

In: Kuhlmann, Sabine/ Bogumil, Jörg/ Wollmann, Hellmut (Hrsg.): Leistungsmessung und -vergleich in Politik und Verwaltung. Wiesbaden: VS Verlag für Sozialwissenschaften. 2004, S. 94-120.

Sabine Kuhlmann

Interkommunaler Leistungsvergleich in Deutschland: Zwischen Transparenzgebot und Politikprozess

1 Hintergrund und Problemaufriss

Die Messung und Bewertung öffentlichen Handelns über Leistungskennzahlen und -indikatoren ist keine Erfindung des New Public Management (NPM), sondern besitzt in Staat und Verwaltung eine lange Tradition. Versuche zur Messung und Bewertung öffentlicher Aufgabenerfüllung reichen, etwa im Fall von Großbritannien und den USA, bis ins ausgehende 19. Jahrhundert zurück¹ und erlebten mit der Etablierung der Evaluation als Analyse- und Informationsverfahren bei der Politikvorbereitung und Wirkungskontrolle in den 1960er und 70er Jahren international einen bemerkenswerten Aufschwung (Wollmann 2001: 14; ferner Wollmann in diesem Band). Mit dem Aufkommen der NPM-Bewegung zeichnet sich jedoch nunmehr eine Entwicklung ab, wonach Leistungsmessungen² und -vergleiche in mehrfacher Hinsicht an Bedeutung zu gewinnen scheinen (Pollitt/Bouckaert 2000: 87, De Bruijn 2002; Pitarelli/ Monnier 2000: 9 ff.; Wollmann 2003), denn im internationalen Kontext ist zu beobachten, dass:

- Leistungsmessungen, aufbauend auf neuen Informationstechnologien und Verfahren des Finanzmanagements, in einer zunehmend größeren Anzahl von Politikfeldern und Verwaltungsbereichen praktiziert werden³, was auf eine *extensivere* Nutzung entsprechender Verfahren hindeutet,

¹ So wurden beispielsweise in Großbritannien und in den USA schon Ende des 19. Jahrhunderts Kennzahlensysteme angewendet, um die Leistung von Lehrern in staatlichen Schulen zu messen (Pollitt/Bouckaert 2000: 87).

² Leistungsmessung (*Performance Measurement*) wird hier als Sammelbegriff für die Messung der Leistungsfähigkeit öffentlichen Verwaltungshandelns über Kennzahlensysteme und Performanzindikatoren (*performance indicators – PIs*) verwendet. Die Messung schließt dabei sowohl Input- und Output- als auch Outcome-Größen ein (vgl. Nullmeier 2001: 384). Zur Unterscheidung von *Performance Measurement* und *Performance Management* sowie innerhalb des Letzteren zwischen *internen* und *externen* Varianten siehe Wollmann in diesem Band.

³ Als Beispielländer, in denen die Ausweitung von Performance Measurement auf neue Politikbereiche zu beobachten war, wären hier zu nennen USA (obligatorische Verwendung von PIs in allen Bundesbehörden auf der Grundlage des Government Performance and Results Act), Australien (wo bereits seit 1980 und verstärkt noch einmal seit dem Antritt der neo-konservativen Howard-Regierung PI-Systeme breite Anwendung finden; vgl. Halligan 2003), Neuseeland (umfassende Anwendung von PI-Systemen im Rahmen der Strategic Results Areas und Key Results Areas, vgl. Halligan 2003), Niederlande (enge Verknüpfung von Performance Measurement und Budgetprozess; vgl. Leeuw 2003), Großbritannien und Frankreich (wo Performance Measurement jeweils in engem Zusammenhang mit Qualitätsmanagement praktiziert wurde – etwa UK Citizen's Charter von 1991 und Public Service Charter in Frankreich; vgl. Pollitt/Bouckaert 2000: 87f.).

- Leistungsmessungen über ihre (traditionelle) *monitoring*- Funktion hinaus mittlerweile verstärkt als Steuerungsinstrumente in verschiedenen Entscheidungsbereichen eingesetzt werden⁴, woraus sich eine *intensivere* Nutzung solcher Verfahren ablesen lässt,
- Leistungsmessungen nicht mehr ausschließlich der Verbesserung der *verwaltungsinternen* Steuerungsfähigkeit dienen, sondern in zunehmend herangezogen werden, um politische Mandatsträger, Verwaltungsklienten und die lokale Öffentlichkeit über Verwaltungsleistungen zu informieren⁵, was sich als „*Externalisierung*“ von Performance Measurement beschreiben ließe.

Vor dem Hintergrund dieser sich international abzeichnenden Entwicklung soll im folgenden Beitrag versucht werden, die aktuell in *Deutschland* zu beobachtende Reformbewegung des interkommunalen Leistungsvergleichs zu verorten, zu analysieren und zu bewerten. Dabei wird in vier Schritten vorgegangen

- Zunächst sollen die *Akteure und Arenen*, die den interkommunalen Leistungsvergleich in Deutschland bestimmen, identifiziert und die *Steuerung* von Kennzahlenvergleichen näher betrachtet werden (Abschnitt 2)
- Sodann wird der Versuch unternommen, die *Ergebnisse* und – soweit ersichtlich – *Wirkungen* des interkommunalen Leistungsvergleichs abzuschätzen, wobei auf die in der Evaluationsforschung gebräuchliche Unterscheidung zwischen institutionellen Veränderungen und Performanzveränderungen zurückgegriffen wird (Abschnitt 3)
- Schließlich soll angesichts dessen, dass interkommunale Vergleichsprojekte zunehmend auf Mobilisierungs- und Umsetzungsprobleme stoßen, aber auch an konzeptionellen Mängeln leiden, nach den Grenzen und Barrieren von Leistungsmessungen in Politik und Verwaltung gefragt werden (Abschnitt 4), wobei zwischen *Konzipierungs- und Methodenproblemen* (4.1), *Implementationskosten* (4.2) und Besonderheiten *politischer Rationalität* (4.3) unterschieden wird.
- Indem abschließend noch einmal Rekurs auf die oben skizzierten Thesen einer *Extensivierung*, *Intensivierung* und *Externalisierung* von Leistungsmessungen im öffentlichen Sektor (Pollitt/Bouckaert 2000: 87) genommen wird, soll der interkommunale Leistungsvergleich in Deutschland einer resümierenden Bewertung unterzogen und sollen Perspektiven für die Zukunft aufgezeigt werden (Abschnitt 5).

2 „Vergleichslandschaft“ und Steuerung von Leistungsvergleichen in Deutschland

Gemessen an der Zahl der seit Beginn der 90er Jahre ins Leben gerufenen Projekte, die auf Leistungsmessungen und -vergleiche mittels Kennzahlen und Indikatoren abzielen, und angesichts der großen Anzahl von Kommunen, die mittlerweile an entsprechenden Projekten beteiligt ist, scheint auf den ersten Blick kein Zweifel daran zu bestehen, dass sich der interkommunale Leistungsvergleich als Reformmethode etabliert hat. Nachdem die Bertelsmann Stiftung 1990 mit ihrem Projekt „Grundlagen einer leistungsfähigen

⁴ Beispielsweise bauen nicht nur substanzielle Politikentscheidungen in verschiedenen Policy-Feldern, sondern vielfach auch Entscheidungen über personalpolitische Fragen (Beförderung usw.) oder Budgetallokationen auf Leistungsmessungen anhand von PI-Systemen auf.

⁵ So enthalten die amtlichen Veröffentlichungen heute deutlich mehr *Performance*-Informationen als dies noch in den 80er Jahren der Fall war (beispielhaft hier die jährlich erscheinenden Next Steps agencies reviews (UK), die National Health Service performance indicators booklets (UK), die GPRA reports (USA), die Berichte des Schwedischen Finanzministeriums über die Produktivität des Öffentlichen Sektors (Swedish Ministry of Finance 1997)). Teilweise genießen bestimmte *Performance*-Informationen, wie beispielsweise die britischen „league tables“ der staatlichen Schulen, eine außerordentlich hohe Medienpräsenz und entsprechende politische Aufmerksamkeit (und – teils – Überbewertung, vgl. Pollitt/Bouckaert 2000: 89f.).

Kommunalverwaltung“ den Startschuss gegeben hatte, (siehe Tabelle 1; vgl. auch Schuster 2001, 2003), folgte die KGSt 1996 mit der Gründung des als „internes Informationssystem der Kommunen“ angelegten IKO-Netzes. Dieses umfasste im Jahr 2001 immerhin 4% aller westdeutschen Kommunen und, bezogen auf das Jahr 2002, bundesweit 748 Kommunen (einschließlich Doppelmitgliedschaften) in 73 Vergleichsringen (Kuhlmann 2003a; vgl. auch Korte in diesem Band)⁶. Weitere Impulse kamen wiederum von der Bertelsmann Stiftung, die im Jahr 2000 die Vergleichsprojekte „kik“ und „kompass“ gründete (Tabelle 1; vgl. auch Tebbe in diesem Band). Neben und teils – z.B. in Bayern und im Saarland - als bewusste Alternative zu diesen überregionalen Vergleichsprojekten der KGSt und der Bertelsmann Stiftung haben sich seit Mitte/ Ende der 90er Jahre zudem in fast allen Bundesländern interkommunale Vergleichsprojekte etabliert⁷, die teils von den kommunalen Landesverbänden moderiert, mitunter auch (temporär) vom Land bezuschusst (Bsp. Bayern, Nordrhein-Westfalen, Sachsen, Schleswig-Holstein), teils in der Eigenregie der Kommunen durchgeführt werden (vgl. unten Tabelle 2). Diese landesspezifischen Vergleichsprojekte, die in einigen Fällen (z.B. Bayern, Sachsen und Schleswig-Holstein) das Ergebnis von Erfahrungsaustausch und Konzepttransfer zwischen den kommunalen Landesverbänden sind, umfassen nach einer ersten Schätzung bereits knapp ein Drittel der deutschen Kreisverwaltungen. In den beiden untenstehenden Tabellen⁸ wird ein knapper Überblick über

⁶ Insgesamt sind seit der Gründung des IKO-Netzes (1996 bis 2002) 127 Vergleichsringe gebildet worden und haben 651 Kommunen an Vergleichsringen teilgenommen, wobei diese Zahlen auch jene Vergleichsringe und Kommunen enthalten, die die Vergleichsarbeit nicht fortgesetzt haben (IKO-Netz der KGSt, Rundbrief 1/2003: 1).

⁷ Lediglich in den Bundesländern Hessen, Rheinland-Pfalz und Thüringen waren zum Erhebungszeitpunkt (Ende 2003) keine landesspezifischen Vergleichsprojekte außerhalb von KGSt und Bertelsmann Stiftung vorzufinden.

⁸ Die Tabellen 1 und 2 basieren auf folgenden Quellen: Kienbaum Unternehmensberatung GmbH 1996; Nordrhein-Westfälischer Städte- und Gemeindebund 1996; Adamaschek 1998; Innenministerium Baden-Württemberg 2001; Consens_ GmbH Hamburg 2002; Niedersächsischer Städte- und Gemeindebund 2002;

die wichtigsten überregionalen sowie länderspezifischen Projekte in der Bundesrepublik gegeben, die sich mit interkommunalen Leistungsvergleichen befassen.

Niedersächsischer Landkreistag 2002; Sächsischer Landkreistag/Sächsischer Städte- und Gemeindebund 2003; Kuhlmann 2003a (m.w.N.) sowie Telefoninterviews mit Vertretern der kommunalen Landesverbände vom 11.2.2003 (Niedersachsen - SGB, Sachsen); 8.7.2003 (Brandenburg); 16.12.2003 (Bayern, Sachsen-Anhalt); 17.12.2003 (Nordrhein-Westfalen, Saarland, Schleswig-Holstein); 18.12.2003 (Rheinland-Pfalz, Niedersachsen - LKT) und kommunalen Akteuren vom 9.4.2002 (Stadt Fürstenwalde), 10.4. 2002 (Stadt Hennigsdorf), 15.1.2003 (Stadt Düsseldorf), 29.1.2003 (Stadt Heidelberg), 10.2.2003 (Stadt Kamen), ferner: <http://www.bay-landkreistag.de/lkt/mitteilungen/m/150103d.htm>; <http://www.consens-info.de>; <http://www.kik-net.de>; <http://www.kompass-modellkommunen.de>. Unbeschadet der somit für die Bundesrepublik flächendeckend vorliegenden Rechercheergebnisse wird nicht der Anspruch auf Vollständigkeit erhoben. Beispielsweise enthalten die Tabellen nicht die Vergleichsprojekte, die unter dem Dach von kommunalen Fachverbänden (z.B. Verband kommunaler Unternehmen, Verband kommunale Abfallwirtschaft, Deutscher Volkshochschulverband, Abwassertechnische Vereinigung) durchgeführt werden. Ferner sei darauf hingewiesen, dass in der Darstellung „reine“ Qualitätswettbewerbe bzw. Qualitätsvergleiche (bspw. nach dem CAF-Modell, vgl. Löffler 2002; Van Dooren/Van de Walle 2002; Saatweber in diesem Band) keine Berücksichtigung gefunden haben. Auch sind nicht alle von den kommunalen Landesverbänden moderierten Vergleichsprojekte enthalten (z.B. Landkreistag Baden-Württemberg).

Tabelle 1 : Bundesweite/ Länderübergreifende Projekte interkommunaler Leistungsvergleiche in Deutschland (Stand: 2003)

Vergleichsprojekt	Trägerinstitution/ Initiative	Start (Ende)	Zahl der Vergleichskommunen	Zahl d. Vergleichs- ringe	Zahl der Aufgaben- bereiche	Anzahl definierter Kennzahlen	Vergleichs- ebene
„Grundlagen einer leistungsfähigen Kommunalverwaltung“	Bertelsmann Stiftung; Deutscher Beamtenbund	1992 (1998)	„Kernring“ aus 5 Städten ⁹ ; weitere 150 Teilnehmer-kommunen	ca. 30	8	Bsp.: Kennzahlenset Sozialwesen mit 221 Kennzahlen (für 16 Produkte)	Produkte
IKO-Netz	KGSt	1996	Ca. 400 Kommunen (2001)	73 (2002)	25 (2002)	Bsp.: Kennzahlenset Sozialhilfe mit 80 Kennzahlen (für 2 Produkte)	Produkte
kik	Bertelsmann Stiftung	2000 (vorauss. 2004)	Ca. 50 Kommunen	20 (2003)	9	Bsp.: Kennzahlenset Sozialhilfe mit 14 Kennzahlen	Produkte/ teils Wirkungen
kompass	Bertelsmann Stiftung	2000 (vorauss. 2004)	8 Kommunen	1 (2003)	5	Bsp.: „Erste Steuerungsebene“: 140 Kennzahlen in 6 Zielfeldern	Wirkungen
Benchmarking für soziale Dienstleistungen	Sozialämter/ überörtl. Sozialhilfeträger und con_sens Consulting	1995/1999 ¹⁰	84 deutsche Kommunalverwaltungen ¹¹	5 (2003)	2 (Arbeitsmarkt- und Sozialpolitik)	Kennzahlenset „Hilfe zum Lebensunterhalt“ mit ca. 40 Kennzahlen (zzgl. Unterkennzahlen)	Produkte/ Wirkungen / teils Prozesse

⁹ Zwischenzeitlich waren noch 3 weitere Städte (Böblingen, Potsdam, Dessau) hinzugekommen, die aber bis 1997 alle den Vergleichsring wieder verließen.

¹⁰ Zunächst wurde das Vergleichsprojekt der 16 großen Großstädte unter der Moderation der Kienbaum Unternehmensberatung GmbH 1995 gegründet, welches 1999 von con_sens Consulting übernommen wurde. Seit diesem Zeitpunkt kamen weitere Benchmarking-Projekte hinzu.

¹¹ Hier ist auch der (landesspezifische) Benchmarking-Kreis der 18 Städte in NRW enthalten.

Tabelle 2 : Länderspezifische Projekte interkommunaler Leistungsvergleiche in Deutschland (Stand: 2003)

Bundesland	Vergleichsprojekt	Trägerinstitution/ Initiative/ Begleitung	Start (Ende)	Zahl der Vergleichs- kommunen	Zahl d. Vergleichs- ringe	Zahl der Aufgaben- bereiche
Baden- Württemberg	AG Produktkennzahlen/ Vergleichsringe Baden-Württemberg	Kommunen und Städtetag Baden-Württemberg	2002	Ca. 40 Kommunen	10 (2003)	10
Bayern	Innovationsring Bayerische Landkreise	Bayerischer Landkreistag (Startphase vom Freistaat Bayer bezuschusst)	1996	10 Landkreise	1	3 (7 Produkte)
Brandenburg	Vergleichsring „Kreisfreie Städte Brandenburg“	IKO-Netz der KGSt und Städte- und Gemeindenbund Brandenburg	2003	4 Kreisfreie Städte	1	8
Mecklenburg- Vorpommern	Controlling im Aktionsprogramm zur Integration von Sozialhilfeempfängern in den allgemeinen Arbeitsmarkt	Land MV (Versorgungsamt Rostock); Landkreise/ kreisfreie Städte; Trägergesellschaft des Landes mbH (Controlling-Durchführung)	2001	12 Landkreise, 6 kreisfreie Städte	1	1
Niedersachsen	(1) Niedersächsischer Kosten- und Leistungsvergleich	Niedersächsischer Städte- und Gemeindebund/ Gemeinden/ Diekwisch Consulting	1994	28 kleinere/ mittlere Gemeinden (2002)	1	12
	(2) Kennzahlenvergleiche der Niedersächsischen Landkreise	Landkreistag Niedersachsen	1999	37 Landkreise und Region Hannover	keine	36
Nordrhein- Westfalen	Nordrhein-Westfälischer Kosten- und Leistungsvergleich ¹²	Nordrhein-Westfälischer Städte- und Gemeindebund/ Gemeinden/ Diekwisch Consulting	1995-96	24 Gemeinden	1	12
Saarland	Interkommunale Vergleichsringe Saarland	Landkreistag Saarland	1999	6 Landkreise	6 (1 VR je Bereich)	6
Sachsen	Innovationsringe Sachsen	Sächsischer Landkreistag/ Sächsischer Städte- und Gemeindetag (80% Zuschuss vom Land)	2002	19 Kommunen	5	5
Sachsen- Anhalt	Vergleichsring Bauordnungsämter ¹³	Landkreistag Sachsen-Anhalt; Landkreise	2002-03	21 Landkreise	1	1 (6 Produkte)
Schleswig- Holstein	Innovationsring Schleswig-Holstein (Leistungsvergleich noch in Vorbereitung)	Landkreistag Schleswig-Holstein; Landkreise (Zuschuss vom Land)	2002	11 Landkreise	1	3

¹² Mangels Resonanz bei den Kommunen (nur 24 von 358 Gemeinden hatten sich beteiligt) wurde der Leistungsvergleich vom nordrhein-westfälischen Städte- und Gemeindebund in den Folgejahren nicht wieder angeboten.

¹³ Die vom Landkreistag Sachsen-Anhalt moderierten interkommunalen Vergleichsprojekte laufen jeweils ein Jahr und beziehen sich auf wechselnde Aufgabenfelder. So gab es in den Jahren vor 2002 bereits Vergleichsprojekte in den Bereichen KfZ und Kitas, an denen ebenfalls alle 21 Landkreise beteiligt waren.

Aus den Tabellen wird ersichtlich, dass wir es bei der Etablierung von Performance Measurement und Leistungsvergleichen, wie generell bei der Modernisierung des Öffentlichen Sektors in Deutschland, nicht mit einem zentralen federführenden Akteur, sondern mit einer *Vielzahl von Akteuren und Arenen* zu tun haben, worin die Vielgestaltigkeit des dezentral-fragmentierten Politik- und Verwaltungssystems der Bundesrepublik ihre Widerspiegelung findet (vgl. Wollmann/Schröter 1997: 188; siehe auch Wollmann in diesem Band). Die Vergleichsprojekte werden – durchweg auf freiwilliger Basis - von unterschiedlichen Akteuren, zu unterschiedlichen Zeitpunkten, mit unterschiedlichen Instrumentierungen sowie teilweise gezielt in lokaler Eigenregie und Abgrenzung zu staatlichen Kontrollinstanzen (Fachaufsicht) ins Leben gerufen¹⁴. Sie stellen im Kern eine *Selbstorganisation der Kommunen* dar, in der sich die deutsche Politik- und Verfassungstradition der kommunalen Selbstverwaltung widerspiegelt. Diesem dezentral-fragmentierten Vorgehen können als markante Gegenbeispiele die zentralistisch veranlassten und weithin als staatliche Kontrollinstrumente genutzten Leistungsvergleiche unter den britischen Kommunen gegenüber gestellt werden, die seit den 80er Jahren unter Federführung der eigens zu diesem Zweck eingerichteten Nationalen Audit Commission durchgeführt werden (vgl. Bovaird/Löffler 2001; ferner Wollmann sowie Wegener in diesem Band)¹⁵. Dagegen lassen sich in den skandinavischen Ländern, aber auch in den USA¹⁶ eher solche Leistungsvergleiche vorfinden, die dezentral initiiert und durchgeführt werden und damit eher Ähnlichkeiten mit den im deutschen Kontext gebräuchlichen, auf Freiwilligkeit und Selbstorganisation beruhenden bottom up-Strategien aufweisen (vgl. Strid sowie Askim in diesem Band).

Ähnlich wie in den skandinavischen Ländern kommt auch in Deutschland den kommunalverbandlichen Akteuren eine besondere „Promotorenrolle“ bei der Etablierung von Leistungsvergleichen zu. Insbesondere die KGSt, aber auch die kommunalen Landesverbände und andere Non Profit Organisationen (vor allem der Bertelsmann-Stiftung; vgl. Tebbe in diesem Band) haben die Verbreitung von Performanzindikatoren und Vergleichssystemen maßgeblich vorangetrieben. Darüber hinaus konnten sich auch zunehmend private Beratungsunternehmen im Bereich interkommunaler Leistungsvergleiche etablieren und sich entsprechende „Marktsegmente“ erschließen. Dagegen nehmen die Landesregierungen derzeit noch überwiegend eine lediglich (finanziell) unterstützende, zum Teil moderierende Funktion im Hinblick auf lokale Vergleichsprojekte wahr (z.B. Innenministerien von Baden-Württemberg und Sachsen). Neuerdings gibt es jedoch vermehrt Hinweise darauf, dass die Landesregierungen im Verhältnis zu den Kommunen zunehmend verwaltungspolitisch aktiv und auch gesetzgeberisch tätig werden, was etwa aus den Gesetzesänderungen im Haushaltsrecht, die sich in einigen Bundesländern (z.B. NRW, Hessen) in der Vorbereitung befinden, ablesbar ist. Diese im Zeichen der Haushaltskrise verfolgten Novellierungen könnten in den kommenden Jahren bewirken, dass der traditionelle Input-Haushalt durch den

¹⁴ Beispielsweise wurde vom Landkreistag in Sachsen-Anhalt das bewusst eigenständige und von der Landesregierung (auch finanziell) unabhängige Vorgehen im interkommunalen Leistungsvergleich damit begründet, dass man sich nicht von den Aufsichtsbehörden abhängig machen und kontrollieren lassen möchte (Telefoninterview vom 16.12.2003).

¹⁵ Allerdings gibt es neuerdings auch unter den britischen Kommunen in Zusammenarbeit mit ihrem Verband „Local Government Association“ Versuche, dem zentralistischen Vorgehen der Regierung durch die Bildung eigener „benchmarking clubs“ entgegenzutreten (vgl. Wollmann 2004 sowie in diesem Band).

¹⁶ Das 1995 gestartete US-amerikanische *Comparative Performance Measurement Program* wurde – zunächst 44 Städte und Counties mit jeweils über 200.000 Einwohnern umfassend – von der International City/County Management Association (ICMA), einer Selbstorganisation der Gemeinden, ins Leben gerufen. Seit 1997 nehmen auch eine Reihe mittlerer und kleinerer Gebietskörperschaften am Projekt teil (vgl. *International City/County Management Association – ICMA, About the ICMA Center for Performance Management*, Internet-Seite: <http://www.icma.org/abouticma/programs/performance/about.cfm>, 1999; zitiert nach Schuster 2001a: 213).

Produkthaushalt abgelöst wird (hierzu Budäus 2004) und sich die Kommunen auf dieser Grundlage zunehmend veranlasst, wenn nicht in die Pflicht genommen sehen, Leistungsvergleiche auf Produktbasis durchzuführen¹⁷.

3 Ergebnisse und Wirkungen von Leistungsvergleichen

Um die Ergebnisse und den Nutzen von Leistungsmessungen und -vergleichen zu ermitteln, empfiehlt es sich, auf die in der Wirkungsforschung gebräuchliche Unterscheidung von „Institutionen-“ und „Performanzevaluation“ zurückzugreifen, die insbesondere auf Fragen der Institutionen- und Verwaltung(reform)politik angewendet wird. Während bei der „Institutionenevaluation“ der Implementationsstand bestimmter Reforminstrumente – in unserem Fall des Leistungsvergleichs und Benchmarkings - im Mittelpunkt steht und – methodisch gesprochen – als abhängige Variable konzipiert wird, dreht sich die „Performanzevaluation“ um die Frage, ob und wie sich diese neu etablierten Verfahren auf das Verwaltungshandeln (*performance*) auswirken bzw. ob und inwieweit Veränderungen im Verwaltungshandeln – als abhängige Variablen – auf die besagten institutionellen und instrumentellen Innovationen zurückzuführen sind (Wollmann 2000: S. 200f.).

3.1 Institutionelle und instrumentelle Veränderungen

Was die Frage nach den *institutionellen und instrumentellen Veränderungen* anbelangt, so fällt als Erstes auf, dass die interkommunalen Vergleichsprojekte in einer Vielzahl von Kommunen den Anstoß dafür gegeben haben, Verfahren und Instrumente zur laufenden Beobachtung und „(Selbst)Evaluation“ des Verwaltungshandelns zu etablieren, durch die die Entwicklung, Institutionalisierung und Verbreitung von Leistungsindikatoren wesentlich gefördert worden ist¹⁸. Eine Vielzahl kommunaler Akteure hat sich im Verlauf der Vergleichsprojekte mit Methoden und Verfahren strukturierter Datenerhebung und -auswertung vertraut machen können, die nunmehr einen Platz im „institutionellen Gedächtnis“ der Verwaltungen hat. Vor dem Hintergrund der vielfach, nicht zuletzt von Seiten der Wissenschaft, geäußerten Kritik am „Evaluationsdefizit“ in der öffentlichen Verwaltung können diese Aktivitäten der Kommunen durchaus als ein bemerkenswerter Schritt in Richtung auf die Institutionalisierung einer „Evaluationsfunktion“ angesehen werden (Wollmann 2000: 212ff.).

Darüber hinaus hat es im Kontext von Leistungsvergleichen vielfältige – wenn auch überwiegend kleinteilig-inkrementelle - Veränderungen in den *Organisations-, Ablauf-* und zum Teil *Personalstrukturen* gegeben¹⁹, die von der Schnittstellenoptimierung über die Verbesserung der IT-Ausstattung bis hin zur Umstellung der Personalplanung und Mitarbeiterschulung reichen. Durch den interkommunalen Erfahrungsaustausch in den Vergleichsringen wurde zudem die Verbreitung des *Bürgeramtsmodells* gefördert, was daraus ersichtlich wird, dass sich viele Kommunen, die am Leistungsvergleich mitwirkten, aufgrund

¹⁷ So etwa die Prognose des (z. Zt. noch nicht im Leistungsvergleich aktiven) Thüringischen Gemeinde- und Städtebundes, wonach die bevorstehenden landesrechtlichen Normierungen im kommunalen Haushaltsrecht einen klaren Impuls für produktbezogene interkommunale Leistungsvergleiche mit sich bringen werden (Telefoninterview vom 16.12.2003)

¹⁸ Im IKO-Netz der KGSt lagen Anfang 2003 über 50 Kennzahlensysteme zu den verschiedensten kommunalen Aufgabenbereichen vor. Vergleichswerte auf der Basis der Kennzahlensysteme waren zu Beginn des Jahres 2002 für mehr als 10 IKO-Netz-Vergleichsringe über die IKON-Vergleichsdatenbank abrufbar.

¹⁹ Eindeutig NPM-bezogene umfassende Strukturveränderungen, etwa durch Hierarchieabflachung und Dezentralisierung, sind dagegen nur in Ausnahmefällen als Reaktion auf Vergleichsaktivitäten zu beobachten (Schuster 2003).

der positiven Erfahrungen der anderen beteiligten Kommunen, ebenfalls entschlossen, Bürgerämter einzurichten²⁰.

Die Durchführung von Vergleichsprojekten hat den Aufbau *interkommunaler Netzwerke*, die – neben indikatorengestützten („harten“) Leistungsvergleichen – einen qualitativen („weichen“) Erfahrungsaustausch zwischen den kommunalen Akteuren ermöglichen, begünstigt und vorangetrieben. Der Austausch von „Anwenderwissen“ über sektoral organisierte Netzwerke stellt einen wesentlichen, wenn nicht – aus Sicht der Akteure – überhaupt den wichtigsten Nutzen interkommunaler Leistungsvergleiche dar²¹. Gerade im Zusammenhang mit der aktuellen – unter dem Stichwort „Governance“ (vgl. Benz u.a. 2003; Klenk/Nullmeier 2003) firmierenden - Diskussion um neue Formen von Selbstorganisation durch Netzwerkbildung, die jenseits von Markt und Hierarchie angesiedelt sind, verdienen diese Bemühungen der Kommunen, sich zu Vergleichsringen und Netzwerken zusammenzuschließen als innovative Form von *local governance* besondere Beachtung (Heinelt 2003).

Hat der interkommunale Leistungsvergleich damit einerseits durchaus für Bewegung auf der lokalen Ebene in Deutschland gesorgt, bleibt andererseits festzuhalten, dass er bislang nur selten auf „echtes“ Benchmarking hinausläuft und sich überwiegend als „reiner Zahlenvergleich“ darstellt. Bekanntlich besteht die Idee des Benchmarking-Ansatzes vor allem darin, auf der Basis von „Betriebsvergleichen“ *best practice*-Lösungen zu identifizieren und damit gerade die Unterschiede zwischen Organisationen für institutionelles Lernen nutzbar zu machen. Im derzeitigen Leistungsvergleich der deutschen Kommunen geht es jedoch eher darum, möglichst einheitliche Organisationsstrukturen und normierte Standards für Leistungsmessungen auf der Ebene von Produkten (Outputs) und Ressourcen (Inputs) zu entwickeln²². Da organisatorische oder prozessuale Unterschiede einheitliche Messungen und standardisierte Zahlenvergleiche erschweren (siehe hierzu auch Abschnitt 4.1), werden sie in den derzeit praktizierten Leistungsvergleichen eher als hemmend und störend angesehen. Außerdem fällt auf, dass ungeachtet der beachtlichen Fortschritte, die die Evaluationsforschung in den 80er Jahren hinsichtlich der Indikatorenbildung und Wirkungsmessung erzielen konnte (vgl. etwa Hellstern/Wollmann 1984) im interkommunalen Leistungsvergleich auf diese Vorarbeiten kaum zurückgegriffen wird und stattdessen neuere Reformansätze oftmals hinter den Ergebnissen früherer Anläufe zurückbleiben

3.2 Auswirkungen auf das Verwaltungshandeln

Es scheint auf der Hand zu liegen, dass interkommunale Leistungsvergleiche letztlich darauf zielen (sollten), die Leistungsfähigkeit (*performance*) lokaler Politik und Verwaltung zu verbessern. Umso überraschender ist vor diesem Hintergrund der Befund, dass sich nur

²⁰ Mit Blick auf das Bertelsmann-Grundlagen-Projekt war sogar von einer „regelrechten Welle von Neugründungen solcher Bürgerbüros“ (z.B. in Borken, Celle, Herford, Pforzheim, Werne) die Rede (Adamaschek 1997: 121). Mittlerweile verfügen in Nordrhein-Westfalen 85% der Städte ab 20.000 Einwohner und in Baden-Württemberg 78% der Städte über Bürgerämter (vgl. Bogumil u.a. 2003). Allerdings sei darauf hingewiesen, dass der interkommunale Leistungsvergleich die Etablierung von Bürgerbüros vermutlich lediglich gefördert, nicht aber also solche ausgelöst hat.

²¹ Zu dieser Vermutung verhält sich auch die Feststellung konsistent, dass sich eine überwältigende Mehrheit (80%) der in einer IKO-Netz Kundenbefragung angeschriebenen Teilnehmer-Kommunen von der Mitarbeit in den Vergleichsringen eine Unterstützung beim interkommunalen *Erfahrungsaustausch* erwartet und in diesem das zentrale Teilnahmemotiv erblickt (siehe Korte in diesem Band).

²² Eine bemerkenswerte Ausnahme bildet hier das von der Fa. Con_sens moderierte Projekt, in dem explizit das Benchmarking – und nicht (nur) der quantitative „Zahlenvergleich“ - im Vordergrund steht (vgl. Hollenrieder in diesem Band).

wenige der beteiligten Kommunen tatsächlich aus dem Leistungsvergleich „Performanzverbesserungen“ versprechen²³. Insbesondere diejenigen Verwaltungen, die mittlerweile aus den Vergleichsprojekten ausgestiegen sind, bemängeln, dass bislang signifikante Veränderungen und ein sichtbarer Nutzen der Kennzahlenvergleiche auf der Ebene des praktischen Verwaltungshandelns kaum zu erkennen seien, so dass der damit verbundene Aufwand als nicht gerechtfertigt erscheint (hierzu unten Abschnitt 4.2) Versucht man, sich anhand der vorliegenden empirischen Hinweise ein präziseres Bild über die Auswirkungen des interkommunalen Leistungsvergleichs auf das Verwaltungshandeln und die Leistungsfähigkeit von Politik und Verwaltung zu verschaffen, so könnten die folgenden „Performanzfelder“ unterschieden werden (vgl. auch Bogumil/Kuhlmann 2004; Kuhlmann 2004):

Input-Veränderungen (Kosteneinsparungen, Effizienzgewinne): Einerseits wird vielfach konstatiert, dass das Kostenbewusstsein in den Fachämtern deutlich gestiegen sei, was vor allem darauf zurückgeführt wird, dass Kosten und Leistungen transparenter sind und auf der Basis vorliegender Vergleichswerte überhaupt erst sinnvoll bewertet werden können. Auch gibt es einige Hinweise darauf, dass die am interkommunalen Leistungsvergleich beteiligten Kommunen gewisse Kostenersparnisse zu verzeichnen haben²⁴. Andererseits fällt bei näherem Hinsehen auf, dass die konstatierten Effizienzgewinne in den Teilnehmerkommunen häufig aus einmaligen, punktuell anfallenden Mehreinnahmen resultieren, die kaum geeignet scheinen, eine längerfristige und nachhaltige Steigerung der Wirtschaftlichkeit des Verwaltungshandelns auszulösen. Auch sei daran erinnert, dass bislang, soweit ersichtlich, nicht der Versuch unternommen worden ist, die mit der Konzipierung, Implementation, Fortschreibung und Evaluation von Leistungsvergleichen verbundenen *Transaktions- und Opportunitätskosten*²⁵ (vgl. Abschnitt 4.2) zu messen und ins Verhältnis zum erreichten Nutzen zu setzen. Es kann plausibel argumentiert werden, dass eine solche Analyse die reformbezogene Kosten-Nutzen-Bilanz, vor allem in Hinsicht auf Effizienzkriterien, erheblich verschieben, resp. verschlechtern würde.

Output- und Prozessveränderungen (Servicequalität, Kundenfreundlichkeit): Zum einen lässt sich zeigen, dass es bei den an interkommunalen Leistungsvergleichen beteiligten Kommunalverwaltungen zu messbaren Verbesserungen im Serviceniveau gekommen ist (vgl. Schuster 2001: 216 ff.), die partiell auf die Anreizwirkung (Wettbewerbssurrogat) der Vergleichsprojekte zurückzuführen sind. So waren in den fünf Städten des „Kernrings“ der Einwohnermeldeämter im Bertelsmann-Projekt sichtbare Niveauverbesserungen in den Leistungsparametern „Öffnungszeiten“ und „Wartezeiten“ zu verzeichnen. Zwischen 1995 und 1997 waren die Wartezeiten in den beteiligten Kommunen dieses „Kernringes“ durchweg rückläufig, und 1997, also kurz vor Abschluss des Projekts, gab es in keiner Kommune mehr

²³ Ausweislich einer IKO-Netz-Kundenbefragung aus dem Jahr 2003, in der 697 teilnehmende Kommunen angeschrieben wurden (Rücklaufquote 51,5%) versprechen sich nur 20% der „IKO-Netz-Kunden“ Verbesserungen in der Wirtschaftlichkeit, 8% eine Steigerung der Bürgerzufriedenheit, 12% eine Verbesserung der Organisation und gerade 3% eine Erhöhung der Mitarbeiterzufriedenheit (vgl. Korte in diesem Band).

²⁴ Als Beispiele aus dem „Bertelsmann-Vergleichsprojekt“ wären etwa zu nennen: Mehreinnahmen von jährlich 5.600 DM in der Stadt Pforzheim durch Aufstellung eines Passbildautomaten im Bürgerzentrum; Einsparung von zwei Stellen durch Auflösung des Schreibbüros im Landkreis Osnabrück; Reduzierung des Unterhaltungsaufwandes bei Spielplätzen in der Stadt Gütersloh durch Festsetzung einer Mindestgröße bei Spielplätzen und Verkauf der kleinen Grundstücke von „Minispielplätzen“ (vgl. Bertelsmann Stiftung <http://www.leistungsvergleich.de/info/2.htm>; Recherche vom 9.7.2002).

²⁵ Opportunitätskosten fallen dadurch an, dass die Verwaltungsbediensteten für den Leistungsvergleich Arbeitszeit aufwenden müssen, während der sie nicht für andere Aufgaben zur Verfügung stehen (vgl. Pollitt 1995: 139).

Wartezeiten von über 10 Minuten²⁶. Andererseits zeigte sich, dass Städte mit mittlerer bis guter Platzierung eher geneigt sind, bei ihrem bisherigen Serviceniveau zu bleiben (Bsp. Castrop-Rauxel, Gütersloh), so dass mithin eine beachtliche „Unbeweglichkeit im Mittelfeld“ (ebd.: 220) festzustellen war. Im Bertelsmann-Grundlagen-Projekt wurde phasenweise – etwa im Bereich der Wartezeiten – sogar eine Verschlechterung des Leistungsniveaus beobachtet und zudem konstatiert, dass die „objektiven“ Serviceverbesserungen zum Teil nicht mit der subjektiven Kundenzufriedenheit korrelierten²⁷. Darüber hinaus gibt es Hinweise darauf, dass Wirtschaftlichkeitsgewinne und Einsparungen (etwa durch Stellenabbau) häufig auf Kosten der Leistungsqualität „erkauft“ werden²⁸. Dies dürfte nicht zuletzt damit zusammenhängen, dass sich der Leistungsvergleich aufgrund des Bestrebens der lokalen Akteure, die aktuelle Verwaltungsmodernisierung für Haushaltseinsparungen zu nutzen, überwiegend auf quantifizierbare und monetarisierbare Größen, wie z.B. Bearbeitungsdauer, bezieht, wohingegen inhaltliche, policy-spezifische Qualitätsaspekte ausgeblendet bleiben. Mit Blick auf die rechtsstaatlich noch wenig geübten ostdeutschen Verwaltungen scheint zudem der Umstand problematisch, dass die rechtliche Qualität des Verwaltungshandelns als Performanzkriterium (rechtliche Korrektheit der Entscheidungen, „Prozessanfälligkeit“ usw.) im Zusammenhang mit interkommunalen Leistungsvergleichen kaum thematisiert wird (vgl. Kuhlmann 2003b).

Outcomes und System-Veränderungen (politische Steuerung, Mitarbeitermotivation): Die „Erfolgsbilanz“ des interkommunalen Leistungsvergleichs im Hinblick auf die Forderung, Kosten-, Leistungs- und Wirkungsinformationen für eine verbesserte (evidenz-basierte) politisch-strategische Steuerung nutzbar zu machen, muss in mehrfacher Hinsicht besonders kritisch ausfallen. Zum einen fällt auf, dass Leistungsvergleiche fast ausschließlich aus der Verwaltung heraus initiiert und von der Verwaltung durchgeführt werden. Auch die Nutzung von Vergleichsinformationen erfolgt überwiegend *binnenadministrativ*. Nennenswerte Versuche, neben den Verwaltungsakteuren auch die Kommunalpolitiker oder gar die lokale Bürgerschaft und gesellschaftliche Akteure in Vergleichsprojekte einzubeziehen, sind lediglich in Anfängen zu verzeichnen und zudem mit erheblichen Mobilisierungsproblemen verbunden (hierzu unten Abschnitt 4.3)²⁹. Zum andern ist festzustellen, dass sich kommunale Leistungsmessungen und -vergleiche derzeit fast ausschließlich auf der Ebene von „Produkten“ (Outputs) bzw. der für deren Erstellung erforderlichen *Ressourcen* (Inputs) bewegen, wohingegen die aus Sicht der Kommunalpolitiker interessanten *Wirkungen* des Verwaltungshandelns (*Outcomes*, z.B. Nachhaltigkeit, Verteilungsgerechtigkeit, Lebensqualität usw.) kaum eine Rolle spielen (vgl. auch Brüggemeier sowie Reichard in diesem Band)³⁰. Vor diesem Hintergrund bieten Vergleichsdaten bislang nur wenig

²⁶ Siehe auch die vielfältigen Städtebeispiele bei Adamaschek (1998) und unter <http://www.leistungsvergleich.de/info/2.htm>.

²⁷ So verschlechterte sich in Pforzheim im Zeitraum 1994 bis 1996 die Kundenzufriedenheit ungeachtet dessen, dass in dieser Zeit die Öffnungszeiten von 26,5 auf 28,5 Wochenstunden erweitert worden waren (Schuster 2003).

²⁸ Beispielhaft hierfür sei die Stadt Kamenz erwähnt, die zur Verringerung des Zeitaufwands bei Eheschließungen den personellen Betreuungsaufwand verringerte und auf die Nutzung von Musterreden bei Trauungen zurückgegriffen hat (vgl. Adamaschek 1998: 102).

²⁹ Das von der Bertelsmann Stiftung im Jahr 2000 initiierte Projekt Kompass (vgl. Tabelle 1), in dem es darum geht, den Erfolg und die Auswirkungen kommunalpolitischen Handelns auf die „Lebensqualität“ in der Kommune zu bestimmen, bildet hier ein Ausnahmbeispiel. In diesem Projekt werden die Kommunalpolitik und lokale Bürgerschaft - sei es durch Befragungen, sei es auch durch institutionelle Einbindung - in den Prozess der (strategischen) Zielbildung und Indikatorenentwicklung einbezogen.

³⁰ Diese Feststellung deckt sich auch mit den internationalen Beobachtungen zu Performance Measurement und Programmevaluation. So resümieren Pollitt/Bouckaert in ihrer ländervergleichenden NPM-Evaluation: „(...) the availability of such outcome measures, confidently linked to programme interventions, is the exception rather than the rule. At the level of broad programmes of management reform we know of not a single study

Anknüpfungspunkte für Entscheidungen im Rahmen eines strategischen Controllings oder einer wirkungsorientierten Steuerung, da diese die Messung von Wirkungen anhand von Outcome-Indikatoren voraussetzen. Wie aus der Evaluationsforschung bekannt, liegt dieses Defizit nicht zuletzt darin begründet, dass Wirkungsanalysen auf besondere methodische Konzipierungs-, Zuordnungs- und Kausalitätsprobleme stoßen (vgl. Wollmann 2000), da Outcome-Werte zumeist „multikausale“ Größen sind, bei deren Zustandekommen der kommunale Beitrag nicht eindeutig festzustellen ist. Da der Ausbau der institutionellen Architektur von Berichtswesen und Controlling damit vor allem der Stärkung der inneradministrativen Informations- und Steuerungskapazität gedient hat, während die Kommunalvertretungen davon weitgehend abgeschnitten blieben (Wollmann 2001: 33), hat sich im Ergebnis die Handlungsschwäche der Kommunalvertretung gegenüber der Verwaltung eher verschärft und haben die (politisch-parlamentarischen) Steuerungsdefizite noch zugenommen.

Nicht-intendierte Effekte und „Performanzverschlechterungen“ hat der interkommunale Leistungsvergleich auch im Bewertungskriterium der Mitarbeiterzufriedenheit hervorgebracht. Zunehmend werden kritische Stimmen laut, die insbesondere auf gravierende *Motivationseinbrüche* bei den Mitarbeitern aufmerksam machen³¹. Da mit den Kennzahlenmessungen und Fortschreibungen von Vergleichsdaten eine erhebliche, zusätzlich zum Tagesgeschäft zu leistende Mehrarbeit anfällt, ist die Motivation der Beschäftigten und ihre Bereitschaft, entsprechende Vorhaben zu unterstützen, vor allem in den kommunalen Fachämtern markant rückläufig, deren Arbeitsvolumen ohnedies schon die verfügbaren knappen Ressourcen übersteigt (etwa Sozialämter). Dieser Motivationsverlust wird dadurch noch zusätzlich befördert, dass sich die einzelnen Mitarbeiter durch die Aufschreibungen kontrolliert fühlen (Schuster 2001: 217) und mithin von „formalem Hierarchie-Druck“ die Rede ist, da die Arbeit in Vergleichsprojekten vielfach intern top-down gesteuert und auf eine breite Einbindung und Informierung der Mitarbeiter, die die Leistungserhebungen durchführen müssen, verzichtet wird (Anke u.a. 1999: 43)³². Vor diesem Hintergrund verwundert es wenig, dass vielerorts eher eine rückläufige denn steigende Mitarbeiterzufriedenheit festzustellen ist³³, zumal auch eine Reihe von Maßnahmen, die auf Effizienzsteigerung (z.B. durch Personaleinsparungen) und Erhöhung der Kundenzufriedenheit (z.B. durch Verlängerung der Öffnungszeiten) gerichtet sind, fast naturgemäß in einem Spannungsverhältnis zum Ziel der Mitarbeiterorientierung stehen.

4. Probleme und Grenzen von Leistungsvergleichen in Politik und Verwaltung

Haben interkommunale Leistungsvergleiche einerseits, wie die offenkundige Resonanz, auf die die Vergleichsprojekte in den Kommunen stoßen, es nahelegt, mittlerweile durchaus ihren Platz in der deutschen Reformlandschaft, gibt es andererseits eine Reihe von Hinweisen, die zu eher skeptischen Prognosen hinsichtlich der längerfristigen Aufrechterhaltung und Fortführung solcher Vergleichsvorhaben Anlass geben.

from our ten countries that convincingly links the actions taken with a set of positive and safely attributable final outcomes“ (Pollitt/Bouckaert 2000: 111).

³¹ Siehe zur Problematik der Beschäftigtenbeteiligung auch Herbing in diesem Band.

³² In der Stadtverwaltung Vauberg wurden im Ergebnis einer zweijährigen Projektlaufzeit (Bertelsmann-Grundlagen-Projekt) „tiefe Enttäuschung“, „Burn out-Symptome“ und durchweg negative Erwartungen der Mitarbeiter bezüglich einer Weiterführung des Leistungsvergleichs festgestellt (Anke u.a. 1999: 43).

³³ Beispielhaft hier die „Bertelsmann-Kommune“ Gütersloh, in der sich die anfänglich hohe Mitarbeiterzufriedenheit (Schulnote 2,08 im Jahr 1993) im Laufe des Vergleichsprojekts drastisch verschlechterte (Schulnote 4,6 im Jahr 1997, vgl. Schuster 2003).

Zum einen wird das auf den ersten Blick ermutigende Bild des interkommunalen Leistungsvergleichs in Deutschland dadurch getrübt, dass die Bereitschaft der Kommunen, an Vergleichsprojekten teilzunehmen bei weitem nicht flächendeckend ist und markante regionale Unterschiede, insbesondere im Ost-West-Vergleich erkennen lässt. Dies wird besonders deutlich, wenn man den Anteil *ostdeutscher Kommunen* bzw. Vergleichsringe (VR) an der Gesamtzahl der Teilnehmer im IKO-Netz der KGSt betrachtet, der die begrenzte, um nicht zu sagen fast fehlende Nachfrage der ostdeutschen Kommunalverwaltungen an einer Mitwirkung im IKO-Netz widerspiegelt. Während Anfang des Jahres 2000 noch immerhin 5% (=4 VR) aller (insgesamt 79) IKO-Netz Vergleichsringe von ostdeutschen Kommunen gebildet worden waren, sank ihr Anteil bis Ende 2002 auf 3%, was einer Anzahl von nur noch 2 Vergleichsringen, die im Land Brandenburg ansässig sind, entspricht (vgl. Kuhlmann 2003a m.w.N.). Von den Ende 2001 insgesamt 672 am IKO-Netz beteiligten Kommunen (einschließlich Doppelmitgliedschaften) waren ganze 94% aus Westdeutschland (=633 Kommunen), wohingegen nur 39 Kommunen, also 6%, ostdeutscher Herkunft waren (ebd.). Dies weist schlaglichtartig auf die Probleme und Hemmnisse bei der Initiierung und Institutionalisierung des interkommunalen Leistungsvergleichs in den neuen Bundesländern hin.

Zum andern deutet manches darauf hin, dass sich auch die Bereitschaft, den interkommunalen Leistungsvergleich längerfristig aufrechtzuerhalten und fortzuschreiben vielfach in Grenzen hält. Wiederum sind es vor allem die ostdeutschen Verwaltungen, die sich besonders häufig von den Vergleichsprojekten verabschieden³⁴. Aber nicht nur im IKO-Netz dürfte die Quote der „*Aussteiger*“, die nach Ablauf der „Probezeit“ die Projekte verlassen, beachtlich sein, sondern auch die von den kommunalen Landesverbänden ins Leben gerufenen Vergleichsprojekte treffen auf zum Teil erhebliche Mobilisierungsprobleme und es ist vielerorts bereits vom „Einschlafen“ des interkommunalen Leistungsvergleichs die Rede³⁵. Führt man sich die Umsetzungsprobleme und Kosten, die (vor allem auf lange Sicht) mit interkommunalen Leistungsmessungen und -vergleichen verbunden sind, so lassen sich eine Reihe von Erklärungen für diese Entwicklung finden.

4.1 Konzipierungs- und Datenprobleme

Eine mögliche Erklärung für die Umsetzungsschwierigkeiten, auf die die Vergleichsprojekte in der kommunalen Praxis stoßen, dürfte in den – generell für Evaluationen typischen – Konzipierungs- und Datenproblemen zu sehen sein, welche bei der Ziel- und Indikatorenbildung sowie bei der Erhebung und beim Vergleich der Daten auftreten. Zum einen besteht bei interkommunalen Vergleichsprojekten die Schwierigkeit, *einheitliche Kennzahlensysteme* für eine Vielzahl von institutionell sehr heterogenen Verwaltungseinheiten zu definieren. So erschweren unterschiedliche Ämter- und

³⁴ Beispielhaft hier der Austritt von 13 thüringischen Landkreisen und 4 brandenburgischen Städten aus dem IKO-Netz. Zwischen März 2000 und Dezember 2001 reduzierte sich in mehr als der Hälfte der im IKO-Netz vertretenen kommunalen Aufgabenbereiche (19 von 34) die Zahl der teilnehmenden Kommunen, so dass es in dieser Zeitspanne zu einem Gesamtrückgang von 11% (84 Kommunen) kam (Kuhlmann 2003a m.w.N.).

³⁵ So wörtlich die Einschätzung des Geschäftsführenden Direktors beim Landkreistag Rheinland-Pfalz (Telefoninterview vom 16.12.2003). Ähnlich auch die Einschätzung des Hessischen Städtetages, nach dessen Auskunft der überwiegende Teil der Städte, die sich noch bis Ende der 90er Jahre in Vergleichsringen engagiert hatten, mittlerweile aus den entsprechenden Projekten ausgetreten ist, so dass die nach wie vor im IKO-Netz aktiven hessischen Kommunen eher Ausnahmefälle darstellen (Telefoninterview vom 16.12.2003). Auch in Nordrhein-Westfalen wurde das vom Städte- und Gemeindebund (unter Mitwirkung von Diekwisch Consulting) koordinierte interkommunale Vergleichsprojekt mangels Interesses auf Seiten der Städte und Gemeinden 1996 eingestellt (Telefoninterview vom 17.12.2003).

Sachgebietsstrukturen und Zuständigkeiten in den einzelnen Aufgabenbereichen, aber auch der Umstand, dass die beteiligten Kommunen unterschiedlich weit fortgeschritten sind, was die Umsetzung von Modernisierungsmaßnahmen, etwa im Zusammenhang mit der Einführung des Neuen Steuerungsmodells, anbelangt, die Festsetzung einheitlicher Kennzahlensets³⁶. Vor diesem Hintergrund sind die Bemühungen der lokalen Akteure – häufig bestärkt durch ihre externen Projektmoderatoren – vor allem darauf gerichtet, die Vergleichbarkeit von Kennzahlen und Messwerten sicherzustellen, indem Leistungen standardisiert und Messverfahren und -instrumente so weit wie möglich vereinheitlicht werden (siehe oben Abschnitt 3.1)³⁷. Mittlerweile mehren sich vor diesem Hintergrund die kritischen Stimmen, welche dieser gegenüber lokalen Besonderheiten und Problemstrukturen geradezu „blinden“ Neigung zur Vereinheitlichung und Standardisierung die Forderung entgegenstellen, solche Unterschiede systematisch zu berücksichtigen und produktiv zu nutzen.

Betrachtet man die „frühen“ Kennzahlensets der Bertelsmann Stiftung und der KGSt (vgl. oben Tabelle 1), so stechen die hohe Zahl von Kennzahlen und Subkennzahlen, der insgesamt beachtliche Detaillierungsgrad und die tiefgegliederte Struktur der Indikatorensysteme besonders ins Auge, worin sich das Bemühen der Akteure widerspiegelt, ein möglichst vollständiges Bild kommunalen Handelns mittels Kennzahlen wiederzugeben. Diese Kennzahlensets erwiesen sich für Steuerungszwecke in der Verwaltung als kaum geeignet, geschweige denn taugten sie als Informationsbasis für politische Entscheidungen. So ist man mittlerweile in einigen Vergleichsprojekten auch dazu übergegangen, anstelle einer Myriade von steuerungsungeeigneten Kennzahlen und Subkennzahlen nur noch einige aussagefähige und steuerungsrelevante „Kernkennzahlen“ zu definieren und eher einfache „*abgespeckte*“ Kennzahlensets zu verwenden (z. B. kik-Projekt; vgl. Tebbe, ferner Reichard in diesem Band)³⁸.

Als letzter Punkt sei noch erwähnt, dass vielfach auch die Datenerhebung und -aggregation selbst auf Schwierigkeiten stößt, so dass der Prozess der Kennzahlenmessung und -überprüfung häufig für nicht gesichert gehalten wurde (Adamaschek 1997). Da die Daten oft in Eile von Einzelpersonen zusammengetragen, scheinbar gleiche Kennzahlen vollkommen unterschiedlich erhoben und auffällige Nachlässigkeiten bei der Datenerhebung festgestellt worden sind (Schuster 2003), seien mithin erhebliche *Zweifel an der Datenqualität* angebracht. Als Konsequenz ging man (z. B. im Bertelsmann-Grundlagen-Projekt) bisweilen dazu über, verstärkte interne und externe Qualitätskontrollen der Daten – als eine Art Meta-Performance Measurement – durchzuführen, was natürlich vor allem mit Blick auf die sich dadurch noch einmal potenzierenden Transaktions-, Opportunitäts- und Kontrollkosten (vgl. hierzu den folgenden Abschnitt 4.2) als besonders prekär erscheinen muss.

³⁶ Beispielsweise steht die Definition einheitlicher Kostenkennzahlen dann vor Problemen, wenn einige Kommunen des Vergleichsrings bereits die Kosten- und Leistungsrechnung eingeführt haben, während andere noch nach „klassischer“ Kameralistik arbeiten. Auch sehen manche Akteure länderübergreifende Kennzahlenvergleiche angesichts zu unterschiedlicher institutionell-rechtlicher Voraussetzungen von vornherein zum Scheitern verurteilt (bspw. Landkreistag Saarland, Telefoninterview vom 17.12.2003).

³⁷ Beispielsweise war aus Nordrhein-Westfalen zu hören, dass vorrangig mit dem Ziel, Verwaltungseinheiten und deren Leistungen vergleichen zu können, die gesamte Organisation der Landesarchive vereinheitlicht werden soll, was sachlichen und inhaltlichen Erwägungen vollkommen entgegensteht. Zum Leistungsvergleich auf Länderebene (Baden-Württemberg) vgl. auch Seyfried in diesem Band.

³⁸ Während im Bertelsmann-Grundlagen-Projekt im Bereich des Sozialwesens noch mit 221 Kennzahlen operiert worden ist, stützt sich das Bertelsmann-kik-Projekt zur Leistungsmessung in der Sozialhilfe nur noch auf 14 Schlüsselkennzahlen (siehe Tabelle 1).

4.2 Implementations- und Transaktionskosten

Aus einer institutionenökonomischen Sicht dürfte die mancherorts begrenzte Motivation zur Teilnahme am interkommunalen Leistungsvergleich, aber vor allem auch die im Verlauf der Vergleichsprojekte häufig schwindende Bereitschaft zur dauerhaften Fortführung der Kennzahlenmessungen daraus zu erklären sein, dass die mit der Konzipierung, Implementierung und längerfristigen Fortschreibung der Indikatorensysteme und Vergleichsdaten verbundenen Kosten den – noch weitgehend als ungesichert erscheinenden – Gewinn dieser Unternehmungen übersteigen. Kurz gesagt, ist es aus Sicht vieler kommunaler Akteure zu einer deutlichen Schieflage im Kosten-Nutzen-Verhältnis interkommunaler Leistungsvergleiche gekommen, die vielfach den Anlass dazu gegeben hat, sich von den Vergleichsprojekten zu verabschieden³⁹.

Der „Löwenanteil“ an den Gesamtaufwendungen, die für Leistungsmessungen und -vergleiche erforderlich sind, entfällt ohne Zweifel auf die *Personalkosten*⁴⁰, die sich aus den arbeitszeitlichen Aufwendungen der Verwaltungsbeschäftigten für die Konzipierung und Implementierung der Kennzahlensysteme, für arbeitstägliche Aufschreibungen zu Bearbeitungs- und Durchlaufzeiten sowie für das Führen von (häufig manuellen) Strichlisten zur Zahl bearbeiteter Anträge oder antragsbezogener Verrichtungen usw. ergeben⁴¹. Aus Sicht der Kommunen erscheinen dabei, abgesehen von den zunächst einmalig anfallenden Einführungs- und Implementationskosten, vor allem diejenigen Aufwendungen als problematisch, die dauerhaft, d.h. bei einer längerfristigen Unterhaltung und Aktualisierung der Datenbestände und laufenden (teils unterjährigen) Berichterstattung entstehen und die auf längere Sicht personelle und zeitliche Kapazitäten binden. Zwar wird in diesem Zusammenhang von einigen Projektträgern argumentiert, dass sich der zu leistende Aufwand durch technische Unterstützung und „Digitalisierung“ beträchtlich vermindern werde und ohnedies als Bestandteil eines später mit der Einführung von Kosten- und Leistungsrechnung, Berichtswesen und Controlling zu erbringenden Arbeitsaufwandes zu sehen sei (Korte/Pook 2002: 17). Diese Argumentation lässt aber zum einen außer Acht, dass die technische Unterstützung von Kennzahlenmessungen in den Verwaltungen noch bei Weitem keine Selbstverständlichkeit ist und von einer automatisierten Fortschreibung wohl strukturierter Datenbestände „per Mausklick“ in vielen Kommunen noch keine Rede sein kann. Zum anderen wird übersehen, dass der interkommunale Leistungsvergleich eben nur zum Teil – und nach Einschätzung vieler Kommunalakteure in einem viel zu geringen Ausmaße – für das interne Controlling der einzelnen Verwaltungen genutzt werden kann, also in der Regel einen beachtlichen zusätzlichen Aufwand darstellt⁴². Gerade den ostdeutschen Kommunen, die

³⁹ Ein Projektteilnehmer des IKO-Netzes formulierte dieses Problem zugespitzt mit den Worten: „*Es war viel Aufwand und wir haben nichts geschafft*“ (Telefoninterview vom 15.1.2003, vgl. FN 8).

⁴⁰ Außerdem sind die Beraterhonorare (Bsp. IKO-Netz: 1.500 bis 5.000 EUR für bis zu 2 Jahre Projektlaufzeit) sowie die Kosten für neue IT-Anschaffungen (Bsp. IKO-Netz: jährlich 200 bis 1.300 EUR) und Kommunikationslizenzen (Bsp. IKO-Netz: jährlich 200 bis 1.300 EUR) zu berücksichtigen (vgl. Kuhlmann 2003a m.w.N.).

⁴¹ Beispielhaft geschätzt durch eine befragte Fachbereichsleiterin (vgl. FN 8) einer mittelgroßen Stadt (27.000 Einwohner) mit einem zehn Mitarbeiter umfassenden Sozialamt, das sich über zwei Jahre am Leistungsvergleich der Sozialämter im KGSt IKO-Netz beteiligte, liegen diese „Opportunitätskosten“ (berechnet für die zweijährige Projektdauer) bei ca. 45.500 EUR, d.h. bei zwei Jahresgehältern für eine Vollzeitstelle im mittleren bis gehobenen Verwaltungsdienst.

⁴² So die ausgesprochen kritische Einschätzung von Talkenberg (Interviewprotokoll vom 19.3.1996), der seinerzeit Projektverantwortlicher im Bertelsmann-Grundlagen-Projekt in der Stadtverwaltung Potsdam war; ähnlich aber auch die Einschätzungen der Fachbereichsleiterin Soziales in der Stadt Hennigsdorf mit Blick auf den IKO-Netz-Vergleichsring „Sozialämter in Brandenburg“ und des Projektverantwortlichen der Stadt Düsseldorf im IKO-Netz Vergleichsring „Stadtplanung Großstädte“ (vgl. FN 8).

zusätzliche Ressourcen und Handlungsspielräume (*organizational slack*) für die Vergleichsarbeit kaum mobilisieren können, erscheint der interkommunale Leistungsvergleich vor diesem Hintergrund häufig eher als ein exotisches Instrument, für das im Verwaltungsalltag kein Platz ist.

4.3 Politische Rationalität

Weiter oben wurde bereits darauf hingewiesen, dass Benchmarking und interkommunale Leistungsvergleiche als Verfahren interner Selbstbewertung und -kontrolle in den weiteren Zusammenhang von Evaluation und Wirkungsanalyse einzuordnen sind (vgl. Pitarelli/Monnier 2000: 9 ff.; Wollmann 2000 sowie in diesem Band). Berücksichtigt man neben den oben skizzierten methodisch-konzeptionellen und durchführungsbezogenen „Fallstricken“ die Handlungsrationalität in konkurrenzdemokratisch verfassten Systemen, so müssen sich Leistungsmessungen und –vergleiche vor allem der politischen Frage stellen, welche Akteure, unter welchen Bedingungen und aus welchen Motiven heraus ein Interesse an Erfolgskontrollen und Wirkungstransparenz haben könnten. Bei der Beantwortung dieser Frage stehen sich bis heute zwei widerstreitende Positionen gegenüber:

Aus einer (eher „pessimistischen“) Perspektive, die – den Prämissen der ökonomischen Theorie der Politik folgend (vgl. Downs 1957) – das Streben der politischen Akteure nach Stimmenmaximierung und Machterhalt in den Mittelpunkt rückt, kann es im politischen Raum kaum ein Interesse an „objektiven“ Erfolgskontrollen geben, da diese das Risiko mit sich bringen, dass nach außen hin sichtbar wird, um wieviel und aus welchen Gründen ein gesetztes Ziel verfehlt worden ist (Bogumil 2003: 22). Dieser Position zufolge handeln Politiker, vor allem unter wettbewerbsdemokratischen Bedingungen, ähnlich wie Unternehmen und sind in erster Linie an „Marktsicherung“ interessiert, während inhaltliche Politikziele oder eine Orientierung am Gemeinwohl nachrangig sind. Einer solchen „ökonomischen“ Handlungsrationalität entspricht es auch, politische Ziele allgemein und unpräzise zu formulieren, um sich Spielräume für politische Manöver zu erhalten und dem politischen Gegner gegenüber möglichst wenig Angriffsfläche zu bieten. Die Chancen, vermittels Leistungs- und Wirkungsmessungen zu einem präzisen Bild des Zielerreichungs- und Wirkungsgrades politisch-administrativen Handelns zu gelangen, erscheinen vor diesem Hintergrund eher begrenzt (vgl. auch Brüggemeier sowie Longo/ Plamper am Beispiel von Italien in diesem Band), was sich in den beschriebenen Mobilisierungs- und Implementationsproblemen, auf die entsprechende Initiativen im kommunalen Raum treffen, widerspiegelt. Dass sich Kennzahlenvergleiche und Leistungsmessungen dennoch in vielen Bereichen als Reforminstrumente etabliert haben, erklärt sich dieser Argumentation zufolge vor allem aus dem Interesse der Politikakteure nach vorzeigbaren Politikerfolgen (Reichard 2001: 84) und daraus, dass sie als „Kennzahlenshow“ (vgl. Nullmeier in diesem Band) genutzt werden können, um bestimmte Politikstrategien öffentlichkeitswirksam zu rechtfertigen und symbolische Politik zu betreiben⁴³. Soweit jedoch schlechte

⁴³ Ein anschauliches Beispiel für eine solche Strategie bietet die aktuelle Diskussion um Leistungsmessungen und -vergleiche in Berlin. Dort erklärt sich das derzeit erstaunlich hohe (politische) Interesse an Vergleichsdaten vor allem daraus, dass mit diesen nach außen hin nachgewiesen, dass Berlin im Hinblick auf relevante Leistungsindikatoren und Ressourcenparameter zumindest nicht schlechter als andere Städte abschneidet und dass alles unternommen worden ist, um aus eigener Kraft aus der Haushaltsnotlage herauszukommen. Dieser Nachweis wird als die Voraussetzung dafür angesehen, dass der Bund (im Ergebnis der Verfassungsklage) dem Land Berlin zukünftig bei der Bewältigung der Haushaltskrise finanziell unter die Arme greift (Expertengespräch mit Manfred Passutti, Leiter der Projektgruppe „Modernisierung der Verwaltung“ bei der Senatsverwaltung für Finanzen Berlin, Sitzungsprotokoll des Projektseminars „Verwaltungsmodernisierung in Berlin“ vom 28.1.2003).

„Erfolgsbilanzen“ in Aussicht stehen, die die Wiederwahl gefährden könnten, ist den lokalen Mandatsträgern, insbesondere der Mehrheitsfraktionen, und leitenden Beamten aus nachvollziehbaren Gründen kaum an Leistungs- und Wirkungstransparenz gelegen (Banner 2001)⁴⁴.

In diese Argumentation fügt sich auch die Beobachtung ein, dass politisches Handeln nicht strikt ziel- und wirkungsorientiert, strategisch vorausschauend und nach einer rational-umfassenden Methode mit klaren Zielen, Bewertungskriterien und Ergebnistransparenz erfolgt, sondern eher einem iterativen, schrittweise probierenden Vorgehen des „Sich Durchwurstelns“ entspricht (Lindblom 1975: 162), das von den (kontingenten) Handlungskalkülen der Akteure abhängig ist. Anstelle einheitlicher Bewertungsmaßstäbe können die Auffassungen relevanter Akteure über Ziele und Bewertungskriterien stark divergieren⁴⁵; die Zielauswahl erfolgt oft erst im administrativen Entscheidungsprozess (Stichwort: Behörde als „Clearingstelle“), das Bestreben, politischen Konsens zu finden, ist häufig wichtiger als klare politische Ziele zu formulieren; ad hoc-Anpassungen an situative Problemlagen sind aus Sicht der Politikakteure in der Regel „rationaler“ sind als langfristige Planungen usw. usf. Das teilweise zu beobachtende Scheitern von Leistungs- und Wirkungsmessungen erklärt sich vor diesem Hintergrund aus den Umsetzungsproblemen, auf die das „rationale Politikmodell“ mit seinem Anspruch auf „Evidenzbasierung“, Planung und Steuerung stößt⁴⁶.

Dagegen lehnt sich eine andere – eher „optimistische“ – Position stärker an Argumente der vergleichenden Policy-Forschung, insbesondere der Parteiendifferenztheorie, an und erklärt ein grundsätzliches Interesse der Politikakteure an Wirkungskontrollen aus ihrem Streben nach inhaltlichen (Policy-)Problemlösungen und Programmumsetzung (Schmidt 1995: 580 ff.). Aus dieser Sicht haben die Politikakteure durchaus ein Interesse an Ergebnistransparenz vor allem deshalb, weil sie ihr Handeln (auch) mit programmbezogenen Verpflichtungen und der Durchsetzung *inhaltlicher Politikziele* verbinden, die ja gerade eine unerlässliche Voraussetzung für die Wiederwahl sind. Denn erst, wenn das Tun und Lassen eines Amtsinhabers den Erwartungen seiner Klientel näherungsweise gerecht wird, scheint seine Wiederwahl gesichert. Eine solche Orientierung an bestimmten „Policy-Werten“ und inhaltlichen Politikzielen vorausgesetzt, dürften die politischen Amtsträger ein ureigenes Interesse an Informationen über den Zielerreichungs- und Wirkungsgrad ihres Handelns hegen, nicht zuletzt, um Programmentscheidungen ggf. rechtzeitig korrigieren zu können. Diese Politik- und Programmorientierung vorausgesetzt, können sich Leistungsmessungen und -vergleiche gerade im Zuge von Cutback Management und Haushaltskonsolidierung als Alternativen zum „Rasenmäherprinzip“ und als geeignete Instrumente erweisen, um die Ressourcenallokation innerhalb des gegebenen Budgets zu rationalisieren und „evidenzbasiert“ (vgl. Sanderson 2002) zu legitimierbaren Budgetkürzungen zu gelangen (Derlien 1997: 7). Mit anderen Worten: „Programme, die gute Resultate hervorbringen,

⁴⁴ „(...) the politician's prime goal is to be re-elected rather than to respect technical evidence“ (Cook zitiert nach Sanderson 2002).

⁴⁵ Es sei daran erinnert, dass die Bewertungskriterien öffentlichen Handelns, nach denen dieses „gemessen“ und evaluiert werden kann, vielfältig und widerstreitend sind, da im Bereich der öffentlicher Aufgabenerfüllung – im Gegensatz zum privaten Unternehmen – die negativen Folgekosten des Wirtschaftens nicht externalisiert, sondern abgefangen, verarbeitet und internalisiert werden. Für die öffentliche Leistungserbringung bedeutet dies, dass etwa das Bewertungskriterium hoher Effektivität (beispielsweise bei der Prävention von Jugendkriminalität) im Konflikt mit der Forderung nach einem möglichst wirtschaftlichen Aufgabenvollzug stehen kann, ohne dass dieser Konflikt (definitiv) aufgehoben werden kann (vgl. Bogumil 2003).

⁴⁶ „There is nothing a politician likes so little as to be well informed; it makes decision-making so complex and difficult“ (Keynes zitiert in Sanderson 2002: 5).

werden ausgedehnt; diejenigen, die sich als unfruchtbar erweisen, werden fallengelassen oder drastisch modifiziert“ (Weiss 1974: 20).

Die Leistungsmessungen und -vergleiche in den Kommunen müssen sich grundsätzlich beiden „Rationalitätskriterien“ stellen und stehen damit vor der doppelten (und ambivalenten) Herausforderung, einerseits Transparenz in den politischen Entscheidungsprozess zu bringen, an der aus Sicht einer inhaltlichen Politikorientierung auch ein grundsätzliches Interesse besteht. Andererseits müssen sie das Bestreben der politischen Akteure, sich unter den Bedingungen der Konkurrenzdemokratie zu behaupten und damit ihre Neigung im Blick behalten, das Ausmaß an Durchsichtigkeit im Politikprozess nach dessen Nützlichkeit für die eigene politische Profilierung und Wiederwahl auszurichten. Die Diskussionen um die Frage, inwieweit der interkommunale Leistungsvergleich stärker an die Öffentlichkeit gebracht und Vergleichsergebnisse auch nach außen hin transparent gemacht werden (sollten), ist hierfür symptomatisch (vgl. auch Schuster in diesem Band). So besteht einerseits, insbesondere in den Fachbereichen und -ämtern ein fachlich-technisches (Policy-) Interesse an Leistungsdaten und -vergleichen, während andererseits die Anonymisierung solcher Vergleichsdaten (Verzicht auf Namen und Einwohnerzahl der Kommune bei der Berichterstattung) nicht nur gegenüber der Öffentlichkeit, sondern auch gegenüber den lokalen politischen Vertretungsorganen vielfach zur Bedingung für eine Teilnahme am Leistungsvergleich gemacht wird (Bsp. Landkreise Niedersachsen, Saarland, Sachsen-Anhalt). Während man in den Verwaltungen noch (zähneknirschend) bereit ist, Leistungsdaten etwa über Intranet intern oder über ein interkommunales Berichtswesen sektoralbezogen, also nur für die beteiligten Fachämter zugänglich und bekanntzumachen⁴⁷, wird in der Befürchtung, die Vergleichswerte könnten „politisiert“ und in Form von Ranking-Listen medienwirksam „missbraucht“ werden, auf eine generelle Veröffentlichung verzichtet⁴⁸. Gerade vor dem Hintergrund des politischen und budgetären Drucks, die Kennzahlenwerte zur Rechtfertigung weiterer Personalkürzungen zu nutzen, und der sich somit verstärkenden (berechtigten) Angst der Beschäftigten vor dem Arbeitsplatzverlust, erweist sich die Frage der Veröffentlichung von Vergleichsdaten und ihre mögliche politische Funktion im Cutback Management als ein ausgesprochen „wunder Punkt“⁴⁹ der Verwaltungsmodernisierung.

5 Zusammenfassung und Ausblick

Interkommunale Leistungsvergleiche und Selbstbewertungen der öffentlichen Leistungserbringung durch die lokalen politisch-administrativen Akteure sind in Deutschland zweifelsohne zu einem dominanten Instrument der Öffentlichen Sektor Modernisierung avanciert. Sie reißen sich somit in den international zu beobachtenden Trend einer zunehmenden *Extensivierung* von Performance Measurement ein, der sich markant dadurch auszeichnet, dass in einer wachsenden Zahl öffentlicher Verwaltungen und Politiksektoren auf Verfahren von Leistungsmessung und Kennzahlenvergleichen zurückgegriffen wird, um

⁴⁷ So sprach der Vertreter des Saarländischen Landkreistages von einer wahren „Seelenmassage“, der es bedurft hatte, um die Verwaltungsmitarbeiter zu überzeugen, die Vergleichsdaten ins *Intranet* zu stellen (Telefoninterview vom 17.12.2003, vgl. FN 8).

⁴⁸ Beispielsweise haben im Kennzahlenvergleich der Niedersächsischen Landkreise (siehe Tabelle 2) nur die Fachämter in den betreffenden Aufgabenbereichen einen Überblick über die landesweiten Vergleichsdaten, kennen also beispielsweise die Vergleichswerte ihres Nachbarkreises in dem für sie relevanten Aufgabengebiet. Dagegen bekommen Politik und Öffentlichkeit nur anonymisierte Datenbestände zur Verfügung (Telefoninterview vom 18.12.2003; vgl. FN 8).

⁴⁹ So wörtlich der Vertreter des Saarländischen Landkreistages (Telefoninterview vom 17.12.2003); ähnlich der Gesprächspartner aus dem Landkreistag Sachsen-Anhalt (Telefoninterview vom 16.12.2003).

die Ergebnisse und Wirkungen politisch-administrativen Handelns sichtbar, interorganisatorisch vergleichbar und für künftige Politikentscheidungen nutzbar zu machen. Andererseits deuten die wachsenden Umsetzungsprobleme und „Ermüdungserscheinungen“ bei Vergleichsprojekten auch auf die Grenzen dieses Reformansatzes hin. Vor diesem Hintergrund und angesichts der eher begrenzten Anreize für die kommunalen Akteure, sich längerfristig einem nicht-marktlichen Wettbewerb zu unterziehen, wurde bisweilen auch die Forderung erhoben, Leistungsvergleiche und die Veröffentlichung von Vergleichsergebnissen zukünftig rechtsverbindlich vorzuschreiben, etwa durch entsprechende Regelungen in den Kommunalverfassungen (vgl. Banner 2001: 63ff.). Dieser Argumentation könnte jedoch mit Blick auf die (nicht durchweg positiven) ausländischen Erfahrungen entgegengehalten werden, dass im Grundsatz der Freiwilligkeit und Dezentralität gerade eine besondere Stärke der kommunalen Vergleichsprojekte in Deutschland zu erblicken ist. Denn dieser lässt eine vergleichsweise hohe Reliabilität und „Glaubhaftigkeit“ der Daten erwarten, wohingegen beispielsweise für die britische Citizen’s Charter-Initiative und ihre neueren Nachfolgeprojekte (Best Value und Comprehensive Performance Assessment) konstatiert wurde, dass die Kommunen ihre Daten häufig „schönfärben“, wenn nicht „frisieren“, um – zumal bei in Aussicht stehenden positiven oder negativen Sanktionen – von den externen Auditoren möglichst gut benotet zu werden (vgl. Löffler 1998: 75ff.; Wollmann in diesem Band)⁵⁰.

Während eine Reihe von ausländischen Beispielen belegt, dass Leistungsmessungen zunehmend genutzt werden, um die Entscheidungs- und Steuerungsfähigkeit in Politik und Verwaltung zu erhöhen (*Intensivierung*), zeigen sich beim Leistungsvergleich der deutschen Kommunen auffällige Defizite hinsichtlich der Steuerungsfunktion von Leistungsindikatoren und Vergleichsdaten. Bislang sind interkommunale Leistungsvergleiche kaum über die traditionelle Monitoring-Funktion von Performance Measurement hinausgekommen und werden in der Regel nicht für Zwecke einer informierteren politisch-strategischen Steuerung (z.B. Budget- und Programmentscheidungen) eingesetzt. Versuche, Leistungsinformationen für Kontrakte, Zielvereinbarungen oder Leistungsprämien nutzbar zu machen, sind bislang kaum zu erkennen.

Angesichts dessen, dass der interkommunale Leistungsvergleich in Deutschland in erster Linie als „*exekutive Selbstevaluation*“ und interne Leistungseinschätzung für die Verwaltungen auf der operativen Handlungsebene erfolgt, kommt der Informierung der politischen Mandatsträger und der lokalen Bürgerschaft bislang eine nachrangige Bedeutung zu. Im Unterschied zur international zu beobachtenden *Externalisierung* von Performance Measurement, wonach dieses nicht nur zur Verbesserung der verwaltungsinternen Steuerungsfähigkeit, sondern zunehmend eingesetzt wird, um Politiker und Verwaltungsklienten über Verwaltungsleistungen zu informieren, spielen die lokale Bürgerschaft und die Kommunalvertretungen als mögliche Initiatoren und Adressaten von Leistungsmessungen in Deutschland bislang eine ausgesprochen untergeordnete Rolle. Einerseits ist dadurch, dass die Vergleichsdaten für die lokale Öffentlichkeit bislang nicht verbindlich zugänglich gemacht werden, die „Kontrollfunktion“ des Bürgers hinsichtlich der

⁵⁰ Darüber hinaus ist nicht zu übersehen, dass solche Regulationen lokaler Verwaltungspolitik geeignete Instrumente zur staatlichen Kontrolle des kommunalen Politik- und Verwaltungshandelns darstellen und – zumal angesichts der prekären Finanzsituation – genutzt werden können, um die lokalen (insbesondere finanziellen) Handlungsspielräume zu restringieren. Außerdem sei daran erinnert, dass solche Eingriffe „von oben“ auch die in weiten Teilen vorhandene – und im Vergleich zu den anderen föderalen Ebenen sehr ausgeprägte – Reformmotivation und -bereitschaft sowie „Experimentierfreudigkeit“ der Kommunen beeinträchtigen könnten, da Reformmaßnahmen dann als „aufoktroiert“ und als „alien impositions“ (Pollitt/Bouckaert 2000: 48) empfunden werden, was ihre „Überlebensfähigkeit“ von vornherein begrenzt.

Verwaltungsleistungen nach wie vor ausgesprochen gering ausgeprägt. Andererseits kann aber auch von einer Stärkung der kommunal-parlamentarischen Kontrollfunktion durch Leistungsmessungen und -vergleiche derzeit noch keine Rede sein.

Mit Blick auf die Kommunalverfassungsreformen der 90er Jahre ließe sich spekulieren, dass die veränderten politisch-institutionellen Rahmenbedingungen in den deutschen Kommunen, insbesondere die flächendeckende Einführung des direktgewählten Bürgermeisters und die damit einhergehende Ausbildung eines *lokalen Präsidialsystems*, zu einer Stärkung der Evaluationsfunktion im politischen Prozess führen könnten (vgl. am Beispiel der Schweiz Janett 2000). Denn diese „Verfassungskonstellation“ könnte – zumal bei „Kohabitationen“ – bewirken, dass die „klassische“ Gewaltenteilungslinie (Bürgermeister/Verwaltung – Rat) erkennbar wird, wonach die Kommunalvertretung als Ganze ihre Kontrollfunktion wahrnimmt⁵¹. Wirksamkeitsuntersuchungen, Leistungsmessungen und -vergleiche könnten ihr hierfür geeignete Instrumente zur Verfügung stellen. Mit Blick darauf, dass die bisherigen Modernisierungsanläufe auf der lokalen Ebene den Handlungs- und Machtspielraum der Kommunalvertretung fast durchweg geschmälert haben⁵², könnten „parlamentarische Evaluationen“ und Leistungsmessungen eine Möglichkeit bieten, diese Kontroll- und Machtverluste ein Stück weit auszugleichen.

⁵¹ Damit würden die (parlamentarische) Gewaltenteilung zwischen Verwaltungschef und Mehrheitsfraktion auf der einen und „lokal-parlamentarischer“ Minderheit auf der anderen Seite verschwimmen, Parteiprofile „vernebeln“ und der Parteienwettbewerb im lokalen Entscheidungssystem weniger dominant sein.

⁵² Auch durch die Möglichkeiten des Bürgerbegehrens, den Ausbau kooperativer Demokratieformen (Auftraggeberrolle) und die Privatisierung kommunaler Unternehmen hat die Kommunalvertretung Machtverluste erleiden müssen (Bogumil u.a. 2004).

Literatur

- Adamaschek, B., 1997: Leistungssteigerung durch Wettbewerb in deutschen Kommunen – der interkommunale Vergleich. In: Naschold, F./Oppen, M./Wegener, A. (Hrsg.): Innovative Kommunen. Stuttgart, S. 107-124.
- Adamaschek, B. (Hrsg.), 1998: Interkommunaler Leistungsvergleich. 100 spürbare Erfolge. Gütersloh.
- Anke, T./Grabowski, J./Wetzel, R., 1999: Interkommunaler Leistungsvergleich – Kritische Erfolgsfaktoren. In: Adamaschek, B./Baitsch, C. (Hrsg.): Interkommunaler Leistungsvergleich – Kritische Erfolgsfaktoren. Gütersloh, S. 19-88.
- Banner, G., 2001: Politik und Transparenz: Benchmarking als Steuerungsinstrument. In: Hofmeister, A. (Hrsg.): Bewerten – Vergleichen – Gewinnen. Self-Assessment – Benchmarking – Quality Awards. Schriftenreihe der Schweizerischen Gesellschaft für Verwaltungswissenschaften (SGVW), Band 41. Bern, S. 37-70.
- Benz, A., 2003: Einleitung: Governance – Modebegriff oder nützliches sozialwissenschaftliches Konzept? In: Benz, Arthur u.a.: Governance – Eine Einführung. Polis-Heft. Hagen, S. 13-33.
- Bogumil, J., 2003: Politische Rationalität im Modernisierungsprozess. In: Schedler, K./Kettinger, D. (Hrsg.): Modernisieren mit der Politik. Bern u.a., S. 15-42.
- Bogumil, J./Holtkamp, L./Schwarz, G., 2003: Das Reformmodell Bürgerkommune. Leistungen – Grenzen – Perspektiven. Berlin.
- Bogumil, J./Holtkamp, L./Kißler, L., 2004: Modernisierung lokaler Politik – Auswirkungen auf das kommunale Entscheidungssystem. In: Jann, W./Bogumil, J./Bouckaert, G./Budäus, D./Holtkamp, L./Kißler, L./Kuhlmann, S./Reichard, C./Wollmann, H.: Statusreport Verwaltungsreform – eine Zwischenbilanz nach zehn Jahren. Berlin, S. 64-74.
- Bogumil, J./Kuhlmann, S., 2004: 10 Jahre kommunale Verwaltungsmodernisierung. Ansätze einer Wirkungsanalyse. In: Jann, W./Bogumil, J./Bouckaert, G./Budäus, D./Holtkamp, L./Kißler, L./Kuhlmann, S./Reichard, C./Wollmann, H.: Statusreport Verwaltungsreform – eine Zwischenbilanz nach zehn Jahren. Berlin, S. 51-63.
- Bovaird, T./Löffler, E., 2001: Moving from excellence models of local service delivery to benchmarking ‚good local governance‘. In: International Review of Administrative Science. Vol. 68 (2002), S. 9-24.
- Budäus, Dietrich, 2004: Modernisierung des öffentlichen Haushalts- und Rechnungswesens. In: Jann, W./Bogumil, J./Bouckaert, G./Budäus, D./Holtkamp, L./Kißler, L./Kuhlmann, S./Reichard, C./Wollmann, H.: Statusreport Verwaltungsreform – eine Zwischenbilanz nach zehn Jahren. Berlin, S. 75-86.
- Con_sens GmbH Hamburg, 2002: Benchmarking der mittelgroßen Großstädte in Nordrhein-Westfalen. Kennzahlenvergleich 2000. Hilfe zur Arbeit (HzA). Bericht vom 31. Januar 2002.
- De Bruijn, H., 2002: Managing Performance in the Public Sector. London/New York.
- Derlien, U., 1997: Die Entwicklung von Evaluation im internationalen Kontext. In: Bussmann, W./Klöti, U./Knoepfel, P. (Hrsg.): Einführung in die Politikevaluation. Basel/Frankfurt am Main, S. 4-12.

- Downs, A., 1957: *An Economic Theory of Democracy*. New York.
- Halligan, J., 2003: Public-sector reform and evaluation in Australia and New Zealand. In: Wollmann, H. (Ed.): *Evaluation in Public Sector Reform. Concepts and Practice in International Perspective*. Cheltenham etc., S. 56-79.
- Heinelt, H., 2003: Governance auf lokaler Ebene. In: Benz, A. u.a.: *Governance – Eine Einführung*. Polis-Heft. Hagen, S. 33-49.
- Hellstern, G.-M./Wollmann, H. (Hrsg.), 1984: *Evaluierung und Erfolgskontrolle in Kommunalpolitik und -verwaltung*. Basel u.a.
- Innenministerium Baden-Württemberg 2001: *Produkte – Ziele – Kennzahlen. Kommunaler Produktplan Baden-Württemberg*. Schriftenreihe des Innenministeriums Baden-Württemberg, Heft 7. Stuttgart.
- Janett, D., 2000: *Parliamentary Evaluation in Switzerland and its Effects on Legislative and Administrative Action*. Paper presented at the 4th ESS Conference October 12-14th 2000, Lausanne.
- Kienbaum Unternehmensberatung GmbH, 1996: *Projektbericht zum interkommunalen Kennzahlenvergleich für die Hilfe zum Lebensunterhalt*.
- Klenk, T./Nullmeier, F., 2003: *Public Governance als Reformstrategie*. Düsseldorf.
- Korte, R./Pook, M., 2002: *Leistungsvergleiche haben die Bewährungsprobe bestanden. Austausch von Erfahrungen bietet den Kommunen Verbesserungspotentiale*. In: *Innovative Verwaltung* 7-8/2002, S. 14-17.
- Kuhlmann, S., 2003a: *Benchmarking auf dem Prüfstand: Kosten, Nutzen und Wirkungen interkommunaler Leistungsvergleiche in Deutschland*. In: *Verwaltungsarchiv* (2003)1, S. 99-126.
- Kuhlmann, S., 2003b: *Rechtsstaatliches Verwaltungshandeln in Ostdeutschland. Eine Studie zum Gesetzesvollzug in der lokalen Bauverwaltung*. Opladen.
- Kuhlmann, S., 2004: *Evaluation lokaler Verwaltungspolitik: Umsetzung und Wirksamkeit des Neuen Steuerungsmodells in den deutschen Kommunen*. (Erscheint in PVS 3/2004).
- Leeuw, F. L., 2003: *Evaluation and New Public Management in the Netherlands*. In: Wollmann, H. (Ed.): *Evaluation in Public Sector Reform. Concepts and Practice in International Perspective*. Cheltenham etc., S. 104-117
- Lindblom, C., 1975: *Inkrementalismus: die Lehre vom „Sich-Durchwursteln“*. In: Narr, W.-D./Offe, C. (Hrsg.): *Wohlfahrtsstaat und Massenloyalität*. Köln, S. 161-177.
- Löffler, E., 1998: *Verwaltungsmodernisierung im internationalen Vergleich – Messkriterien und Implementationsstrategien in Deutschland, Großbritannien und den USA*. Stuttgart u.a.
- Löffler, E., 2002: *Waiting for Godot: The Contribution of the European Common Assessment Framework to Leadership Development in UK Local Government*. In: *Local Governance*, 2002.
- Niedersächsischer Städte- und Gemeindebund, 2002: *Interkommunaler Kosten- und Leistungsvergleich 2002*. Honnover.
- Niedersächsischer Landkreistag, 2002: *Allgemeine Grundsätze zu Kennzahlenvergleichen*. Stand: Oktober 2002. (ohne Ort)

- Nordrhein-Westfälischer Städte- und Gemeindebund, 1996: Interkommunaler Kosten- und Leistungsvergleich 1996. Berichtsband. (ohne Ort).
- Nullmeier, F., 2001: Kennzahlen und Performance Measurement. In: Blanke, B./Bandemer, S. v./Nullmeier, F./Wewer, G. (Hrsg.): Handbuch zur Verwaltungsreform. 2. Auflage. Opladen, S. 383-391.
- Pitarelli, E./ Monnier, E., 2000: Benchmarking: the missing link between evaluation and management? Paper presented at the 4th ESS Conference October 12-14th 2000, Lausanne.
- Pollitt, C., 1995: Justification by Works or by Faith? Evaluating the New Public Management. In: Evaluation (1995)2, S. 133-154.
- Pollitt, C./Bouckaert, G., 2000: Public Management Reform. A Comparative Analysis. Oxford.
- Reichard, C., 2001: Strategisches Management in der Kernverwaltung. In: Eichhorn, P./Wiechers, M. (Hrsg.): Strategisches Management für Kommunalverwaltungen. Baden-Baden, S. 80-91.
- Sächsischer Landkreistag/Sächsischer Städte- und Gemeindebund, 2003: Kommunale Verwaltungsmodernisierung Sachsen. Abschluss der Projektphase II. (Ohne Ort).
- Sanderson, I., 2002: Evaluation, Policy Learning and Evidence-Based Policy Making. In: Public Administration (2002)1, S. 1-22.
- Schmidt, M.G., 1995: Policy-Analyse. In: Mohr, A. (Hrsg.): Grundzüge der Politikwissenschaft. München/Wien, S. 567-605.
- Schuster, F., 2001: Benchmarking als Ersatz für Wettbewerb. Können interkommunale Leistungsvergleiche ein Motor für Veränderungen sein? In: Edeling, T./Jann, W./Wagner, D. (Hrsg.): Reorganisationsstrategien in Wirtschaft und Verwaltung. Opladen, S. 201-228.
- Schuster, F., 2003: Der interkommunale Leistungsvergleich als Wettbewerbssurrogat. Berlin.
- Van Dooren, W./Van de Walle, S., 2002: Self promotion by self assessment? A study of the motives underlying the use of the Common Assessment Framework (CAF) in Belgian public agencies. In: Local Governance, 2002, S. 235-244.
- Weiss, C. H., 1974: Einführung. In: Küchler, M. (Hrsg.): Evaluierungsforschung. Methoden zur Einschätzung von sozialen Reformprogrammen. Opladen, S. 19-30.
- Wollmann, H., 2000: Evaluierung und Evaluierungsforschung von Verwaltungspolitik und -modernisierung – zwischen Analysepotenzial und -defizit, in: Stockmann, Reinhard (Hrsg.), Evaluationsforschung. Opladen, S. 195-233.
- Wollmann, H., 2001: Zwischen Management- und Politiksystem: Die Kommunalverwaltung der 90er Jahre auf einer Modernisierungswelle. In: Edeling, T./Jann, W./Wagner, D. (Hrsg.): Reorganisationsstrategien in Wirtschaft und Verwaltung. Opladen, S. 15-57.
- Wollmann, H. (Ed.), 2003: Evaluation in Public Sector Reform. Concepts and Practice in International Perspective. Cheltenham etc.
- Wollmann, H., 2004: Reformen der kommunalen Politik- und Verwaltungsebene in Großbritannien, Schweden und Frankreich. Ansätze, Verläufe und Ergebnisse. In: Jann, W./Bogumil, J./Bouckaert, G./Budäus, D./Holtkamp, L./Kißler, L./Kuhlmann, S./Reichard, C./Wollmann, H.: Statusreport Verwaltungsreform – eine Zwischenbilanz nach zehn Jahren. Berlin, S. 36-50.

Wollmann, H./Schröter, E., 1997: Public Sector Reforms in Germany: Whence and Where? A Case of Ambivalence. In: Administrative Studies/ Hailinnon Tutmikus, 1997, no. 3, S. 184-200.