

Benchmarking in föderalen Systemen¹

Alan Fenna

Seit ein paar Jahren kann ein rasch wachsendes Interesse an der Nutzung des Benchmarkings zur Verbesserung der politischen Leistung in föderalen Systeme beobachtet werden. Es handelt sich dabei um eine neue, zweifellos bedeutende Entwicklung, die sich aber noch im Anfangsstadium befindet. Sollte es an klaren Anzeichen dafür gefehlt haben, so änderte sich dies im August 2009 mit der Aufnahme des Artikels 91d in das deutsche Grundgesetz. „Bund und Länder können zur Feststellung und Förderung der Leistungsfähigkeit ihrer Verwaltungen Vergleichsstudien durchführen und die Ergebnisse veröffentlichen.“² Wie weit die gesetzliche bzw. praktische Auswirkung einer solch mehrdeutigen Bestimmung reichen wird, ist nicht klar. Klar ist jedoch ihre symbolische Bedeutung: Benchmarking ist nun ein anerkanntes Instrument des modernen Föderalismus.

Dabei handelt es sich allerdings um alles andere als ein einfaches Thema. Nur ein Jahr nach der Inkraftsetzung von Artikel 91d in Deutschland kündigte die neue Regierung Großbritanniens an, dass sie die *Audit Commission* auflösen werde. Dieses Prüfungsamt hatte in England seit zwei Jahrzehnten die Hauptverantwortung für den Großteil der Leistungskontrollen und des Benchmarking der lokalen Gebietskörperschaften getragen.³ Das Aufeinandertreffen dieser beiden Ereignisse gibt eine Vorstellung davon, wie komplex dieses Thema ist: Wenn

¹ Deutsche Übersetzung von Alan Fenna, *Benchmarking in Federal Systems*, Occasional Paper, Forum of Federations (January 2011).

² Art 91d des Grundgesetz für die Bundesrepublik Deutschland, eingefügt im Rahmen der Änderung vom 1. August 2009.

³ „Eric Pickles to disband Audit Commission in new era of town hall transparency“, Ministry of Communities and Local Government, London 2010. Der Beschluss gilt nur für England. Schottland und Wales haben Ihre eigenen entsprechenden Einrichtungen.

Benchmarking in Großbritannien, wo es ein maßgeblicher Motor der Reformen war, in Frage gestellt wird, was erwartet es in einem föderalen Kontext?

Es treffen in diesem Zusammenhang zwei Dinge aufeinander: eine besondere Staatsform und eine besondere Form der Unternehmensführung. Beide sind an sich bereits komplexe Themen. Die Frage ist, inwieweit sie sich ergänzen können. Wie kompatibel ist Benchmarking mit den Prinzipien des Föderalismus? In welchem Maß kann Benchmarking bestehende föderale Arrangements „aufwerten“, indem es entweder eine überlegene Form der intergouvernementalen Beziehungen oder bessere materielle Resultate für die Bürger ermöglicht? Bei der Beantwortung dieser Fragen muss man zwangsläufig die verschiedenen Formen berücksichtigen, die Föderalismus und Benchmarking annehmen können.

Dieser Abschnitt bietet mittels vier Aussagen einen Überblick über das Problem:

1. Benchmarking ist eine folgerichtige, aber auch herausfordernde Methode zur Führung und Verwaltung des öffentlichen Sektors. Es tritt in verschiedenen Formen auf und bringt eine Anzahl Risiken bzw. Beschränkungen mit sich.
2. Föderalismus ist eine sehr spezifische Staatsform, die auf bekannten Normen basiert und bestimmte Vorteile verspricht. Bedeutende Unterschiede in der jeweiligen praktischen Umsetzung des Föderalismus erschweren jedoch direkte Vergleiche.
3. Im Prinzip besteht eine starke Affinität zwischen Föderalismus und bestimmten Arten des Benchmarking.
4. In der Praxis ist zu beobachten, dass die Anwendung des Benchmarking in föderalen Systemen je nach Zielsetzung verschiedene Formen annimmt: Wer führt es durch? Wie hierarchisch ist die Umsetzung? Welches Gleichgewicht wird zwischen quantitativen und qualitativen Messgrößen gewahrt?

Die Herausforderungen des Benchmarking im öffentlichen Sektor

„Benchmarking“ ist ein Begriff, der ziemlich frei verwendet wird und zumindest zwei etwas unterschiedliche Bedeutungen annimmt, wobei eine anspruchsvoller ist als die andere. In einem weiteren Sinn kann man unter Benchmarking eine vergleichende Leistungsmessung verstehen. In einem engeren, spezifischeren Sinn kann man unter Benchmarking die Nutzung vergleichender Leistungsmessung als ein Werkzeug zur Bestimmung und Anwendung

leistungsfähigerer bzw. wirkungsvollerer Praktiken verstehen.⁴ Im ersten Sinne ist es eine Beurteilungsmethode, im zweiten ein Lern- und Regulierungswerkzeug.

Archetypen: externes und internes Benchmarking

Benchmarking fand seinen Anfang im privaten Sektor, der uns das klassische Modell davon liefert, worum es bei dieser Praktik geht. Die erste Form des Benchmarking bietet einer einzelnen Firma die Möglichkeit, die Leistung einiger Aspekte des eigenen Betriebs im Vergleich zu führenden Unternehmen in anderen Sektoren zu bewerten, und zeigt ihr anhand dieses Vergleichs, wie sie ihre Praktiken verbessern kann. Das Begutachtete und der Begutachter sind im Grunde ein und dasselbe. Ein solches „externes“ Benchmarking erfolgt freiwillig bzw. selbstgesteuert und ist ausschließlich lernorientiert.

Es gibt jedoch noch ein weiteres Modell aus dem privaten Sektor, das „interne“ Benchmarking: Dabei stellt die Geschäftsführung den einzelnen Abteilungen einer Firma Benchmarking-Anforderungen, um mittels internen Wettbewerbs Optimierungen herbeizuführen. Mit dieser zweiten Form des Benchmarking werden die Marktkräfte imitiert, die intern durch die Gründung des Unternehmens erst einmal verdrängt worden waren. In dieser Version decken sich das Begutachtete und der Begutachter nicht, und ersteres unterliegt Sanktionen, die von letzterem verhängt werden. Was die einzelnen Geschäftsbereiche betrifft, ist diese Version weder selbstgesteuert noch wissensorientiert; sie ist vielmehr hierarchisch strukturiert und zwingend. Sie setzt den Akzent auf Leistungskontrolle und schließt eine Art von Sanktion ein. Eine solche interne Leistungskontrolle entspricht sowohl der weiteren als auch der engeren Auslegung des Begriffs Benchmarking.

Vom privaten zum öffentlichen Sektor

Der offensichtliche Wert eines Leistungsvergleichs, einer Bestimmung des besten Verfahrens und eines Engagements in den Bereichen Lernen und Optimierung macht Benchmarking sowohl für den öffentlichen als auch privaten Sektor attraktiv. „Was gemessen wird, wird auch erreicht“ heißt es nach dem viel zitierten Argument von Osborne und Gaebler, und wenn Organe des öffentlichen Sektors zu messen beginnen, was sie tun, werden sie Möglichkeiten

⁴ Watson, Gregory H.: A Perspective on Benchmarking, Benchmarking: an international journal 1 (1994), S. 5-10.

finden, es besser zu tun.⁵ Wenn Regierungen anhand von Leistungsmessungen jene Dinge beleuchten, die letztendlich zählen – was Regierungen *erreichen*, im Unterschied zu dem, was sie lediglich *tun* – werden sie vermutlich auch einen Weg finden, mehr zu erreichen.

Auch wenn sich der öffentliche und private Sektor in einigen Belangen ähneln, sind sie sich jedoch in anderen sehr unähnlich. In zwei sehr wichtigen Punkten sind die Unterschiede sogar fundamental. Erstens sind Regierungen und ihre verschiedenen Organe keine gewinnorientierte Unternehmen, die sich in einem geschäftlichen Konkurrenz- und Überlebenskampf am Markt befinden. Zweitens ist es ihr Sinn und Zweck, nicht lediglich Produkte (*Output*) herzustellen, sondern gesellschaftliche Wirkungen bzw. Ergebnisse (*Outcomes*) zu erzielen. Deshalb hat der öffentlich Sektor weder die gleiche *Notwendigkeit* noch verfügt er über die gleiche *Kapazität*, Benchmarking zu betreiben, wie der private Sektor.

Überwindung des Motivationsproblems

Das erste dieser Probleme, das Fehlen intrinsischer Motivation, ist in gewisser Weise genau der Grund für die Einführung des Benchmarking – wie es auch der Grund des internen Unternehmens-Benchmarking war. Leistungskontrollen und die Auferlegung von Benchmarking-Auflagen sind Surrogate des öffentlichen Sektors für die Kräfte des Marktes. Diese können von einer einzelnen Behörde ausgehen, die ihre Leistung verbessern möchte – externes Benchmarking –, aber angesichts der geringen intrinsischen Motivation und der großen Schwierigkeiten ist dies wohl eher die Ausnahme als die Regel. In der Wirklichkeit führt der geringere Motivationsgrad dazu, dass den Organen des öffentlichen Sektors solche Anforderungen gestellt werden müssen. Die zweite Form des Benchmarking, das interne Benchmarking, ist deshalb im Bereich des öffentlichen Sektors wahrscheinlicher.

Internes Benchmarking funktioniert mittels Sanktionen. Im privaten Sektor manifestieren sich diese in Form von Entscheidungen zur Kapitalallokation. In diesem Sinne ist es ein Zwangsinstrument. Im öffentlichen Sektor können Sanktionen eine Reihe von Formen annehmen, von denen zwei besonders herausragen. Eine, dem Beispiel aus dem privaten Sektor

⁵ Osborne, David/Gaebler, Ted: Reinventing Government: how the entrepreneurial spirit is transforming the public sector, Addison-Wesley, Reading MA 1992, S. 146

folgenden, beruht auf finanziellen Strafen und Belohnungen. Finanzielle Strafen haben jedoch ihre Nachteile. Dazu gehört die reale Möglichkeit, dass die Behebung unzureichender Leistungen den Einsatz von mehr und nicht von weniger Mitteln bedingt. Ein Unternehmen mit mehreren Standorten kann es sich erlauben, Standorte, die die Erwartung nicht erfüllen, eingehen zu lassen; Regierungen dürfen dies nicht. Daraus resultiert die Anziehungskraft, die von einer ganz anderen Sanktionsform ausgeht: dem politischen Instrument des *öffentlichen Anprangerns*. Dabei gibt es ein externes Publikum, das, wie angenommen wird, erreicht werden kann und auf eine Weise antworten wird, die den erwünschten Sanktionseffekt hat.

Jede Form von Sanktion schafft jedoch auch Anreize für ein Verhalten, das den Absichten des Benchmarking-Systems entgegensteht. Das ehrliche Bestreben, die in den Messkriterien erwünschten Ergebnisse zu erbringen, kann die Form eines „teaching to the test“ annehmen. In solchen Fällen wird in der Bemühung, die Messvorgaben zu erreichen, die allgemeine Zielsetzung vergessen. Unredliches Bestreben nimmt die Form eines Spieles an, wobei der gewünschte *Hinweis* auf Resultate durch Manipulation erzeugt wird, ohne zwangsläufige Verbesserung bei den *tatsächlichen* Resultaten. Je höher die Einsätze, desto größer ist wahrscheinlich die Tendenz zu nicht beabsichtigtem Verhalten, sowohl in ihrer ehrlichen als auch unehrlichen Form.

Messprobleme

Die zweite Herausforderung, die besonders den öffentlichen Sektor betrifft – der seinen Fokus auf die erzielten Ergebnisse (Outcomes) und nicht so sehr auf die erzeugten Produkte (Outputs) legt – ist weniger leicht zu bewältigen. Private Unternehmen beurteilen ihre Erfolge nach ihrem Output, der in jedem Fall einen Geldwert hat. Es steht nicht zur Debatte, was letztlich das Ziel ist. Die Privatwirtschaft beschäftigt sich nicht damit, welche Wirkung sie haben könnte. Würde sie es tun, würden viele allgemein verfügbare Güter und Dienstleistungen nicht mehr produziert werden. Die Regierung produziert auch Produkte, diese sind aber nur ein Mittel zum Zweck, nämlich bestimmte wirtschaftliche oder gesellschaftliche Probleme anzugehen. Das Endziel sind *Ergebnisse (Outcomes)* und die sind schwierig zu messen, zuzuordnen und zu steuern. Für viele Ergebnisse bestehen soziale Indikatoren oder können soziale Indikatoren entwickelt werden, jedoch mit unterschiedlicher Schwierigkeit, besonders für Ergebnisse mit langfristigem

Zeithorizont. Die Schulen sollten Kinder mit erkenn- und testbaren kognitiven Fähigkeiten hervorbringen. Diese sind aber bis zu einem gewissen Grad nur ein Hinweis auf ein Ergebnis bzw. ein Zwischenergebnis. Letztlich sollten die Schulen Bürger hervorbringen, die sich längerfristig als fähige Wirtschaftsakteure und gut angepasste Mitglieder der Gesellschaft erweisen. Der Trend zu ergebnisorientiertem Benchmarking im öffentlichen Sektor spornt dazu an, die Forschung zu den sozialen Indikatoren, und die entsprechenden Investitionen, zu erhöhen.⁶

Auch wenn sich Ergebnisse ohne Weiteres messen lassen, können sie unter Umständen nicht so einfach durch politische Mittel beeinflusst werden. Welchen Faktoren messen wir Leistung zu? Angesichts der oft widerspenstigen und „boshaften“ Natur politischer Herausforderungen können unter Umständen auch größte Anstrengungen keine beeindruckenden Ergebnisse erzielen. Und schließlich bestehen im Gegensatz zum privaten Sektor legitime Meinungsunterschiede darüber, welche Ergebnisse der öffentliche Sektor in manchen Bereichen anstrebt.

Natürlich kann viel Nutzen daraus gezogen werden, die Outputs des öffentlichen Sektors und die dabei erzielte Leistung zu messen. Angesichts der Komplexität vieler Outputs des öffentlichen Sektors bestehen auch hier nicht unbedeutende Probleme. Das Argument der Befürworter des Benchmarking lautet, die Schaffung eines solchen Systems führe zu und fördere eine fortschreitende Verbesserung der Daten. Das kann wohl kaum bestritten werden. Die Erfahrung Großbritanniens mit Leistungskontrollen, die von quantitativen Indikatoren abhängen, hat uns jedoch unter anderem gezeigt, dass unter Umständen bedeutende qualitative Messgrößen durch die Maschen schlüpfen, was dazu führen kann, dass ziemlich irreführende Schlussfolgerungen gezogen werden.⁷

⁶ vgl. dazu beispielsweise: Atkinson, Tony/Cantillon, Bea/Marliier, Eric/Nolan, Brian: Social Indicators: the EU and social inclusion, Oxford University Press, Oxford 2002; OECD (Hrsg.): Statistics, Knowledge and Policy: key indicators to inform decision making, Organisation for Economic Co-operation and Development, Paris 2005.

⁷ Nothing But the Truth? Audit Commission, London 2009.

Was den Föderalismus betrifft

Es muss auf vier Punkte zum Föderalismus hingewiesen werden, denen bei dieser Frage besondere Bedeutung zukommen: die Besonderheit des Föderalismus als Methode der Staatsführung; seine mutmaßlichen Vorteile; seine operative Komplexität; und seine Vielfältigkeit.

1. Die Besonderheit des Föderalismus

Föderalismus ist erstens nicht einfach ein formalisiertes System einer mehrstufigen Staatsführung, sondern eine besondere Form der konstitutionalisierten Machtteilung, wobei Souveränität in gewisser Hinsicht und bis zu einem gewissen Grad gemeinsam ausgeübt wird und die Kompetenzen auf zwei Regierungsebenen, nämlich die Zentralregierung und die Regierungen der Gliedstaaten, verteilt werden.⁸ Er beruht auf drei wesentlichen Grundsätzen. Erstens genießen die zwei Ebenen verfassungsrechtlich geschützte Autonomie; keiner der beiden Ebenen kann unilateral den Status oder die Rolle des anderen ändern. Zweitens tragen die Gliedstaaten in einem sinnvollen Maße Verantwortung für lokale Angelegenheiten. Und drittens werden Entscheidungen für Angelegenheiten, die alle betreffen, national und nicht lokal getroffen. Zusammen genommen sind diese letzten beiden Prinzipien dem Subsidiaritätsprinzip der Europäischen Union ähnlich: der Regel, dass eine Aufgabe von der jeweils untersten Regierungsebene ausgeführt werden sollte, die diese Aufgabe in geeigneter Weise ausüben kann.

Aus diesen Grundsätzen ergeben sich zumindest zwei logische Konsequenzen. Einerseits sind die Landesregierungen einer Föderation nicht sich selbst untereinander und auch nicht einer weiteren nationalen Gemeinschaft gegenüber rechenschaftspflichtig, sondern in erster Linie ihrer eigenen politischen Gemeinschaft. Es liegt nicht an der nationalen Gemeinschaft, lokale Gemeinschaften aufgrund „schlechter“ Politik zu bestrafen oder sich über sie hinwegzusetzen. Andererseits muss die Pflege der Beziehungen zwischen den zwei Regierungsebenen eines Bundesstaates den Prinzipien des gegenseitigen Respekts folgen.

⁸ vgl. Hueglin, Thomas/Fenna, Alan: Comparative Federalism: a systematic inquiry, Broadview Press, Peterborough Ontario 2006.

2. Die (vermeintlichen) Vorteile des Föderalismus

Wenn es das ist, was als Föderalismus bezeichnet wird, dann ist die zweite Frage: Was ist der Zweck des Föderalismus? Warum gibt es Föderalismus und warum sollten wir versuchen, ihn aufrechtzuerhalten?

Dem Föderalismus werden gegenüber einem Einheitsstaat seit langem bestimmte Vorteile zugeschrieben. Traditionell ist man der Meinung, der Föderalismus sei ein Schutz für legitime Unterschiede und ermögliche es, Politik auf die lokalen Bedürfnisse und Vorlieben zuzuschneiden. Der Raum für solche Unterschiede wurde jedoch über das letzte Jahrhundert stark eingeeengt und das Interesse daran hat abgenommen. Es sind andere Vorteile, die nun größere Aufmerksamkeit erzielen. Die wichtigsten dieser Vorteile sind, dass durch die Vermehrung der Orte, wo politische Entscheidungen getroffen werden, die Möglichkeiten, politische Erfahrungen zu sammeln, erweitert werden – der sogenannte *Laborföderalismus* – und dass den Bürgern die Möglichkeit gegeben wird, die Leistung ihrer Regierung mit den Regierungen anderer Gebietskörperschaften zu vergleichen, auch *Wettbewerbsföderalismus* genannt. Dies sind jedoch theoretische bzw. hypothetische Vorteile. Ob das konstruktive Potenzial des Labor- oder Wettbewerbsföderalismus umgesetzt wird – oder in einem Maß umgesetzt wird, das die unvermeidbaren Nachteile der Kompetenzverteilung wettmacht – ist eine andere Frage.

3. Die Komplexität des Föderalismus

Der dritte Punkt betrifft die Beziehung zwischen Theorie und Praxis und im Besonderen die große Kluft, die oft zwischen dem Föderalismus der Theorie und dem tatsächlich existierenden Föderalismus besteht.

Föderationen haben sich in hoch komplexe, vertrackte Gebilde politischer und administrativer Verflechtungen entwickelt, die nur sehr annäherungsweise dem typischen Idealmodell entsprechen. Dies gilt vor allem für die Anglo-Föderationen, in denen eine aus dem achtzehnten oder neunzehnten Jahrhundert stammende verfassungsrechtliche Gewaltenteilung den

modernen Bedingungen angepasst werden musste.⁹ Die Konsequenz ist eine Vielzahl von Politikbereichen, in denen traditionelle lokale Zuständigkeitsbereiche der Beteiligung, Steuerung oder dem Einfluss der Zentralregierung unterliegen. Ein besonders deutliches Beispiel ist Australien. In Australien übt die Bundesregierung, der *Commonwealth*, seinen Einfluss auf eine Vielzahl von Politikbereichen aus, die verfassungsmäßig im Kompetenzbereich der Staaten liegen. In einem Arrangement, das man manchmal auch „kooperativer Föderalismus“ nennt, behalten die Staaten normalerweise die administrative Verantwortung für die Bereitstellung von Dienstleistungen bei, sind jedoch an gewisse Weisungen des *Commonwealth* gebunden.¹⁰ Dies wird mit einer Reihe von Mechanismen erreicht, überwiegend jedoch durch an Bedingungen gebundene Zuschüsse.¹¹ Die Verfügbarkeit dieses Instruments wurde durch eine starke Tendenz zu „vertikalen fiskalischen Ungleichgewichten“ in den Anglo-Föderationen ermöglicht. Kanada und die USA haben ein geringeres vertikales fiskalisches Ungleichgewicht. Während jedoch die Regierung Kanadas die Möglichkeit, ihre Zuschüsse an Bedingungen zu knüpfen, selten nutzt, sind in den USA Transferzahlungen mehrheitlich mit Bedingungen verbunden.¹²

Traditionellerweise liegen Sozialfürsorge, Bildung und Gesundheitswesen in den meisten Föderationen in der Zuständigkeit der lokalen Behörden, werden jedoch mit der Zeit mehr oder weniger stark und in einer oder der anderen Form „nationalisiert“. Dies geschieht aus einer Vielzahl von sich gegenseitig stärkenden Gründen. Dazu gehört die Tatsache, dass viele Zuständigkeitsbereiche nun eine nationale Dimension angenommen haben oder dass ihnen jetzt eine nationale Dimension zugeschrieben wird, die ihnen zuvor fehlte. Bildung beispielsweise wurde traditionsgemäß als eine Angelegenheit von fast ausschließlich lokaler

⁹ Fenna, Alan: The Malaise of Federalism: comparative reflections on Commonwealth–State Relations, in: *Australian Journal of Public Administration* 66 (2007), S. 298–306.

¹⁰ vgl. Williams, John M./MacIntyre, Clement: *Commonwealth of Australia*, in: *Distribution of Powers and Responsibilities in Federal Countries*, hrsg. v. A. Majeed, R. L. Watts und D. M. Brown, McGill–Queen's University Press, Montreal und Kingston 2006.

¹¹ vgl. Fenna, Alan: *Commonwealth Fiscal Power and Australia Federalism*, in: *University of New South Wales Law Journal* 31 (2008), S. 509-29; Morris, Alan: *Commonwealth of Australia*, in: *The Practice of Fiscal Federalism: comparative perspectives*, hrsg. v. A. Shah, McGill–Queen's University Press, Montreal und Kingston 2007.

¹² vgl. Boadway, Robin: *Canada/Fox, William: United States of America*, in: *The Practice of Fiscal Federalism: comparative perspectives*, hrsg. v. A. Shah, McGill–Queen's University Press, Montreal, Kingston 2007.

Bedeutung betrachtet. Seit ein paar Jahren wird sie jedoch aufgrund der für Produktivität und Innovation zunehmenden Bedeutung des „menschlichen Kapitals“ als integraler Bestandteil der wirtschaftlichen Vitalität einer Nation gesehen.

4. Die Vielfalt föderaler Systeme und Erfahrungen

Der vierte und letzte Punkt ist, dass jede tatsächlich bestehende Föderation einen ganz eigenen Charakter besitzt. Jede ist auf eine sinnvolle Weise einzigartig. Eine Verallgemeinerung ist schwierig. Außerdem gibt es wenige Beispiele und darum ist uns ein vergleichendes Glätten mit großen Stichproben vorenthalten. Sogar ein Vergleich der Erfahrungen der fünf wichtigsten westlichen Föderationen, Australien, Deutschland, Kanada, der Schweiz und der USA, bereitet Schwierigkeiten. Diese Schwierigkeiten erhöhen sich, wenn man die Europäische Union (EU) einbezieht, eine wichtige im Werden begriffene Föderation.

Die drei Anglo-Föderationen haben einige wichtige Gemeinsamkeiten. Sie beruhen im Kern alle auf der gesetzlichen Gewaltenteilung, wobei die zwei Ebenen – jedenfalls vordergründig – innerhalb ihrer zugewiesenen Bereiche umfassende Befugnisse in den Bereichen Politikformulierung, Umsetzung und Verwaltung ausüben. Darin unterscheiden sie sich vom deutschen Modell, das die Befugnis zur Politikformulierung in vielen Bereichen der Zentralregierung zuweist und die Verantwortung für die Umsetzung und Verwaltung den Ländern überträgt.¹³ Als logische Konsequenz dieses Ansatzes wird im deutschen Modell die Repräsentanz der Gliedstaaten, also der Länder, – über den Bundesrat – im legislativen Prozess der Zentralregierung erhöht.¹⁴

Da die australischen Gründerväter beim Entwurf und der Gestaltung der Verfassung weitgehend dem amerikanischen Beispiel folgten, ist die Ähnlichkeit der australischen und der

¹³ Kramer, Jutta: Federal Republic of Germany, in: Constitutional Origins, Structure, and Change in Federal Countries, hrsg. v. J. Kincaid und G. A. Tarr. McGill–Queen's University Press, Montreal und Kingston 2005; Schneider, Hans-Peter: The Federal Republic of Germany, in: Distribution of Powers and Responsibilities in Federal Countries, hrsg. v. A. Majeed, R. L. Watts und D. Brown, McGill–Queen's University Press, Montreal und Kingston 2006.

¹⁴ Oeter, Stefan: Federal Republic of Germany, in: Legislative, Executive, and Judicial Governance in Federal Countries, hrsg. v. K. Le Roy und C. Saunders, McGill–Queen's University Press, Montreal und Kingston 2006.

amerikanischen Föderation auf dem Papier sehr groß.¹⁵ Aber sogar zwischen den Anglo-Föderationen sind die Unterschiede bedeutend. Von Bedeutung sind:

- die Tatsache, dass die USA sich einer präsidentialen Regierungsform mit Gewaltentrennung bedient, während Kanada und Australien parlamentarische Systeme sind;
- die größere Anzahl Gliedstaaten des amerikanischen Systems;
- die Besonderheit von Kanada als einer Föderation mit einer englischsprachigen Mehrheit und einer französischsprachigen Minderheit, die sich in einer der größten Provinzen zentriert;
- und die Abwesenheit bedeutender grundlegender föderaler Unterschiede in Australien. Sie reichen vom fiskalisch stark dezentralisierten Kanada bis zum fiskalisch hoch zentralisierten Australien.

Wenn wir dann noch Deutschland, die Schweiz und die EU einbeziehen, erhöht sich die Vielfalt beträchtlich – und sie würde noch weiter zunehmen, würden wir weitere, weniger konventionell strukturierte Föderationen wie Spanien oder Belgien berücksichtigen. Fazit: Die Vergleiche müssen stark kontextualisiert werden.

Welcher Nutzen bietet Benchmarking dem Föderalismus?

Oberflächlich betrachtet haben Föderalismus und Benchmarking etliche Ähnlichkeiten. Bei beiden geht es darum, vielfache Erfahrungen auszuwerten, um herauszufinden, wie etwas besser gemacht werden kann. Formalisiertes Benchmarking bietet unter Umständen Möglichkeiten, das politische Lernpotenzial des Föderalismus zu nutzen.

Regierungen als lernende Unternehmen

Beim originalen externen Modell des Benchmarking aus dem privaten Sektor, bei dem eine unabhängige Firma als Lernübung, die es ihr erlaubt, bestmögliche Verfahrenselemente zu

¹⁵ Saunders, Cheryl: *Commonwealth of Australia*, in: *Constitutional Origins, Structure, and Change in Federal Countries*, hrsg. v. J. Kincaid und G. A. Tarr, McGill-Queen's University Press, Montreal und Kingston 2005.

integrieren, eine vergleichende Bewertung ihrer Leistung veranlasst, sind die Ähnlichkeiten zu den föderalen Systemen sofort ersichtlich. Wir könnten uns eine Föderation vorstellen, bei der die Gliedstaaten sich wie Unternehmen verhalten, die Verbesserungsmöglichkeiten suchen, und dazu ihre eigene Leistung ständig anhand der Leistungen der Regierungen der übrigen Gliedstaaten beurteilen und die gewonnenen Erkenntnisse integrieren. In diesem Idealfall stärkt sich jeder Akteur selbst, ohne das Rad jedes Mal neu erfinden zu müssen.

Die Wirklichkeit sieht anders aus. Einige der dazu beitragenden vielfältigen Gründe sind:

- Die Optimierungsfreudigkeit der Regierungen ist suboptimal.
- Es ist nicht immer einfach, Leistung zu messen und die besten Verfahrensweisen zu bestimmen.
- Politikziele sind oft wertgetränkt.
- Aufgrund des Drucks der Wahlen lassen sich Regierungen nicht auf Selbstbewertungen ein.
- Die Mechanismen für einen die Zuständigkeitsbereiche übergreifenden Lernprozess sind unter Umständen unzulänglich.
- In vielen Bereichen wirkt ein homogenisierender Einfluss der Zentralregierung.

So gesehen könnte die Einführung von Benchmarking-Praktiken und –Anforderungen den nötigen Impuls und Mechanismus für auf Wettbewerb beruhende Verbesserungen und das Sammeln von politischem Wissen liefern. Regierungen, die freiwillig Benchmarking-Vereinbarungen eingehen – „Benchmarking-Clubs“ beitreten –, können einen Rahmen schaffen, in dem eine systematischere Evaluation, größeres Experimentieren und erweitertes Lernen stattfindet. Angesichts der politischen Wirklichkeit wird der Fokus dabei wahrscheinlich eher auf Aspekten der Dienstleistungsbereitstellung als auf politischen Rahmenbedingungen liegen.

Da die Wahrscheinlichkeit, dass subnationale Regierungen ganz von selbst Benchmarking betreiben, begrenzt ist, liegen mindestens zwei mögliche Alternativen nahe. Erstens besteht die

Möglichkeit, dass das Benchmarking von einer unabhängigen, nichtstaatlichen Organisation durchgeführt wird. Das hat den Vorteil, dass dank deren Unabhängigkeit die Möglichkeit einer neutralen Bewertung gegeben ist. Diese Alternative bietet den weiteren Vorteil, dass sie aufgrund ihres nichtstaatlichen Charakters keinen Einfluss auf die Machtverhältnisse im föderalen System hat. Sie hat jedoch auch Schwächen, die die Kehrseite der Vorteile bilden: Solche Organisationen sind dem Wohlwollen der Regierung ausgeliefert, von dem sie Informationen einholen möchten. Sie haben kein formelles Druckmittel. Die andere Alternative besteht darin, dass die Aufgabe von einem Organ wahrgenommen wird, die wirkliche Autorität besitzt, der Zentralregierung.

Welche Rolle hat die Zentralregierung?

Im klassischen föderalen Modell, wenigstens im Modell der Anglo-Föderationen, gibt es nur wenig legitimen Handlungsspielraum für die Zentralregierung bei der Überwachung der Tätigkeiten der Gliedstaaten – das Benchmarking einbegriffen. Die subnationalen Regierungen sind für ihre Leistung nicht der Bundesregierung oder der nationalen Gemeinschaft gegenüber rechenschaftspflichtig, sondern ihren eigenen Wählern. Es gibt jedoch eine Anzahl Gründe, diese Struktur zu lockern.

Am allerwichtigsten ist die animierende und fördernde Rolle, die die Zentralregierung übernehmen kann, indem sie in den Bereichen, in denen die Gliedstaaten weiterhin dominieren, zwischen diesen für einen optimalen Grad an Benchmarking sorgt. Da sowohl der Labor-Föderalismus als auch der Wettbewerbsföderalismus fast unvermeidlich rare Güter sind, fällt der Zentralregierung die konstruktive Rolle desjenigen zu, der zu Experimenten ermutigt, komparative Leistungsmessungen fördert und koordiniert und Lernmöglichkeiten schafft.¹⁶

Es gibt auch verschiedene von der Zentralregierung geleitete und von ihr mehr oder weniger stark finanziell unterstützte Programme, die in Bereichen subnationaler Zuständigkeit operieren. Benchmarking stellt in diesen Kontexten eine alternative Koordinierungsmethode dar, mit der sich überregulierte und zu detaillierte Kontrollen möglicherweise durch eine Reihe von Anreizen ersetzen lassen, die das in den Mittelpunkt stellen, worum es in der Politik

¹⁶ wie beispielsweise dargelegt in Metzenbaum, Shelley K.: Strategies for Using State Information: measuring and improving program performance, IBM Center for the Business of Government, University of Maryland, 2003.

letztlich geht: Ergebnisse. Im Bemühen, erfolgreichere und leistungsfähigere Mittel zum Zweck der Dienstleistungsbereitstellung zu finden, könnten Experimentiergeist und Lernen möglicherweise dadurch gefördert werden, dass der Fokus von den Ressourcen und Produkten (Input und Outputs) auf die Ergebnisse (Outcomes) verlagert wird.

Was finden wir vor?

Wir können feststellen, dass in föderalen Staaten viele verschiedene Erfahrungen mit Benchmarking gemacht wurden, obgleich das Benchmarking in den meisten Staaten bislang nur lückenhaft genutzt wird. Dabei handelt es sich bei einigen um eine Veranschaulichung des föderalen Ethos und bei anderen um eine Weiterführung nichtföderaler Praktiken mit neuen Mitteln. Bei allen stellt sich die Frage, wie effizient sie sind und wie das Messen von Leistung in politisches Lernen übersetzt werden kann. Eine Anzahl der Fälle befinden sich noch in einem frühen Anfangsstadium. Es können drei allgemeine Ansätze erkannt werden:

1. eine Überwachung durch unabhängige Organisationen;
2. eine zwangsweise, hierarchische Überwachung;
3. und ein kollegiales Benchmarking.

Unabhängige Überwachung

In mehreren Staaten wird die Leistungskontrolle der Gliedstaaten von nichtstaatlichen Organisationen oder Einrichtungen durchgeführt. In den USA ist es das Pew Center, das für eine periodische „Benotung“ der Staaten zuständig ist. Unter Verwendung einer Zwölfpunkteskala und im Stile eines Zeugnis oder einer Rangliste wird eine zusammenfassende Bewertung präsentiert, und die Informationen werden der Öffentlichkeit auf einer Webseite zugänglich gemacht. Bei der letzten Bewertung standen die Staaten Utah, Virginia und Washington ganz oben auf der Liste, New Hampshire erhielt die schlechteste Note und landete auf dem Schlussplatz. Bewertungskriterien sind der Zustand der Finanzen, das Personalwesen, die Infrastrukturbereitstellung und die Informationstechnologie. Das Pew Center wiederum hebt in seiner Leistungsbewertung die Nutzung von Leistungsmessung durch die Regierungen der Bundesstaaten als ein wesentlicher Erfolgsfaktor hervor. In Deutschland wird eine ähnliche Seite | 14

Bewertung durch die privatwirtschaftliche Bertelsmann Stiftung durchgeführt. Sie legt den Schwerpunkt insbesondere auf die Finanzleistung.¹⁷ In der Schweiz erstellt die BADAC (Datenbank über die Schweizer Kantone und Städte), ein einer Hochschule angegliedertes Institut, Leistungsvergleiche zu einer Reihe finanzieller und Regierungs- und Verwaltungsindikatoren.¹⁸ Sie bemüht sich zudem, ihre Bewertungen über einen Webauftritt größtmöglich verfügbar zu machen.

Eine solche unabhängige Überwachung hat offensichtliche Vorteile und Nachteile. Aufgrund ihrer Unaufdringlichkeit und ihres desinteressierten, die Zuständigkeitsbereiche übergreifenden Fokus auf Stärken und Schwächen stehen sie in komplettem Einklang mit dem Föderalismus und steuern einen Grad der vergleichenden Leistungsbeurteilung bei, der ansonsten nicht gegeben wäre. Er sollte sowohl zu Labor- als auch Wettbewerbsföderalismus beitragen. Aber welchen Einfluss hat diese Art der Praxis? Unabhängige Überwachungen bedienen sich weitgehend frei verfügbarer Daten. Dies kann dazu führen, dass Messungen nicht aufgrund ihrer Bedeutung oder Aussagekraft gemacht werden, sondern einfach, weil die Daten existieren und verfügbar sind. Und die Regierungen schenken unter Umständen den Befunden keine große Beachtung, da die Verantwortung hierfür nicht bei ihnen liegt.

Zwangsüberwachung

Am anderen Ende stehen die Fälle, bei denen die Zentralregierung, genauso wie Großunternehmen mit ihren Geschäftsbereichen, internes Benchmarking als Mittel verwendet, um über Sanktionen Leistungsverbesserungen herbeizuführen. Dies kann die Weiterführung traditioneller Zentralisierungstendenzen föderaler Systeme in neuer Form bedeuten, wobei die Zentralregierung spezifische verfassungsmäßige Befugnisse oder ihre finanzielle Überlegenheit nutzt, um eine faktische Änderung in der Gewalten- und Kompetenzverteilung zu erzielen, oder aber die Anwendung einer neuen Regierungs- und Verwaltungsmethode auf die zwischen

¹⁷ vgl. beispielsweise: Berthold, Norbert/Kullas, Matthias/Müller, Andreas: Die Bundesländer im Standortwettbewerb, Bertelsmann Stiftung, Gütersloh 2007; Wagschal, Uwe/Wintermann, Ole/Petersen, Thieß: Konsolidierungsstrategien der Bundesländer: Verantwortung für die Zukunft, Bertelsmann Stiftung, Gütersloh 2009.

¹⁸ Datenbank der Kantone und Städte (BADAC) am Hochschulinstitut für öffentliche Verwaltung (IDHEAP) in Lausanne.

nationalen und subnationalen Regierungen bestehende Aufteilung der Funktionen. Keine der beiden Versionen ist frei von Schwierigkeiten.

In der Schweiz scheint sich die Einführung von Leistungsvergleichen zwischen dem Bund und den Kantonen im Bereich der Arbeitsvermittlung infolge Widerstands gegen die Sanktionierungsregelungen größtenteils als erfolglos zu erweisen.¹⁹ Obwohl in Australien die Schulen in den Zuständigkeitsbereich der Staaten fallen, hat das Commonwealth die Website „My School“ gestartet, auf der nach Schulen geordnet die Ergebnisse der Schüler bei nationalen Prüfungen veröffentlicht werden. Das markanteste Beispiel für Zwangs-Benchmarking kommt jedoch aus den USA.

Der *No Child Left Behind Act* (NCLB)²⁰ der US-Regierung aus dem Jahr 2001 bewirkte eine bedeutende Verlagerung der Kontrolle über die Primar- und Sekundarschulen. Traditionell besaßen die Staaten nahezu alleinige Zuständigkeit für diesen Bereich.²¹ Damit wurden die Staaten explizit in die ehrgeizige Leistungsmanagement-Gemeinschaft aufgenommen, die der Kongress 1993 mit der Verabschiedung des Government and Performance Results für die Bundesbehörden ins Leben gerufen hatte. NCLB wurde unilateral erarbeitet und den Staaten als eine Erweiterung der traditionellen Methode der zweckgebundenen Zuschüsse auferlegt, um den Einfluss des Kongresses auf Angelegenheiten auszudehnen, die in der Zuständigkeit der Staaten liegen. Der Widerstand war groß, und viele Kommentatoren betrachten die erzielten Erfolge als klein. Sie sagen, die Gründe dafür lägen im Föderalismus und in den Schwierigkeiten, per Leistungsmessung zu regieren.²²

¹⁹ Hilbert, Christopher: Implementation of Performance Measurement in Public Employment Services in Switzerland, in: The Evaluation of Active Labour Market Policies: measures, public private partnerships and benchmarking, hrsg. v. J. De Koning, Edward Elgar, Cheltenham 2007.

²⁰ Ein Gesetz, um mithilfe von Rechenschaftspflicht, Flexibilität und Wahlfreiheit die Leistungsunterschiede an den öffentlichen Schulen der USA auszugleichen.

²¹ vgl. Manna, Paul: School's In: federalism and the national education agenda, Georgetown University Press, Washington D.C. 2006. McGuinn, Patrick J.: No Child Left Behind and the Transformation of Federal Educational Policy, 1965–2005, University Press of Kansas, Lawrence, Kansas 2006. Wong, Kenneth: Toward Federalizing Education Policy, Publius 40 (2009), S. 226–233.

²² vgl. beispielsweise Radin, Beryl A.: Challenging the Performance Movement: accountability, complexity, and democratic values, Georgetown University Press, Washington DC 2006. Ravitch, Diane: Time to Kill 'No Child Left Behind', Education Week.

Kollegiales Monitoring und Benchmarking

Zwischen diesen beiden Extremen, der unabhängigen und der hierarchischen Überwachung, gibt es ein drittes Modell: eine Art kollegiales Monitoring, dessen Durchführung auf intergouvernementalen Vereinbarungen und in einigen Föderationen auf der Zusammenarbeit zwischen den Gliedstaaten beruht. Die Zentralregierung ist normalerweise auch beteiligt, spielt jedoch eine eher unterstützende Rolle. Die Zielgruppen bilden in erster Linie die Regierungen selbst, oder aber die Gemeinschaft in einem breiteren Sinne. Das führende Beispiel ist Australiens *Report on Government Services* (ROGS), der nun bereits seit 15 Jahren veröffentlicht wird. Ein von Vertretern aller Regierungen bestehender Lenkungsausschuss legt den Rahmen für das Leistungsmonitoring fest und beaufsichtigt die Veröffentlichung des Berichtes. Die *Productivity Commission*, eine unabhängige Forschungseinrichtung der Regierung des Commonwealth, fungiert dabei als Knotenpunkt: Sie dient als Sekretariat, trägt die Daten zusammen und erstellt den Bericht. Zwar erfasst der ROGS nicht alle Tätigkeiten, an deren Durchführung die Landesregierungen beteiligt sind, aber er behandelt ein breites Spektrum von öffentlichen Dienstleistungen, die einen wesentlichen Bestandteil der Tätigkeiten der Landesregierungen ausmachen.

Eine wirkliche Entsprechung zum ROGS gibt es in keinem anderen Bundesstaat. Aber es lassen sich ähnliche Einrichtungen finden, die in spezifischen Bereichen tätig sind. In der Schweiz beispielsweise fördert das Bundesamt für Raumentwicklung kantonale Leistungskontrolle im Bereich der Nachhaltigkeitspolitik. Wie es scheint, beschränkt sich die Rolle der Bundesregierung dabei auf die Förderung der Datenqualität und damit deren Nutzwert. Im Besonderen fördert sie die Vergleichbarkeit der auf lokaler Basis erzeugten Daten. In Kanada agiert die Bundesregierung über das *Canadian Institute for Health Information* (CIHI) als Knotenpunkt für eine ähnliche Initiative im Bereich des Gesundheitswesens.

Die Frage, die sich zu diesen verschiedenen Beispielen kollegialen Benchmarkings stellt, ist: Wie effektiv fließt deren zunehmend raffiniertere Erzeugung und Anhäufung von Leistungsdaten zurück in politisches Lernen und in eine Verbesserung der Leistungsbereitstellung in den einzelnen Zuständigkeitsbereichen? In anderen Worten: Inwieweit übersetzt sich Leistungsüberwachung in echtes Benchmarking? Der Fokus liegt in allen drei hier erwähnten

Fällen auf quantitativen Indikatoren und vergleichender Leistungsmessung; Mechanismen für qualitatives Lernen sind abwesend oder nur zweitrangig. Dies bringt uns zur *offenen Koordinierungsmethode* (OKM) der europäischen Union.

Die EU hat die OKM als Methode der Politikkoordinierung entwickelt, die in Bereichen Anwendung findet, die nicht in ihre Zuständigkeit fallen.²³ Zur Unterscheidung vom „Hard Law“, das die EU über ihre „Direktiven“ ausübt, wird die OKM als „Soft Law“ bezeichnet. Bedeutende Politikbereiche, die nicht in die Zuständigkeit der EU fallen, sind die sozialstaatlichen Einrichtungen, die Bildungssysteme und die Arbeitsmarktprogramme – also grob gefasst die Sozialpolitik. Ursprünglich galten diese in Bezug auf das Hauptziel der EU, der Förderung der wirtschaftlichen Dynamik mittels wirtschaftlicher Integration, als nebensächlich, werden jetzt jedoch als maßgebliche Faktoren der Wirtschaftsleistung betrachtet.

Im Rahmen der im Jahr 2000 verkündeten Lissabon-Strategie wird die EU durch die Einführung von Leistungsmessungen und Benchmarking-Rahmen, die Durchführung von Evaluationen und *Peer-Reviews* und der Förderung gegenseitigen Lernens Verbesserungen in diesen Politikfeldern anstreben. Die OKM sollte starke und komplementäre quantitative und qualitative Messgrößen haben, freiwillig sein und „kontextualisiertes Lernen“ fördern. Damit ist ein Lernen gemeint, das die in den verschiedenen Zuständigkeitsbereichen herrschenden unterschiedlichen Bedingungen und unterschiedlichen kulturellen und institutionellen Ordnungen berücksichtigt. Großer Wert wird auf eine ausführliche *Peer-Review-Praxis* gelegt, die ihr Augenmerk auf die qualitative Messgröße richtet, die für ein zwischen den Zuständigkeitsbereichen erfolgreiches Lernen erforderlich ist.

Die Unzufriedenheit mit der beschränkten Wirkung der OKM führte nach ein paar Jahren zur Empfehlung, zu einer Methode des öffentlichen Anprangerns und Rühmens in der Form von Ranglisten zu wechseln, mit der die Mitgliedstaaten auf aggressivere Art zur Übernahme von besten Praktiken überredet werden könnten.²⁴ Diese Empfehlung wurde zurückgewiesen, und es besteht wenig Grund zur Annahme, dass sie erfolgreich gewesen wäre. Obwohl die OKM die

²³ vgl. beispielsweise Tholoniati, Luc: The Career of the Open Method of Coordination: lessons from a 'soft' EU instrument, *West European Politics* 33 (2010), S. 93–117.

²⁴ Kok, Wim: Facing the Challenge: the Lisbon strategy for growth and employment (2004).

föderalen Prinzipien der Zusammenarbeit, des gegenseitigen Respekts, der unabhängigen Rechenschaftspflicht und der Optimierung durch gegenseitiges Lernen verkörpert, bleibt umstritten, welche substantielle Auswirkung sie hat.²⁵

Schlussfolgerungen

Föderalismus und Benchmarking verbindet eine tentative, exploratorische Beziehung. Teils beruht diese auf gutgläubigen Versuchen, Teile des Potenzials, das der Föderalismus als lernorientierte Regierungs- und Verwaltungsordnung bietet, auszuschöpfen, teils spiegelt sie anhaltende Zentralisierungsdynamiken wider. Logik und auch Indizien deuten darauf hin, dass ein externes, auf beste Praktiken ausgerichtetes Benchmarking sowohl im öffentlichen Sektor im Allgemeinen als auch in föderalen Systemen im Besonderen ein Gut mit beschränkter Verfügbarkeit ist. Den Zentralregierungen scheint deshalb bei der Förderung von Zusammenarbeit dieser Art eine wichtige Rolle zuzukommen. Benchmarking bietet unter Umständen auch eine administrativ und substantiell überlegene Alternative zu mehr auf Direktiven basierenden Methoden zentralisierter Politikformulierung in föderalen Systemen. Die Erzeugung verlässlicher und wirklich indikativer Daten, die Zuordnung von Produkten (*Outputs*) und Wirkungen (*Outcomes*), die Bestimmung und Umsetzung von Verfahrensoptimierungen und die Anwendung von Sanktionen ist jedoch mit echten Herausforderungen verbunden. Aufgrund der Art dieser Herausforderungen muss kooperativen und kollaborativen Prozessen Priorität beigemessen werden.

²⁵ Heidenreich, Martin/Zeitlin, Jonathan (Hgg.): *Changing European Employment and Welfare Regimes: the influence of the open method of coordination on national reforms*, Routledge, London 2009. Kröger, Sandra: *The Effectiveness of Soft Governance in the Field of European Anti-Poverty Policy: operationalization and empirical evidence*, *Journal of Comparative Policy Analysis* 11 (2009), S. 197–211.