁷⁴ (oder später die Auswirkungen derselben auf Migrantinnen und Migranten) zum Gegenstand. Auch diese werden regelmäßig durch das Bundesministerium für Arbeit und Soziales ausgewertet, das die Unterrichtung des Bundestags durch die Bundesregierung vorbereitet.

2.6.2.4 Evaluation der kommunalen Option (§ 6c SGB II)

Von der *laufenden Wirkungsforschung* nach § 282 SGB III bzw. § 55 SGB II war die *vergleichende Begleitforschung* gem. § 6c SGB II zu trennen. Diese betraf [nur] die Frage der Bewährung der beiden möglichen Modelle der Trägerschaft der Grundsicherung für Arbeitsuchende: entweder bei den Arbeitsämtern (später: Arbeitsagenturen) oder bei den Landkreisen oder kreisfreien Städten (sog. „Optionskommunen“).

Denn die Zusammenlegung von Arbeitslosenhilfe und Sozialhilfe für erwerbsfähige Sozialhilfeempfänger zur Grundsicherung für Arbeitslose (ALG II) durch das „Hartz IV-Gesetz“ verlangte einen umfassenden organisatorischen Wandel: Während die Sozialhilfe bislang von den Kommunen und die Arbeitslosenhilfe von den Arbeitsämtern (für die Bundesanstalt für Arbeit) ausgezahlt wurden, erforderte das neue, gemeinschaftliche Finanzierungs- und Fördersystem ein Angebot ‚aus einer Hand‘. Es hätte nahe gelegen, die Trägerschaft für diese Leistungen bei der Bundesanstalt/Bundesagentur für Arbeit bzw. bei den für sie handelnden Arbeitsämtern/Arbeitsagenturen anzusiedeln. Nach § 44b SGB II a. F. konnten die Arbeitsagenturen mit den Kommunen aufgrund privatrechtlicher oder öffentlich-rechtlicher Verträge sogenannte „Arbeitsgemeinschaften“ einrichten. Allerdings war es dem Deutschen Landkreistag über die CDU/CSU ge-

74 *IAQ/FIA/Genda*: Bewertung der SGB II-Umsetzung aus gleichstellungspolitischer Sicht. Jahresbericht 2007 des Gender-Projekts – Kurzfassung –, Duisburg/Essen/Marburg 2007, download unter http://www.bmas.de/SharedDocs/Downloads/DE/PDF-Publikationen/for-schungsbericht-f364.pdf?__blob=publicationFile [letzter Zugriff am 26.06.2013].

lungen, ein „Optionsmodell“ durchzusetzen, nach dem bis zu 69 Landkreise oder kreisfreie Städte (analog zu der Zahl der Sitze im Bundesrat) die alleinige Trägerschaft der Grundsicherung für Arbeitssuchende gem. § 6 Abs. 1 SGB II beantragen konnten. Beide Modelle („Arbeitsgemeinschaften“, ARGEn) und „Optionskommunen“ (offiziell: „zugelassene kommunale Träger“, zkT) traten damit in Wettbewerb.

Das Modell war Ergebnis eines Kompromisses zwischen der Bundesregierung und den Regierungen der Länder. Es war auf sechs Jahre befristet (1. Januar 2005 bis 31. Dezember 2010). Auf der Grundlage eines nach § 6c SGB II a. F. vorgeschriebenen gesetzlichen Auftrags, den Wettbewerb zwischen den Modellen der Aufgabenwahrnehmung wissenschaftlich evaluieren zu lassen und die Ergebnisse dem Bundestag Ende 2008 vorzulegen, wollte das Parlament über die Fortführung des Optionsmodells entscheiden. Laut § 6c SGB II a. F. war für die Festlegung der Methodik der Evaluierung das Bundesarbeitsministerium in Zusammenarbeit mit den 16 Bundesländern unter Einbeziehung der drei kommunalen Spitzenverbände, der Bundesagentur für Arbeit sowie der Bundesvereinigung der Deutschen Arbeitgeberverbände und des Deutschen Gewerkschaftsbunds verantwortlich. Zur Einigung auf die konkreten Modalitäten hatte sich im Januar 2005 ein Arbeitskreis Evaluation konstituiert; mit der Gesamtkoordination dieser Evaluation wurde das Institut für Sozialforschung und Gesellschaftspolitik (ISG) beauftragt.⁷⁵

Ziel der Forschung war es, die beiden Modelle der Aufgabenwahrnehmung ARGE und zkT zu vergleichen und mit den Ergebnissen eine empirisch abgesicherte Informationsbasis für die weitere Ausgestaltung der organisatorischen Umsetzung der Grundsicherung für Arbeitssuchende zu schaffen.⁷⁶ Dazu sollte der Aktivierungsprozess in beiden Formen der Aufgabenwahrnehmung beobachtet und dabei auch die Organisationsvielfalt innerhalb der Gruppe der ARGEn und zkT berücksichtigt werden.⁷⁷ In beiden Modellten sollte – unter Berücksichtigung der Maßnahmenvarianz – der gesamte Aktivierungsprozess bezogen auf die Frage „Wer macht es besser?“ untersucht werden, und zwar anhand folgender drei Zielgrößen: Integration in ungeforderte Erwerbsarbeit, Erhalt bzw. Verbesserung der Beschäftigungsfähigkeit, soziale Stabilisierung.⁷⁸

75 Baethge-Kinsky/Bartelheimer/Wagner (Fn. 17), S. 8.

76 Baethge-Kinsky/Bartelheimer/Wagner (Fn. 17), S. 8.

77 Ebd.

78 Vgl. den Endbericht des *Bundesministeriums für Arbeit und Soziales*: Evaluation der Experimentierklausel nach § 6c SGB II – Vergleichende Evaluation des arbeitsmarktpolitischen Erfolgs der Modelle der Aufgabenwahrnehmung „Zugelassener kommunaler Träger“ und „Arbeitsgemeinschaft“, Berlin 2008, http://www.sozialpolitik-aktuell.de/tl-files/sozialpolitik-aktuell/_Politikfelder/Arbeitsmarkt/Dokumente/Gesamtbericht%20%C2%A7%206c%20SGB%20II.pdf [letzter Zugriff am 26.06.2013], s. a. BT-Drs. 16/11488.

Die Evaluation wurde in vier Untersuchungsfelder aufgeteilt, die an vier Forschungskonsortien vergeben wurden:

- Untersuchungsfeld 1: Deskriptive Analyse und Matching⁷⁹,
- Untersuchungsfeld 2: Implementations- und Governanceanalyse⁸⁰,
- Untersuchungsfeld 3: Wirkungs- und Effizienzanalyse⁸¹,
- Untersuchungsfeld 4: Makroanalyse und regionale Vergleiche⁸².

Auf der Basis der Evaluationsergebnisse sollte entschieden werden, wer künftig für die Grundsicherung für Arbeitssuchende zuständig sein sollte.

2.7 Evaluationsergebnisse

Hinsichtlich der Evaluationsergebnisse, die die weitere Debatte beeinflussten, seien nun einige schlagwortartig herausgegriffen, die für die hier behandelte Thematik besonders interessant sind.

- 79 *Institut für angewandte Wirtschaftsforschung (IAW)/Zentrum für Europäische Wirtschaftsforschung (ZEW)*: Endbericht an das Bundesministerium für Arbeit und Soziales, Evaluation der Experimentierklausel nach § 6c SGB II – Vergleichende Evaluation des arbeitsmarktpolitischen Erfolgs der Modelle der Aufgabenwahrnehmung „zugelassener kommunaler Träger“ und „Arbeitsgemeinschaft“, Untersuchungsfeld I: „Deskriptive Analyse und Matching“, Tübingen 2008, download unter http://www.bmas.de/SharedDocs/Downloads/DE/PDF-Publikationen/forschungsbericht-f385.pdf?jsessionid=2030D3258C48630C3269003BACA328C5?__blob=publicationFile [letzter Zugriff am 26.06.2013].
- 80 *Wissenschaftszentrum Berlin/infas/Fachhochschule Frankfurt am Main*: Evaluation der Experimentierklausel nach § 6c SGB II – Vergleichende Evaluation des arbeitsmarktpolitischen Erfolgs der Modelle der Aufgabenwahrnehmung „Zugelassener kommunaler Träger“ und „Arbeitsgemeinschaft“, Untersuchungsfeld 2: Implementations- und Governanceanalyse, download unter <http://doku.iab.de/externe/2009/k090216f20.pdf> letzter Zugriff am 26.06.2013].
- 81 *Zentrum für Europäische Wirtschaftsforschung (ZEW)/Institut Arbeit und Qualifikation (IAQ)/TNS Emnid*: Evaluation der Experimentierklausel nach § 6c SGB II – Vergleichende Evaluation des arbeitsmarktpolitischen Erfolgs der Modelle der Aufgabenwahrnehmung „Optierende Kommune“ und „Arbeitsgemeinschaft“, Untersuchungsfeld 3: „Wirkungs- und Effizienzanalyse“, Erster Bericht durch den Forschungsverbund Mannheim/Gelsenkirchen/Bielefeld 2007, download unter http://www.bmas.de/SharedDocs/Downloads/DE/PDF-Publikationen/forschungsbericht-f362.pdf?__blob=publicationFile [letzter Zugriff am 26.06.2013].
- 82 *Institut für Wirtschaftsforschung (Ifo)/Institut für angewandte Wirtschaftsforschung (IAW)*: Evaluation der Experimentierklausel nach § 6c SGB II, Feld 4: Makroanalyse und regionale Vergleiche, Studie im Auftrag des Bundesministeriums für Arbeit und Soziales, München/Tübingen 2007, download unter http://www.sh-landkreistag.de/media/custom/100_23337_1.PDF [letzter Zugriff am 26.06.2013].

2.7.1 Personal-Service-Agenturen

Mit dem „Ersten Gesetz für moderne Dienstleistungen am Arbeitsmarkt“ („Hartz I“) wurde es möglich, private *Personal-Service-Agenturen* (PSAn) für „vermittlungsorientierte Arbeitnehmerüberlassung“ bei den Arbeitsämtern (ab dem „Hartz III-Gesetz“: den Agenturen für Arbeit) einzurichten (vgl. § 37c SGB III a. F., dieses geschah durch den Abschluss von Verträgen zwischen den Arbeitsämtern/Arbeitsagenturen und den im Sinne des Arbeitnehmerüberlassungsgesetz erlaubten tätigen Verleihern)⁸³. Nach Abs. 1 der zum 1. Januar 2003 in Kraft getretenen Neufassung der Bestimmung hatte jedes Arbeitsamt bzw. ab 2004 jede Arbeitsagentur die Einrichtung mindestens einer PSA sicherzustellen. Ursprünglich konnte sich die Agentur für Arbeit namens der Bundesagentur an Verleihunternehmen beteiligen (Abs. 3) beziehungsweise solche gründen, wenn keine Verträge zustande kamen (Abs. 4). Die den PSAn zugrunde liegende Idee war es, dass diese Agenturen Arbeitssuchende einstellen und diese entweder im Wege der Arbeitnehmerüberlassung nach dem Arbeitnehmerüberlassungsgesetz (insbesondere der Leiharbeit) an Dritte vermitteln oder sie in den dazwischen liegenden Zeiten weiterqualifizieren sollten. Dafür sollten die PSAn mit den Arbeitsämtern (ab 2004: Agenturen für Arbeit, diese im Namen der BA/Bundesagentur für Arbeit) nach Durchführung eines Vergabeverfahrens Verträge schließen, für die auch ein Honorar vereinbart werden konnte (Abs. 2).

Die hinter der Einrichtung der PSAn liegenden Anschauungen waren es zum einen, dass Private „besser“ vermitteln könnten als die recht schwerfälligen Arbeitsämter. Zum anderen bestand die – durchaus nachvollziehbare – Idee, dass die Arbeitssuchenden durch die Einstellung in die PSA ein sozialversicherungs-

83 § 37c SGB III (Personal-Service-Agentur) lautete in der bis 2008 geltenden Fassung: „(1) Jedes Arbeitsamt hat die Einrichtung mindestens einer Personal-Service-Agentur sicherzustellen. Aufgabe der Personal-Service-Agentur ist insbesondere, eine Arbeitnehmerüberlassung zur Vermittlung von Arbeitslosen in Arbeit durchzuführen sowie ihre Beschäftigten in verleihefreien Zeiten zu qualifizieren und weiterzubilden. (2) Zur Einrichtung von Personal-Service-Agenturen schließt das Arbeitsamt namens der Bundesanstalt mit erlaubt tätigen Verleihern Verträge. Für die Verträge mit den Personal-Service-Agenturen gilt das Vergaberecht. Kommen auf diese Weise Verträge nicht zu Stande, ist das ursprüngliche Vergabeverfahren aufzuheben und ein neues Vergabeverfahren über denselben Leistungsgegenstand durchzuführen. Das Arbeitsamt kann für die Tätigkeit der Personal-Service-Agenturen ein Honorar vereinbaren. Eine Pauschalierung ist zulässig. [...] (3) Sind Verträge nach Absatz 2 nicht zu Stande gekommen, kann sich das Arbeitsamt namens der Bundesanstalt an Verleihunternehmen beteiligen. [...]. (4) Kommt auch eine Beteiligung nach Absatz 3 nicht zu Stande, kann das Arbeitsamt namens der Bundesanstalt eigene Personal-Service-Agenturen gründen. Absatz 3 Satz 2, 3 und 4 gilt entsprechend. (5) In den Fällen des Absatzes 3 oder 4 hat das Arbeitsamt mindestens einmal jährlich zu prüfen, ob zu einem späteren Zeitpunkt Verträge nach Absatz 2 geschlossen werden können.“

pflichtiges Beschäftigungsverhältnis eingehen würden, das ihnen wegen der Zahlung der Beiträge in die Renten-, Kranken-, Arbeitslosenversicherung letztlich zugute kommen würde.

Die Evaluation der PSA-Einrichtung durch die Evaluatoren und insbesondere auch das BMAS fiel recht ernüchternd aus: So habe sich die Einrichtung von PSA als untaugliches Arbeitsmarktinstrument erwiesen. Einige PSA-Beschäftigte seien sogar später in den Arbeitsmarkt integriert worden als vergleichbare Arbeitslose. Darüber hinaus führe ein Beschäftigter in einer PSA zu durchschnittlichen Mehrkosten von 5700 Euro, denen keine spürbare Verbesserung seiner Beschäftigungschancen gegenüberstehe.⁸⁴

Das den Evaluationsbericht für das BMAS erstellende Projektkonsortium hatte neben den soeben zitierten Effekten auch das Auftreten von Mitnahmeeffekten privater Anbieter kritisiert. Das BMAS fasste die Gründe hierfür folgendermaßen zusammen: „Allerdings beenden PSA-Beschäftigte *aufgrund* dieser Tätigkeit *später* als vergleichbare Arbeitslose ihre Arbeitslosigkeit bzw. PSA-Beschäftigung durch eine Integration in Erwerbstätigkeit. Wesentlich hierfür ist der sog. Lock-In-Effekt, d.h. dass mit dem Übergang in eine PSA die Vermittlungs- und die Eigensuchaktivität in Bezug auf reguläre Beschäftigung zum Erliegen kommen.“⁸⁵ Es kam daher aufgrund der Expertengutachten zu dem Schluß „PSA verschlechtern also die Eingliederungschancen der Teilnehmer/innen“.⁸⁶ Nicht weiter thematisiert wurde allerdings der Umstand, dass die PSAn nicht die Klientel der ‚klassischen‘ privaten Arbeitsvermittler bzw. Zeitarbeitsfirmen hatten, die sich ihre Klienten aussuchen konnten. Auch wurde der Umstand, dass durch die Einstellung bei einer PSA aus sozialpolitischen Gründen unmittelbar ein sozialversicherungspflichtiges Beschäftigungsverhältnis entstehen sollte, nicht weiter gewürdigt.

2.7.2 Arbeitsbeschaffungsmaßnahmen

Unter „Arbeitsbeschaffungsmaßnahmen“ (ABM), die auch schon vor 2003 existierten, verstand man kurz befristete Tätigkeiten, die in der Regel keine oder eine nur geringe Qualifikation voraussetzten. Die entsprechenden ‚Stellen‘ fanden

84 Bundesministeriums für Arbeit und Soziales: Die Wirksamkeit moderner Dienstleistungen am Arbeitsmarkt, Bericht 2006 des Bundesministeriums für Arbeit und Soziales zur Wirkung der Umsetzung der Vorschläge der Kommission Moderne Dienstleistungen am Arbeitsmarkt (*ohne Grundsicherung für Arbeitsuchende*), download unter http://www.bmas.de/SharedDocs/Downloads/DE/PDF-Publikationen/f356-hartz-bericht-langfassung.pdf?__blob=publicationFile [letzter Zugriff am 26.6.2013], insbesondere S. 87f.

85 Ebd., S. 87f.

86 Ebd.

sich meist bei kommunalen oder gemeinnützigen Trägern. Allerdings hatte sich durch den Einsatz von ABM ein sog. „zweite[r] Arbeitsmarkt“ entwickelt, in dem es „immer mehr um ‚Maßnahmen‘ und immer weniger um alternative Beschäftigung“ ging, so dass die dort beschäftigten Personengruppen aus diesem „temporäre[n] Auffangbecken“ nicht mehr herauszukommen drohten.⁸⁷ Nach der Wiedervereinigung wurden ABM mit einigen Besonderheiten in den neuen Bundesländern eingesetzt, um „die Folgen der massiven Deindustrialisierung abzu-dämpfen und rasch für neue (Übergangs-)Beschäftigung zu sorgen.“⁸⁸ Die Träger, sogenannte „Beschäftigungsförderungsbetriebe“ oder „-gesellschaften“ beschäftigten keinesfalls nur gering Qualifizierte, sondern wurden zu wichtigen Sozialunternehmern z. B. im kulturellen, sozialen oder ökologischen Bereich. Allerdings wurde diese Form der Beschäftigungspolitik in Ostdeutschland schon bis 2004 deutlich zurückgefahren und passte sich nach und nach den westdeutschen Verhältnissen an.⁸⁹ Ab 2004 war es aufgrund des „Hartz IV-Gesetzes“ nicht mehr möglich, sich durch eine ABM-Beschäftigung einen neuen Anspruch auf Arbeitslosengeld zu erarbeiten. Dadurch wurden die ABM-Beschäftigungen unattraktiver. Zudem traten die sog. „Ein-Euro-Jobs“ (offiziell „Arbeitsgelegenheiten mit Mehraufwandsentschädigung“) in Konkurrenz zu den ABM-Maßnahmen.

Die Evaluation dieser arbeitsmarktpolitischen Instrumente durch die Forschungskonsortien fiel differenziert aus. Ihnen wurde auf der einen Seite bescheinigt, dass sie oftmals ge- oder missbraucht wurden, um die Arbeitslosenstatistik zu ‚frisieren‘, da Teilnehmer einer Arbeitsbeschaffungsmaßnahme nicht als Arbeitslose in der Statistik geführt wurden. Bemängelt wurde auch, dass der genannte „zweite Arbeitsmarkt“, der mit diesem Instrument geschaffen wurde, zu Lohndumping führen könne, da mithilfe der Maßnahmen unterstützte Arbeit unter den realen Kosten und damit unter den marktüblichen Preisen angeboten werden könne, was wiederum Unternehmen im ersten Arbeitsmarkt die Wettbewerbsgrundlage entziehen würde. Gleichzeitig erhielten die Teilnehmer der Maßnahme für diese Arbeit kein hinreichendes Einkommen, sondern seien letztlich weiterhin (ergänzend) auf staatliche Transferleistungen angewiesen.

Positiv wurde angeführt, dass die ABM und die sie tragenden Beschäftigungsgesellschaften „multiple Zielkonstellationen“ verfolgen würden, die sie nicht nur zu einem „schlichte[n] Vollzugsorgan der Arbeitsmarkt- und Sozial(hilfe)politik“⁹⁰ machten, sondern weit in die Zivilgesellschaft hineinreichten:

87 Bode (Fn. 56), S. 20.

88 Ebd.

89 Ebd.

90 Bode (Fn. 56), S. 22.

„Es geht *immer auch* um Armutsbekämpfung, um die Erweiterung des (lokalen) und sozialen Dienstleistungsangebots, um die Förderung der (regionalen) infrastrukturellen und wirtschaftlichen Entwicklung oder die Vertretung spezifischer Klientele bzw. sozialpolitischer Ziele. Dies geschieht durch die Mitwirkung in politischen Netzwerken und in aktiver – z. T. öffentlicher – Auseinandersetzung mit der lokalen Sozialpolitik bzw. der nationalen Gesetzgebung. Die Träger nehmen Einfluss auf örtliche Sozial- und Arbeitsämter, bemühen sich um das Design neuer Integrationsprogramme und versuchen, die Perspektiven spezifischer Zielgruppen zu thematisieren.“⁹¹

Das BMAS kam in seiner Gesamtevaluation allerdings – bei aller Konzedierung positiver sozialpolitischer Auswirkungen von ABM – zu einer negativen Einschätzung des Instruments.⁹²

2.7.3 Ergebnisse der fortlaufenden Querschnittsevaluation

Die fortlaufenden Querschnittsevaluationen, die eine Betrachtung sämtlicher arbeitsmarktpolitischer Instrumente unter einer spezifischen Fragestellung zum Inhalt hatten, beschäftigten sich beispielsweise mit den Effekten dieser Instrumente auf gleichstellungs- und familienpolitische Fragen. Dabei erwies es sich, dass insbesondere die Einführung des Instruments der „Minijobs“ sowie die Veranlagung von SGB II-Empfängern in „Bedarfsgemeinschaften“ nachteilige Auswirkungen auf die Einkommensstruktur in den Familien hatten. Dies geschah zum einen durch die beobachtete Umwandlung sozialversicherungspflichtiger Arbeit im Niedriglohnsektor, in dem üblicherweise zumeist Frauen tätig sind, in sog. „Minijobs“. Zum anderen war ein Trend zur „Verdrängung“ von Frauen aus dem Berufsleben festzustellen, da das Gehalt *beider* Ehepartner (oder in fester gemeinschaftlicher Verbindung Lebender) bei Bedarfsgemeinschaften bei den Zuverdienstmöglichkeiten nach ALG II angerechnet wird. Da üblicherweise die Frauen diejenigen sind, die den schlechter bezahlten Beruf ausüben, gaben sie oftmals ihren Beruf auf, um den gemeinschaftlichen Zuverdienst zu senken.

Es zeigte sich ebenfalls, dass viele Maßnahmen im Bereich auf die leicht wieder vermittelbaren Kunden (vor allem Akademiker/innen) konzentriert wurden (sog. „*creaming*“-Effekt⁹³). Die ‚Schere‘ zwischen leicht und schwerer Vermittelbaren schien daher immer weiter ‚auseinanderzugehen‘.

Darüber hinaus erwies sich das Konzept des „Forderns und Förderns“ als zu anspruchsvoll. Es scheint sich nur bei denjenigen positiv auszuwirken, die über

91 Bode (Fn. 56), S. 22, Hervorhebungen im Original.

92 Bundesministeriums für Arbeit und Soziales (Fn. 84), S. 87 („negative Integrationswirkung“).

93 Unter „*creaming*“-Tendenzen werden in der Arbeitsmarktpolitik „Neigungen der Träger, sich aus Kostengründen zuvorderst mit der am einfachsten (in den Normalarbeitsmarkt) zu reintegrierenden Klientel zu befassen“ verstanden, vgl. Bode (Fn. 56), S. 32.

gewisse (Bildungs-)Voraussetzungen verfügen. Ähnliche Ergebnisse zeigten sich bei jüngeren Querschnittsevaluationen bezüglich der Auswirkung der arbeitsmarktpolitischen Instrumente auf Migrantinnen und Migranten.

2.7.4 Evaluation der kommunalen Option (§ 6c SGB II)

Die bereits begonnene Evaluation der unterschiedlichen Organisationsformen aufgrund der „Experimentierklausel“ erbrachte erste Ergebnisse, abhängig von der jeweiligen Organisationsform. Es zeichnete sich eine vorsichtige Tendenz zur Zusammenarbeit in sog. „Arbeitsgemeinschaften“ (ARGEn) ab. In die Evaluation derselben, die auch durch eine unabhängige Begleitforschung im Auftrag des Deutschen Landkreistages⁹⁴ u. a. analysiert worden war, kam das Urteil des Bundesverfassungsgerichts zu den ARGEn vom 20. Dezember 2007 hinein und bestimmte die rechtlichen Rahmenbedingungen neu. Bestimmte Gestaltungsoptionen, die sich aufgrund der Evaluationsergebnisse angeboten hätten, schieden danach aus. Denn das Urteil des Bundesverfassungsgerichts hatte den „Grundsatz eigenverantwortlicher Aufgabenwahrnehmung“ der verschiedenen Verwaltungsträger als Ausfluss des grundgesetzlichen Prinzips der grundsätzlich verbotenen „Mischverwaltung“ herausgearbeitet, gegen den die Einrichtung und Arbeitsweise der ARGEn verstoße:

„Arbeitsgemeinschaften gemäß § 44b SGB II widersprechen dem Grundsatz eigenverantwortlicher Aufgabenwahrnehmung, der den zuständigen Verwaltungsträger verpflichtet, seine Aufgaben grundsätzlich durch eigene Verwaltungseinrichtungen, also mit eigenem Personal, eigenen Sachmitteln und eigener Organisation wahrzunehmen.“⁹⁵

2.8 Umsetzung der Evaluationsempfehlungen

2.8.1 Kommunale Option (ARGEn)

Der Richterspruch machte mithin unmöglich, was sich ersten Evaluationsergebnissen zufolge bewährt hatte. Er war von einer sehr grundsätzlichen Vorstellung des grundgesetzlichen Gebots der Trennung der Verwaltungs- und Finanzierungskompetenzen im Bundestaat geprägt. Die Lösung wurde über die Etablie-

94 Joachim Jens Hesse: Evaluation der Aufgabenträgerschaft nach dem SGB II. Erster Zwischenbericht. Untersuchung im Auftrag des Deutschen Landkreistages, Berlin 2005, download unter http://www.internationales-institut.de/content/publikationen/Zwischenbericht_SGB_II_Text.pdf [letzter Zugriff am 26.06.2013].

95 Bundesverfassungsgericht, Urteil des Zweiten Senats vom 20. Dezember 2007 – 2 BvR 2433/04, 2 BvR 2434/04 –, BVerfGE 119, 331.

nung einer neuen „unechten“ oder „fakultativen“ Gemeinschaftsaufgabe⁹⁶ im Bundesstaat gesucht, nämlich die Schaffung von Art. 91e GG. Dessen Abs. 1 ermöglicht das Zusammenwirken von Bund und Ländern bzw. den nach Landesrecht zuständigen Gemeinden und Gemeindeverbänden „bei der Ausführung von Bundesgesetzen auf dem Gebiet der Grundsicherung für Arbeitsuchende“ und sieht eine entsprechende Kooperation sogar als Regelfall vor, von der nach Abs. 2 S. 1 der Bund eine Ausnahme machen konnte, in dem er zulässt, dass „eine begrenzte Anzahl von Gemeinden und Gemeindeverbänden auf ihren Antrag und mit Zustimmung der obersten Landesbehörde die Aufgaben nach Absatz 1 allein wahrnimmt“.

Trotz dieser fundamentalen externen Einwirkung in den Evaluationsprozess wurde gleichwohl ein Bericht zu den unterschiedlichen Bewirkungsformen erstellt, und die Erkenntnisse der Stärken und Schwächen wurden nach der Grundgesetzänderung einfachgesetzlich berücksichtigt.⁹⁷ Beispielsweise wurde die im Bereich der Grundsicherung für Arbeitsuchende eingeleitete Verlagerung der Entscheidungskompetenz auf die dezentrale Ebene der „Jobcenter“ für die Agenturen für Arbeit fortentwickelt. Der mit der Einführung des Vermittlungsbudgets eingeschlagene Weg der Zusammenfassung von mehreren Instrumenten und individuellen Förderleistungen wurde fortgesetzt.⁹⁸ Insbesondere wurde die Zusammenarbeit der Jobcenter mit den Personalserviceagenturen und weiteren Trägern der beruflichen Weiterbildung mithilfe sog. „Vermittlungsgutscheine“ erleichtert.⁹⁹

2.8.2 Personal-Service-Agenturen

Obwohl die Bilanz der Forschungskonsortien sowie des zusammenfassenden Berichts des BMAS bezüglich des Instruments der Personal-Service-Agenturen, wie unter 2.7.1 dargestellt, keinesfalls positiv ausgefallen war, wurde dieses Instrument in modifizierter Form beibehalten.¹⁰⁰ So entfiel zum 1. Januar 2006 die 2003 eingeführte Verpflichtung der Arbeitsämter (ab 2004 Arbeitsagenturen),

96 Begriff nach *Margrit Seckelmann*: „Konvergenz und Entflechtung im Wissenschaftsföderalismus von 1998 bis 2009, insbesondere in den beiden Etappen der Föderalismusreform“, in: dies./Stefan Lange/Thomas Horstmann (Hrsg.), *Die Gemeinschaftsaufgaben von Bund und Ländern in der Wissenschafts- und Bildungspolitik: Analysen und Erfahrungen*, Baden-Baden 2010, S. 61-86, 74, dort insbesondere Fn. 54.

97 *Deutscher Bundestag*: Gesetzentwurf der Bundesregierung „Gesetz zur Verbesserung der Eingliederungschancen am Arbeitsmarkt“ vom 24.6.2011, Drs. 17/6277.

98 Ebd., S. 2.

99 Ebd., S. 78.

100 *Bundesministeriums für Arbeit und Soziales* (Fn. 84), S. 87 f.

die Einrichtung mindestens einer PSA sicherzustellen. Ebenso entfiel die gesetzlich vorgesehene Möglichkeit der Arbeitsämter (ab 2004 Arbeitsagenturen), sich namens der BA bzw. Bundesagentur an privaten Verleihunternehmen zu beteiligen beziehungsweise solche zu gründen, wenn keine Verträge zustande kamen, da sie nicht genutzt wurde.

2.8.3 Arbeitsbeschaffungsmaßnahmen

Aufgrund der vorwiegend negativen Gesamteinschätzung des BMAS bezüglich der positiven Auswirkungen von Arbeitsbeschaffungsmaßnahmen wurde von Seiten des Bundestags beschlossen, dieses Instrument auslaufen zu lassen. Seit dem 1. Januar 2009 fielen die Arbeitsbeschaffungsmaßnahmen endgültig aus dem SGB II heraus. Jedoch haben die Beschäftigungsförderungsbetriebe keinesfalls ihren Sinn verloren, sondern unterliegen einem tief greifenden Wandel hin zu managerialistisch geführten Sozialunternehmen, die – insbesondere in Folge der „Hartz-Reformen“ – zu „Auftragnehmer[n] eng geschnittener Integrationsmodule (wie Berufsvorbereitung im Schnellverfahren, Profiling und Vermittlung, kurzfristige ‚Resozialisierungsmaßnahmen‘)“ aufgrund wettbewerblich ausgeschriebener Vergabeverfahren wurden und/oder zugleich in den „subventionierten Arbeitnehmerverleih“ eingestiegen sind.¹⁰¹ Im Rahmen der zweiten „Instrumentenreform“ von 2011/12 (nach 2008/09¹⁰²) wurde das Instrument der Arbeitsbeschaffungsmaßnahmen (ehemals §§ 260 bis 271 SGB III) aus dem SGB III herausgenommen.¹⁰³ Seit dem Inkrafttreten des „Gesetzes zur Verbesserung der Eingliederungschancen am Arbeitsmarkt“ zum 1. April 2012 werden keine neuen ABM mehr gefördert.

3. Fazit: Erfolgsbedingungen von Evaluationen und von „Politiklernen“?

3.1 Die Interpretation der Evaluationsergebnisse

Versucht man nun, aus den Erfahrungen mit der Umsetzung der Ergebnisse der Evaluationen der „Hartz-Gesetze“ Grundsätzliches für die Nutzung von Evaluationsergebnissen durch die Politik abzuleiten, so lässt sich wohl Folgendes sagen:

101 Bode (Fn. 56), S. 9.

102 Gesetz zur Neuausrichtung der arbeitsmarktpolitischen Instrumente vom 21.12.2008, BGBl. I/2008, S. 2917.

103 Vgl. den Art. 2 Nr. 19 des Gesetzes zur Verbesserung der Eingliederungschancen am Arbeitsmarkt vom 20. Dezember 2011, BGBl. I/2011, S. 2854.

Die Analyse des Umgangs von Politik und Verwaltung mit den Evaluationsergebnissen zu den „Hartz-Reformen“ erweist zum einen, dass sich Belege dafür finden lassen, dass sich Regierung und Verwaltung lediglich die Reputation der Wissenschaftler und die dadurch gewonnene *zusätzliche* Legitimation bei der Umsetzung ohnehin geplanter Reformen zu eigen gemacht haben. Zum anderen lassen sich aber auch Anhaltspunkte dafür finden, dass sie sich mit den *Ergebnissen und Empfehlungen selbst beschäftigt* haben. Für beides spricht, dass beim „Gesetz zur Neuausrichtung der arbeitsmarktpolitischen Instrumente“ von 2009 ein expliziter Bezug auf die Hartz-Evaluationen genommen wurde: Ausweislich der Gesetzesbegründung¹⁰⁴ soll „das Reformkonzept der Arbeitsmarktreformen“ der Vorgängerregierung grundsätzlich weitergeführt, aber um die Erkenntnisse aus den Evaluationen angereichert werden.¹⁰⁵ Mit dem 2009 eingeführten Instrument des „Eingliederungsgutscheins“ wurde sogar direkt eine Empfehlung aus den Evaluationen umgesetzt.¹⁰⁶

Allerdings ist die Bilanz der Umsetzung der Empfehlungen differenziert zu sehen. Denn als besonderes Problem der Evaluationen erwies sich die Entkopplung der politischen Zielformulierung von den eher kleinteilig-instrumentell ausgerichteten wissenschaftlichen Evaluationsaufträgen. So herrschte nicht nur über die zu überprüfenden Ziele (Effizienz oder soziale Aspekte?) Unsicherheit als auch über die Bewertungskriterien und die Indikatorenbildung.

Der Umstand, dass diese Unklarheit bestand, verschaffte auf der einen Seite den Evaluatoren Freiräume, die sie (etwa bei der Analyse der gleichstellungs- und familienpolitischen Auswirkungen der Instrumente) für sich zu nutzen wussten. Insoweit lässt sich anhand der „Hartz-Evaluationen“ durchaus eine Selbstbestätigung der Unabhängigkeit wissenschaftlicher Politikberatung feststellen. Auf der anderen Seite führten die genannten Unsicherheiten aber dazu, dass die „Hartz-Evaluationen“ äußerst komplexe und – wie anhand der eingangs erwähnten Zitate verdeutlicht – äußerst differenzierte Ergebnisse für verschiedene Instrumente und ebensolche Personengruppen im Rahmen komplexer Ursache-Wirkungs-Zusammenhängen zeitigten (wobei die unterschiedlichen Evaluationen für sich genommen von hoher Qualität waren). Da Arbeitsmarktpolitik durch komplexe Ursache-Wirkungs-Zusammenhänge gekennzeichnet ist, kann eine isolierte Instrumentenbetrachtung zwar entpolitisierend und konfliktdämpfend wirken, zugleich aber kausalanalytisch problematisch sein.

104 *Deutscher Bundestag*: Gesetzentwurf der Bundesregierung „Entwurf eines Gesetzes zur Neuausrichtung der arbeitsmarktpolitischen Instrumente“ vom 8.11.2008, Drs. 16/10810, Begründung, insbesondere S. 24.

105 Ebd., S. 22.

106 Vgl. ebd., S. 22.

Das Folgeproblem bestand darin, zu ermitteln, welche Handlungsempfehlungen für Regierung und Verwaltung aus derartig komplexen Befunden abzuleiten seien. Hier erwies es sich als problematisch, dass viele der positiven (und durchaus beabsichtigten) Auswirkungen der „Hartz-Reformen“ sich erst langfristig, also deutlich nach der Evaluation, zeigten¹⁰⁷, während die negativen Nebeneffekte (etwa gleichstellungspolitischer Art) schon relativ zeitnah zu Tage traten. Daher kann die Frage nach den Möglichkeiten und den sinnvollen Betrachtungszeiträumen für eine Analyse der Langfristwirkung von Instrumenten und Langfristorientierung einer fundierten Wirkungsforschung noch keinesfalls als beantwortet angesehen werden. Diese konnte in den hier untersuchten Evaluationen (noch) nicht vorgenommen werden, die sich eher mit einer Begleitforschung zur (kurzzeitmässigen) arbeitsmarktpolitischen Instrumentendebatte beschäftigen.

3.2 Umsetzung durch Politik und Verwaltung?

Eine weitere Frage ist, inwieweit die extern generierten Handlungsempfehlungen auch von Regierung und Verwaltung aufgegriffen wurden. Die Analyse der „Hartz-Evaluationen“ ergibt, dass dann, wenn sich ein spezifisches Gelegenheitsfenster bietet, also etwa aufgrund der Unsicherheiten mit den gewählten Instrumenten, tatsächlich ein Bedarf an Evaluation empfunden wird. Weiter zeigt sich, dass dann, wenn Politik und Verwaltung Vertrauen in die Expertise der Evaluatoren haben, die evaluationsüblichen organisationsinternen Widerstände minimiert werden können. Aus den genannten Gründen zeigte sich bei diesen Evaluationen die gesetzesvorbereitende und -beobachtende (Ministerial-)Verwaltung als relativ offen gegenüber den Ergebnissen der Begleitforschung.

Allerdings boten die relativ komplexen Ergebnisse ihrerseits wiederum Interpretations- (und damit Handlungs-)Spielräume für die Politik. Diese bediente sich daher äußerst selektiv aus den Evaluationsergebnissen. Denn aufgrund des Fehlens eines Gesamtkonzepts bei der „Hartz-Evaluation“ sowie der Kleinteiligkeit der Fragenstellungen waren die Interpretation und die Gesamtbewertung des Systems äußerst umstritten. Externe Ereignisse wie das Urteil des Bundesverfassungsgerichts zu den sog. „ARGen“¹⁰⁸ machten die Sache nicht leichter.

107 Vgl. Brenke/Zimmermann (Fn. 37) sowie die in Fn. 38 genannten Aufsätze.

108 BVerfGE 119, 331 (vgl. Fn. 95).

3.3 Die „Hartz-Reformen“: Erfolgreicher oder erfolgloser Versuch der Implementation eines „lernenden arbeitsmarktpolitischen Systems“?

Daher ließe sich auf den ersten Blick zu dem Ergebnis kommen, dass der Versuch, im Rahmen der „Hartz-Evaluationen“ ein „lernendes arbeitsmarktpolitisches System“ einzurichten, eher ein Fehlschlag gewesen sei, der durch einen hohen Ressourceneinsatz bei eher uneindeutigen Ergebnissen und vielen offenen Fragen geprägt gewesen sei. Denn die späteren Änderungen der SGB II und III verliefen teilweise abgekapselt von der arbeitsmarktpolitischen Wirkungsforschung, viele Änderungen wurden ohne Bezug hierzu vorgenommen. Insbesondere fehlte bei den Evaluationen, die durch die Projektkonsortien im ersten Arbeitspaket vorgenommen wurden, eine kritische Erörterung der grundsätzlichen politischen Programmziele der „Hartz-Gesetze“ – sie war vom Gesetzgeber auch nicht gewollt.

Der Begriff des „Politiklernens“ wird – je nach wissenschaftstheoretischem oder kognitionswissenschaftlichen Zugang – sehr unterschiedlich definiert. Gemeinsam ist allen Definitionen aber wohl, dass dieser Begriff zumindest die „Veränderung von Überzeugungen, Wahrnehmungen und/oder Einstellungen“ im Rahmen politischer Prozesse umfasst.¹⁰⁹ Differenziert man nach *Arten des Lernens*, so lässt sich feststellen, dass wohl in erster Linie von einem *inkrementellen Lernen* auf der Ebene einzelner arbeitsmarktpolitischer Instrumente (Mikroebene) wie der Organisation der Leistungsverwaltung („Job-Center“, Mesoebene) gesprochen werden kann: So gingen Erkenntnisse aus der Wirkungsforschung zu Vermittlungsgutscheinen, Eingliederungszuschüssen, Minijobs und Bedarfsgemeinschaften durchaus in die weitere Arbeitsmarktpolitik ein.

Es stellt sich jedoch die Frage, ob darüber hinaus auch von einem *grundsätzlichen Lernen* ausgegangen werden kann. Dafür könnte sprechen, dass gewisse Elemente aus den „Hartz-Evaluationen“ auf Dauer gestellt wurden, wie etwa das Gleichstellungsmonitoring zum SGB II. Allerdings ist diese Einbindung responsiver Elemente in das Recht gemessen am Großprojekt der „Hartz-Reformen“ und ihrer Evaluationen relativ gering. Grundsätzliches Lernen setzt auch entsprechende Rahmenbedingungen voraus. So wies bereits Fritz Scharpf 1982 darauf hin, dass man sich für die Implementation eines innovativen arbeitsmarktpolitischen Programms – und diese Worte sind auch jetzt gültig – „drei Bedingungen gewünscht hätte, die erst ein optimales Experimentieren erlaubt hätten: Zeit

109 Nils Bandelow: Politisches Lernen: Begriffe und Ansätze im Vergleich, in: Klaus Schubert/ders. (Hrsg.), Lehrbuch der Politikfeldanalyse 2.0, München 2009, S. 313-347, 314, vgl. auch ders.; Lerntheoretische Ansätze in der Policy-Forschung, in: Matthias Leonhard Maier/Achim Hurrelmann/Frank Nullmeier/Tanja Pritzlaff/Achim Wiesner (Hrsg.), Politik als Lernprozess. Wissenszentrierte Ansätze in der Politikanalyse, Opladen 2003, S. 98-121.

für Lernprozesse, Ermessensspielräume für die Arbeitsämter und Anreize zu selektivem Vorgehen“, denn:

„Jede Verwaltung, die mit neuartigen, von den bisherigen Bedingungen und Verfahrensweisen abweichenden Programmen konfrontiert wird, braucht Zeit, um ihre eigene Vorgehensweise zu entwickeln, zu erproben und zu standardisieren. Dies gilt insbesondere dann, wenn damit neuartige Zielsetzungen verfolgt werden.“¹¹⁰

Und das, was Scharpf hier zur experimentellen Politik ausgeführt hat, gilt umgekehrt auch für die Evaluationen komplexer (z. B. arbeitsmarktpolitischer) Instrumente: Sie brauchen Zeit.

Daraus ergibt sich für das „politische Lernen“, dass auch dieses sich in langfristigen Zyklen vollzieht, die durch einen durchaus steinigen, aber wohl vorhandenen, Weg von der *Evaluation* über das *Lernen* zur *Weiterentwicklung der Arbeitsmarktpolitik* gekennzeichnet sind. Es handelt sich um einen Prozess des stetigen „Thematisieren[s], Anregen[s], Gestalten[s] und Durchführen[s]“¹¹¹, der in erster Linie von der Suche nach Mitstreitern¹¹² bzw. Koalitionspartnern¹¹³ geprägt ist. Experimentierklauseln wie diejenige zur „kommunalen Option“ können ein derartiges „Lernen“ institutionalisieren,¹¹⁴ wenn nicht zuvor externe Ereignisse dieses Lernen (zumindest auf direktem Wege) erschweren. Gleichwohl bleibt auch in den letztgenannten Fällen ein mittelbares Lernen, eine „lernende Beobachtung“¹¹⁵, möglich. Denn die Akteure in Politik und Verwaltung (und natürlich auch der Wissenschaft) lagern die vorgeschlagenen Modelle und die Kritik in ihren „Wissensspeichern“¹¹⁶ ab, die zu einem „Organisationsgedächtnis“¹¹⁷

110 Fritz W. Scharpf: Die Implementation des Sonderprogramms im Vergleich, in: ders./Dietrich Garlicks/Friederike Maier/Hans Maier, Implementationsprobleme offensiver Arbeitsmarktpolitik. Das Sonderprogramm der Bundesregierung für Regionen mit besonderen Beschäftigungsproblemen, Frankfurt/New York 1982, S. 271-310, 304.

111 Wolfram Lamping: Lernfähigkeit und Lernresistenz der Politik – Zur Bedeutung dieser Frage aus politikwissenschaftlicher Sicht, in: Dieter Schimanke/Andrea Fischer/Mathias Bucksteeg (Hrsg.), Wie lernt Politik? Voraussetzungen, Formen und Erfolge. Ein Werkstattbericht, Münster/New York/München/Berlin 2006, S. 17-39, 30.

112 Vgl. u. a. Marian Döhler: Lernprozesse in Politiknetzwerken, in: Werner Bussmann (Hrsg.), Lernen in Verwaltungen und Policy-Netzwerken, Chur/Zürich 1994, S. 39-54.

113 Nils Bandelow: Lernende Politik. Advocacy-Koalitionen und politischer Wandel am Beispiel der Gentechnologiepolitik, Berlin 1999.

114 Hartmut Borchert: Lernen des Gesetzgebers durch Experimentierspielraum für Verwaltungen, in: Hermann Hill/Hagen Hof (Hrsg.), Wirkungsforschung zum Recht II, Verwaltung als Adressat und Akteur, Baden-Baden 2000, S. 421-436.

115 Begriff (allerdings mit etwas anderer Zielrichtung) nach Martin Eifert: Lernende Beobachtung des Verwaltungsrechts durch das Verfassungsrecht, in: Michael Bäuerle/Philipp Dann/Astrid Wallrabenstein (Hrsg.): Demokratie-Perspektiven. Festschrift für Brun-Otto Bryde zum 70. Geburtstag, Tübingen 2013, S. 355-366.

116 Margrit Seckelmann: Benchmarking (Art. 91 d GG) und Evaluation – lernende Verwaltung?, in: Hermann Hill/Utz Schliesky (Hrsg.), Innovationen im und durch Recht. Evolution des Rechts- und Verwaltungssystems II, Baden-Baden 2010, S. 201-223.

von Politik und Verwaltung werden können. Werden ähnliche Fragestellungen in der Folgezeit wieder thematisiert, so kann es durchaus sein, dass zuvor – aus politischen Gründen, wegen der mangelnden aktuellen Durchsetzbarkeit oder bestimmter Trends – verworfene Modelle wieder aus der Schublade geholt oder bislang ausgeklammerte Grundsatzfragen gestellt werden.

Es kann indes nicht die Aufgabe von Evaluationen sein, politisch ausgesparte Themen nachträglich auf die Agenda zu setzen. Was Evaluationen aber durchaus können, ist die Unklarheit (arbeitsmarkt-)politischer Ziele und die Probleme, die sich aus deren Umsetzung ergeben, in das Bewusstsein zu rufen (Zielkonflikte, Instrumentenwahl, Effektivitäts- und Effizienzfragen etc.).

Der Umgang mit Evaluationsergebnissen indes muss in Deutschland noch ‚geübt‘ werden, um eine evidenzbasierte Politikkonzeption und -durchführung wie etwa in Großbritannien¹¹⁸ zu erreichen. Zu sehr noch ist dieser von den Gesetzmäßigkeiten des Parteienwettbewerbs und der oben erwähnten Angst vor der „Abhakliste für die Opposition“¹¹⁹ geprägt. Es kommen die Bedingungen der Massenmedien hinzu, die Fachwissen und „Infotainment“ gerne vermischen und zu Skandalisierungen neigen.¹²⁰ Die spezifische deutsche¹²¹ Kopplung zwischen Parteienwettbewerb und Evaluationen bzw. Performanzmessungen gilt es aufzubrechen, will man das Instrument der Evaluationen auch in Deutschland fruchtbar machen. Jörg Bogumil hat insoweit einen neuen „Umgang[] mit Fehlern in der öffentlichen Verwaltung“ angemahnt, der den Lernaspekt über die Angst vor dem Versagen stellt.¹²²

Noch weiter zu erforschen wäre, wie man die unglückliche Verbindung von Parteienwettbewerb im deutschen Bundesstaat (wo sehr häufig Wahlen stattfinden) und Evaluationen auflösen oder zumindest weiter lockern kann. Eine Möglichkeit ist es, und die wurde sowohl in der „Hartz-Kommission“ selbst wie auch bei den Evaluationen erprobt, möglichst viele *stakeholder* im Vorfeld zu integrieren. Eine andere ist es, möglichst breit anerkannte Expertise einzuwerben –

117 Henning Schridde: Verwaltungskultur, Change Management und lernende Organisation, in: Bernhard Blanke/Stephan von Bandemer/Frank Nullmeier/Götttrik Wewer (Hrsg.), Handbuch zur Verwaltungsreform, 3. Auflage 2005, S. 216-225, 217; vgl. auch den „Klassiker“ Chris Argyris/Donald A. Schön: Die Lernende Organisation, Stuttgart 1999.

118 Vgl. dazu Sanderson (Fn. 22).

119 Vgl. Fn. 25.

120 Katja Patzwald: Was leistete Politikberatung in den rot-grünen Arbeitsmarktreformen?, in: Klaus Dieter Wolf (Hrsg.), Staat und Gesellschaft – fähig zur Reform?, Baden-Baden 2007, S. 211-228, 224.

121 Siehe Ferdinand Schuster, Nicht wie Feuer und Wasser: Leistungsvergleiche und Öffentlichkeit – ein Zwischenruf, in: Sabine Kuhlmann/Jörg Bogumil/Helmut Wollmann (Hrsg.), Leistungsmessung und -vergleich in Politik und Verwaltung. Konzepte und Praxis, Wiesbaden 2004, S. 357-360, 357.

122 Bogumil (Fn. 21), S. 398.

auch das wurde bei den „Hartz-Evaluationen“ erprobt, in denen das IAB zwar eine bedeutende, aber keineswegs die einzige Rolle spielte. Allerdings können beide Lösungen, die auch kombiniert werden können oder sogar sollen, Evaluationen recht lange dauern lassen und für komplexe Sachverhalte und Fragen ggf. uneindeutige Ergebnisse produzieren. Aber das muss die Politik aushalten – und hier beginnt Politik! Daher führen Evaluationen und Performanzmessungen – bei allen Problemen – jedenfalls nicht zu einer ‚Entpolitisierung‘ – im Gegenteil!¹²³

Vielleicht aus diesem Grunde hat im Rahmen der letzten „Instrumentenreform“ der Gedanke der Zielsteuerung über Kennzahlenvergleiche (§ 48a) und der Zielvereinbarungen (§ 48b)¹²⁴ Eingang in das SGB II gefunden. Diese sind nicht unproblematisch,¹²⁵ aber sie sind ein Schritt auf dem Wege zu einer Verstehtigung politischen Lernens und damit auch – mit allen Abstrichen – einem weniger „blinden“¹²⁶ Umgang mit der Funktionsweise gesetzlicher Steuerung.

123 So auch die Arbeiten von *Veit* (Fn. 10).

124 Vgl. dazu <http://www.gib.nrw.de/service/downloads/zielsteuerung-im-sgb-ii-in-nrw> [zuletzt eingesehen am 26.06.2013].

125 Vgl. zu Leistungsvergleichen u. a. *Margrit Seckelmann*: „Informationen durch *performance measurement* – die Leistungsvergleiche nach Art. 91 d GG, Karlsruher Schriften zum Informationsrecht, Karlsruhe 2012, *dies.*: Wettbewerb per Grundgesetz? Die Leistungsvergleiche nach Art. 91d GG, DVBl. 2010, S. 1311-1314.

126 *Luzius Mader*: *L'évaluation législative. Pour une analyse empirique des effets de la législation*, Lausanne 1985, S. 176: “[L]’évaluation, législative me semble constituer un enrichissement pour le droit, une source d’informations qui rend moins aveugle à la réalité.”

Zum Einfluss von Evaluationen auf die politische Entscheidungsfindung

Gottfried Konzendorf

1. Einleitung und Zielsetzung

Es ist eine Trivialität, dass politischen Entscheidungen in modernen, demokratisch verfassten Gesellschaften in aller Regel komplexe Beteiligungsprozesse vorausgehen. An ihnen sind im föderalen System Deutschlands häufig staatliche Akteure unterschiedlicher Ebenen sowie gesellschaftliche Gruppen beteiligt. Oftmals spielen auch supra-staatliche Akteure und supra-staatliche Vorgaben (z.B. EU-Verordnungen, EU-Richtlinien) eine gewichtige Rolle im Entscheidungsprozess. Die unterschiedlichen Akteure bringen ihre Erkenntnisse und Interessen in den Entscheidungsprozess ein. Im Vorfeld dieser Einbringung befassen sie sich weitgehend autonom mit dem zur Entscheidung anstehenden Gegenstand. Die Managementlehre hat – unter Rückgriff auf den Neurophysiologen und Kybernetiker Warren McCulloch – für diese dezentralen Entscheidungs- und Machtprozesse den zutreffenden und anschaulichen Begriff der Heterarchie eingeführt. Die Evaluation¹ ist dabei nur eine, wenn auch besondere Form der Befassung mit einem Thema. Ihre Besonderheit besteht darin, dass wissenschaftliche Standards der Erkenntnisgewinnung (vgl. z. B. Deutsche Gesellschaft für Evaluation) zu beachten sind. Dennoch sind die Ergebnisse der Evaluation im politischen Entscheidungsprozess nur ein Faktor unter anderen. Es ist offen, ob und wie die Ergebnisse von Evaluationen im Einzelfall in eine politische Entscheidung einfließen.

In diesem Artikel sollen Faktoren identifiziert werden, die den Einfluss von Evaluation auf die politische Entscheidungsfindung erklären können. Angestrebt wird die Entwicklung einer Theorie zum Einfluss von Evaluation auf die politische Entscheidungsfindung. Dabei kann es sich – nach derzeitigen Erkenntnissen – nur um eine Theorie mittlerer Reichweite (Merton 1968) handeln; es ist davon auszugehen, dass die Zusammenhänge – aufgrund des gesellschaftlichen Wandels – an Raum und Zeit gebunden sind. Aufgrund der vorliegenden Erfahrungen

1 Der Begriff Evaluation wird in diesem Aufsatz – wie häufig in der Literatur – sowohl für den Prozess der Analyse und der Bewertung eines Sachverhalts als auch für die Ergebnisse dieses Prozesses benutzt; das je Gemeinte erschließt sich aus dem Kontext der Verwendung.

wird die Theoriebildung auf das politische System Deutschlands in der erstreckten Gegenwart begrenzt. Bei einer solchen Theoriebildung kann es nicht darauf ankommen, für jeden Einzelfall alle Erklärungsfaktoren aufzuführen. Vielmehr sollen solche Faktoren herangezogen werden, mit denen statistische Effekte erklärt werden können. Es geht also um die Entwicklung einer Theorie, die allgemein und einfach, aber nicht ganz genau bei der Analyse und Vorhersage von Einzelfällen ist (vgl. Weick 1985, S. 54 ff.). Eine zunächst diskursiv-qualitative und anschließend quantitative Überprüfung der hier explorativ formulierten Thesen zu Faktoren, die den Einfluss von Evaluation auf die politische Entscheidungsfindung erklären können, wird angestrebt.

2. *Wissenschaftliche Grundlagen*

Die Policy-Forschung befasst sich seit langem mit dem politischen Entscheidungsprozess. In den Arbeiten von Charles E. Lindblom zum Inkrementalismus (Lindblom 1959) sowie von Karl Popper zum Piecemeal Engineering (Popper 1965) wird hervorgehoben, dass politische Prozesse nicht umfassenden Planungsansätzen folgen, sondern sich in kleinen Schritten, eben inkremental vollziehen. Vor allem Lindblom zeigt die Unmöglichkeit einer politischen Gesamtplanung auf. Gelegentlich wird eine Übertragung der Argumente Lindbloms auf die Evaluation einzelner politischer Maßnahmen vorgenommen. Diese Übertragung erscheint jedoch ungerechtfertigt (vgl. Konzendorf 2011, S. 143). Im politischen Entscheidungsprozess sind Analysen und Bewertungen von avisierten Maßnahmen durchaus üblich und entscheidungsrelevant.

Im Policy-Cycle, der auf das Input-Output-Modell von David Easton zurückzuführen ist (Easton 1965), werden die Aktivitäten des Entscheidungsprozesses in einer logischen Abfolge beschrieben. Dieses Modell veranschaulicht, wie ein politischer Prozess idealerweise aussehen könnte. Diese logische Abfolge ist in politischen Entscheidungsprozessen jedoch nur ausnahmsweise anzutreffen. Darauf reflektieren sowohl das Garbage-Can-Modell von Michael Cohen, James March und Johan Peder Olsen (Cohen u.a. 1972), das Advocacy-Coalition-Framework von Paul A. Sabatier (Sabatier 1988) als auch das Multiple-Stream-Modell von John W. Kingdon (Kingdon 1995). Diese Ansätze betonen, dass der politische Prozess – anders als es der Policy-Cycle nahelege – nicht seriell ablaufe, vielmehr fänden Einzelereignisse des Prozesses unabhängig voneinander statt.

Das Multiple-Stream-Modell von Kingdon (Kingdon 1995) ist eine Weiterentwicklung des Garbage-Can-Modells. Stärker als in diesem werden im Multiple-Stream-Modell die Problemanalyse sowie die Entwicklung und Analyse von

Lösungsalternativen beachtet. Damit wird der Qualität im politischen Entscheidungsprozess eine höhere Bedeutung als im Garbage-Can-Modell zuerkannt.

Neben diesen Ansätzen sind die Forschungen zur Wissensnutzung (z. B. Caplan et. al. 1975, Weiss 1977, Wollmann 2008) und zur Politikberatung (z. B. Konzendorf 2002) für die hier behandelte Fragestellung nach dem Einfluss der Evaluation auf den politischen Entscheidungsprozess bedeutsam. Die einschlägigen Ansätze lassen sich drei Modellen der Politikberatung zuordnen:

1. dem wissenschaftszentriert-technokratischen Modell
2. dem dezisionistischen Modell
3. dem pragmatischen Modell.²

Während der Politiker – überspitzt formuliert – im wissenschaftszentriert-technokratischen Modell zu einem Vollzugsorgan der wissenschaftlichen Intelligenz verkümmert, bleibt im dezisionistischen Modell die Entscheidung des Politikers letztlich – trotz aller wissenschaftlichen Beratung – konkurrierenden weltanschaulichen Wertordnungen und Glaubensmächten verhaftet. Vielfach wird im Kontext des dezisionistischen Modells hervorgehoben, dass die wissenschaftlichen Erkenntnisse von politischen Akteuren instrumentell und taktisch genutzt werden. Nach dem pragmatischen Modell der Politikberatung findet eine kritische Interaktion zwischen Wissenschaft einerseits und Politik und Verwaltung andererseits statt.

„Anstelle einer strikten Trennung zwischen den Funktionen des Sachverständigen und des Politikers tritt im pragmatistischen Modell gerade ein kritisches Wechselverhältnis, das eine ideologisch gestützte Ausübung von Herrschaft nicht etwa nur einer unzuverlässigen Legitimationsbasis entkleidet, sondern im ganzen der wissenschaftlich angeleiteten Diskussion zugänglich macht und dadurch substanziell verändert. Weder ist der Fachmann, wie es im technokratischen Modell vorgestellt wird, souverän geworden gegenüber den Politikern, ...; noch behalten diese, wie das dezisionistische Modell unterstellt, außerhalb der zwingend rationalisierten Bereiche der Praxis ein Reservat, in dem praktische Fragen nach wie vor durch Willensakte entschieden werden müssen“ (Habermas 1978, S. 126 f.).

In Deutschland ist – nach Hellmut Wollmann – „das Interesse an der (wissenschaftlichen) Erforschung der Verwendung von Evaluationsergebnissen bislang bemerkenswert gering“ (Wollmann 2012, S. 7) geblieben. Im Anschluss an einen von der DFG geförderten Projektverbund fanden nur vereinzelt Folgeuntersuchungen statt (vgl. Wollmann 2012, S. 6).

2 Gelegentlich wird dieses Modell auch als pragmatistisches Modell bezeichnet.

3. Erste Erwartungen (Hypothesen) und Befunde

Fragt man nach dem Einfluss von Evaluationen auf die politische Entscheidungsfindung, so ist zunächst ein hoher Einfluss der Evaluation zu erwarten. Dies vor allem, da Evaluationen versprechen, Unsicherheiten bezüglich der Wirkungen, Kosten, Nebenwirkungen und anderer Wirkungskriterien von politischen Maßnahmen zu reduzieren. Dies gilt selbst, wenn nicht von einer kausalen Wirkungskette der Evaluation ausgegangen wird, nach der aus A zwangsläufig B folgt, sondern eine komplexe Vorstellung zugrunde gelegt wird. Wie Gerd Biesta ausführt, sind moderne Gesellschaften offene Systeme, in denen permanenter Wandel stattfindet und Informationen im Allgemeinen sowie Evaluationsergebnisse im Besonderen eine Interpretation durch soziale Akteure erfahren. Der Übertragbarkeit einer evidenzbasierten Erkenntnis (Evaluation) auf neue Entscheidungssituationen sind daher enge Grenzen gesetzt. In offenen Systemen gilt, aus A folgt B oder B' oder eine Variation davon. Um unter diesen komplexen Bedingungen nützliche und praxisrelevante Erkenntnisse liefern zu können, ist die Wissenschaft darauf angewiesen, den Adressaten einer Maßnahme zu verstehen. Das heißt, es ist zu fragen, warum eine Maßnahme wirkt und wie Maßnahmen zur Wirkung gebracht werden können (Biesta 2012). Ein solcher Evaluationsansatz muss das Verständnis, die Interpretation und die Akzeptanz der von einer politischen Maßnahme adressierten sozialen Akteure einbeziehen. Unabhängig davon, ob eine kausale oder komplexe Erkenntnis-Wirkungs-Annahme zugrunde gelegt wird, besteht zunächst die Vermutung einer positiven Rezeption von Evaluationsergebnissen durch politische Entscheidungsträger. Allerdings steht dieser Erwartung ein eher ernüchternder Befund entgegen.

Vielfach wird von einer Instrumentalisierung von Evaluationsergebnissen gesprochen. Politik – so wird argumentiert – bediene sich der analytischen Hilfestellung der Wissenschaft in instrumenteller Absicht zur „Erreichung bestimmter politisch (bereits vor-) entschiedener Ziele“ (Wollmann 2008, S. 353). Dies umfasse auch die Möglichkeit, dass politisch bereits getroffene Entscheidungen nachträglich legitimiert werden bzw. für politische Entscheidungen Zeit gewonnen werde. Der politischen Phantasie scheinen hier kaum Grenzen gesetzt.

Aus wissenschaftlicher Perspektive sei eine adäquate und zeitlich unmittelbare Rezeption und Umsetzung wissenschaftlicher Erkenntnisse in politische Entscheidungen „ein verhältnismäßig rarer Ausnahme- und ‚Glücksfall‘“ (Wollmann 2008, S. 354). Dieser Befund ändere sich jedoch, sobald die Möglichkeit mittel- und längerfristiger, in ihrem Ablauf komplexer und interaktiver Lern- und „Sicker“-Prozesse ins Auge gefasst würde. Unter dieser Perspektive „dürften vielfach durchaus handlungsrelevante, allerdings eher „sicherhaft“-inkrementelle denn umstürzende Nutzungs- und Lernprozesse und -ergebnisse sichtbar werden“ (Wollmann 2008, S. 21).

Zudem wird eine Kolonialisierung der Wissenschaft durch die Politik beobachtet. Angewandte Wissenschaft, also auch die Evaluationsforschung und besonders die Auftragsforschung, sei einem „strukturellen, bis in ihre Erkenntnisprämissen reichenden ‚Sog‘ ... ausgesetzt ... , der die Gefahr einer ‚Kolonialisierung‘ wissenschaftlicher Forschung durch Politik- und Praxisinteressen in sich“ (Wollmann 2008, S. 344) berge.

4. *Direkte Nutzung von Evaluation: Einflussfaktoren*

Im Weiteren soll aus pragmatischen Gründen auf die Analyse mittel- und längerfristiger Einflussnahmen der Evaluation verzichtet werden. Eine Analyse des Einflusses von Evaluation auf eine politische Entscheidungsfindung scheint unter der Rahmenbedingung mittel- und längerfristiger Sickereffekte methodisch kaum kontrollierbar. Stattdessen soll – trotz des ernüchternden Befundes – gefragt werden, welche Faktoren den direkten Einfluss von Evaluation auf den politischen Entscheidungsprozess beeinflussen. Aus praktischer Sicht ist damit die Frage verbunden, wie der Einfluss von Evaluation auf die politische Entscheidungsfindung erhöht werden kann.

Neben Ansätzen der Policy-Forschung und der Wissensnutzungsforschung fließen in die Analyse dieser Faktoren die Erkenntnisse aus Fallstudien sowie durch teilnehmende Beobachtung gewonnene Erfahrungen ein. Die Erklärungsfaktoren werden im Weiteren thesenartig eingeführt. Die Thesen werden kurz erläutert, Beispiele werden angeführt.

- I. These: Eine im Sinne der Fragestellung erfolgreiche Evaluation muss sich ko-evolutiv in politische und gesellschaftliche Veränderungsprozesse einpassen (Faktor: Ko-evolutive Einpassung der Evaluationsergebnisse).

Politik und Gesellschaft befinden sich in einem fortlaufenden Veränderungsprozess. In diesem Veränderungsprozess emergieren paradigmatisch gestützte Phasen mit neuen Handlungsmustern und Strukturen gesellschaftlicher Organisation sowie Übergangsphasen (bewegte Ordnung) (Böhret / Konzendorf 1997). Allerdings sind nicht Identität und Morphogenität die kennzeichnenden Prinzipien der Paradigmen, sondern evolutive Selbstähnlichkeit (Fraktalismus; vgl. Böhret / Konzendorf 1997, S. 213). Das bedeutet, dass innerhalb eines für einige Zeit handlungsleitenden Paradigmas durchaus Handlungsspielräume für Politik und Evaluation bestehen. Eine im Sinne der hier verfolgten Fragestellung erfolgreiche Evaluation muss sich in die bewegte Ordnung ko-evolutiv einpassen. Das heißt keineswegs, dass die Evaluation sich einem politischen Diktat (Kolonialisierung) unterordnen muss, um rezipiert zu werden. Vielmehr muss sie gesell-

schaftliche Probleme und aktuelle Lösungsmuster aufgreifen und innerhalb oder am Rand dieser Muster Lösungen für konkrete Problemlagen aufzeigen.

Ein Beispiel: Die Brain-Waste-Studie (Englmann u.a. 2007), in der die Praxis der Anerkennung ausländischer Qualifikationen in Deutschland analysiert wird, hat deutlich erkennbaren Einfluss³ auf das zum 1. April 2012 in Kraft getretene Gesetz zur Verbesserung der Feststellung und Anerkennung im Ausland erworbener Berufsqualifikationen („Anerkennungsgesetz“). Die Rezeption der Brain-Waste-Studie in der politischen Entscheidungsfindung steht im engen Zusammenhang zur „Neu“-Ausrichtung der Ausländerpolitik in Deutschland. Mit dem Wechsel zur Rot-Grünen Bundesregierung im Jahre 1998 hat sich Deutschland dazu bekannt, ein Einwanderungsland zu sein. In Folge dessen traten am 1. Januar 2000 erleichterte Einbürgerungsbestimmungen vor allem für Ausländerkinder in Kraft. Das Abstammungsprinzip (*Jus sanguinis* – „Recht des Blutes“) wurde vom Geburtsrecht (*Jus soli* – „Recht des Bodens, Landes“), wonach die Staatsangehörigkeit vom Geburtsort bzw. -land abgeleitet wird, abgelöst. Das Inkrafttreten des Gesetzes zur Steuerung und Begrenzung der Zuwanderung und zur Regelung des Aufenthalts und der Integration von Unionsbürgern und Ausländern (Zuwanderungsgesetz) im Januar 2005 markiert einen weiteren wichtigen Eckpunkt der Neuausrichtung in der Ausländerpolitik, die ihre Fortsetzung darin fand, dass die Große Koalition von CDU/CSU und SPD im Jahre 2005 das Thema Integration von Migranten zu einem Schwerpunkt (vgl. Meier-Braun 2006, S. 208) ihrer Politik bestimmte.

In diese Neuausrichtung der Ausländerpolitik passte sich die Brain-Waste-Studie ko-evolutiv ein. Die politische Neuausrichtung bildete den Nährboden für die politische Rezeption der Studie (vgl. Mobou Dzuno 2011, S. 105).

Voraussetzung für die handlungsrelevante ko-evolutive Einpassung von Evaluationen in die Entscheidungsfindung sind freilich Freiräume für die Erkenntnisgewinnung und für aus diesen abgeleitete Empfehlungen. Falls diese Freiräume etwa durch eine strikte Positionierung der Politik eingeengt sind, besteht für die Evaluation kaum eine Chance der Einflussnahme. Die Erfahrung zeigt, dass die Durchführung und Rezeption von Evaluationen kaum gelingt, wenn mögliche Handlungsspielräume durch strikte politische Vorgaben eliminiert werden.

3 Dass die Brain-Waste-Studie Einfluss auf die Entscheidungsfindung gehabt hat, bedeutet nicht, dass ihre Forderungen und Anregungen eins zu eins umgesetzt worden seien.

- II. These: Die Evaluation muss empirisch-gestützte Informationen und praktikable Lösungsvorschläge für den politischen Entscheidungsprozess bereitstellen (Faktor: Hohe Qualität (Evidenz, Relevanz) der Informationen)

Nach Kingdon steigt die Chance, ein Thema im politischen Raum zu positionieren (Agenda-Setting), wenn das zu lösende Problem hinreichend genau analysiert und beschrieben ist (vgl. Kingdon 1995, S. 93). Zudem besteht die Anforderung, dass die Lösungsvorschläge praktikabel, finanzierbar, zukunftssträftig und in der Fachcommunity akzeptiert sind (vgl. Kingdon 1995, S. 143 f.). Um diesen Anforderungen an das politische Agenda-Setting zu genügen, können Evaluationen einen gewichtigen Beitrag leisten. Sie können die bestehende Praxis empirisch analysieren und Lösungsvorschläge hinsichtlich ihrer Folgen abschätzen. Erfolgreiche Beispiele hierfür sind die Brain-Waste-Studie sowie das Gutachten „Modernisierung der Verwaltungsbeziehungen von Bund und Ländern“ (Präsident des Bundesrechnungshofes 2007), das der Präsident des Bundesrechnungshofes als Bundesbeauftragter für Wirtschaftlichkeit der Verwaltung in die „Kommission zur Modernisierung der Bund-Länder-Finanzbeziehungen“ (Föderalismuskommission II) eingebracht hat. Dieses Gutachten hat in der Kommission aufgrund seiner empirischen Analysen der bestehenden Verwaltungsbeziehungen von Bund und Ländern sowie aufgrund seiner praktikablen Lösungsvorschläge erhebliche Aufmerksamkeit erfahren.

Die Anforderung an die Qualität von Evaluationen implizieren mit Blick auf die politische Praxis umgekehrt, dass Informationen, die den Anforderungen nicht entsprechen im politischen Entscheidungsprozess ausgeblendet werden. An dieser Stelle sei der Hinweis auf die politische Nutzlosigkeit etlicher obligatorischer Berichts- und Dokumentationspflichten sowie auf allzu simple und standardisierte Evaluationsverfahren erlaubt. Diese führen zu Datenfriedhöfen, die zwar einen hohen Aufwand verursachen, aber kaum politischen Nutzen erzeugen. Im Hinblick auf vielfach anzutreffende Bestrebungen zur Verallgemeinerung und Standardisierung von Evaluationsverfahren wird erfahrungsbasiert vermutet, dass das Was (z.B. Ziele, Praktikabilität, Kosten, Nebenwirkungen etc.) einer politiknahen Evaluation weitgehend verallgemeinerbar ist (Böhret / Konzendorf 2001). Das Wie, also die speziellen Anforderungen an die Datenerhebung und ihre Auswertung, ist aber im hohen Maße vom Einzelfall abhängig. Im Prozess der Evaluation sind dabei oftmals neue, überraschende Fragestellungen zu beachten, für die eine vorgefertigte Evaluationsschablone keine Verfahren bereithalten kann.

III. These: Die Evaluationsergebnisse müssen rechtzeitig in den politischen Entscheidungsprozess eingebracht werden (Faktor: Rechtzeitige Informationsbereitstellung)

In formaler Hinsicht ist diese Anforderung an eine erfolgreiche Evaluation eine Banalität. Ergebnisse einer Evaluation müssen selbstverständlich rechtzeitig, z. B. vor einer regierungsinternen oder parlamentarischen Befassung, vorliegen. Nur dann können sie im politischen Entscheidungsprozess beachtet werden. Allerdings geht diese Anforderung über diesen formalen Aspekt hinaus. Rechtzeitig bedeutet mit Blick auf den Einfluss von Evaluationen z.B. auch, rechtzeitig vor einer öffentlichen parteipolitischen Positionierung. Die Logik des politischen Systems erschwert die Rücknahme einer einmal erfolgten parteipolitischen Festlegung. Daher verengen sich die Einflusschancen für Evaluation nach einer solchen Positionierung erheblich.

Im Einzelfall kann der „kritische“ Zeitpunkt für die Einbringung der Evaluationsergebnisse schwer kalkulierbar sein. Er kann z.B. von zufälligen katastrophalen oder medialen Ereignissen abhängen. Aus Sicht der Evaluation ist – um überraschende Ereignisse möglichst zu vermeiden – eine zeitliche Verzahnung zwischen politischer Öffentlichkeitsarbeit und Evaluationsprozess anzustreben.

IV. These: Die Herstellung von (Teil-)Öffentlichkeit ist ein gewichtiger Faktor für den Einfluss von Evaluation im politischen Entscheidungsprozess (Faktor: (Teil-)Öffentlichkeit / Transparenz)

Publikationen und (teil-)öffentliche Events haben für den Einfluss von Evaluation im politischen Entscheidungsprozess verschiedene Funktionen. So dienen sie dazu, die durchgeführten Analysen und die aus den Analysen abgeleiteten Empfehlungen und Problemlösungsvorschläge einer kritischen Reflektion zu unterziehen (vgl. Kingdon 1995, S. 113). Zudem reduzieren sie die Möglichkeit, dass ein Thema, Evaluationsergebnisse und ggf. zwischen verschiedenen Akteuren bereits abgestimmte Positionen von der politischen Agenda verschwinden. Die Bedeutung dieser (teil-)öffentlichen Events bestätigt sich bei der Analyse verschiedener Beispiele. So wurde – im Zusammenhang der bereits erwähnten Brain-Waste-Studie und dem Anerkennungsgesetz – im Juni 2009 die Fachtagung „Brain Waste – Anerkennung gestalten“ durchgeführt. Auch bei der Verankerung des Themas Leistungsvergleiche im Grundgesetz (Art. 91 d GG) erwiesen sich teil-öffentliche Veranstaltungen (Tagung: Evaluation und Benchmarking in Politik und Verwaltung im November 2007) als unterstützend. Wie bedeutsam (teil-)öffentliche Aufmerksamkeit und Transparenz sind, zeigt sich gerade in Fällen der politischen Ausblendung von Evaluationsergebnissen. Beispielsweise wurden die Ergebnisse der Evaluation einer neuen Organisationsstruktur politisch nicht rezipiert, obwohl sie gemeinsam mit betroffenen Organisationen erarbeitet wurden und dadurch praxisrelevante Bedeutung hatten. Da in

diesem nicht öffentlich zugänglichen Fallbeispiel auf öffentliche Events, Publikation und somit auf Transparenz verzichtet wurde, kann vermutet werden, dass die Ausblendung der Evaluationsergebnisse aus dem politischen Entscheidungsprozess auf diesen Faktor zurückzuführen ist.

V. These: Eine breite Akteursbasis ist für den Einfluss von Evaluation auf den politischen Entscheidungsprozess hilfreich (Faktor: breite Akteursbasis)

Politische Diskurse finden in voneinander getrennten Politikfeldern statt; ein politikfeldübergreifender Austausch ist die Ausnahme (vgl. z. B. Cohen u.a. 1972). Das heißt, jedes Politikfeld analysiert seine Probleme und entwickelt eigene Problemlösungsstrategien. Evaluationen finden somit im Normalfall auch nur politikfeldbezogene Aufmerksamkeit.

Im politischen Entscheidungsprozess kommt es jedoch in aller Regel darauf an, Bündnispartner für die Durchsetzung einer Position zu finden. Diese Partner können durchaus in anderen Politikfeldern gefunden werden, wenn z. B. durch Evaluationen nachgewiesen wird, dass ein Problemlösungsansatz Dysfunktionalitäten in verschiedenen Politikfeldern zu reduzieren oder aufzuheben verspricht. Auch hier kann das Thema Anerkennung ausländischer Qualifikationen als Beispiel angeführt werden. Die Politikfelder „Integration von Migranten“ und „Fachkräftemangel“ haben sich zunächst unabhängig voneinander entwickelt. In beiden Politikfeldern wurden Studien durchgeführt und Lösungsansätze diskutiert. Dabei wurde deutlich, dass das Thema Anerkennung ausländischer Qualifikationen eine relevante Problemlösungsstrategie für beide Politikfelder darstellen kann. Das Anerkennungsthema bildete somit eine Schnittstelle der beiden Politikfelder und führte zu einer befristeten Kopplung (strategische Allianz) von zwei unabhängigen Akteursgruppen (vgl. Mobou Dzuno 2011, S. 48 ff.). Diese Kopplung der Akteursbasis, die auf Evaluationserkenntnisse und abgeleitete Empfehlungen zurückführbar ist, erhöhte die politische Aufmerksamkeit für das Thema und somit die Chance seiner Umsetzung deutlich. Als Randbedingung kann vermutet werden, dass sich die Durchsetzungschancen für ein Thema erhöhen, wenn es wirtschaftliche Interessen aufgreift. Auch dies ist im Fallbeispiel gegeben.

VI. These: Interaktion zwischen Politik, Verwaltung und Evaluation stützt den Einfluss der Evaluation auf die politische Entscheidung (Faktor: Interaktion)

Eine Interaktion zwischen Politikern, Fachleuten aus Verwaltung und gesellschaftlichem Umfeld (Verbände, Zivilgesellschaft etc.) sowie Evaluatoren hilft, den Einfluss der Evaluation auf die Entscheidungsfindung zu erhöhen. Grundsätzlich kann in diesen Interaktionen kommunikatives oder zweckrationales Handeln erfolgen. Im Falle kommunikativen Handels basieren die Pläne zur Verfolgung eines Ziels auf der Grundlage einer gemeinsamen Situationsdefinition

(vgl. Habermas, 1981, Bd. 2, S. 193). Daraus ergeben sich selbstverständlich größere Handlungsfreiheiten und Einflusschancen der Evaluation als im Falle zweckrationalen Handelns. Aber auch im Falle einer zweckrationalen Ausrichtung der Interaktion besteht durchaus die Chance einer erheblichen Einflussnahme der Evaluation auf die Entscheidungsfindung, da die Mittel zur Erreichung eines vorgegebenen Zieles analysiert und bewertet werden.

Unabhängig davon, welche Ausrichtung der Evaluation vorliegt, erhöht die Interaktion zwischen den genannten Akteuren die Chance des Einflusses der Evaluation. Hierfür lassen sich sowohl auf der Input-Seite der Evaluation als auch auf ihrer Output-Seite praktische Gründe anführen. So können auf der Input-Seite durch den Austausch der Akteursgruppen besonders praxisrelevante Fragestellungen in eine Evaluation aufgenommen werden. Zudem dürften z. B. Hinweise zu relevanten Akteursgruppen hilfreich sein, um die Qualität und Akzeptanz der Evaluation steigern. Auf der Output-Seite ermöglicht die Interaktion die Einspeisung der Evaluationsergebnisse in die „geregelten“ Entscheidungsabläufe in Politik und Verwaltung.

5. Faktoren im Überblick und Ausblick

Gruppiert man die identifizierten Faktoren, so lassen sich eher datenorientierte Faktoren von prozessorientierten unterscheiden. Zu den datenorientierten Faktoren zählen:

- I. Ko-Evolutive Einpassung der Ergebnisse und Empfehlungen der Evaluation in gesellschaftliche Veränderungsprozesse
- II. Hohe Qualität (Evidenz, Relevanz) der Information
- III. Rechtzeitige Informationsbereitstellung.

Zu den eher prozessorientierten Faktoren zählen:

- IV. Herstellung von (Teil-)Öffentlichkeit / Transparenz
- V. Herstellung einer breiten Akteursbasis
- VI. Interaktion zwischen Evaluatoren, Politikern und Fachleuten aus Verwaltung sowie gesellschaftlichem Umfeld.

Mit Blick auf praktische Entscheidungsprozesse und die avisierte Theoriebildung wird vermutet, dass für eine erfolgreiche Rezeption und Umsetzung von Evaluationsergebnissen nicht zwingend alle Faktoren erfüllt werden müssen. Doch dürfte ein hoher Erfüllungsgrad bezüglich sämtlicher Faktoren den Einfluss der Evaluation auf die politische Entscheidungsfindung deutlich steigern.

Es wird weiter vermutet, dass zwischen den Faktoren Wechselbeziehungen bestehen. Dabei dürfte dem Faktor Interaktion eine besondere Funktion zukommen. So dürfte die Interaktion zwischen den drei genannten Akteursgruppen zum Beispiel Einfluss haben auf die Qualität und Relevanz der Evaluationsergebnisse.

se. Durch die Interaktion kann die Evaluation auf praxis- und entscheidungsrelevante Aspekte konzentriert werden. Auch die Herstellung von (Teil-)Öffentlichkeit dürfte durch gemeinsame Aktivitäten von Evaluatoren mit Akteuren der anderen beiden Gruppen leichter gelingen. Insbesondere dürfte die (teil-)öffentliche Aufmerksamkeit steigen, wenn politische Akteure in die Aktivitäten einbezogen sind. Die Herstellung von (Teil-)Öffentlichkeit dürfte wiederum die Erweiterung der Akteursbasis beeinflussen. Und ein Austausch zwischen den Akteuren macht die rechtzeitige Einbeziehung der Evaluationsergebnisse in politische Entscheidungen und Aktivitäten wahrscheinlicher – zumindest im „Normalfall“.

Falls die hier explorativ vorgetragenen Thesen einem diskursiven Prozess und einer quantitativen empirischen Überprüfung standhalten, kann das hier thesenartig vorgestellte Erklärungsmodell zum Einfluss von Evaluation auf die politische Entscheidungsfindung an das pragmatische Modell der Politikberatung von Jürgen Habermas anknüpfen. Dabei müssen allerdings neue Entwicklungen aufgenommen werden. So hat sich die politische Entscheidungsfindung durch den technologisch bedingten, rasanten Strukturwandel der Öffentlichkeit (Habermas 1962) in der jüngeren Vergangenheit ebenso gewandelt wie das Verhältnis von Politik, Verwaltung, Verbänden und Zivilgesellschaft. Die Evaluation muss sich in diesen Wandel ko-evolutiv einpassen. Weder das wissenschaftsorientiert-technokratische noch das dezisionistische Modell der Politikberatung scheinen diese Entwicklungen angemessen aufnehmen zu können. Das pragmatische, interaktive Modell scheint hingegen passgerecht. Dies nicht zuletzt, weil es die Möglichkeit eröffnet, durch gemeinsame Situationsdefinitionen der betroffenen Akteure komplexe Ursache-Wirkungsmechanismen aufzunehmen. Erfolgreiche Evaluationen dürften demnach „verstehende“ Evaluationen sein, in denen die Adressaten von politischen Maßnahmen als soziale Akteure und Subjekte begriffen werden.

Literatur:

- Biesta, Gerd*, Evidence, one more time: Towards a critical role for evaluation research, Jahrestagung der Deutschen Gesellschaft für Evaluation am 20.9.2012 in Potsdam, unveröffentlichter Vortrag 2012.
- Böhret, C. / Konzendorf, G.*, Ko-Evolution von Gesellschaft und funktionalem Staat – Ein Beitrag zur Theorie der Politik, Opladen / Wiesbaden, 1997.
- Böhret, C. / Konzendorf, G.*, Handbuch Gesetzesfolgenabschätzung (GFA) – Gesetze, Verordnungen, Verwaltungsvorschriften, Baden-Baden, 2001.
- Caplan, N. / Morrison, A. / Stambaugh, R.*, The Use of Social Science Knowledge in Policy Decisions at the National Level, Ann Arbor, 1975.

- Cohen, M. D. / March, J. G. / Olsen, J. P., A garbage can model of organizational choice, in: Administration Science Quarterly, 17, 1972.
- Deutsche Gesellschaft Für Evaluation, DeGEval-Standards, <http://www.degeval.de/degeval-standards>, 20.11.2012.
- Easton, D., A Framework for Political Analysis, Englewood Cliffs, 1965.
- Englmann, Bettina/Müller, Martina, Brain Waste. Die Anerkennung von ausländischen Qualifikationen in Deutschland, Augsburg, 2007.
- Habermas, Jürgen, Strukturwandel der Öffentlichkeit, Darmstadt und Neuwied, 1962.
- Habermas, Jürgen, Theorie des kommunikativen Handelns, Bd. 2, Frankfurt am Main, 1981.
- Habermas, Jürgen, Verwissenschaftliche Politik und öffentliche Meinung, in: Technik und Wissenschaft als Ideologie, Frankfurt am Main 1968, (9. Aufl.) 1978.
- Kingdon, J. W., Agendas, Alternatives, and Public Policies, New York, 1995.
- Konzendorf, G., Gesetzesfolgenabschätzung, in: Blanke, Bernhard / Nullmeier, Frank / Reichard, Christoph / Wewer, Göttrik (Hrsg.), Handbuch zur Verwaltungsreform, Wiesbaden, (4. Aufl.) 2011.
- Konzendorf, G., Politikberatung, in: Greiffenhagen, Martin / Greiffenhagen, Sylvia (Hrsg.), Handwörterbuch zur politischen Kultur der Bundesrepublik Deutschland, (2. Aufl.) 2002.
- Lindblom, The Science of Muddling-Through. Public Administration Review, Jg. 19, 1959.
- Meier-Braun, in: Landeszentrale für politische Bildung Baden-Württemberg (Hrsg.), Der Bürger im Staat, 56. Jg, Heft 4, Zuwanderung und Integration, 2006.
- Merton, R. K., Social Theory and Social Structure, [1949] 1968.
- Mobou Dzuno, Edith, Anerkennung ausländischer Qualifikationen, unveröffentlichte Diplomarbeit, Potsdam 2011.
- Popper, Karl, Das Elend des Historizismus, Tübingen, 1965.
- Präsident des Bundesrechnungshofes als Bundesbeauftragter für Wirtschaftlichkeit der Verwaltung, Modernisierung der Verwaltungsbeziehungen von Bund und Ländern, Kommissionsdrucksache 055 der Kommission zur Modernisierung der Bund-Länder-Finanzbeziehungen“ (Föderalismuskommission II), 2007.
<http://www.kulturrat.de/dokumente/brh-gutachten.pdf> vom 23.11.2012.
- Riescher, Gisela, Zeit und Politik, in: Nohlen, Dieter / Schultze, Rainer-Olaf (Hrsg.), Lexikon der Politikwissenschaft, Bd. 2, München, 3. Aufl., 2005.
- Sabatier, Paul A., An advocacy coalition framework of policy change and the role of policy-oriented learning, in: Policy Sciences, Volume 21, Issue 2-3, 1988.
- Weick, Karl E., Der Prozeß des Organisierens, Frankfurt am Main, 1985.
- Weiss, Carol H., Using Social Research in Public Policy Making, Lexington M.A.: D.C. Heath, 1977.
- Wollmann, Hellmut, Evaluation der Nutzung von Evaluationsergebnissen in Politik und Verwaltung. Konzepte, Ansätze, Beispiele, 2012. [discussion paper für Arbeitskreis Evaluierung und Verwaltung der Deutschen Gesellschaft für Evaluation, Jahrestagung 2012 am 19./21. September 2012 in Potsdam]
<http://amor.cms.hu-berlin.de/~h0598bce/html/publikationen.html>
- Wollmann, Hellmut, Kontrolle in Politik und Verwaltung: Evaluation, Controlling und Wissensnutzung, in: Schubert, Klaus / Bandelow, Nils C. (Hrsg.), Lehrbuch der Politikfeldanalyse, 2008.

Erfahrungen mit Performance-Informationen in der Kommunalverwaltung: Datennutzung, Routinen und Soziale Kognition

Alexander Kroll/Anne-Kathrin Meier

Dieser Beitrag verfolgt zwei Ziele. Zum einen soll mittels einer Umfrage dokumentiert werden, wie das Performance Management auf kommunaler Ebene funktioniert und welche Erfahrungen die mittleren Führungskräfte damit gemacht haben. Zum anderen wird vertiefend der Frage nachgegangen, warum Performance-Informationen in einigen Fällen häufiger zu Steuerungszwecken verwendet werden als in anderen. Dazu werden zwei alternative Erklärungen diskutiert, von denen eine auf die Bedeutung von organisationalen Routinen und die andere auf Mechanismen der Sozialen Kognition abstellt. Erste vorläufige Befunde zeigen, dass letztere Hypothese besser geeignet ist, um die (Nicht-)Verwendung von Performance-Daten durch Führungskräfte zu erklären als der bloße Fokus auf Routinen.

1. Einleitung

Das Thema „Performance Management“ hat in den letzten Jahren in der Verwaltungsforschung und -praxis zunehmend an Bedeutung gewonnen. Sowohl in Deutschland als auch in anderen OECD-Staaten hat die öffentliche Verwaltung Anstrengungen unternommen, um Veränderungen von Effizienz, Qualität und Wirkungen ihrer Leistungen und Produkte abzubilden und auf dieser Basis zu verbessern (Bogumil u.a. 2007, Pollitt/Bouckaert 2011, Proeller/Siegel 2009). Neu an diesem Ansatz sind zwei Dinge: Erstens soll der Fokus bewusst auf die Ergebnisseite des Verwaltungshandelns gelenkt werden, also weg von der reinen Steuerung über den Ressourceneinsatz. Zweitens sollen über einen langen Zeitraum, basierend auf Zielen und Indikatoren, Performance-Informationen („Kennzahlen“) erhoben werden, die systematische Soll-Ist-Vergleiche ermöglichen (Hatry 2006, Van Dooren u.a. 2010).

Erste Erfahrungen haben gezeigt, dass die Anwendung eines Performance-Management-Systems nicht automatisch zu Verbesserungen der Verwaltungsleistungen führt (Bogumil u.a. 2007, Kroll 2012a, Radin 2006). So können systematisch produzierte Kennzahlen ungenutzt bleiben, wenn sie eine geringe Validität oder Steuerungsrelevanz besitzen. Sie können sogar dysfunktionale

Wirkungen entfalten, wenn sie „Gaming“ und „Cheating“ befördern oder den Fokus von Führungskräften nur auf die Bereiche lenken, in denen Ziele einfach zu erreichen bzw. Erfolge einfach zu messen sind (Hood 2006, Radin 2006).

Kennzahlensysteme, die zu fundamentalen Performance-Verbesserungen führen sowie Informationssysteme, die völlig scheitern und sogar die Verwaltungs-Performance verringern, stellen zwei extreme empirische Phänomene dar. Sie sind möglich – wir gehen aber eher davon aus, dass sich die meisten Fälle irgendwo zwischen diesen Polen befinden (zwei Studien, die dies mittels Large-N-Samples zeigen, sind Kelman/Friedman 2009, Wenger u.a. 2008). Unser Beitrag nimmt deshalb keine „Best-“ bzw. „Worst-Practice“-Perspektive ein, sondern zielt auf die zahlreichen mittleren Fälle ab. Unser Fokus liegt dabei auf der Kommunalverwaltung, speziell den kreisfreien Städten und den Bezirken der Stadtstaaten, da diese in den letzten Jahren die intensivsten Erfahrungen mit dem Kennzahlenwesen gemacht haben. Unser Forschungsinteresse gilt zwei Fragen:

Wie ist das Performance Management in den kreisfreien Städten und Stadtstaatenbezirken ausgestaltet und welche Erfahrungen wurden bislang gemacht?

Wie kann erklärt werden, warum performance-basierte Steuerung in einigen Fällen besser funktioniert als in anderen?

Zur Beantwortung dieser Fragen werden wir zunächst das Performance Management in den Städten empirisch beschreiben. Der theoretische Fokus dieses Beitrags zielt auf die Frage, inwieweit Unterschiede bei der Datenverwendung durch Berichts- und Lernroutinen oder besser durch kognitiv-psychologische Faktoren, wie die Einstellung der Führungskräfte, soziale Norm oder Verhaltensabsicht, erklärt werden können. Unser Beitrag basiert auf den Ergebnissen einer Befragung, an der sich mehrere hundert Amtsleiter¹ verschiedenster Städte beteiligt haben.

Zunächst werden wir einen ausschnitthaften Überblick darüber geben, was wir über das Performance Management auf kommunaler Ebene bereits wissen. Anschließend erläutern wir die Methodik unserer Umfrage. Der erste Ergebnisteil wird einen Überblick über die Kennzahlensteuerung in den kreisfreien Städten geben: Berichtsrouinen, Datennutzung und erste Erfahrungen der Amtsleiter mit dieser Steuerung. Im Anschluss daran legen wir den Fokus auf mögliche Erklärungen einer unterschiedlich intensiv ausgeprägten Steuerung mit Performance-Daten und enden mit einem vorläufigen Fazit.

1 „In diesem Text wird der Einfachheit halber nur die männliche Form verwendet. Die weibliche Form ist selbstverständlich immer mit eingeschlossen.“

2. *Performance Management in der Kommunalverwaltung*

In den vergangenen Jahren wurde eine Reihe empirischer Studien durchgeführt, die alle die Steuerung mit Kennzahlen zumindest in Bereichen tangiert haben (Bogumil u.a. 2007, Burth/Hilgers 2012, DST/PWC 2011, KGSt 2010). Zunächst stellen diese Studien fest, dass Kennzahlen im Alltag Kommunalverwaltung eine Rolle spielen und somit das Performance Management auch in der deutschen Verwaltung angekommen ist. Bevor allerdings noch auf einige Spezifika der Kennzahlenarbeit eingegangen wird, soll es zunächst darum gehen, das Performance Management in den Kommunen etwas allgemeiner zu charakterisieren.

Im internationalen Vergleich, z.B. mit Kommunen der anglo-amerikanischen Staaten, können die deutschen Städte als späte Reformer und ihre Performance-Messsysteme als relativ einfach entwickelt klassifiziert werden. Kuhlmann (2010) charakterisiert das Performance Management der Kommunen als „bottom-up“-orientiert, freiwillig und bürokratisch. Dies bedeutet u.a., dass positive oder negative Sanktionen allenfalls lose an die organisationale Performance gekoppelt sind. So bleiben laut KGSt (2010, S. 109) schlechte Performance-Ergebnisse in 74% der befragten Kommunen, die bereits mit Kennzahlen arbeiten, folgenlos. Lediglich in 26% der Fälle werden Zielabweichungen überhaupt diskutiert. Obwohl die Kommunalverfassungen in vielen Fällen vorschreiben, dass Kennzahlen zu erheben sind, bleibt es letztlich oft den Kommunalverwaltungen selbst überlassen, um welche Daten es sich dabei handelt und an wen diese berichtet werden. Trotz verstärkter Performance-Orientierung in der Kommunalverwaltung kann deren dominantes Steuerungsparadigma insgesamt noch immer als tendenziell „weberianisch“, hierarchisch und durch rechtliche Vorschriften geprägt beschrieben werden (Kuhlmann u.a. 2008). Dennoch scheinen in der Mehrheit der Kommunen Ziele und Kennzahlen als Steuerungsgrößen an Gewicht zu gewinnen. In einer KGSt-Umfrage aus dem Jahr 2010 (S. 90) gab nur eine Minderheit (42%) der befragten Kommunen an, dass sie noch nicht oder erst sehr eingeschränkt mit Zielen und Kennzahlen arbeiten.

Der Idealtypus Performance Management besteht aus zwei Elementen: 1.) Der Dezentralisierung von Entscheidungsbefugnissen mit dem Ziel, mehr Autonomie für die Fachleute zu schaffen, wenn es darum geht, auf welche Weise gesetzte Ziele erreicht werden sollen. 2.) Der Etablierung eines Performance-Monitorings mit dem Ziel, gute „Performer“ zu belohnen und schlechte zu sanktionieren bzw. generell zu überprüfen, inwieweit die Fachleute ihre Ziele erreichen. In der Kommunalverwaltung scheint sich eine Entwicklung zu vollziehen, die beide Seiten berücksichtigt. So hatten im Jahre 2007 ca. 60% (Bogumil u.a. 2007, S. 51) aller Kommunen zumindest teilweise eine Budgetierung und dezentrale

Fach- und Ressourcenverantwortung umgesetzt, im Jahre 2011 waren dies mit Bezug auf die Budgetierung bereits 79% (DST/PWC 2011, S. 27).

Weitere Informationen zur Ausgestaltung des Performance-Controllings lassen sich aus der Studie von Bogumil u.a. (2007) entnehmen: Mehr als 40% (S. 51) der Kommunalverwaltungen besitzen zumindest in Teilbereichen eine Kosten- und Leistungsrechnung sowie ein Berichtswesen. Beides sind zentrale Elemente des Performance Managements. Mehr als 55% (S. 55) der Kommunalverwaltungen nutzen entweder innerhalb der Verwaltung oder im Außenverhältnis zu Politik oder externen Leistungserbringern das Instrument des Kontraktmanagements. Kontrakte zwischen der Verwaltungsspitze und den untergeordneten Einheiten bestehen in 28% der Kommunen (ebd.). Eine vertiefende Untersuchung der Bauämter zeigte, dass für 74% aller Ämter Performance-Daten Bestandteil ihrer Zielvereinbarungen sind und von 83% der Ämter regelmäßig berichtet werden (S. 55).

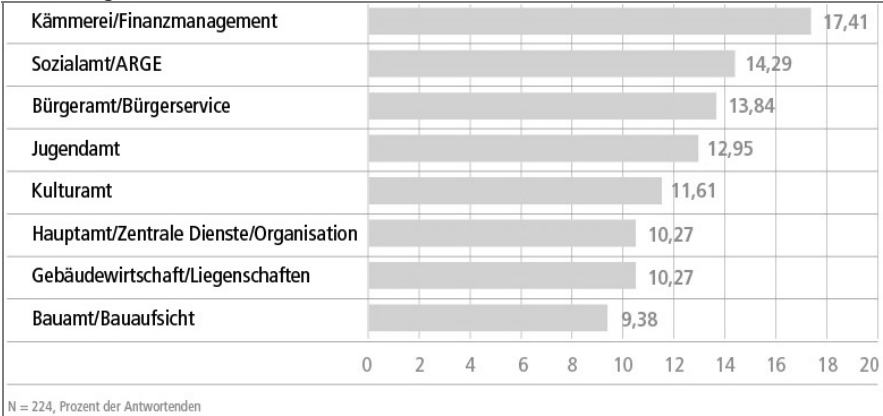
Dieser Abschnitt hat gezeigt, dass wichtige Elemente des Performance Managements bereits Eingang in die Arbeitspraxis der Kommunalverwaltung gefunden haben. Ziele und Kennzahlen spielen auch jenseits von reinen Finanzgrößen eine zunehmend wichtigere Rolle für Bürgermeister und Amtsleiter. Was wir nun in diesem Beitrag genauer untersuchen möchten, sind die Ausgestaltung der Performance-Management-Systeme, die bislang gemachten Erfahrungen der mittleren Führungsebene sowie die Unterschiede bei der Verwendung der Daten in Entscheidungsprozessen.

3. *Methodik der Umfrage*

Die im Folgenden dargestellten Ergebnisse zu den Erfahrungen mit PerformanceManagement-Systemen basieren auf einer von Dezember 2010 bis März 2011 durchgeführten Online-Befragung in deutschen Kommunalverwaltungen. Die Zielgruppe der Befragung waren die mittleren Führungskräfte (Amtsleiter sowie Fachbereichsleiter) aller deutschen kreisfreien Städte (inklusive Bezirke der Stadtstaaten). Die mittleren Führungskräfte wurden in bisherigen Umfragen häufig außen vor gelassen, allerdings nehmen sie als Bindeglied zwischen der Führungsebene und den Beschäftigten eine äußerst wichtige Rolle für die Umsetzung von Verwaltungsreformen ein. Darüber hinaus nehmen Beschäftigte der mittleren Führungsebene Managementfunktionen wahr, durch die sie einen stärkeren Einblick in die (operative) Arbeit mit Kennzahlen haben als dies bei gewählten Exekutivpolitikern der Fall ist. Die kreisfreien Städte wurden als Basis der Befragung gewählt, da diese per Definition eine ähnliche Aufgabenstruktur aufweisen und daher gut vergleichbar sind.

Für die Befragung wurden insgesamt 954 Amtsleiter aus acht gewählten Bereichen in den kreisfreien Städten angeschrieben. Der Rücklauf aus den einzelnen Bereichen ist in der folgenden Abbildung dargestellt.

Abbildung 1: Rücklauf nach Bereichen



Die Rücklaufquote betrug bereinigt 39%.² Unsere Stichprobe weist keinen signifikanten Unterschied von der Grundgesamtheit auf. Somit sind Bundesländer, in denen sich viele kreisfreie Städte befinden, auch in unserer Befragung stärker repräsentiert. Es stammen z.B. zusammen 40 Prozent der Befragten aus Bayern und Baden-Württemberg, während die ostdeutschen Bundesländer (zusammen 12 Prozent), Hessen (4 Prozent) oder Schleswig-Holstein (3 Prozent) eine wesentlich geringere Teilnehmeranzahl aufweisen. Darüber hinaus kennzeichnet sich die Stichprobe durch folgende Merkmale: Die Führungskräfte sind durchschnittlich 51 Jahre alt, männlich (77 Prozent), besitzen einen (Fach-) Hochschulabschluss (90 Prozent), den die Hälfte im Bereich Verwaltung erworben hat. Durchschnittlich arbeiten die Befragten seit 9 Jahren auf ihrer aktuellen Position und sind seit 23 Jahren in derselben Stadtverwaltung tätig. Ein Großteil der Befragten arbeitet in Städten mit mehr als 100.00 Einwohnern.

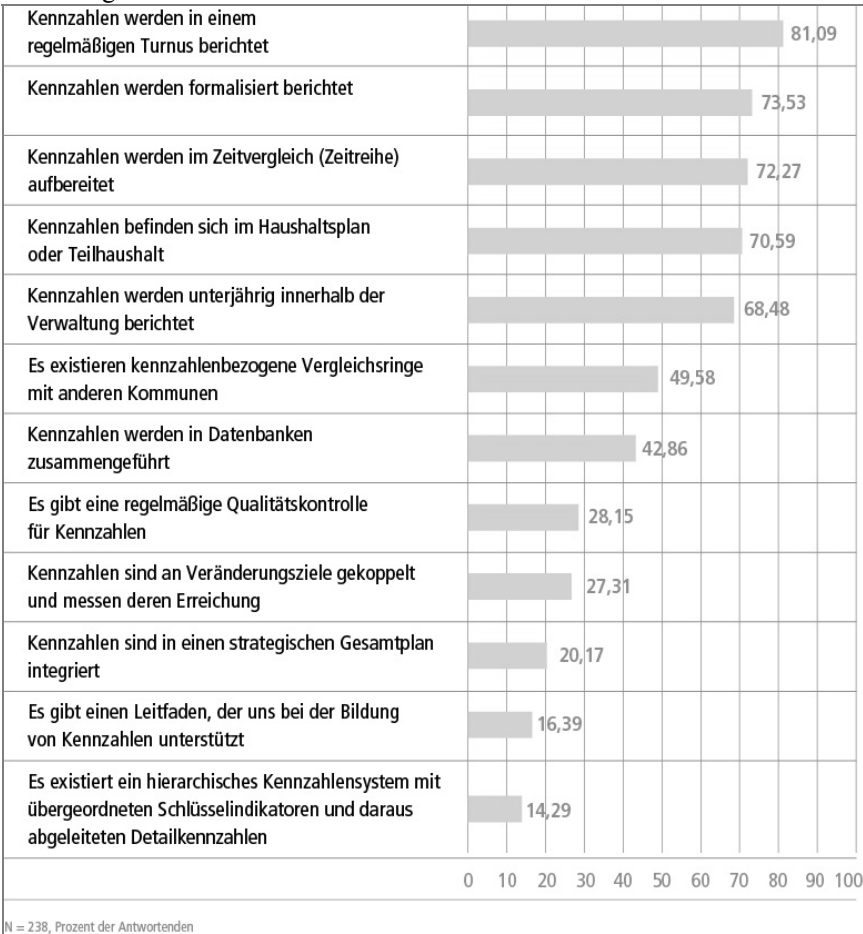
2 Eine Non-Response-Analyse (n=136) hat ergeben, dass 36,6% der angeschriebenen Führungskräfte nicht an der Umfrage teilgenommen haben, weil für ihre Ämter keine Kennzahlen zur Verfügung stehen. Um die Arbeit mit Kennzahlen untersuchen zu können, ist es allerdings notwendig, dass Kennzahlen zumindest erhoben werden. Diese 36,6% (hochgerechnet auf alle Nicht-Teilnehmer) hätten bereits im Vorfeld von der Grundgesamtheit ausgeschlossen werden müssen, methodisch waren sie aber erst im Nachgang identifizierbar. Für die Berechnung der bereinigten Rücklaufquote wurde die Grundgesamtheit um diese Zahl korrigiert.

4. Ergebnisse der Umfrage

Im folgenden Kapitel soll ein Überblick darüber gegeben werden, wie verbreitet die Nutzung von Kennzahlen in den kreisfreien Städten ist und wie ausgereift die Kennzahlensysteme sind.

Zunächst soll ein Überblick über die Mess- und Berichtsroutinen in den kreisfreien Städten gegeben werden. In Bezug auf eine regelmäßige Erhebung von Kennzahlen und deren Aufbereitung zeigt Abbildung 2 ein recht positives Bild.

Abbildung 2: Mess- und Berichtsroutinen



Über 80% der Befragten geben an, dass Kennzahlen in ihrer Verwaltung regelmäßig berichtet werden. Darüber hinaus werden Kennzahlen in über 70% der

befragten Städte im Haushaltsplan abgebildet und im Zeitvergleich dargestellt, wodurch Trendanalysen ermöglicht werden. Zudem werden die Kennzahlen in der Hälfte der kreisfreien Städte auch zum Leistungsvergleich mit anderen Kommunen in Vergleichsringen genutzt.

Nichtsdestoweniger wird dieses positive Bild durch mehrere Punkte getrübt, in denen die Kennzahlenentwicklung in den kreisfreien Städten noch als wenig ausgereift beschrieben wurde. Hierzu gehört insbesondere ein „roter“ Faden bei der Bildung von Kennzahlen und Indikatoren. Ein hierarchisches Kennzahlensystem mit übergeordneten Schlüsselindikatoren und daraus abgeleiteten Detailkennzahlen existiert lediglich in 14% der befragten Städte. Auch die Integration in den Gesamtplan der Verwaltung ist eher schlecht entwickelt, sodass Kennzahlen eher bereichsintern erhoben und ausgewertet werden als eine verwaltungsumfassende Zielrichtung zu verfolgen. Ebenso ist die Kopplung der Kennzahlen an Veränderungsziele mit einer Verbreitung von 27% noch nicht gut ausgereift.

4.1 Nutzung unterschiedlicher Kennzahlentypen

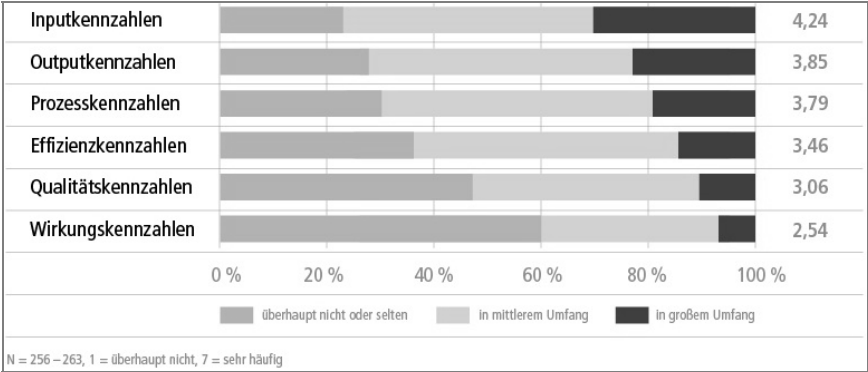
In der Befragung wurde zudem erfasst, welche Kennzahlentypen in den Verwaltungen vorliegen und genutzt werden.

Bei der *Erhebung* von Kennzahlentypen stellte sich heraus, dass Input- und Output-Kennzahlen überdurchschnittlich häufig erhoben werden. Als Inputkennzahlen werden Informationen über den Ressourceneinsatz (z.B. Vollzeitäquivalente oder getätigte Aufwendungen) verstanden, während Outputkennzahlen Leistungs- oder Mengeninformationen (z.B. Anzahl erbrachter Leistungen) abbilden. Aufgrund des traditionellen Berichtens von Finanzdaten ist die Dominanz von Inputkennzahlen im Vergleich zu anderen Kennzahlentypen nicht überraschend. Allerdings ist die Erhebung von Outputkennzahlen vergleichsweise hoch. Dies könnte zum einen daran liegen, dass einige Outputkennzahlen, wie z.B. Fallzahlen, vergleichsweise einfach erhoben werden können. Zum anderen könnte aber auch die lange Tradition des Berichtens von Fallzahlen in einigen Bereichen der kommunalen Verwaltungen (z.B. Jugend oder Soziales) eine längere Tradition haben, ohne dass dies unter dem Titel der Kennzahlensteuerung diskutiert wurde. Die Erhebung von Prozess- und Effizienzkennzahlen folgt noch verhältnismäßig kurz hinter den bereits genannten Kennzahlentypen. Hierunter werden zum Beispiel Bearbeitungszeiten pro Beschäftigtem (Prozesskennzahlen) bzw. Verhältniszahlen wie der Kostendeckungsgrad oder die Produktkosten (Effizienzkennzahlen) verstanden. Am seltensten werden in den kreisfreien Städten Qualitäts- und Wirkungskennzahlen erhoben. Qualitätskennzahlen geben Auskunft darüber, wie die Qualität der öffentlichen Leistungen eingeschätzt wird. Hierzu zählt unter anderem die Kundenzufriedenheit. Wirkungskennzahlen

zeigen wiederum beabsichtigte Verhaltensänderungen bei bestimmten Zielgruppen einzelner Verwaltungseinheiten an. Das kann z.B. die Entwicklung von Neuansiedlungen in einem Stadtgebiet oder die Entwicklung der Jugendkriminalität einschließen. Da die Verwaltung in vielen Fällen nur mittelbar zur Verbesserung der Wirkung beitragen kann und Wirkungen nur schwer zurechenbar sind, ist die geringe Erhebung von Wirkungskennzahlen nachvollziehbar. Nichtsdestotrotz liegt hier dringender Entwicklungsbedarf, da es nur mit ausgereiften Wirkungskennzahlen möglich ist, den realen Effekt von öffentlichen Leistungen zu bestimmen und Verwaltungsmaßnahmen zu optimieren.

Die Kernherausforderung bei der Steuerung über Kennzahlen ist allerdings die tatsächliche *Nutzung* der beschriebenen Kennzahlentypen. Anhand der reinen Erhebung von Kennzahlentypen lassen sich noch keine sicheren Schlüsse in Bezug auf deren Nützlichkeit und Relevanz für die Verwaltungssteuerung ableiten. Daher wird im nächsten Abschnitt genauer darauf eingegangen, welche Kennzahlentypen in den kreisfreien Städten im Verwaltungsalltag verwendet werden. Abbildung 3 stellt die Nutzung der beschriebenen Kennzahlentypen dar. Die Abbildung zeigt einen deutlichen Zusammenhang zwischen der Erhebung und Nutzung von Kennzahlen. Input- und Outputkennzahlen werden von der großen Mehrheit der Befragten in großem oder mittlerem Umfang genutzt. Gleiches gilt, in leicht abgeschwächter Form, auch für die Prozesskennzahlen. Was allerdings bei den Qualitäts- und Wirkungskennzahlen deutlich auffällt, ist der hohe Anteil derjenigen, die diese Kennzahlentypen überhaupt nicht oder nur selten nutzen. Während dieser Anteil bei Inputkennzahlen nur leicht über 20% liegt, umfassen die seltenen Verwender bzw. die Nichtnutzer bei den Qualitäts- und Wirkungskennzahlen einen Anteil von fast 50% (Qualitätskennzahlen) bzw. 60% (Wirkungskennzahlen). Dementsprechend verdeutlicht Abbildung 3 einen klaren Zusammenhang zwischen der Erhebung und Nutzung von Kennzahlen. Je häufiger Kennzahlen vorliegen, desto eher werden sie auch zur Steuerung genutzt.

Abbildung 3: Nutzung von Kennzahlentypen



Insgesamt zeigt die Übersicht zur Nutzung der Kennzahlentypen in großem oder mittlerem Umfang aber ein recht positives Bild. Die überwiegende Mehrheit nutzt Input-, Output-, Prozess- und Effizienzkennzahlen zumindest in mittlerem Umfang. Lediglich bei den Qualitäts- und Wirkungskennzahlen bilden die Nichtnutzer eine gleichgroße bzw. größere Gruppe als die Nutzer. Da allerdings Leistungs- und Finanzdaten in der öffentlichen Verwaltung eine wesentlich längere Tradition haben und so auch tradierte Entscheidungsrouinen entstanden sind, kann auch vermutet werden, dass die Steuerungsrelevanz der beiden wenig genutzten Kennzahlentypen zunimmt, wenn deren Bedeutung in der Verwaltung steigt und sie häufiger erhoben werden.

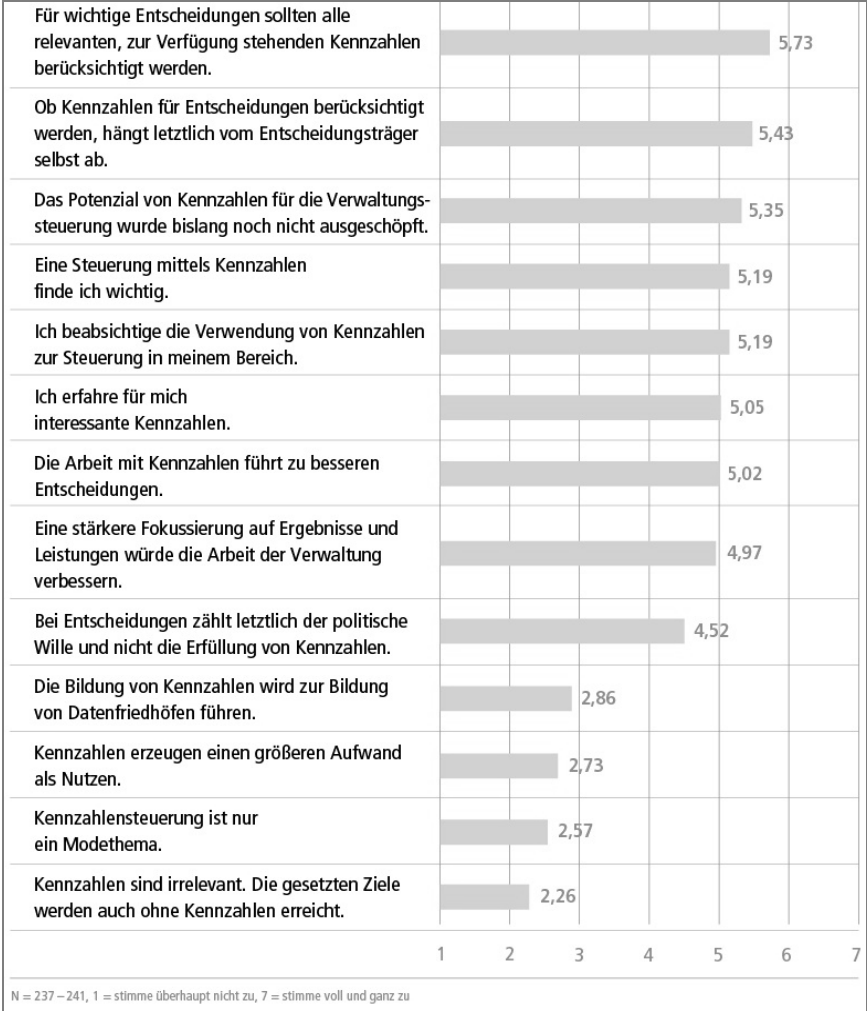
4.2 Einstellung gegenüber Kennzahlen

Ein weiterer wichtiger Faktor zur Steuerungsrelevanz von Kennzahlen ist die Einstellung der Führungskräfte gegenüber der Kennzahlennutzung. Diesbezüglich wurde im Rahmen der Befragung ein Meinungsbild erstellt, welches Auskunft über den Willen der Führungskräfte zur Kennzahlennutzung und ihre Einschätzung zur Relevanz der Kennzahlen gibt.

Dabei werden vor allem zwei wichtige Ergebnisse deutlich: Zum einen sind die meisten Führungskräfte von der Nützlichkeit der Kennzahlensteuerung überzeugt. Sie sind der Meinung, dass Kennzahlen für wichtige Entscheidungen berücksichtigt werden sollten, unter anderem da sie eine stärkere Fokussierung auf Ergebnisse ermöglichen. Zum anderen wird aber auch deutlich, dass das Potential von Kennzahlen noch nicht ausgeschöpft wurde und die Verwendung stark vom Entscheidungsträger abhängt. Dementsprechend zeigt Abbildung 4 ein zweiseitiges Bild: die befragten Führungskräfte machen zwar deutlich, dass sie selbst Kennzahlen verwenden und sich dafür interessieren, aber die Nutzung

ihrer Meinung nach letztlich stark vom persönlichen Interesse und dem politischen Willen abhängt.

Abbildung 4: Einstellung gegenüber Kennzahlen



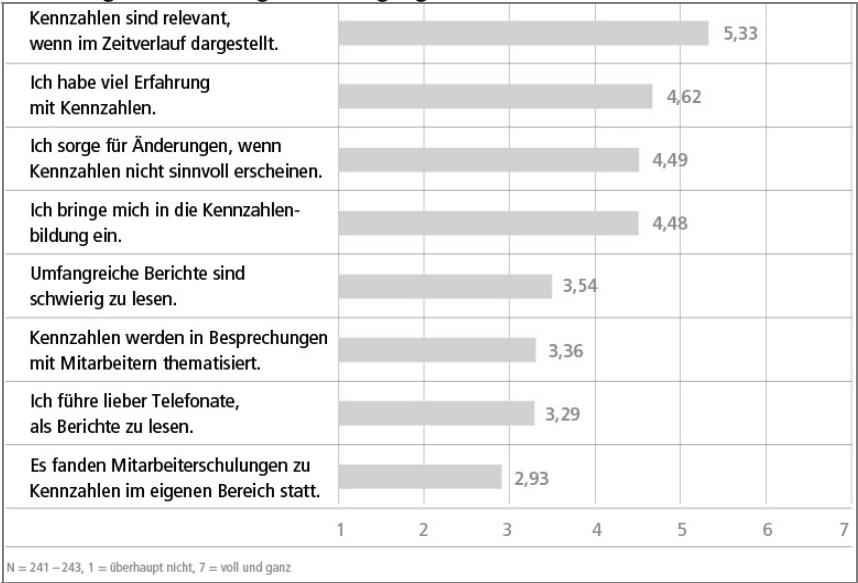
Insgesamt verdeutlicht die Befragung zur Einstellung der Führungskräfte gegenüber Kennzahlen aber eine positive Aussicht auf die Steuerung mit Kennzahlen, denn nur wenige Befragte gehen davon aus, dass es sich bei der Kennzahlensteuerung nur um ein aktuelles Modethema handelt, das schnell wieder an Relevanz verliert. Darüber hinaus ist die Mehrheit der Befragten auch davon überzeugt, dass der Nutzen der Kennzahlen höher ist als der damit verbundene Aufwand.

4.3 Erfahrungen im Umgang mit Kennzahlen

Über die Einschätzung der Kennzahlenrelevanz hinaus wurden die Führungskräfte auch nach ihren eigenen Erfahrungen mit der Kennzahlennutzung gefragt.

Dabei entstand ein recht positives Bild, wenn man die Aussagen über die eigene Nutzung ansieht. So ist mehr als die Hälfte der Befragten der Meinung, viel Erfahrung mit Kennzahlen zu haben. Im Großen und Ganzen bestätigten die Führungskräfte, dass sie sich in die Kennzahlenbildung einbringen und auch für Änderungen sorgen, wenn ihnen Kennzahlen nicht sinnvoll erscheinen.

Abbildung 5: Erfahrungen im Umgang mit Kennzahlen



In Abbildung 5 wird allerdings auch deutlich, dass die Befragten die Mitarbeiter in ihrem eigenen Bereich nur wenig in das Thema Kennzahlen einbeziehen. So stimmten die Befragten eher nicht zu, wenn es darum geht, Kennzahlen in Besprechungen mit Mitarbeitern zu thematisieren. Auch Mitarbeiterschulungen in ihrem Bereich fanden nur selten statt. Somit wird deutlich, dass zukünftig auch verstärkt der Fokus auf die Einbeziehung der Mitarbeiter in die Kennzahlenthematik gefördert werden muss. Solange nur die Führungskräfte selbst Kennzahlen nutzen und an der Weiteroptimierung interessiert sind, bleiben Kennzahlen ein Thema der höheren Verwaltungsebene und dringen nicht in tiefere Ebenen durch. Nur mit der konsequenten Thematisierung und Nutzung der Kennzahlen, auch im Rahmen von regulären Mitarbeiterbesprechungen, kann sichergestellt

werden, dass Kennzahlen zu einer wichtigen Informationsquelle in der kompletten Verwaltung werden.

5. *Die Verwendung von Performance-Daten*

Nachdem wir bereits gesehen haben, dass verschiedene Typen von Kennzahlen eine unterschiedliche Relevanz für die Steuerung besitzen, wird dieser Teil der Frage nachgehen, wie erklärt werden kann, warum einige Amtsleiter ihre Performance-Daten regelmäßiger bei der Entscheidungsfindung heranziehen als andere. Zunächst soll dazu geklärt werden, was wir überhaupt unter der „Verwendung“ der Daten verstehen.

Performance-Daten können für verschiedenste Zwecke verwendet werden. Neben der bereits angesprochenen dysfunktionalen Nutzung („Gaming“, „Cheating“), können sie passiv (um gegenüber Dritten Rechenschaft abzulegen) oder auch politisch (als Evidenz für politische Argumente, z.B. in Budgetverhandlungen) genutzt werden (Moynihan 2009). Die Verwendung, die in diesem Beitrag im Mittelpunkt steht, ist allerdings die zweckmäßige, normativ intendierte Nutzung von Performance-Informationen, z.B. für Soll-Ist-Vergleiche, für die Zuteilung von Ressourcen oder für das „trial-and-error“-Lernen.

Es gibt bereits international ein gutes Dutzend empirischer Studien, die diese Art der normativ erwünschten Verwendung von Performance-Daten untersucht haben (für einen Überblick siehe Kroll 2012b, Kroll/Küchler-Stahn 2009). In diesen Studien wurden verschiedenste Faktoren gefunden, die die Verwendung von Kennzahlen befördern oder behindern könnten. Wir werden in den folgenden Abschnitt den Fokus v.a. auf zwei Einflussfaktoren legen – Routinen und kognitive Mechanismen.

5.1 Theoretische Erklärungen

Wenn es in der öffentlichen Verwaltung darum geht, Akteursverhalten zu verändern, wird dies oft durch den Erlass neuer formaler Regeln und Vorschriften versucht. Die zugrundeliegende Idee dabei ist, organisationale Routinen zu etablieren, die das gewünschte Verhalten befördern oder, sofern dies möglich ist, sogar vorschreiben (Becker 2005, Levitt/March 1988). Diese Logik wurde auch mit Bezug auf Performance-Messsysteme angewendet. Um die Kennzahlennutzung zu befördern, werden Mess- und Berichtsroutinen definiert und formalisiert. Beispiele hierfür sind interne Vorschriften der Stadtverwaltungen, Vorgaben in den Kommunalverfassungen oder, international betrachtet, gesetzliche

Vorgaben wie der Governance Performance and Results Act oder das Program Rating Assessment Tool in den USA.

Auch wenn einige Studien zeigen konnten, dass die Etablierung dieser Routinen nicht zwangsläufig auch zu einer intensivierten „zweckmäßigen“ Nutzung der Kennzahlen führt (Moynihan/Lavertu 2012), gibt es auch zahlreiche Beispiele für den Erfolg von Berichtsroutinen, wenn diese in der Lage waren, eine gewisse Qualität, Relevanz und Regelmäßigkeit der Datenproduktion sicherzustellen (für Verweise siehe Kroll 2012b).

Eine andere Art von Routinen, die die Nutzung von Performance-Daten befördern könnten, sind sogenannte Lernroutinen. Solche Routinen bieten den Mitarbeitern der öffentlichen Verwaltung kontinuierlich die Gelegenheit, bewusst die eigene Arbeit und die Wirkungen dieser zu reflektieren und zu hinterfragen (Agyris/Schön 1996, Dosi u.a. 2009). Es gibt empirische Belege dafür, dass solche Lernforen, in denen Führungskräfte und Mitarbeiter gemeinsam Performance-Daten interpretieren und diskutieren, die Steuerung mit Kennzahlen befördern können (Moynihan 2008; Moynihan/Lavertu 2012).

Ein alternativer Erklärungsansatz kommt aus der Sozialpsychologie. Dieser erklärt menschliches Verhalten nicht durch Routinen, sondern mittels sozialer Kognition. Hierzu werden Entscheidungen als das Ergebnis kognitiver Informationsverarbeitung verstanden, bei der die Wahrnehmung der sozialen Umwelt eine wichtige Rolle spielt (Bandura 1986). Eine zentrale Annahme in diesem Forschungsfeld ist, dass jedem Verhalten eine konkrete Verhaltensabsicht zu Grund liegt, und es steht die Frage im Mittelpunkt, wie verschiedene Verhaltensabsichten erklärt werden können (Fishbein/Ajzen 2010).

Als wichtige Faktoren, die dabei helfen, menschliches Verhalten zu erklären, werden in dieser Literatur die Einstellung zu einem Verhalten sowie die soziale Norm (die Einstellung und Erwartungen wichtiger sozialer Bezugspersonen) genannt. Mit Bezug auf die Verwendung von Performance-Informationen bedeutet dies, dass dieses Verhalten davon abhängig sein könnte, welche generelle Einstellung die Führungskräfte gegenüber einer kennzahlen-basierten Steuerung besitzen und wie die soziale Norm innerhalb der Stadtverwaltungen ausgeprägt ist (Kroll 2012b).

5.2 Erste empirische Befunde

In diesem Abschnitt präsentieren wir bereits einige empirische Ergebnisse und gehen der Frage nach, inwieweit die Verwendung von Performance-Daten die Folge von Routinen bzw. psychologisch-kognitiver Faktoren ist.

Zur Identifikation wichtiger Einflussfaktoren haben wir die statistischen Zusammenhänge zwischen verschiedenen Aussagen der Umfrage untersucht. Dazu

wurden Modelle getestet, die das Ziel verfolgen, die Varianz im Antwortverhalten zur Verwendung von Kennzahlen zu erklären. Mit Hilfe von Regressionsanalysen, welche den Einfluss von Drittvariablen kontrollieren, sind wir zu folgenden vorläufigen Ergebnissen gekommen.

Die Etablierung von Berichtsroutinen (siehe Abb. 2) hat keinen signifikanten Einfluss auf die Nutzung der produzierten Daten. Relevanter ist dagegen die Entwicklung von Lernroutinen, bei denen wir mit 90%iger Wahrscheinlichkeit sagen können, dass sie die Steuerung mit Kennzahlen positiv befördern. Auf der Seite der kognitiven Faktoren sehen die Ergebnisse eindeutiger aus. Die Einstellung der Führungskraft (eine positive Bewertung der Items aus Abb. 4) sowie die Einstellung wichtiger sozialer Bezugsgruppen (Vorgesetzte, Kollegen, Kommunalpolitiker) zeigen starke positive Effekte, der wir uns zu mehr als 95% sicher sein können.

Diese ersten Befunde deuten darauf hin, dass die kognitiven Faktoren eine große Rolle spielen und Führungskräfte mit Performance-Daten vor allem dann steuern, wenn Sie vom Mehrwert dieses Instrumentes überzeugt sind bzw. wenn sich das Gros der Führungskräfte offen gegenüber diesem Instrument zeigt. Es scheint nicht auszureichen, lediglich formale Vorschriften zu machen sowie Mess- und Berichtsroutinen zu etablieren.

6. Fazit

Im Ergebnis dieses Beitrags zeigt sich, dass das Thema „Kennzahlen“ in den kreisfreien Städten angekommen ist. Alle Amtsleiter, die an der Umfrage teilgenommen haben, gaben an, dass in ihrem Bereich Kennzahlen vorliegen. Auch von den Führungskräften, die nicht an unserer Umfrage teilgenommen haben, berichteten nach einer erneuten Nachfrage nur 36 Prozent, dass ihre Nichtteilnahme darin begründet ist, dass ihnen keine Kennzahlen vorliegen.

Wenn von Kennzahlen in den kreisfreien Städten die Rede ist, sind vornehmlich Finanzdaten und einfache Leistungszahlen gemeint. An nächster Stelle folgen Effizienz- und Prozesskennzahlen, erst dahinter werden Qualitäts- und Wirkungsdaten genannt. Diese Reihenfolge ähnelt allerdings dem Entwicklungsstand anderer Länder. Potenzial liegt vor allem bei dem weiteren Ausbau der Qualitäts- und Leistungsdaten, jenseits von einfachen Mengenangaben und Fallzahlen.

Dieser Beitrag zeigte zudem den Bedarf auf, die Messsysteme weiterzuentwickeln. Weniger als die Hälfte aller Befragten arbeitet bereits auf Basis von „integrierten“ bzw. „reifen“ Messsystemen. Weniger als 30 Prozent der Amtsleiter gaben an, dass eine Qualitätskontrolle für Kennzahlen existiert, weniger als 20 Prozent, dass Kennzahlen in strategische Gesamtpläne integriert sind. Demnach droht für einen Großteil der Städte die Gefahr, dass die Nutzung von Kenn-

zahlen als Instrument für die effektivere Steuerung der öffentlichen Verwaltung „auf halbem Wege“ stehen bleibt und sich das erhoffte Potenzial nicht entfalten kann. Gleichzeitig sollte berücksichtigt werden, dass entsprechende technische Verbesserungen nicht automatisch zu einer erhöhten Verwendung der produzierten Daten führen, sondern vor allem auch kognitive Faktoren eine wichtige Rolle spielen.

In Bezug auf die „normativ-intendierte“ Nutzung von Kennzahlen hat sich die Einstellung der potenziellen Datennutzer bzw. ihres direkten Umfelds als zentral herausgestellt. Auch Lernroutinen haben einen positiven Effekt. Kennzahlen können sehr wahrscheinlich eine Steuerungswirkung entfalten, wenn sie bewusst und aktiv von den Amtsleitern für ihren jeweiligen Bereich angepasst und nutzbar gemacht werden. Eine kritische Masse an Promotoren unter den Kollegen („sozialer Druck“) ist dabei hilfreich, die Amtsleiter entsprechend zu motivieren. Soll die Kennzahlennutzung in Zukunft weiter forciert werden, scheint es deshalb unbedingt notwendig, die Führungskräfte vom Ansatz einer kennzahlenorientierten Steuerungsphilosophie zu überzeugen, anstatt lediglich bei den technischen Faktoren anzusetzen. Nur wenn die Amtsleiter die ihnen vorliegenden Kennzahlen für ihre individuellen Steuerungszwecke als zweckmäßig erachten, werden sie diese auch bei der Entscheidungsfindung berücksichtigen und dafür sorgen, dass die Daten aktuell gehalten werden. Vorgaben von der Politik oder Berichtspflichten im Budgetplan scheinen dagegen weniger wichtig, wenn es darum gehen soll, die Nutzung von Kennzahlen auf Seiten der Amtsleiter zu befördern.

Literatur

- Argyris, Chris und Donald Schön. 1996. Organizational learning II: Theory, method, and practice. Reading, Mass: Addison-Wesley Pub.
- Bandura, Albert. 1986. Social foundations of thought and action. A social cognitive theory. Englewood Cliffs, NJ: Prentice-Hall, Inc.
- Becker, Markus. 2005. A framework for apply organizational routines in empirical research: Linking antecedents, characteristics, and performance outcomes of recurrent interaction patterns. *Industrial and Corporate Change* 14(5): 817–46.
- Bogumil, Jörg, Stephan Grohs, Sabine Kuhlmann und Anna Ohm. 2007. Zehn Jahre Neues Steuerungsmodell. Bilanz der Umsetzung. Berlin: Edition Sigma.
- Burth, Andreas und Dennis Hilgers. 2012. Welchen Mehrwert bringt die kommunale Doppik? Ergebnisse einer bundesweiten Kämmerer-Befragung. *Verwaltung & Management* 18(3): 141–146.
- Dosi, Giovanni, Richard Nelson und Sidney Winter (Eds.). 2009. The nature and dynamics of organizational capabilities. Oxford: Oxford Univ. Press.
- DST/PWC (Deutscher Städtetag/ PricewaterhouseCoopers AG). 2011. Evaluierung der Reform des kommunalen Haushalts- und Rechnungswesens. DST/PWC.

- Fishbein, Mark und Icek Ajzen.* 2010. Predicting and changing behavior: The reasoned action approach. New York, NY: Psychology Press.
- Hatry, Harry.* 2006. Performance measurement: Getting results. Washington, D.C: Urban Institute Press.
- Hood, Christopher.* 2006. Gaming in targetworld: The targets approach to managing British public services. *Public Administration Review* 66(4): 515–21.
- Kelman, Steven und John Friedman.* 2009. Performance improvement and performance dysfunction: An empirical examination of distortionary impacts of the emergency room wait-time target in the English National Health Service. *Journal of Public Administration Research and Theory* 19(4): 917–46.
- KGSt (Kommunale Gemeinschaftsstelle für Verwaltungsmanagement). 2010. Stand der Einführung des neuen Haushalts- und Rechnungswesens Ergebnisse einer bundesweiten Umfrage 2010. KGSt-Materialien Nr. 6/2010. Köln: KGSt.
- Kroll, Alexander.* 2012a. Does Performance Information Use Increase Organizational Performance? Examining an Implicit Assumption. Paper prepared for the European Group of Public Administration (EGPA) Conference 2012, Bergen, Norwegen
- Kroll, Alexander.* 2012b. Why Public Managers Use performance Information: Concepts, Theory, and Empirical Analysis. Dissertation, Universität Potsdam.
- Kroll, Alexander und Nicole Küchler-Stahn.* 2009. Performance Management in der öffentlichen Verwaltung – Zwischen Idealismus und Pragmatismus. Ein erweiterter Literaturbericht. *Der moderne Staat*, Heft 2/2009: 483–498.
- Kuhlmann, Sabine.* 2010. Performance Measurement in European local governments: a comparative analysis of reform experiences in Great Britain, France, Sweden and Germany. *International Review of Administrative Sciences* 76(2): 331–45.
- Kuhlmann, Sabine, Jörg Bogumil und Stephan Grohs.* 2008. Evaluating administrative modernization in German local governments: Success or failure of the “New Steering Model”? *Public Administration Review* 68(5): 851–63.
- Levitt, Barbara und James March.* 1988. Organizational Learning. *Annual Review of Sociology* 14: 319–40.
- Moynihan, Donald.* 2009. Through a glass, darkly: Understanding the effects of performance regimes. *Public Performance & Management Review* 32(4): 592–603.
- Moynihan, Donald.* 2008. The dynamics of performance management: Constructing information and reform. Washington, D.C: Georgetown Univ. Press.
- Moynihan, Donald und Stéphane Lavertu.* 2012. Does involvement in performance management routines encourage performance information use? Evaluating GPRA and PART. *Public Administration Review* 72(4): 592–602.
- Pollitt, Christopher und Geert Bouckaert.* 2011. Public management reform: A comparative analysis. Oxford.
- Proeller, Isabella und John Siegel.* 2009. Performance Management in der deutschen Verwaltung: Eine explorative Einschätzung. *Der moderne Staat* 2(2): 463–82.
- Radin, Beryl.* 2006. Challenging the performance movement: Accountability, complexity, and democratic values. Washington, DC: Georgetown Univ.
- Van Dooren, Wouter, Geert Bouckaert und John Halligan.* 2010. Performance management in the public sector. London: Routledge.
- Wenger, Jeffrey, Laurence O'Toole und Kenneth Meier.* 2008. Trading speed for accuracy? Managing goal conflict and accommodation in the U.S. unemployment insurance program. *Policy Studies Journal* 36(2): 175–98.

Indikatoren für eine wirkungsorientierte, erfolgreiche Steuerung von Städten und Landkreisen – ein Praxisbericht

Rainer Christian Beutel

Eine wirkungsorientierte, erfolgreiche Steuerung von Städten und Landkreisen besteht aus einer Reihe von Steuerungselementen, die von der KGSt bei der Praxisbegleitung ihrer 1.780 kommunalen Mitglieder mit 75 Millionen Einwohnern in die Beratung eingebracht werden. Selten werden alle diese Elemente gleichzeitig eingesetzt.

Jedes dieser Elemente leistet aber für sich gesehen schon einen wichtigen Beitrag zu einer erfolgreichen Steuerung.

Es geht um folgende Bausteine.

1. Strategie
2. Wirkungsorientierung
3. Kennzahlen in der Zeitreihe und im interkommunalen Vergleich unter besonderer Nutzung der Möglichkeiten des neuen Haushalts- und Rechnungswesens
4. Monitoringsysteme

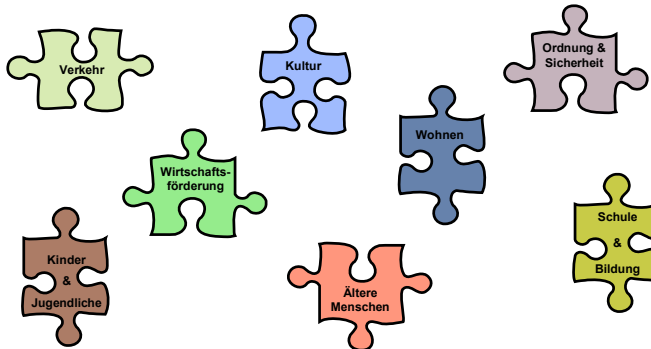
1. Strategie

Die Steuerung über strategische Ziele ist wesentliche Voraussetzung für den Steuerungserfolg. Sie verbessert die Qualität der kommunalen Arbeit. Sie fördert eine gute Stadt- und Kreisentwicklung.

Dadurch, dass Prioritäten und Posterioritäten gesetzt werden, wird ein schoner Einsatz der Ressourcen Personal und Geld gewährleistet. Bei knapper werdenden Fach- und Führungskräften und sich verengenden Finanzspielräumen ist das von großer Bedeutung.

Die strategische Steuerung ermöglicht die im interkommunalen Wettbewerb immer wichtiger werdende Profilbildung. Sie ermöglicht in Handlungsfeldern wie Bildung, Arbeit, Demografie, Integration, in denen die Einbeziehung vieler Akteure auch außerhalb des Rathauses oder Landratsamtes erforderlich ist, dass dies gelingt.

Kommunales Handeln vollzieht sich in vielen Handlungsfeldern ...



... für die es aber keine Insellösungen geben darf!

2. Wirkungsorientierung

Wirkungsorientierte Steuerung ist erforderlich, um mehr für die Bürgerinnen und Bürger der eigenen Stadt oder des eigenen Landkreises zu erreichen. Was für ein Unternehmen der Profit ist, das ist für eine Stadt oder einen Landkreis die Wirkung. Da die Lebensqualität in einer Kommune *das* Kriterium für Erfolg ist, müssen ihre Produkte und Leistungen nach ihrem Wirkungsbeitrag für die Lebensqualität vor Ort beurteilt werden.

Könnte beispielsweise durch ein darauf ausgerichtetes Bündel von Maßnahmen der Notendurchschnitt der Berufsschulabsolventen verbessert werden? Konnten dadurch mehr junge Menschen nach Abschluss ihrer Lehre in feste, nicht prekäre Arbeitsverhältnisse vermittelt werden? Wird das ermittelt, evaluiert und verbessert? Sind wir dadurch als Kommune mehr für die Bürger da, nützen wir ihren Interessen besser, arbeiten wir *nicht* nach dem Grundsatz: Wir sind auch ohne Bürger schwer beschäftigt?

Es geht also ebenso schlicht wie bedeutsam um die Frage: Warum tun wir was im Rathaus oder Landratsamt?

Eine einfache Antwort werden wir noch auf die Frage finden, warum wir den fließenden Verkehr überwachen. Wir wollen die Zahl der Unfälle mit Toten oder Schwerverletzten so weit wie möglich reduzieren.

Ist die Stadt oder der Landkreis aber in der Wirtschaftsförderung oder im Ausländerwesen ebenso auf klare und einheitliche Ziele ausgerichtet?

Ergibt sich aus dem Wirken aller Fachbereiche eine konsistente Botschaft, die Redundanzen und Widersprüche vermeidet?

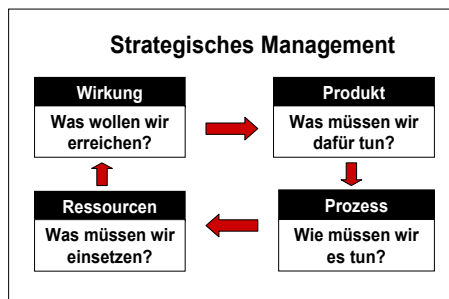
Wissen alle Beteiligten, welche Wirkungen sie mit ihrem Handeln erzielen sollen? Würde jeder Fachbereich dieselben Antworten geben?

Nehmen wir als Beispiel die kommunale Jugendpolitik. Wir haben es bei der hier relevanten Zielgruppe nicht mit gespaltenen Persönlichkeiten zu tun, die 'mal Berufsschüler, 'mal Kunden der Suchtpräventionsstelle, Sportler, Verbraucher, Jugendhausbesucher, Betreute in der Familien- und Erziehungsberatung, Musikschüler oder Führerscheinbewerber sind. Es geht darum, Wirkungsziele für die Jugendpolitik zu entwickeln, die evaluierbar sind und auch evaluiert werden. Beispielsweise, die Zahl der Drogenabhängigen um 30 % zu senken, die Jugendarbeitslosigkeit um 20 % zu vermindern und die Zahl der erfolgreich in Lehrstellen Vermittelten um die Zahl X pro Jahr zu erhöhen.

Wirkungsorientierung



Laufendes Controlling
des Erreichens, eventuell
Anpassungen vornehmen



Instrumentelle Unterstützung durch die KGSt:

Inhouseseminare zur strategischen Steuerung von Städten und Landkreisen

Parallel dazu sind Evaluation, Korrekturen und erneute Messungen unabdingbar. Ebenso interkommunale Leistungsvergleiche, wie sie das KGSt®-IKO-Netz erfolgreich anbietet.

Wer gut steuern will, der muss messen und vergleichen. Sonst weiß er nicht, wo er steht und wo er hin sollte.

3. Kennzahlen

Kennzahlen sind unverzichtbarer Bestandteil erfolgreicher und wirkungsorientierter Steuerung. Nur, wer beispielsweise die Zahl der Arbeitsplätze im Verhältnis zur Einwohnerzahl vor Ort und im interkommunalen Vergleich kennt, wer seinen Ein- und Auspendlerüberschuss ermittelt und vergleicht und dies in den Kontext zur Arbeitslosenquote vor Ort und im regionalen Vergleich stellt, kann seinen Erfolg als Wirtschaftsstandort bewerten und seine Steuerung darauf ausrichten. Nur, wer bei den erzieherischen Hilfen das Verhältnis von Pflegefamilien zur Heimunterbringung ermittelt und unter Berücksichtigung von Sozialstruktur und Jugendhilfedichte interkommunal vergleicht kann im Sinne von Förderung der Jugendlichen und Kostenbewusstsein sein kommunales Handeln fundiert und erfolgreich steuern.

Einen zusätzlichen Schub erfährt die kennzahlengestützte Steuerung durch das neue Haushalts- und Rechnungswesen. Hierdurch werden Ziele und Kennzahlen vermehrt Bestandteil kommunaler Steuerung. Darin stecken erhebliche Chancen für eine verbesserte kommunalpolitische Arbeit und auch für die Haushaltskonsolidierung. Auch kann dadurch die Wirkungsorientierung stärker in den Fokus genommen werden. Das KGSt[®]-IKO-Netz unterstützt dies durch vielfältige interkommunale Leistungsvergleiche.

Ganz im Sinne des früheren Oberkreisdirektors des Kreises Borken, Raimund Pingel, der zutreffend sagte: „Die Bürger interessiert nicht, welches Waschmittel wir verwenden, sondern ob die Wäsche weiß wird.“

4. Monitoringsysteme

Monitoringsysteme, wie sie von der KGSt für die Bereiche Demografie, Integration, Soziales entwickelt wurden, sind bestens geeignete Unterstützungsmethoden für eine wirkungsorientierte, erfolgreiche kommunale Steuerung.

Als Beispiel für ein erfolgreiches Steuern mit Monitoring und dazu gehörenden Indikatoren kann hier das KGSt[®]-Modell „Sozialmonitoring“ dienen.

Beispiele für Indikatoren – das KGSt Sozialmonitoring

KGSt



Seite 10

4.1 Praxisbeispiele strategische Steuerung

Gute Praxisbeispiele für strategische Steuerung finden Sie, ohne dass diese Aufzählung vollständig ist, in der Stadt Ludwigsburg, der Landeshauptstadt Kiel, der Stadt Coesfeld, der Stadt Ahrensburg und der Stadt Emden, der Landeshauptstadt München, der Stadt Osnabrück und der Stadt Heidelberg oder im Landkreis Potsdam-Mittelmark.

Exemplarisch sollen zwei Beispiele etwas näher erläutert werden.

1. Die Stadt Heidelberg hat im Jahre 2007 eine strategische Haushaltsplanung mit einem Produkthaushalt, der Ziele und Kennzahlen zur Grundlage hat, eingeführt.

Der Produkthaushalt auf doppischer Grundlage dient als globale Bewirtschaftungs- und Finanzierungsgrundlage, weist leistungs- und zielorientierte Budgets aus, stellt den Ressourcenverbrauch dar und ermöglicht eine leistungsorientierte Steuerung mit Zielen und Kennzahlen. Damit hat die Stadt Heidelberg finanzielle Spielräume geschaffen und eine klare Auftragsgrundlage zwischen Rat und Verwaltung zur Schaffung einer besseren Steuerung hergestellt.

2. Beispiel:

Praxisbeispiel Strategische Steuerung im Landkreis Potsdam-Mittelmark



Wirkungsorientiert Steuern bedeutet einen Kulturwandel in Politik und Verwaltung...



...wir stehen erst am Anfang!

Verwaltungswettbewerbe und -preise als Trojanisches Pferd der Leistungsmessung?

Benedikt Speer/Marika Gruber

1. Einleitung

Verwaltungswettbewerbe und -preise sind in der Regel mit ausschließlich positiven Assoziationen besetzt: Die prämierten Verwaltungen beteiligen sich, weil sie glauben, dass sie eine vorzeigbare Leistung aufzuweisen haben; verliehene Auszeichnungen dienen nach innen der Motivation und nach außen als Aushängeschild. Auf den ersten Blick scheint daher die Verknüpfung solcher Wettbewerbe mit dem Bild des Trojanischen Pferds nicht wirklich geeignet zu sein, da die entsprechende List der Griechen gegenüber der Stadt Troja für letztere äußerst unangenehme Folgen zeitigte, indem sie ihren Untergang einleitete. Tatsächlich war mit dem Instrument der Verwaltungswettbewerbe jedoch ursprünglich durchaus der (Hinter-)Gedanke der kompetitiven Leistungsfeststellung und -bewertung verbunden, auch wenn dieser Aspekt inzwischen weitgehend vernachlässigt wird. Ziel des Beitrags ist es daher, Spannungsfelder und Ambivalenzen eines auf den ersten Blick harmlosen Mittels zur Identifizierung und – in einem zweiten Schritt – zur Diffusion und Steigerung von Leistungsqualität in öffentlichen Verwaltungen zu untersuchen, wobei der Schwerpunkt auf Erfahrungen aus Österreich liegt.

2. Ausgangslage in Österreich

Die österreichische Verwaltung war bis in die 1990er Jahre ganz überwiegend durch das sogenannte Kelsen/Merklsche Verwaltungsmodell geprägt, das in der Tradition der Wiener Rechtstheoretischen Schule auf demokratische Normerzeugung durch gewählte VolksvertreterInnen und autokratischen Normvollzug durch die Verwaltung abstellte. Die Prämisse einwandfreier Rechtsanwendung stand dabei absolut im Vordergrund, wohingegen der soziale Rahmen und die Umweltbedingungen, unter deren Einfluss die Verwaltungsaufgaben zu erfüllen sind, wenig Beachtung fanden. Dieser Zugang entsprach der allgemeinen Herangehensweise der Wiener Rechtstheoretiker, die „...wie überhaupt den Staat, so auch die gesamte staatliche Verwaltung als Rechtsfunktion...erkennen und die

Lehre von diesem...Gegenstand zur Gänze als Rechtslehre...begründen [wollten]“ (*Merkel* 1927, Vorwort).

Die Frage der praktischen Ausgestaltung der Steuerung von Verwaltung oder gar der Messung von Verwaltungsleistungen stellte sich in diesem theoretischen Rahmen nicht, obgleich sich im Kontext des österreichischen Parteien- und Verbändestaates durchaus eigene korporatistische Funktionsweisen entwickelten, die eine spezifische, mit den idealtypischen Vorstellungen Kelsen und Merkl's nur bedingt zu vereinbarende Verwaltungskultur und -praxis begründeten. Erst mit dem – im europäischen Vergleich verspätet einsetzenden – Beginn der New Public Management-Debatte in Österreich wurden auch hier neue Ansätze diskutiert: neben allgemeinen Einsparungsbestrebungen (Reduktion) und einer verstärkten Einbeziehung der BürgerInnen (Partizipation) ging es nicht zuletzt um erhöhte Arbeitsproduktivität und verbesserte Leistungserstellung (Qualität) (vgl. *Neisser* 1998, 321). Als ein wesentliches Ziel der diskutierten Reformen wurde dabei die Schaffung von Wettbewerbsmechanismen in der öffentlichen Verwaltung gesehen, „wobei neben echtem ‚Markt-Wettbewerb‘ in Form konkurrierender Anbieter auch Wettbewerbssurrogate wie Leistungsvergleiche mittels Kennzahlen, Benchmarking, interne Leistungsverrechnung oder *Qualitätswettbewerbe* im Vordergrund stehen können“ (*Neisser/Meyer/Hammerschmid* 1998, 45; eigene Hervorhebung). Damit hatte die Idee von Verwaltungswettbewerben zumindest in der wissenschaftlichen Betrachtung ihren Niederschlag in Österreich gefunden.

3. (*Qualitäts-*)Wettbewerbe in der öffentlichen Verwaltung

Die Hoffnungen, die sich mit der Einführung von Wettbewerbs-elementen in öffentlichen Verwaltungen verbinden, sind vielfältig: so soll Wettbewerb grundsätzlich dazu beitragen, die Leistungen von Verwaltungen zu steigern und deren Innovationsfähigkeit zu fördern, Ressourcen bestmöglich effizient und effektiv einzusetzen sowie auf Wünsche und Bedürfnisse von Anspruchsgruppen flexibel zu reagieren. In diesem Zusammenhang lassen sich generell zwei Formen von Wettbewerb unterscheiden (vgl. *Nullmeier* 2005, 111 f.):

- *Faktischer Wettbewerb* – Dabei begeben sich Organisationen in einen aktiven Wettbewerb und stehen einander als Marktkonkurrenten gegenüber. Für davon betroffene Verwaltungseinheiten können die Folgen mangelnder Wettbewerbsfähigkeit bis hin zur Auflösung und zum „Marktausscheiden“ reichen, wenn beispielsweise öffentliche Aufgaben an privatwirtschaftliche Konkurrenten übertragen werden.
- *Virtueller Wettbewerb* – Hierbei werden Leistungen verschiedener Verwaltungen nach einheitlichen Maßstäben bewertet und miteinander verglichen.

Diese Form von Wettbewerb bleibt sozusagen virtuell, da sie zu keinen direkt fühlbaren Sanktionen wie etwa der Verdrängung aus dem „Markt“ führt. Im Vordergrund stehen das Lernen und die Leistungsverbesserung aufgrund einer „wechselseitigen systematisierten Beobachtung in einem Vergleichsverfahren...“, daher kann auch von Wettbewerb in der Beobachterperspektive gesprochen werden“ (Nullmeier 2005, 111).

Demzufolge gehören Verwaltungswettbewerbe in die zweite der genannten Kategorien, was auch zum Teil ihre Attraktivität erklären mag, werden sie doch kaum je als harte Instrumente des Wettbewerbs etwa mit der Privatwirtschaft angesehen. Sie sind daher nicht nur in Österreich ein seit den 1990er Jahren beliebtes Innovationsinstrument.

3.1 Hintergrund und Zweck von Qualitätswettbewerben

Eine vertiefende wissenschaftliche Diskussion zu Qualitäts- und Verwaltungswettbewerben existiert allerdings bis heute nur in Ansätzen. So kann eine bereits Ende der 1990er Jahre vorgeschlagene Definition weiterhin Gültigkeit beanspruchen:

„A public quality competition award is defined as a performance measurement instrument which fosters innovation and quality in the public sector by the identification of excellent public organizations by independent panels and with active participation of public agencies“ (Kouzmin u.a. 1999, 129).

Qualitäts- und Verwaltungswettbewerbe sind hier folglich explizit als *Leistungsmessungsinstrument* definiert, über das herausragende öffentliche Verwaltungen durch unabhängige Jurys und unter aktiver Einbeziehung der betreffenden Organisationen identifiziert werden sollen. In der Literatur werden ihnen in diesem Zusammenhang insbesondere zwei Anwendungsbereiche zugeschrieben: zum einen eignen sie sich in Leistungsfeldern, in denen sich ein „natürlicher“ Wettbewerb generell nur schwer organisieren lässt, um Leistungsvergleiche innerhalb und zwischen Verwaltungen zu forcieren (vgl. Nullmeier 2011, 155). Zum anderen können sie überall dort als Hilfsmittel zur Eruierung von erfolgreichen Lösungen und als mögliche Bezugspunkte für andere Verwaltungen dienen, wo einstweilen probate Finanz-, Leistungs- und Qualitätskennzahlen sowie daraus abgeleitete Vergleichsindikatoren noch fehlen (vgl. Kouzmin u.a. 1999, 125 f.).

Der Aspekt der Leistungsmessung wird allerdings regelmäßig bei solchen Wettbewerben nicht ausdrücklich in den Vordergrund gerückt, sondern durch positive Anreize wie das Anbieten einer publikumswirksamen Plattform für Verwaltungen, durch aufwändige Preisverleihungszeremonien und durch öffentliche Anerkennung quasi verschleiert. Den Organisatoren wird daher auch ange-

raten, anstelle des eigentlich angestrebten – kompetitiven – Leistungsvergleichs vor allem kooperative Elemente wie z.B. gegenseitige Lerneffekte hervorzuheben:

„Organizers of quality awards have to stress the cooperative element of awards; namely that every participant ‚wins‘ by learning from others and that winning the award is not terribly important“ (Kouzmin u.a. 1999, 130).

Diesem Muster folgen üblicherweise auch die Festansprachen anlässlich von Preisverleihungen. So wünschte etwa der Landeshauptmann von Oberösterreich bei der Verleihung des sogenannten „Publicus“, dass es diesem Preis gelingen möge „...nicht nur einer breiten Öffentlichkeit die Leistungs- und Weiterentwicklungsfähigkeit des öffentlichen Dienstes deutlich zu machen, er ermöglicht...auch allen Verwaltungseinheiten, von ‚Best Practice‘-Beispielen zu lernen“ (Pühringer 2010, online).

Dadurch werden Verwaltungswettbewerbe allerdings einem nur schwer aufzulösenden Spannungsverhältnis ausgesetzt, indem sie gleichzeitig sowohl kompetitive (Leistungsvergleich, Auszeichnung, Preisverleihung) als auch kooperative (Lernen voneinander, stetige Verbesserung) Elemente in Einklang bringen sollen. Ein Trojanisches Pferd der Leistungsmessung hätte daraus aber nur entstehen können, wenn bei einer vordergründigen Hervorhebung kooperativer Elemente konsequent der eigentlich angestrebte (Hinter-)Gedanke der Leistungsmessung verfolgt worden wäre. In der Praxis trat jedoch die ursprünglich intendierte Leistungsmessung immer weiter zurück und wurde durch die verbreitete Vorstellung verdrängt, dass eine Teilnahme an Verwaltungswettbewerben grundsätzlich freiwillig erfolgen müsse, wobei nicht die Leistungsmessung, sondern in erster Linie die Qualitätsverbesserung im Mittelpunkt stehen solle:

„Dem Ziel der *Qualitätsverbesserung* durch den Vergleich mit anderen Verwaltungen dient auch die *freiwillige* Teilnahme öffentlicher Verwaltungen an Qualitätswettbewerben“ (Schauer 2010, 179, eigene Hervorhebungen).

3.2 Das Faible österreichischer Verwaltungen für Qualitätswettbewerbe

Auch wenn es keine verlässlichen Daten dazu gibt, drängt sich der Eindruck auf, dass österreichische Verwaltungen ausgesprochen gerne an solchermaßen abgeschwächten Verwaltungs- und Qualitätswettbewerben teilnehmen. In einem Land, in dem der Zwang zur Leistungsmessung bislang nur unzureichend verankert ist, kommen sie den VertreterInnen einer eher konsensual geprägten (Verwaltungs-)Kultur entgegen, indem auf einer freiwilligen Basis eigene Leistungserfolge öffentlichkeitswirksam gegenüber Politik, Wirtschaft und BürgerInnen, aber auch gegenüber anderen Verwaltungen herausgestellt werden können. In Zeiten einer auch in Österreich zunehmenden Ressourcenverknappung kann ein gewonnener Verwaltungspreis nach außen hin als Rechtfertigung dienen, um

sich die erforderlichen Mittel für die Aufrechterhaltung qualitätsvoller Dienstleistungen zu sichern. Nach innen kann eine entsprechende Auszeichnung eine hohe Motivationswirkung bei den MitarbeiterInnen entfalten, insbesondere wenn dadurch die Ergebnisse von vorherigen Reformanstrengungen gewürdigt werden. Nicht zuletzt haben österreichische Verwaltungen gerade auch in internationalen Wettbewerben immer wieder Preise und Anerkennungen erzielt, die im Inland eine zusätzliche Legitimationswirkung entfalten.

Die nachfolgende Darstellung internationaler und nationaler Verwaltungs- bzw. Qualitätswettbewerbe kann nicht abschließend sein, soll aber jedenfalls einen Eindruck von der Vielzahl diesbezüglicher Veranstaltungsformate und -inhalte vermitteln:

- *Speyerer Qualitätswettbewerb* – Der Speyerer Qualitätswettbewerb wurde von der Deutschen Hochschule (jetzt: Universität) für Verwaltungswissenschaften Speyer initiiert und fand zwischen 1992 und 2005 insgesamt sieben Mal statt. Als Vorreiter wurde er zum Vorbild für zahlreiche andere Qualitätswettbewerbe im deutschsprachigen Raum. Unter den Preisträgern und bei den Anerkennungen fanden sich regelmäßig auch mehrere österreichische Verwaltungen¹ (vgl. <http://www.dhv-speyer.de/qualitaetswettbewerb>).
- *European Public Sector Award (EPSA)* – In der Nachfolge des Speyerer Qualitätswettbewerbs wird seit dem Jahr 2007 der European Public Sector Award (EPSA) durch das European Institute of Public Administration (EIPA) in Maastricht verliehen. Eines seiner Ziele ist die Förderung von „constructive competition amongst public sector organisations“, d.h. die Förderung des Leistungs- und Wettbewerbsgedankens. Dies soll jedoch in erster Linie durch die Errichtung eines Exzellenznetzwerks erreicht werden, um „conditions for a European learning platform on public administration solutions“ zu schaffen, also durch die Stärkung des kooperativen Elements (EIPA 2009, online). Auch der EPSA-Wettbewerb brachte einige österreichische

1 Ausgewählte Beispiele für österreichische Preisträger sind: 1996: Marktgemeinde Langenrohr (Europas erste zertifizierte Gemeinde); Stadtgemeinde Linz (Organisationsentwicklung). 1998: Marktgemeinde Seekirchen (Innovationen – der Weg zum Erfolg); Stadtgemeinde Spittal a.d. Drau (Die lernende Organisation – Vernetzung als Strategie); Bezirkshauptmannschaft Zell am See (Die BH als Dienstleistungsunternehmen); Stadtgemeinde Wien (Studienbeihilfenbehörde – Verwaltungsreform). 2000: Stadtgemeinde Schwechat/Stadtgemeinde Salzburg (Fremdenbehörde – Bürgerorientierung); Stadtgemeinde Salzburg (E-Government); Gemeinde Pasching (Personalmanagement). 2002: Bundeskellerinspektion Wien (Personalmanagement); Versicherungsanstalt des österreichischen Bergbaues (Personalmanagement); Montanuniversität Leoben – Institut für Wirtschafts- und Betriebswissenschaften (Wissensmanagement).

Preisträger² hervor (über die aktuelle Homepage sind auch die Wettbewerbe 2007, 2009 und 2011 zugänglich: vgl. <http://www.epsa2013.eu>).

- *European eGovernment Awards* – Durch Österreich mitangeregt, wurden die European eGovernment Awards als Auszeichnung der Europäischen Kommission für innovative Projekte der öffentlichen Verwaltung im Bereich der Informations- und Kommunikationstechnologie organisiert. Ziel war es, einen Know-how-Transfer zwischen europäischen Verwaltungen im Bereich von E-Government anzustoßen, zur Verbesserung der Lebensqualität der BürgerInnen beizutragen, das Vertrauen in die Verwaltung zu stärken und den Wettbewerb zwischen europäischen öffentlichen Verwaltungen zu stimulieren (vgl. *European Union* 2009, online). Bis zur Einstellung des Awards 2009 gab es zahlreiche österreichische Preisträger³, welche die europäische Vorreiterrolle des Landes auf dem Gebiet des E-Government bestätigten und international stark rezipiert wurden (vgl. <http://www.donau-uni.ac.at/de/departement/gpa/forschung/projekte/id/11119/index.php>).
- *Österreichischer Verwaltungspreis* – Auf der nationalen Ebene wird der Österreichische Verwaltungspreis vom Bundeskanzleramt verliehen. Mit ihm sollen „zukunftsweisende Projekte, innovative Entwicklungen und erfolgreiche Modernisierungsprozesse“ österreichischer Verwaltungen auf Bundes-, Länder- und kommunaler Ebene identifiziert und ausgezeichnet werden. Der Österreichische Verwaltungspreis stellt eindeutig auf das „Voneinander Lernen“ durch den Austausch von „Best-practices“ und die Übertragung gelungener Entwicklungen auf andere Organisationen ab. Durch diesen kooperativen Ansatz sollen Impulse zur zukunftsorientierten Weiterentwicklung des öffentlichen Sektors gegeben werden (vgl. *Bundeskanzleramt Österreich* 2013, online).
- Auch auf Bundesländerebene gibt es in Österreich Verwaltungswettbewerbe, wie zum Beispiel den (mittlerweile eingestellten) oberösterreichischen Ver-

2 Beispiele für EPSA-Preisträger und Best-Practice-Zertifikate sind: 2011: ÖkoKauf Wien – EcoBuy Vienna (Kategorie "Going Green: Concrete Solutions from the Public Sector"); Stadtgemeinde Wien (Gewerbe-Online); Land Oberösterreich (Optimale Versorgung von Patienten mit Langzeitbeatmung unter qualitativen und wirtschaftlichen Aspekten); Stadtgemeinde Wien (It's all about kids – PPP for a school campus); Stadtgemeinde Wien (Sei dabei); Rat für Forschung und Technologieentwicklung (Policy Counseling 2.0); Stadtgemeinde Wien (Umweltmanagementprogramm der Wiener Stadtverwaltung); Österreichische Energieagentur (Klima:Aktiv – Going Green: Gebäude, Mobilität, Energieeffizienz, erneuerbare Energien).

3 Beispiele für eGovernment Awards-Preisträger und Good Practice-Cases sind: HELP.gv.at; The European Order for Payment Application – EU-OPA (Kategorie „Government fördert den EU-Binnenmarkt; zusammen mit Deutschland“); Amt der Steiermärkischen Landesregierung (Civil Protection Online); Bundeskanzleramt Österreich (ELAKimBund); Agrar Markt Austria – AMA (eAMA).

waltungspreis „Publicus“, den „Verwaltungspreis des Landes Steiermark“ und den „Speedy Offiziales“ des Landes Kärnten, auf den im Folgenden noch ausführlicher eingegangen wird.

3.3 Zwischenfazit

Werden die Aussagen der eher begrenzten Literatur zu Qualitäts- und Verwaltungswettbewerben als Instrument der Leistungsmessung mit den – zugegebenermaßen unsystematischen und punktuellen – Beobachtungen zu deren Realität abgeglichen, so lässt sich feststellen, dass bislang vor allem die Identifizierung und Prämierung von Good Practices freiwillig teilnehmender Verwaltungen im Vordergrund gestanden hat. Auffallend ist zudem, dass es kaum transparente Informationen über Auswahl- und Bewertungskriterien oder über die tatsächlichen Entscheidungsgründe der jeweiligen Preisjurys gibt. Der Aspekt der konsequenten Leistungsmessung durch „virtuellen Wettbewerb“ ist damit mehr als fraglich geworden, das kompetitive Element ist weitgehend zurückgedrängt. Doch auch bezüglich des angestrebten kooperativen Voneinander-Lernens im Rahmen oder als Folge von Qualitäts- und Verwaltungswettbewerben scheinen Zweifel angebracht zu sein: insbesondere ist bedenklich, dass nur in wenigen Fällen nach der Preisverleihung und einer allfälligen Publikation/Verlautbarung der Versuch einer systematischen Nachbereitung oder gar Evaluation der Ergebnisse gemacht worden ist. Eine Ausnahme stellt hier etwa der European Public Sector Award dar, der zumindest eine Art Nachbereitungsstrategie im Rahmen der sogenannten „EPSA Knowledge – Transfer Activities“ verfolgt. So wurden beispielsweise nach der Preisverleihung 2011 verschiedene nationale Workshops und die erste Knowledge-Transfer Conference zum Thema „Cities of Excellence in the Public Sector: The Innovation Front with their New Practices and Formulas“ in Bilbao im Jahr 2012 durchgeführt (vgl. <http://www.epsa2013.eu/en/content/show/&tid=120>). Selbst in diesem Fall sucht man jedoch vergeblich nach Erkenntnissen zur faktischen Wirkung der angestrebten Lern- und Diffusionsprozesse.

Tatsächlich liegen Evaluationen zu den externen und internen Effekten von Qualitäts- und Verwaltungswettbewerben, jedenfalls soweit ersichtlich, bislang überhaupt nicht oder zumindest nicht öffentlich zugänglich vor. Demgegenüber lässt sich sehr wohl konstatieren, dass manche dieser Veranstaltungen inzwischen Mobilisierungsschwierigkeiten haben oder sogar ganz eingestellt werden mussten, wie beispielsweise die European eGovernment Awards. Bei aller Vorsicht ist daher zu konstatieren, dass Qualitäts- und Verwaltungswettbewerbe entgegen den ursprünglich damit verbundenen Zuschreibungen eher als kurzfristig aufleuchtendes Feuerwerk der Verwaltungsinnovation bzw. -präsentation

denn als Trojanisches Pferd der Leistungsmessung beschrieben werden können. Dass dies nicht zwangsläufig so sein muss, zeigt allerdings ein Gegenbeispiel: der „Speedy Offiziales“ des Bundeslandes Kärnten.

4. Der „Speedy Offiziales“: ein „deviant case“ aus Kärnten?

„Speedy Offiziales“⁴ ist die Bezeichnung für einen von der Wirtschaftskammer und der Landesregierung ausgelobten Verwaltungspreis des Bundeslandes Kärnten, der – in seiner ursprünglichen Form – für das schnellste Betriebsanlagengenehmigungsverfahren an Kärntner Bezirksverwaltungsbehörden verliehen wird. Dieser Qualitäts- und Verwaltungswettbewerb stellt insofern einen abweichenden Fall gegenüber den vorherigen Ausführungen dar, als die Teilnahme zum einen mehr oder weniger verpflichtend ist und zum anderen von Anfang an eine externe wissenschaftliche Begleitung durchgeführt wurde. Teilnehmer sind grundsätzlich die Gewerbebehörden der acht Kärntner Bezirkshauptmannschaften sowie der zwei Statutarstädte Klagenfurt am Wörthersee und Villach. Aufbauend auf einem Ranking erhält der Gewinner des „Speedy Offiziales“ eine Siegerurkunde sowie ein Preisgeld, verbunden mit öffentlicher Aufmerksamkeit und entsprechend positiver Berichterstattung. Anhand der Entwicklung der letzten 15 Jahre gewinnt die These, dass solche Wettbewerbe durchaus ein Trojanisches Pferd der Leistungsmessung sein können, an Bedeutung, während sich gleichzeitig der interessante Effekt eines von der Verwaltung initiierten Gegen-trojaners beobachten lässt.

4.1 Vom Verwaltungspreis für Schnelligkeit ...

Der „Speedy Offiziales“ wurde erstmals im Jahr 1998 durchgeführt (zum Folgenden vgl. *Fachhochschule Kärnten, Studienbereich Wirtschaft & Management* 2012); damit fällt sein Ursprung in die Frühphase der Durchführung solcher Wettbewerbe und auch der wissenschaftlichen Befassung mit ihnen. In den Anfangsjahren – der zweite Durchgang fand 2001 statt – lag der Fokus ausschließlich auf der Dauer von Betriebsanlagengenehmigungsverfahren, wobei angesichts der (Mit-)Trägerschaft des Wettbewerbs durch die Wirtschaftskammer der Verdacht nahe liegt, dass hier gezielt eine Einengung auf die möglichst rasche

⁴ Die Verfasser danken den anderen Mitgliedern des Projektteams „Speedy Offiziales“ am Studienbereich Wirtschaft & Management der Fachhochschule Kärnten für die Bereitstellung der Auswertungen, insbesondere dem Rektor und Studienbereichsleiter, FH-Prof. Dr. *Dietmar Brodel*, und der Projektkoordinatorin, Mag.^a *Christina Mundschtz*.

Abwicklung solcher Verfahren im Sinne einer „unternehmensfreundlichen“ Genehmigungspraxis vorgenommen wurde. Der eindeutigen und einseitigen Ausrichtung auf „Kundenbedürfnisse“ folgte auch die Ergänzung des Wettbewerbs durch das Kriterium der „UnternehmerInnenzufriedenheit“ mit dem Betriebsanlagengenehmigungsverfahren im dritten Durchlauf 2005, in dem eine gesonderte Untersuchung zur „Servicequalität“ durchgeführt wurde.

Im Unterschied zu anderen Qualitäts- und Verwaltungswettbewerben stand damit beim „Speedy Offiziales“ von Anfang an der Aspekt der Leistungsmessung – Schnelligkeit von und später Zufriedenheit mit einem Verwaltungungsverfahren – ersichtlich im Vordergrund. Die kaum verhüllt interessengeleitete Fokussierung auf lediglich zwei Aspekte eines sehr komplexen Verwaltungsprozesses, in dem neben der Wirtschaft auch andere Anspruchsgruppen wie zum Beispiel AnrainerInnen tangiert sein können und zahlreiche Schutzvorschriften zu beachten sind, rief jedoch bald einen verwaltungsinternen Widerstand hervor, der sich zudem relativ einfach objektivieren ließ. Nach dem dritten „Speedy Offiziales“ 2005 setzte daher eine erhebliche, wenn auch nicht dokumentierte und öffentlich gemachte Kritik von Seiten der Bezirksverwaltungsbehörden gegenüber dem Amt der Kärntner Landesregierung ein, das Wettbewerbsdesign nachhaltig zu verändern. In Frage gestellt wurden dabei in erster Linie nicht der Wettbewerb selbst oder der Gedanke einer vergleichenden Leistungsmessung, sondern vor allem die Sinnhaftigkeit der isoliert gemessenen Parameter „Verfahrensdauer“ und „Kundenzufriedenheit“ sowie die Art und Weise der angewandten Messmethoden.

4.2 ... zum Verwaltungspreis für Verfahrensqualität

Eine erste Reaktion des Amtes der Kärntner Landesregierung auf die Kritik aus den Bezirks- und Magistratsverwaltungen war die Beauftragung der Fachhochschule Kärnten mit den Erhebungen für den vierten Wettbewerb im Jahr 2008, nachdem zuvor die Wirtschaftsuniversität Wien und die Universität Klagenfurt für die wissenschaftliche Begleitforschung zuständig waren. Um die Vergleichbarkeit mit den vorherigen Studien zu gewährleisten, wurden weiterhin die Kriterien der „Verfahrensdauer“ und der „Kundenzufriedenheit“ erhoben, hinzu kam jedoch die Einbeziehung situativer Umfeldbedingungen und Besonderheiten in den jeweiligen Verwaltungsbezirken, durch welche die genannten Faktoren beeinflusst werden können. Zudem wurden erstmals Hintergrundgespräche mit den VerwaltungsmitarbeiterInnen selbst geführt, die in die Interpretation der Ergebnisse mit einfließen. Dem bislang letzten „Speedy Offiziales“, der Ende 2011 zum fünften Mal in Folge durchgeführt worden ist, lag dann bereits ein erheblich modifiziertes Methodendesign zugrunde. Gemessen werden nunmehr die Rechts-

sicherheit, die Verfahrensdauer sowie die Zufriedenheit der AntragstellerInnen (direkt durch Befragung) und der AnrainerInnen (indirekt über Beschwerden) mit insgesamt 124 per Zufallsstichprobe ausgewählten und bereinigten Genehmigungsverfahren, die im Jahr 2010 abgeschlossen worden sind. Berücksichtigt wurde dabei auch der – möglicherweise verzerrende – Einfluss der Wirtschafts-, Bezirks- und Personalstruktur, der antragstellenden Unternehmensbranchen, der Art und Anzahl der Genehmigungen, der jeweils gewählten Verfahrensart und der erforderlichen Einbeziehung von Sachverständigen. Wesentlich ergänzt wurde die Analyse zudem – vor allem auch aufgrund des weiterhin bestehenden Unmuts gegenüber einem nicht als völlig fair empfundenen „Wettbewerb“ – durch eine systematische Befragung der VerwaltungsmitarbeiterInnen, um deren Fachwissen bezüglich der Besonderheiten von Betriebsanlagengenehmigungsverfahren miteinbeziehen zu können.

Erst aus der Zusammenschau all dieser Parameter ergibt sich die seit 2011 für den Gewinn des „Speedy Offiziales“ ausschlaggebende Verfahrensqualität als Basis für das weiterhin durchgeführte Behördenranking und für die Preisverleihung:

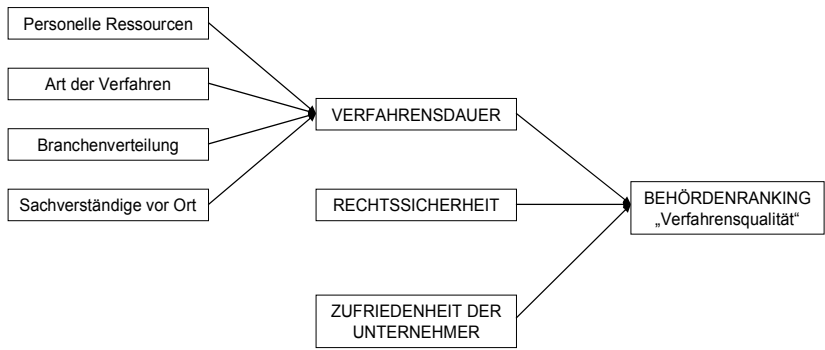


Abb 1: Parameter für das Behördenranking des „Speedy Offiziales“

Der vorliegende Beitrag ist nicht der Ort für eine detaillierte Darstellung der durchaus interessanten Ergebnisse des „Speedy Offiziales“; er soll vielmehr der Diskussion der These von Qualitäts- und Verwaltungswettbewerben als möglichem Trojanischen Pferd der Leistungsmessung dienen.

So war der „Speedy Offiziales“ diesbezüglich zunächst als reines, wenn auch unzulässig vereinfachtes Leistungsmessungsinstrument konzipiert, das in Über-

einstimmung mit Teilen der Literatur eher als Wettbewerb kaschiert wurde. Im Laufe der Zeit gelang es den durch mehr oder weniger sanften Zwang zur Teilnahme verpflichteten Verwaltungen jedoch, über Interventionen und Sachargumente den ursprünglichen Trojaner zu nutzen und in einen Gegentrojaner – hier vor allem gegenüber sachfremden Einflüssen einzelner Interessenvertreter – umzuwandeln.

Dies war in erster Linie durch die schwer abzuweisende Forderung möglich, nicht nur einzelne Aspekte zu isolieren, sondern den gesamten Prozess des Betriebsanlagengenehmigungsverfahrens und die jeweiligen Einflussgrößen in die Betrachtung einzubeziehen. Damit war die ursprüngliche Fokussierung alleine auf die Verfahrensdauer aufgebrochen und diese stellte nur mehr ein Element bei der umfassenden Messung der Verfahrensqualität dar. Tatsächlich konnten die am „Speedy Offiziales“ teilnehmenden Bezirks- und Magistratsverwaltungen zeigen, dass ihre Genehmigungsentscheidungen nicht nur ausgesprochen rechtsicher und bestandskräftig sind. Vielmehr gelang es ihnen auch, die antragstellenden UnternehmerInnen als ihre „Kunden“ in einem sehr hohen Maß zufrieden zu stellen. Gerade im Sinn der von der Wirtschaft geforderten Servicequalität unter dem Stichwort der Verwaltung als Dienstleister erzielten die Kärntner Genehmigungsbehörden sehr hohe Werte:

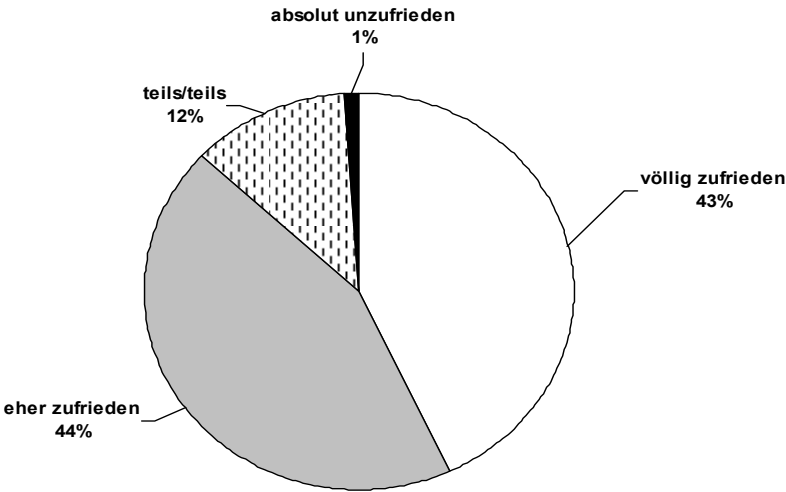


Abb. 2: Zufriedenheit mit dem Gesamtverfahren

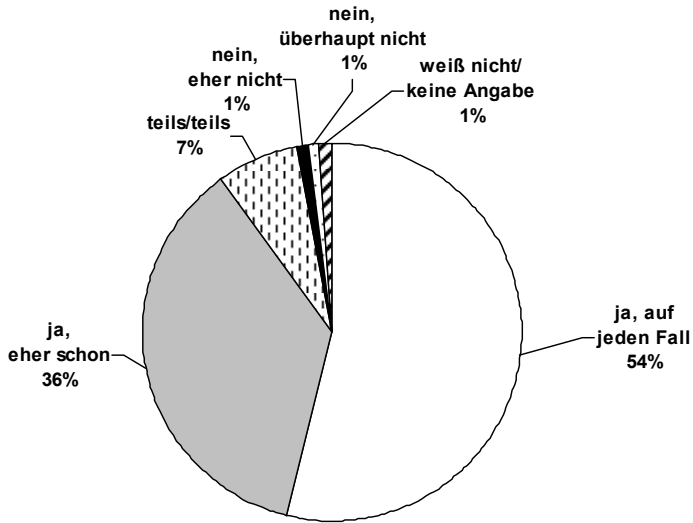


Abb. 3: Behörde als serviceorientierter Dienstleister

Die darüber hinausgehende und im Rahmen des „Speedy Offiziales“ 2011 erstmals systematisch durchgeführte Befragung der MitarbeiterInnen war von diesen selbst gefordert worden und kann gleichfalls als eine Art Gegentrojaner gegenüber den als einseitig empfundenen Versuchen der Landesregierung und der Wirtschaftskammer Kärnten verstanden werden, den kompetitiven Vergleich von Betriebsanlagengenehmigungsverfahren auf wenige, potenziell verzerrende Kriterien zu reduzieren. Die MitarbeiterInnen erhielten dadurch die Gelegenheit, ihre fachlichen Erfahrungen und differenzierten Sichtweisen den Initiatoren und Auftraggebern des „Speedy Offiziales“ rückzuspiegeln. Einer immer noch großen Zufriedenheit mit dem eigenen Arbeitsumfeld und der Möglichkeit des eigeninitiativen Arbeitens stand dabei die vehemente Kritik an Versuchen sachfremder Einflussnahme, zunehmend knapperen Ressourcen und Zielkonflikten zwischen kurzer Verfahrensdauer und rechtmäßiger, neutral-objektiver Verfahrensabwicklung gegenüber. Als besonders problematisch wurde zudem gesehen, dass das Interesse der Fachabteilungen bei der Rekrutierung geeigneter MitarbeiterInnen nur zum Teil Berücksichtigung findet und Einstellungsverfahren eher intransparent und im Ergebnis nicht selten wenig zufriedenstellend ausgestaltet werden.

Neben der Kritik an diesen Missständen, welche sich negativ auf die Verfahrensqualität auswirken, aber von den Genehmigungsbehörden selber kaum zu

beeinflussen sind, waren die MitarbeiterInnen jedoch auch in der Lage, detaillierte Verbesserungsvorschläge zur weiteren Optimierung der Betriebsanlagengenehmigungsverfahren zu geben. So forderten sie zum Beispiel die Entwicklung eines Verfahrensleitbildes unter Einbeziehung aller Anspruchsgruppen, die Einführung eines einheitlichen Verfahrenscontrollings mit einem behördenübergreifenden Qualitäts- und Wissensmanagementsystem, die Verbesserung der Kommunikation mit und der Information von UnternehmerInnen sowie die Ermöglichung einer elektronischen Beantragung und einer Verfahrenskonzentration. Weitere Vorschläge betrafen die Einbeziehung von Sachverständigen – zum Beispiel die vermehrte Ausbildung von UniversalistInnen, die Schaffung bzw. Ausdehnung des Versicherungsschutzes sowie eine Weiterbildungsverpflichtung – und die Verbesserung des Verfahrensablaufes selbst, etwa durch die vermehrte Durchführung von Erstberatungsgesprächen, die Vermeidung von Vorprüfungen, den verwaltungsinternen Austausch fachlicher Expertise bei speziellen Verfahren und die Einführung eines mit den AntragstellerInnen vereinbarten Bescheiddatums.

Durch die Präsentation der Ergebnisse des „Speedy Offiziales 2011“ in einem Workshop und einer nachfolgenden Pressekonferenz unter Beteiligung des zuständigen Landesrates und des Präsidenten der Wirtschaftskammer fanden diese Forderungen öffentlichkeitswirksam Verbreitung. Auch nach der eigentlichen Preisverleihung stehen die entsprechenden Erkenntnisse für eine sachliche, nicht länger einseitig dominierte Diskussion über die weitere Ausgestaltung des Betriebsanlagengenehmigungsverfahrens in Kärnten zur Verfügung. Gleichzeitig ist damit der Weg für eine sinnvolle Fortführung des „Speedy Offiziales“ als praxisrelevantem Qualitäts- und Verwaltungswettbewerb eröffnet, der einen umfassenden Leistungsvergleich ermöglicht.

5. *Resümee*

Die vorhergehenden Ausführungen zeigen, dass Qualitäts- und Verwaltungswettbewerbe häufig nicht die Funktionen als Instrument der Leistungsmessung und als „Wettbewerbssurrogat“ erfüllen können, welche ihnen in der Fachliteratur ursprünglich zugeschrieben worden sind. Im Vordergrund der derzeitigen Praxis stehen wohl eher kurzfristige Anerkennungs-, Legitimations- und Informationsfunktionen für die üblicherweise freiwillig an solchen Veranstaltungen teilnehmenden Verwaltungen. Eine besondere Bedeutung könnte ihnen allerdings als „Türöffner“ in heterogenen, schlecht vernetzten und/oder barrierebehafteten Kontexten zukommen: dies gilt im Speziellen innerhalb der Europäischen Union, wo sie den in einem noch losen Verwaltungsverbund zusammengeschlossenen nationalen Verwaltungen einen angstfreien, zwanglosen und

grenzüberschreitenden Austausch von Good Practices ermöglichen können. Fragen wie Repräsentativität, Leuchtturmwirkungen und Nachhaltigkeit solcher Prozesse sind jedoch noch weithin ungeklärt.

Im Gegensatz dazu zeigt das Beispiel des Verwaltungspreises „Speedy Offiziales“ in Kärnten, dass solche Qualitäts- und Verwaltungswettbewerbe, so wie ursprünglich beabsichtigt, durchaus als echte Trojanische Pferde der Leistungsmessung ausgestaltet werden können. Dieses Vorgehen wird allerdings auf Verwaltungsseite regelmäßig nach einiger Zeit Widerstände hervorrufen, insbesondere wenn die Art der Leistungsmessung methodisch angreifbar ist und die Verwaltung ihre Kritik als konstruktiven Verbesserungsvorschlag einbringt. Im Umkehrschluss gilt damit aber auch: wenn Elemente der Leistungsmessung mit einem Qualitäts- und Verwaltungswettbewerb verbunden werden, kann dies in letzter Konsequenz auch von der Verwaltung selbst genutzt werden, um ihre ansonsten meist schwer zu vermittelnden Anliegen gegenüber der Politik und den externen Anspruchsgruppen zu artikulieren. In diesem Fall ist es gerade die Öffentlichkeitswirksamkeit von Preisverleihungen, welche eine höhere Responsivität für Auftrag, Funktionsweise und Probleme von Verwaltungen erzeugen kann. Hierin liegt ein bedeutender Unterschied von Qualitäts- und Verwaltungswettbewerben gegenüber anderen Instrumenten des Leistungsvergleichs, wie zum Beispiel dem internen und externen Benchmarking von Verwaltungen, denen dieser Öffentlichkeitsfaktor fehlt.

Verwaltungspreise können damit idealerweise ein Mittel unter anderen sein, um ein tieferes Verständnis für die Leistungen und die leistungsbeeinflussenden Kontextbedingungen von Verwaltungen zu generieren. Im besten Fall tragen sie zu einem fundierteren Verständnis für Verwaltungserfordernisse in einer zunehmend komplexer werdenden Umwelt bei; im schlechtesten Fall schaden sie jedenfalls nicht, sondern erzeugen zumindest eine kurzfristige positive Aufmerksamkeit für die ausgezeichneten Verwaltungen.

Quellennachweis

Bundeskanzleramt Österreich (2013): Der Österreichische Verwaltungspreis. Online unter URL: <http://www.bka.gv.at/site/7223/default.aspx> [Stand: 30.04.2013].

EIPA – European Institute of Public Administration (2009): Aims of EPSA. Online unter URL: <http://www.epsa2009.eu/en/content/show/&tid=107> [Stand: 30.04.2013].

European Union (2009): Welcome to the European eGovernment Awards 2009. Online unter URL: <http://www.epractice.eu/awards> [Stand: 30.04.2013].

Fachhochschule Kärnten, Studienbereich Wirtschaft & Management (2012): Ergebnisbericht Speedy Offiziales 2011. Studie im Auftrag der Kärntner Landesregierung zur Dauer und Qualität von Betriebsanlagengenehmigungsverfahren in Kärnten 2010. Villach [unveröffentlicht].

- Kouzman, Alexander/Löffler, Elke/Klages, Helmut/Korac-Kakabadse, Nada* (1999): Benchmarking and performance measurement in public sectors. Towards learning for agency effectiveness. In: *The International Journal for Public Sector Management*, Vol. 12, No. 2, pp. 121-144.
- Merkel, Adolf* (1927): *Allgemeines Verwaltungsrecht*, Wien.
- Neisser, Heinrich* (1998): Die politischen und institutionellen Rahmenbedingungen der Verwaltungsreform. In: *ders./Hammerschmid, Gerhard* (Hrsg.): *Die innovative Verwaltung. Perspektiven des New Public Management in Österreich*. Wien, S. 295-335.
- Neisser, Heinrich/Meyer, Renate/Hammerschmid, Gerhard* (1998): Kontexte und Konzepte des New Public Management. In: *Neisser, Heinrich/Hammerschmid, Gerhard* (Hrsg.): *Die innovative Verwaltung. Perspektiven des New Public Management in Österreich*. Wien, S. 19-55.
- Nullmeier, Frank* (2011): Wettbewerb. In: *Blanke, Bernhard/Nullmeier, Frank/Reichard, Christoph/Wewer, Götz* (Hrsg.): *Handbuch zur Verwaltungsreform*. 4. Aufl., Wiesbaden, S. 150-160.
- Nullmeier, Frank* (2005): Wettbewerb und Konkurrenz. In: *Blanke, Bernhard/Bandemer, Stephan von/Nullmeier, Frank/Wewer, Götz* (Hrsg.): *Handbuch zur Verwaltungsreform*. 3. Aufl., Wiesbaden, S. 108-120.
- Pühringer, Josef* (2010): Festansprache im Rahmen der Verleihung des oberösterreichischen Verwaltungspreises „Publicus“. In: *Raiffeisen Landesbank Oberösterreich: Von den Besten lernen. Oberösterreichs Verwaltungspreis „PUBLICUS“ verliehen*. Online unter URL: http://www.raiffeisen.at/eBusiness/rlbooe_template2/15752112992436962-506082648250998262-697653340150006632-NA-25-EN.html [Stand: 30.04.2013].
- Schauer, Reinbert* (2010): *Öffentliche Betriebswirtschaftslehre – Public Management*. 2. Aufl., Wien.

Die Auswirkungen von Bewertungen auf die öffentliche Verwaltung

Heinz Pansi

1. Allgemeines

- Würde man heute die Auswirkungen des Drucks der leeren Kassen auf die öffentlichen Verwaltungen außer Acht lassen, so käme man vielleicht zur Feststellung, dass die Modernisierung der öffentlichen Verwaltungen aus der Perspektive des Idealismus, aus dem Gedanken der Zukunftsorientierung, der Serviceorientierung, der Effizienz, der Effektivität, der Leistungsstärke, des Erfolges oder des Bestrebens nach einem objektiven Vergleich mit anderen Einrichtungen getragen werde.

Natürlich war in Österreich, aber auch in Deutschland, der Start der jüngsten Modernisierung der öffentlichen Verwaltung vom Gedanken des New Public Managements getragen. Der Gedanke der Trennung von Politik und Management, das Wettbewerbsprinzip, der Unternehmergeist mit einzelwirtschaftlichen Betrachtungen, die innerbetrieblichen Reformen der Strukturen und der Geschäftsprozesse, oder der Gedanke des flexiblen Personaleinsatzes gehen darauf zurück.

- Aufgrund der neuen gesellschaftlichen und der angesprochenen wirtschaftlichen Gegebenheiten ist es notwendig geworden, die Modernisierungskonzepte des Staates zu erweitern und damit auch Maßnahmen anzudenken, wonach die Verwaltung ihre Außenorientierung stärkt, da sie die Aufgaben nicht mehr im Alleingang lösen kann.

Eine Redimensionierung der öffentlichen Verwaltung und damit auch eine Aktivierung der Ressourcen im Umfeld ist ein Ausgangspunkt zu einer neuen Verwaltung.

Politik, Verwaltung, Wirtschaft, Bürger sind die tragenden Elemente einer solchen neuverstandenen Verwaltung, die ihren Weg über das New Public Management zum Good Governance beschreitet, wobei die Verwaltung nicht als Hemmnis, sondern als Aktivator des Umfeldes verstanden werden soll.

Ausgehend von diesen rudimentär dargelegten theoretischen Ebenen des Verwaltungsstrukturgedankens in der Republik Österreich bieten sich verschiedene Möglichkeiten zur Stärkung der Modernisierungsbestrebungen, wozu eben auch Bewertungen und Vergleiche in verschiedenen Formen zählen.

Gerade die Kriterien des Good Governance sind dazu angetan, die Verwaltung aus verschiedenen Perspektiven zu beleuchten, zu bewerten und daraus Rückschlüsse zu ziehen.

2. Die Bezirkshauptmannschaft

Die Bezirkshauptmannschaften (BH) sind monokratisch organisierte erstinstanzliche Landesbehörden. Sie besorgen alle Angelegenheiten der Bundes- und Landesverwaltung, die nicht anderen Behörden zugewiesen sind.

Sie ist die Behörde der allgemeinen staatlichen Verwaltung. Ihre Sprengel überziehen das gesamte Landesgebiet mit Ausnahme einiger Städte mit eigenem Statut, die die Gemeindeverwaltung und die Bezirksverwaltung in sich vereinen!

Die Organisationsstruktur der BH ergibt sich bereits aus dem Jahre 1853. An der Spitze der Bezirkshauptmannschaft steht der Bezirkshauptmann. Dieser ist für die gesamte Geschäftsführung verantwortliche und ist das entscheidende Organ. Damit ist für den Bereich der Bezirkshauptmannschaft das monokratische System lückenlos durchgeführt.

Die Bezirkshauptmannschaften, die mit den Landkreisen in Deutschland vergleichbar sind, haben sowohl die Geschäfte der mittelbaren Bundesverwaltung als auch die der Landesverwaltung zu führen.

Dies bedeutet also, dass überall dort, wo es nicht eigene Bundesbehörden gibt, die Bezirkshauptmannschaft grundsätzlich 1. Instanz bei der Erledigung sämtlicher Angelegenheiten ist. 2. Instanz in fachlicher Hinsicht ist der Landeshauptmann, wobei die 2. Instanz für sämtliche Rechtsmittel gegen getroffene Entscheidungen die unabhängigen Verwaltungssenate/Verwaltungsgerichte sind. Die Zuständigkeiten der Bezirksverwaltungsbehörde betreffen daher (auszugsweise):

- Bau- und Raumordnung
- Veterinärwesen
- Sozialwesen/Jugend und Familie
- Forstwesen/Jagdwesen
- Dienstaufsicht, Personalangelegenheiten
- Führung der sozialen Struktur des Bezirkes
- Führung der schulischen Struktur
- Koordination der Gemeinden
- Leistungsfeststellungen/Disziplinarangelegenheiten
- Katastrophenbekämpfungskoordination
- Schulen
- Gewerbe
- Kfz/Führerscheinrecht
- Verkehr

- Sicherheitspolizei/Dienstbehörde Polizei
- Strafen
- Pass/Fremdenpolizei/Staatsbürgerschaft
- Gesundheitswesen
- Jugendfürsorgeärztlicher Dienst
- Wasserrecht

Bei der Wahrnehmung der öffentlichen Sicherheitsangelegenheiten steht dem Bezirkshauptmann der Wachkörper Bundespolizei zur Verfügung.

Ausgehend von dieser Summe von Kompetenzen, ist es nur logisch, dass Bewertungen für eine allumfassende Basisverwaltung, die im ganzen Bundesgebiet gleich strukturiert ist, sehr weitreichende Konsequenzen haben.

3. *Die Bewertungen in der Praxis*

Bewertungen können nach verschiedenen Perspektiven unterteilt werden. Im gegenständlichen Fall werden die Kundenbewertungen, die Qualitätsbewertungen – mit daraus resultierenden Konsequenzen – sowie das Benchmarklearning (Projekt BH plus) dargestellt.

Dieser Unterteilung geht eine Betrachtung der Bewertungen nach Organisationszielen und nach politischen Zielen voraus. All diese Betrachtungen erfolgen aus dem Blickwinkel der Behörde I. Instanz, der BH, „der Verwaltungseinheit“ in Österreich schlechthin.

Grundsätzlich kann festgestellt werden, dass die Organisationsziele jene Zielsetzungen verfolgen, die für das zielorientierte Arbeiten einer Verwaltungsbehörde entscheidend sind. Es geht dabei um die Entwicklung allgemeiner Standards, um die Erhöhung von Qualität und Wirkung, aber auch um die Ablaufoptimierung von Geschäftsprozessen, um Kostenreduktion und natürlich um einen entsprechenden Imagegewinn.

Getragen sind all diese Ziele natürlich vom Verkaufseffekt, von dem die Politik lebt. Es geht um die Schaffung einer kunden- und dienstleistungsorientierten Verwaltung zur Steigerung der politischen Wirkungsziele und damit auch um die Schaffung eines positiven Politikbegriffes.

Natürlich verfolgt die Politik weniger die klassischen Organisationsziele der Verwaltung, zumal das ein Thema von Fachleuten ist – wohl aber geht es der Politik um das Thema Kostenreduktion!

Die Entstehung von Bewertungen ist stets auf oben dargelegte Ziele begründet.

3.1 Die Kunden/befragung/bewertung

Eine Bewertung, die für die Verwaltungspraxis von großer Bedeutung ist, ist die sogenannte „selbstbestimmte“ Bewertung.

Die Verwaltung und die Politik im Bundesland Kärnten wollte wissen, in welchem Ausmaß die Akzeptanz der Verwaltung in der Bevölkerung vorliegt. Aus diesem Grunde wurde von den Bezirkshauptmannschaften in Kärnten die Kundenbefragung/bewertung im Sinne eines „Qualitätsdialoges mit dem Bürger“ ins Leben gerufen. Da die Politik davon überzeugt war, dass die Verwaltung eine gute ist, hat sie diesen Prozess sehr unterstützt.

Dabei wurde die Ausgangslage festgestellt und Fragen nach den Bedürfnissen der Bevölkerung gegenüber der Verwaltung positioniert. Darauf aufbauend wurden für Parteien-Verhandlungen, aber auch für die „Laufkundschaft“ innerhalb der Behörde Fragebögen entwickelt und aufgelegt, die es zu beantworten galt.

Die Fragen betrafen organisatorische Fragen, Wegleitsysteme, Telefonumleitungen, Öffnungszeiten, Parkplätze aber auch das Erscheinungsbild nach außen. Zudem wurde die Frage positioniert, wie in der öffentlichen – oder besser in der veröffentlichten – Meinung die Position der Bezirkshauptmannschaften gesehen wird.

Bei allen Fragestellungen wurden Benchmarks mit den anderen Bezirkshauptmannschaften gelegt, die Bewertungskriterien wurden festgelegt, die Auswertungen wurden sodann vorgenommen.

3.1.1 Auswirkungen

Die Bewertung der Kundenzufriedenheit ist für die BH Hermagor äußerst zufriedenstellend abgelaufen. Zudem wurden sehr, sehr viele Anregungen, auch für die künftige Tätigkeit, abgegeben. Daraus resultierten regelmäßige Fachinformationen in den regionalen Medien, die Schaffung eines Tages der offenen Tür, die Gründung einer Antidrogenplattform, Partnerschaften mit anderen Bezirkshauptmannschaften wurden dadurch intensiviert, und nicht zuletzt wurde das Zusammengehörigkeitsgefühl innerhalb unseres Bezirkes und des öffentlichen Dienstes gestärkt. Ein Resultat der Umfragebewertungsaussagen war sogar die Abhaltung eines „Balls der Einsatzkräfte“, sowie die Einführung der „Schülergalerie“ in der BH Hermagor, um einen weiteren Beitrag zur engen Bindung zwischen Verwaltung und Bürger zu schaffen.

4. *Die (nicht selbstbestimmte) Qualitätsbewertung am Beispiel des „Speedy Offiziales“*

Mein Vorredner hat zu diesem Thema den theoretischen Unterbau sehr umfassend dargelegt. In der Praxis ging es um die Aussage Dauer, Qualität, Zufriedenheit und galt es die daraus resultierenden Konsequenzen zu beleuchten. So etwa war die Steigerung der Reputation aber auch die Steigerung des Konkurrenzdenkens durch Bekanntgabe eines Rankings der beteiligten Behörden, Ziel dieser Bewertung.

Die angewandte Systematik war im Wesentlichen das Aktenstudium, die Datenerhebung und die Befragung. Wenn man nun diese Form von Bewertung, bei der unter Umständen die Zielsetzungen nur mehr marginal deckungsgleich mit den eingebrachten Zielsetzungen der bewerteten Organisationen sind, betrachtet, besteht die Gefahr, dass neben den offiziell bekannt gegebenen und sicherlich ehrenwerten Zielen der Bewertung, in der Realität eine Gefahr droht:

Die Politik sieht sich insbesondere in Zeiten von Vorwahlkämpfen oder auch in wirtschaftlich schweren Zeiten nicht dazu bemüht, Positives zu verstärken, sondern setzt mit Schnelligkeit ein Zuviel an Personal gleich, mit Langsamkeit eine zu geringe Kompetenz gleich und egal ob etwas zu schnell oder zu langsam ist, die Vertreter der Wirtschaft, die Träger dieses Wettbewerbes in concreto, können in Versuchung kommen oft auf dem Rücken der Verwaltung mit der Politik abzurechnen.

Gerade der Wettbewerb 2012 hat solche Befürchtungen der teilnehmenden Behörden bestätigt. Seitens der Personalabteilung des Amtes der Kärntner Landesregierung wurden genau diese oben dargelegten Positionen ins Treffen geführt, weshalb ich eigentlich sehr froh war, dass meine Behörde nicht die Schnellste war! Seitens der Wirtschaftskammer wurde im Zuge dieses Wettbewerbes, der objektiv bewertet ein hohes Lob für die Verwaltung brachte, überhaupt die Abschaffung der Bezirkshauptmannschaften in Österreich gefordert, um wohl in einem zweiten Schritt jede fachliche Kompetenz ausschließlich bei den Kammern anzusiedeln.

Es ist wohl nachvollziehbar, dass derzeit unsere Bereitschaft, bei solchen Wettbewerben aktiv mit dabei zu sein, als limitiert zu betrachten ist.

5. *Benchlearning der Bezirkshauptmannschaften*

Neben der Kundenbefragung, der Selbstbewertung – dem CAF, oder den Verwaltungswettbewerben, die ja auf eine freiwillige Beteiligung bauen und insoweit von großem Interesse für die teilnehmenden Organisationen sind – erscheint mir der Bereich des Benchlearnings der Bezirkshauptmannschaften als erwäh-

nenswerte Bewertung mit Auswirkungen auf das Leben der öffentlichen Verwaltung!

Das „Benchlearning der Bezirkshauptmannschaft“ wird von fünf Bezirkshauptmannschaften in Österreich getragen. Ziel war es in klar definierten Themenbereichen zusammen zu arbeiten und voneinander zu lernen! Diese Themenbereiche betrafen: Den Bereich der Führungsparameter, die strukturierte Vorgangsweise im Fremdenwesen, die Ethik in der Verwaltung (Stichwort Korruption), die Sicherheitsagenden und die Großschadensereignisse.

Systematisch wurde die Ausgangslage definiert, wurden Zieldefinitionen vorgenommen und wurden nach der Ergebnisfeststellung und Wirkungsfeststellung schließlich die gemeinsamen Standards definiert, die für oben genannten Themenbereiche künftig als Parameter gelten.

Jede Bezirkshauptmannschaft hat im Wesentlichen die Themenführerschaft über ein Thema übernommen, wobei innerhalb von zwei Jahren sehr zielführende Ergebnisse erzielt werden konnten, die innerhalb der Behörden sehr intensiv gelebt werden.

5.1 Beispiel Führung

- Ausgangslage
Ein standardisierter Einsatz von Führungsinstrumenten ist nicht vorgesehen gewesen
- Ziel
CAF-Selbstbewertung in allen teilnehmenden Bezirkshauptmannschaften
- Ergebnis
Der Vergleich der Selbstbewertungen und die Schaffung einer CAF-ModeratorInnenausbildung
- Wirkung
Durch den CAF werden Prozesse eingeleitet und Entwicklungen forciert
- Standard
CAF-Selbstbewertung; Alle 5 Jahre soll unter externer/professioneller Moderation ein jährlicher Qualitätszirkel der Führungskräfte stattfinden.

5.2 Beispiel „Fremdenwesen“

- Ausgangslage
In allen beteiligten Bezirkshauptmannschaften wurde festgestellt, dass unzureichende und unstrukturierte Informationen zum Thema Fremdenwesen ein Problem darstellen.

- Ziel
Die Steigerung der Kundenorientierung und der Effizienz, sowie ein verbessertes Informationsangebot (vor allem für Frauen).
- Ergebnis
Die Erstellung eines mehrsprachigen Informationsfolders
- Wirkung
Die Verbesserung der Kundenorientierung und die Berücksichtigung der Diversity
- Standard
2 Behördenkontakte zur Erteilung aufenthaltsrechtlicher Bewilligungen sollen ausreichen, wobei die Information in der jeweiligen Muttersprache zur Verfügung gestellt werden soll.

Eine Aktualisierung der Informationen soll im Jänner und September jedes Jahres erfolgen.

5.3 Beispiel „Ethik in der Verwaltung“

- Ausgangslage
In der Verwaltung ist das Thema „Ethik in der Verwaltung“ unzureichend thematisiert.
- Ziel
Das Thema der Werthaltungen soll von den Mitarbeitern intensiv wahrgenommen werden und so ein Bewusstsein für ethisches Handeln bei den MitarbeiterInnen geschaffen werden.
Die gemeinsam getragenen Werte sind daher zu definieren.
- Ergebnis
Mehrere positive Werte als gemeinsame Basis, woraus Richtlinien für ethisches Verhalten entstanden sind.
- Wirkung
Ein bewusstes ethisches Verhalten der MitarbeiterInnen.
- Standard
Die Thematisierung von Ethik im jährlichen Mitarbeitergespräch.

5.4 Beispiel „Sicherheitsagenden“

- Ausgangslage
Es wurde in den Arbeitsrunden festgestellt, dass unterschiedliche Zugänge zu den sicherheitsbehördlichen Aufgaben bestehen.

- Ziel
Die Erarbeitung regionaler kriminalpolizeilicher Strategien
- Ergebnis
Sicherheitsbefragung in den Bezirken
Regionale kriminalpolizeiliche Strategien
- Wirkung
Die Steigerung der Sicherheit und die Hebung des Sicherheitsgefühls
- Standard
Die Vorgangsweise betreffend den Prozess der Strategieerstellung wurde vereinheitlicht und so ein inhaltlicher Standard einer kriminalpolizeilichen Strategie geschaffen.

5.5 Beispiel „Großschadensereignis“

- Ausgangslage
Die unterschiedlichen Vorgangsweisen der Katastrophenschutzbehörden bei der Bewältigung von Großschadensereignissen („Katastrophen“)
- Ziel
Die Definition von Mindeststandards für alle Behörden bei der Vorbereitung und Abarbeitung des Szenarios und des Aufbaus des behördlichen Einsatz/Führungsstabes
- Ergebnis
Die Definition von Mindeststandards und die Idealkonstruktion des behördlichen Führungsstabes
- Wirkung
Die Bewältigung von Großschadensereignissen durch koordinierte Führung, sowie optimale Hilfestellung für die Betroffenen im Anlassfall, verbunden mit strukturierter Medienarbeit
- Standard
Die Mindeststandards (Erreichbarkeit, Ausbildung, Ausstattung, Infrastruktur, Personal,) wurden festgemacht.

6. Die Auswirkung von Bewertungen in der Praxis, unter dem Aspekt des „Benchlearning-Projektes“ der BH

6.1 Die Auswirkungen von Bewertungen auf personeller Ebene

Bewertungen stellen für Mitarbeiter jedenfalls

- eine Standortbestimmung dar
- einen Erfahrungsaustausch durch enge Kontakte während der Projektphase dar
- aufgrund der Kontakte und Impulse für neue Ideen, sowie
- den „Blick über den Tellerrand dar“

Unstrittig ist damit also eine Steigerung der Mitarbeitermotivation und eine Steigerung der Leistungsbereitschaft verbunden.

6.2 Auswirkungen von Bewertungen auf die organisatorische Ebene

Hier kann summarisch festgehalten werden, dass es zu

- einer Standortbestimmung für die teilnehmenden Organisation
- zu Innovationen durch Kooperationen,
- zu Kooperationen bei der Leistungserstellung sowie zu
- einem gegenseitigen Nutzen durch den Austausch des Know-how kommt.

Die organisatorische Ebene baut daher ihre Stärke, aufgrund der Stärkung der gemeinsamen Positionen durch wechselseitige Unterstützungen, aus.

Die verbesserte Informationsschiene ermöglicht die Anwendung von „Best Practice“-Beispielen, wodurch

- eine Steigerung der Prozessqualität,
- eine Steigerung der Servicequalität sowie
- die Weiterentwicklung der aktivierenden Verwaltung

möglich wird!

6.3 Die Außenwirkungen des Verhaltens der Organisation

Mit den mit der Bewertung verstandenen Maßnahmen wird die Verwaltung für den Bürger spürbar. Die Behörde wird von der Bevölkerung aktiv (gestaltend) wahrgenommen, wodurch ein positives Image der Organisation (trotz allfälliger negativer rechtlicher Entscheidungen) transportiert werden kann.

Gerade die Bewertungen im Rahmen des „Benchlearning-Projekts“ haben in der Bevölkerung das hohe Maß an Bürgerservice und das hohe Maß an Kundenzufriedenheit offen dargelegt.

Eingedenk obiger Ausführungen muss man festhalten, dass die Darlegung eines klaren Ergebnisses nur ein kleiner Bildausschnitt ist. Das Entscheidende ist das, was hinter der Form der Bewertung liegt, in der Vorbereitung, in der Ausarbeitung, in der Konsequenz.

Als ein Behördenleiter, der sich darüber freuen kann, dass seine Organisation mehrfach sehr erfolgreich aus verschiedensten Formen von Bewertungen und daraus resultierenden Vergleichen hervorgegangen ist, kann ich bemerken, dass die Auswirkungen auf personeller Ebene, bei korrektem Umgang mit dem Ergebnis, was auch eine klare Definition der Vorgangsweise im Rahmen der Zieldefinition erfordert, sehr positive Auswirkungen zeitigt, verbunden mit den dargestellten Auswirkungen auf die organisatorische Ebene bzw. auf die Öffentlichkeit.

7. *Conclusio*

Wilhelm Shakespeare hat einmal gesagt „Miss den Nächsten nicht nach dem eigen Maß“.

Genau das ist das Problem bei Bewertungen. Genau darum geht es bei der Auswirkung von Bewertungen. Damit eine Bewertung für alle Betroffenen die beabsichtigten positiven, personellen, organisatorischen Auswirkungen hat, bedarf es zuvor einer Vereinbarung des Umgangs mit den Resultaten – und bedarf es der Verlässlichkeit.

Auch die Schlussfolgerungen der Politik müssen letztlich im Bewusstsein der Politik betreffend ihre Verantwortung gegenüber der Verwaltung und der Bevölkerung erfolgen – nämlich dergestalt, dass Bewertungen eine „Motivation“ für die Organisation, ihre Mitarbeiter und die Bevölkerung sein müssen!

Podiumsdiskussion

Christoph Emminghaus

Das öffentliche Interesse an öffentlicher Leistungserbringung und insbesondere den hierfür notwendigen finanziellen Mitteln hat seit der Jahrtausendwende weiter zugenommen. Staatsverschuldung, Stabilitätskriterien und Schuldenbremse sind womöglich nur einige jener Schlagworte, welche die unbedingte Notwendigkeit zu stärkerer Ausgabendisziplin andeuten. Hinzu kommt eine Öffentlichkeit, die häufiger über angebliche oder tatsächliche Verschwendung von öffentlichen Geldern informiert wird und gleichzeitig steigende Ansprüche an die Qualität öffentlicher Leistungserbringung formuliert.

1. Entwicklung von Evaluation und Wirkungsmessung

In diesem Spannungsverhältnis erlangen Wirkungskontrollen und Evaluationen von öffentlichen Leistungen eine bislang nicht gekannte Bedeutung und erleben einen ungebrochenen Aufschwung. Festzumachen ist dieser Aufschwung auf ganz verschiedenen Ebenen. So wird in aktuellen Gesetzesinitiativen zunehmend häufiger die (Selbst)verpflichtung zur Evaluation, Überprüfung oder zumindest Berichterstattung mit festgeschrieben. Prominente Beispiele der vergangenen Jahre auf Bundesebene sind die als Hartz-Gesetze in den allgemeinen Sprachgebrauch eingegangenen Reformen am Arbeitsmarkt, die Reform des Zuwanderungsgesetzes mit der Etablierung von Integrationskursen oder zuletzt der Aufbau eines Rechtsanspruchs auf Kinderbetreuung ab dem ersten Lebensjahr. Weitere Beförderer des Aufschwungs von Evaluationen können die unterschiedlichen europäischen Förderprogramme sein, die flächendeckend in ganz Europa auf den verschiedenen administrativen Ebenen begleitet, überprüft und evaluiert werden. Auf kommunaler Ebene können noch Leistungsvergleiche oder Benchmarkings, wie sie etwa die KGSt oder die Bertelsmann Stiftung Anfang des Jahrtausends initiiert haben, als Wiege von Wirkungsmessung und Evaluation interpretiert werden. Und letztlich haben die Ergebnisse großer internationaler Vergleichsstudien mit zum Durchbruch beigetragen – wobei hier die als PISA-Studie bekannt gewordenen Leistungsvergleiche im schulischen Bereich die mit Abstand größte Bekanntheit und Wirkung auf verschiedenen Ebenen entfaltet haben.

Insgesamt vollzieht sich damit in Deutschland eine Entwicklung nach, der in den angelsächsischen Ländern aber auch in Skandinavien einige Jahre früher begonnen hat. Der Trend hin zu evidenzbasierter Politikgestaltung auf Basis empirischer Befunde.

Diese Entwicklung trägt dazu bei, dass sich Evaluationen aktuell als eigene Methodendisziplin herausbilden und erste Studien- und Weiterbildungsgänge entstehen. Dieses ist insofern logisch, als sich in nur zehn Jahren die methodischen Standards rasant weiterentwickelt haben. Was noch vor zehn Jahren ausgeschrieben, durchgeführt und abgenommen wurde, entspräche heute weder mit Blick auf die verfolgten Zielsetzungen und Fragestellungen noch mit Blick auf die methodische Umsetzungen geforderten Standards.

Mit dem Trend hin zu komplexeren und methodisch anspruchsvolleren Evaluationen verändert sich die Landschaft der beteiligten Akteure. Evaluationen sind nun häufiger Teamwork und werden in Kooperation von verschiedenen Evaluatoren oder sogar Evaluationsinstituten durchgeführt. Es ist eben häufig nicht dieselbe Person, welche einen Stakeholderdialog organisieren kann, komplexe Befragen durchführt und die Datensätze später multivariat auszuwerten vermag. Politisch relevante Evaluationen werden heute überwiegend in die Hände von Expertenteams gelegt, die derartigen Anforderungen gewachsen sind und dies durch ihre bisherige fachliche und methodische Arbeit nachweisen können. Parallel ist eine entsprechende Marktentwicklung zu beobachten, dass dort, wo Ergebnisse eine politische Relevanz erlangen sollen, von Seiten der Auftraggeber auch die Bereitschaft existiert, die notwendigen finanziellen Mittel für aufwendigere Untersuchungsdesigns zur Verfügung zu stellen.

Den Bedeutungsgewinn den Evaluationen als Disziplin erlangt haben, lässt sich am leichtesten an der Entwicklung der Gesellschaft für Evaluation e.V. (DeGEval) ablesen, die 1997 als Zusammenschluss von an Evaluationen beteiligten Akteuren und Instituten gegründet wurde. Waren es Anfang des Jahrtausends lediglich 20-30 institutionelle Mitglieder, so ist diese Zahl bis 2012 auf knapp 150 angestiegen. Gerade mit der Etablierung von DeGEval Standards hat die Gesellschaft einen wichtigen Beitrag zur Sicherung von Mindestkriterien geschaffen. Mit Blick auf die sich rasant weiter entwickelnde Disziplin verwundert es nicht, dass diese Standards aktuell überprüft und ggf. einer ersten Revision unterzogen werden sollen.

2. Umgang mit Evaluation und Leistungskontrolle

Sollen Evaluationen kein Selbstzweck bleiben, ist zu allererst zu klären, welche Ziele hiermit konkret verfolgt werden. Dies kann sich im Detail unterscheiden, im Kern stehen jedoch in der Regel Fragen nach den erreichten Ergebnissen

sowie nach Effektivität und Effizienz. Eng damit verbunden ist das Aufzeigen von Wirkungszusammenhängen und ggf. auch von Handlungsalternativen.

Inwieweit Evaluationen es schaffen, derartige Zielsetzungen zu bedienen, hängt von verschiedenen Einflussfaktoren ab. Aus der Analyse zahlreicher Evaluationen kann jedoch als zentral betrachtet werden, inwieweit es gleich zu Beginn des Prozesses gelingt, eine Klärung zwischen Auftraggebern und Evaluatoren bezüglich der konkret verfolgten Zielsetzungen zu schaffen und in der Folge ein hierauf abgestimmtes methodisches Vorgehen zu entwickeln. Gelingt dies nicht, werden auch professionell arbeitende Evaluatoren keine für den Auftraggeber befriedigenden Ergebnisse erzielen können.

Dort wo dies dagegen gelingt eine konkrete Zielsetzung mit einem angemessenen Design zu unterfüttern, können Evaluatoren Teil eines Interventionssystems werden, das zentral zum Erfolg eines Programms oder Projektes beiträgt. Damit unterscheiden sich Evaluatoren (so wie sie heute überwiegend arbeiten) stark von einer fast antiquiert anmutenden Gutachterrolle. Evaluatoren sollen im Rahmen von begleitenden (formativen) Evaluationen Einfluss auf den Erfolg des Vorhabens nehmen und so Teil eines lernenden Systems werden. Lernen und weniger Kontrolle rückt damit in den Mittelpunkt von Evaluationsvorhaben.

Dies zeigt sich auch in Begrifflichkeiten. Wo vor einigen Jahren noch von Benchmarking gesprochen wurde, werden ähnliche Vorhaben heute teilweise als Benchlearning bezeichnet. Es ist davon auszugehen, dass sich dieser Trend fortsetzen wird und der prozessuale Charakter von Evaluationen immer weiter an Bedeutung gewinnt.

3. *Aktuelle Trends*

In der Zukunft sind eine ganze Reihe von weiteren methodischen Entwicklungen beim Einsatz von Evaluationen zu erwarten. So wird weiterhin daran gearbeitet, robustere Instrumente für Kosten-Nutzen Relationen zu entwickeln und fundiertere Wirkungsmodelle zu etablieren. Die Multikausalität von sozialen Prozessen macht dies zu einer häufig nicht unbedingt leichten Aufgabe.

Zumindest dort, wo dies möglich erscheint und die notwendige Relevanz besitzt, werden darüber hinaus künftig häufiger Experimente in Evaluationen zum Einsatz kommen. Mit diesen teilweise als „Goldstandard“ bezeichneten Verfahren wird das Ziel verfolgt, empirisch gesicherte Aussagen über die Wirksamkeit bestimmter Interventionen zu erlangen. Ähnlich dessen, was beispielsweise in den mehrstufigen Zulassungsverfahren für neue Medikamente etablierte Praxis ist, sind in verschiedenen Politikfeldern Evaluationen zu erwarten, die mit verschiedenen Interventionsgruppen oder mit Vergleichsgruppen arbeiten. Es liegt

auf der Hand, dass solche Ansätze nicht nur ungleich teurer sind, sondern teilweise vor der Anwendung zentrale ethische Fragen zu beantworten sind.

Ein anderer Aspekt, der zunehmend stärkere Beachtung erfährt, ist die Rolle des Auftraggebers. Hier werden aktuell verschiedene Ansätze der Selbstevaluation erprobt. Mit Blick auf häufig enge Finanzrahmen erscheint dies nachvollziehbar. Es ist jedoch nicht davon auszugehen, dass sich derartige Ansätze durchsetzen werden. So fehlt Ergebnissen von Selbstevaluationen häufig die gewünschte legitimierende Wirkung. Gleichzeitig sind Evaluationen mittlerweile häufig zu komplex, um diese ohne fundierten Methodenhintergrund durchzuführen. Dort, wo dies versucht wurde, konnten im Ergebnis teilweise auch die erwünschten Kostenersparnisse nicht realisiert werden, da der Aufwand um ein Vielfaches höher war, als ursprünglich geplant.

Wesentlich erfolversprechender erscheinen dagegen Ansätze, die auf einen Aufbau von Evaluationskompetenz bei den Auftraggebern zielen. So kann es mit vertretbarem Aufwand gelingen, Auftraggeber sprachfähig zu machen, ihre Zielsetzungen besser formulieren zu können und den Prozess angemessen begleiten zu können. Dies hat sowohl auf die Qualität der Ergebnisse als auch auf die Effizienz des Mitteleinsatzes bei einer Evaluation einen erwartbar positiven Einfluss.

Weitergehende Entwicklungen sind aktuell im Bereich der Entwicklungszusammenarbeit zu beobachten, ein Politikbereich, der lange Zeit führend im Einsatz von Evaluationen war. Hier wird aktuell mit dem Deutschen Evaluationsinstitut der Entwicklungszusammenarbeit (DEVal) ein staatliches Institut etabliert, das den Einsatz von Evaluationen in diesem Bereich steuern und vorantreiben soll. Dies könnte eine Blaupause für andere Politikfelder werden.

Vor dem Hintergrund der beschriebenen Entwicklung ist insgesamt zu erwarten, dass die Bedeutung von Evaluationen und Wirkungsmessungen in Deutschland weiter zunehmen wird. Jedoch werden sich in den verschiedenen Politikfeldern unterschiedliche Herangehensweisen und methodische Standards etablieren. Vorreiter werden weiter jene Felder sein, die unter hoher öffentlicher Aufmerksamkeit stehen oder ihre Wirkung häufig unter Beweis stellen müssen.

Wir brauchen Bewertungen und ich glaube auch, dass Bewertungen stattfinden, und zwar überall da, wo es eben möglich ist. Das heißt aber noch lange nicht, dass alle bewertet werden wollen. Bewertungen eröffnen eine ganz andere Perspektive, nämlich den Blick von außen, der zugleich Öffentlichkeit und in gewisser Weise Transparenz herstellt. Zugleich ist bei der Durchführung von Bewertungen darauf zu achten, dass man Zeiten und Zeitpunkte mit im Auge hat. Ein Beispiel: Wenn in Krisen schnell politische Entscheidungen getroffen werden müssen – so denken wir jetzt nur einmal an die Finanzkrise –, wird man über Evaluation in einer anderen Art und Weise nachdenken, als wenn man beispielsweise langfristige Straßenplanungen macht. Es ist auch nicht so, dass im politischen Geschäft sich alles nach Rationalitäten bemisst, sondern manchmal sind es eher Wertungen, die sehr unterschiedlich ausfallen können, und manchmal ist es eher ein Gefühl – und das muss gar nicht falsch sein. Ich denke z. B. beim Thema Sicherheit: Natürlich müssen wir Sicherheit gewährleisten, aber die beste Sicherheit nützt nichts, wenn der Bürger sich unsicher fühlt. Also, das Gefühl, sicher leben zu können, ist hier sicherlich auch eine sehr, sehr wichtige Komponente. Die Problematik von Komplexität ist so oft erörtert worden, dass ich das jetzt nicht noch einmal wiederholen möchte. Die Dinge sind im politischen Geschäft nicht eindimensional. Daraus resultieren Schwierigkeiten. Und als letzten Punkt möchte ich noch die Medien erwähnen, die ihre Berichterstattung nicht nur an Sachthemen ausrichten. Das heißt, das Bewertet Werden von außen, das Sichtbarmachen von Bewertungen würde ich als problematischer bewerten wollen als das Stattfinden von Bewertungen. Und deswegen muss auch der Blick darauf gerichtet sein, wie man eigentlich mit Wirkungsmessungen umgeht. Welche Wirkungen hat eine Evaluation? Politik lebt mit Zielkonflikten. Was aus der einen Perspektive richtig ist, muss es aus der anderen noch nicht sein. Auch wenn niemand Negativbewertungen mag, gehört es doch zur Professionalität, Kritik anzunehmen und aus ihr zu lernen. Deswegen im Fazit: Wir brauchen Evaluationen. Ich bin überzeugt, dass es sehr wichtig ist, dass wir Transparenz schaffen, auch nach außen. Dabei sollten wir Klarheit über unsere Ziele haben. Das ist eine Kulturfrage. So ist beispielsweise der angloamerikanische Raum in dieser Hinsicht sicherlich weiter als wir, er geht auch kompromissloser vor. Schließlich bin ich fest überzeugt, dass wir in einem lernenden Prozess sind, und würde, jedenfalls von meiner Seite, sehr begrüßen, wenn dieser mit Qualität und Engagement weiter betrieben wird.

Evaluationen sind eines der Themen, zu denen man sich als Hochschullehrer gleichzeitig als Praktiker äußern kann. An den Hochschulen finden Lehrevaluationen statt, Forschungsevaluationen erfolgen unsystematisch in der alltäglichen Wahrnehmung der „Community“ und systematisch in formalisierten Verfahren, wie etwa durch Berufungskommissionen. Projekte, für die es eine externe Förderung gibt oder für die eine solche beantragt wird, durchlaufen aufwändige Prozesse der Begutachtung. So ist für die Hochschullehrer neben dem Bewerten auch längst das Bewertetwerden eine geradezu alltägliche Erfahrung. Diese Erfahrung bildet die Grundlage der folgenden Überlegungen, bei denen zudem auch die juristische Perspektive zur Sprache kommen soll.

Ausgangspunkt der Überlegungen ist die Beobachtung, dass Entscheidungsträger im Allgemeinen ein originäres Interesse an Datenmaterial haben, wie es durch Evaluationen bereitgestellt wird. Ausnahmen hiervon sind nur dann zu machen, wenn die Ergebnisse dieser Evaluationen dezidiert unwillkommen sind, insbesondere weil sie eine bestimmte Entscheidung schwieriger werden lassen. Ist eine solche Vorprägung in eine bestimmte Richtung nicht erfolgt, so sind durch Evaluationen erlangte Daten schon deswegen nützlich, weil sie den auf sie gestützten Entscheidungen in der externen Wahrnehmung eine besondere Legitimität geben. Diese legitimierende Funktion ist grundsätzlich auch dann gegeben, wenn die Aussagekraft der Informationen, objektiv betrachtet, äußerst gering ist und es begründete Zweifel an der angewendeten Methode gibt. Ranglisten haben einen plakativen Charakter, der im Kern auch nicht dadurch beseitigt wird, dass die genannten Einwände bestehen. Platzierung in bestimmten Tabellen oder auch Symbole, wie etwa Ampelfarben, verfehlen ihre Wirkung auf die Leser in der Regel nicht. Zudem wird in der öffentlichen Diskussion als wenig glaubwürdig angesehen, wer von einem schlechten Platz aus dessen Aussagekraft in Zweifel zieht. Derartige Kritik hört man eben selten von den Spitzenreitern – selbst früheren Kritikern fällt es nicht schwer, sich mit Wettbewerben anzufreunden, die sie selbst gewinnen.

Eine Differenz zwischen der extern wahrgenommenen Legitimität und faktisch fehlender Aussagekraft von Evaluationen hat erhebliche negative Auswirkungen auf die Motivation der Evaluierten. Anders als die Öffentlichkeit sehen

- 1 Eine fundierte Auseinandersetzung mit vielen für die Tagung neu formulierten Gedanken und eine ausführliche Darstellung der hier kurz skizzierten Rechtslage findet sich in dem Beitrag des Verfassers in: Die Verwaltung 44 (2011), 179 ff.

sich nämlich die von einer negativen Evaluation Betroffenen mit großer Wahrscheinlichkeit die methodischen Grundlagen derselben an. Vor diesem Hintergrund gibt es ein erhebliches Interesse an einem adäquaten Rechtsrahmen für Evaluationen. Ein solcher ist bislang allenfalls in ersten Ansätzen erkennbar. Dabei geht es um die Rechtsprechung zu bestimmten Einzelphänomenen. „Evaluation“ ist bislang kein, jedenfalls aber kein klar definierter Rechtsbegriff. Die genannten Einzelphänomene bilden noch kein kohärentes Gesamtbild. Das mag damit zusammenhängen, dass sich die Gerichte – noch dazu Verwaltungs- und ordentliche – mit sehr disparaten Problembereichen zu beschäftigen hatten. Einer dieser Fälle findet sich bei der gerichtlichen Kontrolle dienstlicher Beurteilungen im Beamtenrecht – die wohl älteste Form der Evaluation, die in der deutschen Rechts- und Verwaltungsgeschichte ihren Platz findet. Neuer sind die Auseinandersetzungen über die Bewertungen der Stiftung Warentest, die vor den Zivilgerichten ausgefochten worden sind. Erste Ansatzpunkte für eine entsprechende Dogmatik finden sich jetzt auch in Bezug auf Fragen von Akkreditierungen von Studiengängen.

Neben einem objektiven Bedürfnis nach einer unabhängigen, vor allem also einer gerichtlichen, Überprüfung dürfte in vielen Fällen auch ein subjektives bestehen. Dies gilt spätestens dann, wenn an die Evaluationen Verteilungsentscheidungen geknüpft sind. Der naheliegende Anknüpfungspunkt ist dann der Gleichheitssatz bzw. andere Gleichbehandlungsgebote. Die Beispiele der Bewertungen durch die Stiftung Warentest und der dienstlichen Beurteilungen zeigen aber, dass mitunter schon weit im Vorfeld tatsächlicher negativer Auswirkungen Rechtsschutz gewährt werden kann. Hinsichtlich der Intensität der Kontrolle ist festzustellen, dass der Ausgangspunkt stets die Annahme eines Beurteilungsspielraums der Evaluierenden bzw. der die anzuwendenden Maßstäbe Festlegenden sein muss. Insofern geht es bei der gerichtlichen Überprüfung letztlich um einen Mindeststandard an Rationalität und eine hinreichende Plausibilität der Entscheidung. Die Maßstabsbildung wird dabei bereichsspezifisch stattzufinden haben. Die erheblichen Unterschiede zwischen den Sachverhalten lassen ansonsten allenfalls eine Schaffung von sehr abstrakten Regelungen zu. Je stärker mit Evaluationen auch Verteilungswirkungen verbunden sind, umso eher wird auch die Forderung nach einer gesetzlichen Regelung im Raum stehen.

Der prägende Eindruck nach 20 Jahren Rechnungshoftätigkeit im Kollegium ist der: Wenn wir isoliert über Erfolgskontrolle und Evaluation reden, beschäftigen wir uns mit dem Schwanz des Pferdes. Die Bewegungen des Pferdes werden aber an anderer Stelle gelenkt – nämlich vorne –, der Schwanz kommt nur hinterher. Die drei wesentlichen Punkte, die ich vor diesem Hintergrund noch einmal kurz ansprechen will, sind folgende: Vor Maßnahmebeginn müssen Ziele definiert werden, es muss eine Ausgangslagenbeschreibung mit den Mängeln geben und drittens müssen wir Wirtschaftlichkeitsuntersuchungen endlich ernstern nehmen als bisher. Dazu will ich Ihnen einige Beispiele aus der Tätigkeit des Hamburgischen Rechnungshofs nennen.

Ohne definierte Ziele ist keine vernünftige Evaluation möglich. Frau Lohmann, wenn Sie zu recht den politischen Charakter von Ausgabensteigerungen z.B. für die öffentliche Sicherheit hervorheben, so war dies in Hamburg natürlich auch so, als nach den Wahlen 2002 über 600 neue Stellen für die Polizei bewilligt wurden. Es ist nun ein wesentlicher Unterschied, ob Sie als Ziel definieren, die öffentliche Sicherheit zu verbessern, dann ist nämlich die Kriminalstatistik Erfolgsparameter, oder ob Sie das Sicherheitsgefühl von Teilen der Bevölkerung verbessern wollen, dann müssen nämlich die Bürger mehr Polizei, beritten, zu Fuß oder in Uniform auf ihren Straßen sehen, d.h. also zunächst eher eine Umstellung der Dienstpläne. Selbst für solche hochpolitischen Angelegenheiten ist also eine Zielkonkretisierung einzufordern. Dies kann die Politik nur selten leisten, wie wir heute Morgen auch von Herrn Konzendorf gehört haben. Es ist vielmehr originäre Aufgabe der Verwaltung, die Ziele ihrer Tätigkeit konkret zu definieren oder mindestens zu operationalisieren – und zwar der leitenden Leute in der Verwaltung. Niemand anders sonst kann das. Das gilt nicht nur für die Verwaltung selbst – insbesondere im Zuwendungsbereich, und hier v.a. bei sozialpolitischen Maßnahmen, stellen wir Mängel fest –, sondern auch für Verselbstständigungen wie etwa 15/26er Betriebe und für die öffentlichen Unternehmen und Anstalten. Denn auch sie sind Instrumente in der Hand des Staates zur Erfüllung öffentlicher Zwecke.

Ein weiteres Beispiel dafür, dass unnütze Ausgaben primär und wesentlich dadurch verringert werden, dass die Zügel in die Hand genommen werden, ist die Sportförderung: Es gibt kaum einen Bereich, mit dem die Politik so vernetzt ist. Das heißt, ein Rechnungshof, der die Sportförderung prüfen will, muss sich warm anziehen und sehr genau überlegen, wie er vorgeht. Wir haben zunächst in den 80er Jahren die Sportförderung vorsichtig, tastend und uns orientierend ge-

prüft. Ende der 90er Jahre dann noch einmal mit folgenden Ergebnissen: Erstens skandalträchtige Einzelfälle, grobe Zuwendungsverstöße, Mittelzweckentfremdungen, die die Veränderungsbereitschaft förderten. Zweitens: Es fehlte jede Transparenz, wie viel der Sport insgesamt bekommt. Allgemeines Aufhören. Und drittens: Die Zwecke der Sportförderung müssen konkretisiert werden, denn ohne definierte Zwecke ist keine Erfolgskontrolle sinnvoll möglich und kann auch vom Rechnungshof nicht ersetzt werden. Da flossen über Jahrzehnte Millionen in den Sport mit der Zwecksetzung: „Zur allgemeinen Sportförderung“. Dies ist gleichbedeutend mit dem Verzicht auf jeden Massstab für die weitere Verwendung von Steuergeldern. Deswegen haben wir gesagt, so geht das nicht, ihr müsst das konkretisieren. Es war klar, dass wir uns damit zunächst nicht durchsetzen würden, weil die Verflechtungen zu eng sind und der Anspruch des Sportes, mit möglichst wenig Bindungen weiter gefördert zu werden, groß und verfestigt war. Aber wir haben immerhin erreicht, dass jedenfalls ein engagierter Staatssekretär, d.h. ohne seinen Minister, der Regierung erste Ansätze für eine fachlich/politische Schwerpunktsetzung unterbreitete. Damit war der Boden für einen zielgerichteten Mitteleinsatz und dessen Evaluation ein Stück weit bereitet. Jetzt sind wir auf der nächsten Stufe: Im diesjährigen Jahresbericht finden Sie nach einer erneuten Prüfung die Zusage, auf Basis sportpolitischer Entscheidungen Ziel- und Leistungsvereinbarungen mit konkreten Maßnahmen und Kennzahlen einzuführen. Der Rechnungshof drängt hier mit seinen Mitteln und seinen Maßstäben – Ordnungsmäßigkeit und Wirtschaftlichkeit – erfolgreich darauf, durch Handeln in der konzeptionellen Phase die wirksame und zielgerichtete Verwendung öffentlicher Mittel sicherzustellen.

Ein weiteres „hochpolitisches“ Beispiel – die soziale Integration – finden Sie ebenfalls im diesjährigen Jahresbericht. In einer ersten Prüfung von 2006 hatten wir darauf gedrungen, bei diesem zentralen Thema die hehren politischen Ankündigungen in konkrete und maßnahmebezogene Felder zu differenzieren, um die vorhandenen Mittel wirksam einzusetzen. Hieran fehlte es bis dahin, obwohl in Hamburg über 450.000 Menschen mit ausländischen Wurzeln leben und 2005 eine bundesrechtliche Integrationsverpflichtung geschaffen worden war. Wir stellten fest, dass die vereinzelt Massnahmen nicht mit einander abgestimmt waren, dass weder Bedarfsanalysen noch ein Gesamtkonzept vorlagen und vorhandene Vorarbeiten stecken geblieben waren. Wie soll der Erfolg sozialer Integration evaluiert werden, wenn es keine Bedarfsbeschreibung und Zielkonkretisierung durch die dazu berufenen staatlichen Instanzen gibt? Vor dem Hintergrund der insoweit günstigen politischen Ausgangslage in Hamburg lag das geforderte Konzept bereits nach einem halben Jahr Parlament und Öffentlichkeit vor, mit einer Beschreibung von Handlungsbedarfen für ca. 60 Arbeitsfelder. Erst damit hatten wir die Möglichkeit, in einer Prüfung des letzten Jahres eine erste Evaluation des Erreichten vorzunehmen. Erst jetzt ist das Thema so aufbereitet,

dass es evaluierend weiter bewegt werden kann und muss, d.h. Rechnungshöfe können Evaluationsmöglichkeiten erzwingen, indem sie Rationalität für die konzeptionelle Phase einfordern. Dies ist besonders dann wichtig, wenn von der Politik ein inhaltlich eher vager Handlungsdruck ausgeht und auch weder der Verwaltung noch eingeschalteten Dritten einfällt, was konkret noch zu machen wäre. In derartigen Fällen sollten Programme eingestellt werden, wie wir das bei Mitteln für „innovative Arbeitsmarktförderung“ erreicht haben, die nur zur Hälfte abgeflossen waren und deren andere Hälfte zu 90% für Verwaltungskosten zweckentfremdet worden war.

Ein letztes Beispiel: Der Klimaschutz. Hamburg hat 120 Millionen bereit gestellt, v.a. um den CO₂-Ausstoß um 2 Millionen Tonnen innerhalb von fünf Jahren zu vermindern. Dazu werden insgesamt 388 Maßnahmen durchgeführt, zu denen viel Fotovoltaik gehört. Viele Maßnahmen wirken gar nicht. Und bei der Fotovoltaik kostet eine Tonne Verminderung mit 1.000 Euro ca. 50mal mehr als wenn Sie ein Gebäude vernünftig energetisch sanieren. Von diesen 388 Maßnahmen waren nur 16 % überhaupt für eine Erfolgskontrolle geeignet. Und von den in das Monitoring einbeziehbaren gut 60 Maßnahmen war dann wieder ein großer Teil schlicht unwirtschaftlich. Auch wenn Rechnungshöfe sich an ihre Leisten halten – Ordnungsmäßigkeit und Wirtschaftlichkeit –, können und müssen ihre Feststellungen die eigentlich für Wirksamkeit und Wirtschaftlichkeit entscheidende Phase der Mittelbereitstellung prägen. Ein zweites wesentliches Element dieser Anfangs- und Entscheidungsphase ist die Beschreibung der Ausgangslage, der Datenbasis und des Änderungsbedarfs. Wer mit fremdem Geld Neues schaffen oder Bestehendes ändern will, hat in jeder Hinsicht die Darlegungslast und kann in der Haushaltsordnung und den Verwaltungsvorschriften die Details nachlesen. Nicht nur für Rechnungshöfe gilt, dass sie ihre Forderungen und Empfehlungen hinsichtlich ihrer Eignung begründen und den zugrundeliegenden Änderungsbedarf nach konsensfähigen Maßstäben belegen müssen.

Erst ganz am Schluss, beim Pferdeschwanz sozusagen, stellt sich die die Frage der Evaluation.

Wir haben in den 80er und 90er Jahren viel erreicht, um die Erfolgskontrolle zu stärken. Herr Dieckmann, den Sie aus der Arbeit der Deutschen Sektion ja gut kennen, hat sich hier hohe Verdienste erworben, die Erfolgskontrolle zu einem wirksamen Instrument in Hamburg zu machen. Eine Besonderheit Hamburgs ist der Anfang der 90er Jahre eingeführte § 7 Abs. 4 LHO, wonach bei allen staatlichen Vorhaben die Zielsetzungen zu bestimmen sind sowie die Zielerreichung begleitend und abschließend zu überprüfen ist. Wir haben uns hier mit Hilfe des Parlaments gegen deutlichen Widerstand in der Exekutive (Einheitlichkeit des Haushaltsrechts, Ausreichen von Verwaltungsvorschriften) durchgesetzt und sind zufrieden, dass wir mit dieser gesetzlichen Verpflichtung zur Zielkonkretheit

sierung vor Handlungsbeginn zu einem rationalitätsfördernden Maßstab für Verwaltung und Politik beigetragen haben.

Ein dritter wesentlicher Punkt für das Gelingen eines Vorhabens – und dann auch seiner Evaluation – ist das vorherige Vorliegen einer Wirtschaftlichkeitsuntersuchung. Schon bei Anmeldung zur Mittelfristigen Finanzplanung müssen realistische Annahmen zu Nutzen und Kosten beigelegt werden. In Hamburg haben wir außerdem die Mittelfristige Aufgabenplanung, die eine konkretere Betrachtung für die kommenden 4 Jahre verlangt. Spätestens zum Haushaltsplanentwurf müssen konkrete und vollständige Wirtschaftlichkeitsuntersuchungen beigebracht werden, d.h. ggfs. auch mit den nichtquantifizierbaren Faktoren. Häufig fehlen in Wirtschaftlichkeitsuntersuchungen die Risiken und Bandbreiten v.a. bei Prognosen. Wenn z. B. bei der von Hamburg im nächsten Jahr durchzuführenden Internationalen Garten-Ausstellung die Einnahmen zu prognostizieren sind, so darf Parlament und Regierung nicht nur der für die Finanzierung unproblematischere Höchstbetrag genannt werden.

Wenn es um die maßgebliche Entscheidungsphase geht, so sind auch realistische Folgekosten offenzulegen. Bei der eben genannten Internationalen Gartenschau haben wir z.B. festgestellt, dass für einen großen neuen Park Unterhaltungskosten von 30000 Euro jährlich angesetzt waren, ein unhaltbar niedriger Betrag. Ob die nach unserer Beanstandung sofort auf 1,5 Mio Euro erhöhte Summe reicht, wird die Zukunft zeigen. Die ausschlaggebende Bedeutung der Entscheidungsphase zeigt sich auch bei ganz normalen Bauinvestitionen: In aller Regel werden nur die schlichten Errichtungskosten angesetzt, obwohl ein Gebäude eine Lebensdauer von 60 oder mehr Jahren hat. Wenn es z.B. mit großzügigen Glasfronten ausgestattet wird, steigen die Betriebskosten für Reinigung, Heizung und Kühlung stark an. Ohne diese Lebenszykluskosten, die durchaus 40-50% der Gesamtkosten betragen können, wird in Unkenntnis der wirklichen Kosten entschieden. Dann entsteht vielleicht sehr preiswert ein ansehnliches Gebäude, dessen Unterhalt aber unangemessen teuer ist und später zu Überforderungen oder Ausgabenverzichten an anderer Stelle nötigt. Noch einmal: Über Wirtschaftlichkeit und sachlichen Erfolg öffentlicher Vorhaben wird in der Anfangsphase entschieden.

Schließlich gehört in die Anfangsphase öffentlicher Vorhaben auch hinein, dass Varianten und alternative Handlungsmöglichkeiten dargestellt und abgewogen werden, dass Verantwortung, Zeiträume und Methoden für die Erfolgskontrolle festgelegt werden und dass die Entscheidungsgrundlagen vollständig – insbesondere die Wirtschaftlichkeitsuntersuchung – Regierung und Parlament vorgelegt werden. In Hamburg jedenfalls hat das Parlament kürzlich auf Anregung des Rechnungshofes beschlossen, dass Wirtschaftlichkeitsuntersuchungen auch ihm vorzulegen sind, ehe es über große Maßnahmen beschließt. Sind zu Beginn von Vorhaben die Ziele konkret formuliert, die Änderungsbedarfe dargestellt

und die Wirtschaftlichkeitsuntersuchung vorgelegt, so ist nicht nur die handwerkliche Basis für Evaluation und Erfolgskontrolle gelegt, sondern auch bei Parlament, Regierung und Öffentlichkeit insoweit Resonanz und Interesse zu erwarten.

Autorenverzeichnis

- Rainer Christian Beutel*, Vorstand der Kommunalen Gemeinschaftsstelle für Verwaltungsmanagement, Köln.
- Dr. *Jörg Bogumil*, Universitätsprofessor, Lehrstuhl für Öffentliche Verwaltung, Stadt- und Regionalpolitik, Ruhr-Universität Bochum.
- Dr. *Christoph Emminghaus*, Geschäftsführer der Ramboll Management GmbH, Berlin.
- Marika Gruber*, Mag.^a (FH), Wissenschaftliche Mitarbeiterin und Forschungsreferentin, Fachhochschule Kärnten, Villach.
- Dr. *Norbert Heller*, EU-Kommission, Task Force for Greece, Brüssel.
- Werner Keggenhoff*, Präsident des Landesamtes für Soziales, Jugend und Versorgung Rheinland-Pfalz.
- Dr. *Gottfried Konzendorf*, Professor, Deutsche Universität für Verwaltungswissenschaften Speyer, Bundesministerium des Innern, Berlin.
- Dr. *Alexander Kroll*, Assistant Professor of Public Administration, School of International and Public Affairs, Florida International University (Miami/USA).
- Beate Lohmann*, Ministerialdirektorin, Bundesministerium des Innern, Berlin.
- Dr. *Veith Mehde*, Universitätsprofessor, Lehrstuhl für Öffentliches Recht und Verwaltungswissenschaft, Leibniz Universität Hannover.
- Anne-Kathrin Meier*, Wissenschaftliche Mitarbeiterin, Lehrstuhl für Public und Nonprofit Management, Universität Potsdam.
- Dr. *Jann Meyer-Abich*, ehem. Präsident des Rechnungshofs der Freien und Hansestadt Hamburg.
- Dr. *Markus Müller*, Honorarprofessor für Politik- und Verwaltungswissenschaften an der Zeppelin Universität Friedrichshafen, Abteilungsleiter im Ministerium für Finanzen und Wirtschaft Baden-Württemberg.
- Dr. *Heinz Pansi*, Bezirkshauptmann, Bezirkshauptmannschaft Hermagor, Österreich.
- Dr. *Margrit Seckelmann*, Regierungsdirektorin, Geschäftsführerin, Deutsches Forschungsinstitut für öffentliche Verwaltung, Speyer.
- Dr. *Benedikt Speer*, Professor für Verwaltungsmanagement, MMag., Fachhochschule Kärnten, Villach.
- Dr. *Sylvia Veit*, Wissenschaftliche Mitarbeiterin, Lehrstuhl für Politikwissenschaft, Verwaltung und Organisation II, Universität Potsdam.
- Dr. *Antoinette Weibel*, Universitätsprofessorin, Lehrstuhl für Verwaltungswissenschaft, insbesondere Managementlehre Fachbereich Politik- und Verwaltungswissenschaften, Universität Konstanz.

