

*Sabine Kuhlmann/Gisela Färber/Dorothea Jansen/  
Sabine Kropp/Joachim Wieland*

### Leistungsvergleich und Benchmarking im Öffentlichen Sektor

Projektverbund des Deutschen Forschungsinstituts für Öffentliche  
Verwaltung Speyer (FÖV)

#### 1. Ziele und Aufgabenstellung des Projektverbunds

Leistungsmessung, Leistungsvergleich und Benchmarking sind im internationalen Kontext zu einem vorrangigen Thema bei der Modernisierung des Öffentlichen Sektors geworden (*Pollitt/Bouckaert* 2004, S. 87; *De Bruijn* 2002; *Wollmann* 2003, 2004; *Kuhlmann* 2005, 2007, 2010; *Proeller/Siegel* 2009). Es kann mithin von einem europäischen, wenn nicht globalen Trend des Performance Measurement ausgegangen werden, der sich voraussichtlich stabilisieren oder sogar weiter verstärken wird und bei dem Leistungsvergleiche und Benchmarking eine zunehmend wichtige Rolle spielen (vgl. *Van Dooren/Van de Walle* 2008; *Bouckaert/Halligan* 2008; *De Lancer* 2009; *Pollitt* 2006; *Kuhlmann* 2010). Als prominentes Beispiel für diesen Trend kann die Grundgesetzänderung im Rahmen der Föderalismusreform II des Jahres 2009 gelten, mit der Leistungsvergleiche erstmals einen verfassungsrechtlichen Status erhalten haben (siehe hierzu weiter unten). Der im Mai 2010 am FÖV in Speyer konstituierte Projektverbund „Leistungsvergleich und Benchmarking im Öffentlichen Sektor“ befasst sich aus Sicht unterschiedlicher verwaltungswissenschaftlich orientierter Teildisziplinen (Politikwissenschaft, Rechtswissenschaft, Soziologie, Wirtschaftswissenschaft) mit den rechtlich-normativen Grundlagen, institutionellen Varianten, instrumentellen Methoden, Anwendungsformen und Auswirkungen von Leistungsvergleich und Benchmarking im nationalen sowie internationalen Kontext.

#### Begriffsklärungen

*Leistungsmessung* (Performance Measurement) wird im Rahmen des Forschungsverbunds als die Ermittlung der Leistungsfähigkeit öffentlichen Verwaltungshandelns über Kennzahlensysteme und Performanzindikatoren (*performance indicators – PIs*) verstanden, die sowohl Input-, Output-, Outcome- als auch Prozess- und Struktur-Indikatoren umfasst (*Nullmeier* 2005). Der Begriff des *Performance Management* schließt zusätzlich noch die Steuerung, d.h. die gezielte Einwirkung auf institutionelles Handeln, mittels Performanzinformationen ein. *Leistungsvergleich* bezieht sich auf den indikatoren- und kennzahlenbasierten Vergleich von Leistungen, Strukturen, Prozessen und Wirkungen im öffentlichen Sektor zwischen oder innerhalb von Verwaltungen, zwischen Ebenen oder auch zwischen föderativen Subjekten. Von *Benchmarking* im engeren Sinne ist zu sprechen, wenn es dar-

um geht, beste Leistungen oder Lösungen (*best practices*) zu identifizieren und aus diesen unmittelbar Lehren für die eigene Organisation abzuleiten („Lernen von den Besten“). Benchmarking beinhaltet also per definitionem die normative Komponente der Leistungsbewertung, die im analytisch neutraleren Begriff des „Leistungsvergleichs“ nicht enthalten ist (vgl. Reichard 2002; Nullmeier 2005; Kuhlmann 2007, 2010; Bouckaert/ Halligan 2008, S. 14-26; Proeller/Siegel 2009, S. 456ff.).

## Forschungsziele und -fragen

Obschon eine internationale Reformbewegung des Performance Measurement zu beobachten ist, liegen bislang keine Studien vor, die sich aus interdisziplinärer Perspektive mit der Systematisierung, international und intra-föderal vergleichenden Analyse sowie der Wirkungsabschätzung von Leistungsvergleichen befassen. Der am FÖV Speyer angesiedelte Forschungsverbund beabsichtigt, diese Lücke teilweise zu schließen. Ziel ist es, fördernde und hemmende Faktoren von Leistungsvergleichen und Benchmarking zu identifizieren, wobei sowohl rechtlich-normative, politisch-institutionelle, organisatorische als auch ökonomisch-fiskalische Aspekte berücksichtigt werden. Des Weiteren wird die (methodische) Qualität verschiedener Typen, Verfahren und Instrumente von Performanzmessung und -vergleich untersucht. Zudem sollen die Projektergebnisse Aufschluss über Nutzungsformen und -bedingungen von Vergleichsinformationen in Politik und Verwaltung geben. Ein wichtiger Beitrag des Verbunds wird zudem darin gesehen, empirische Evidenz über Ergebnisse, Effekte und weiterreichende Auswirkungen dieses Reformansatzes zu erhalten und die Frage zu beantworten, ob und wie in öffentlichen Organisationen aus Leistungsinformationen produktiv und/oder pathologisch gelernt wird.

Neben der empirischen Analyse und rechtsnormativen Bewertung von Leistungsvergleichen beinhaltet der Projektverbund eine politik- und rechtsberatende Komponente. Die Forschungsergebnisse sollen genutzt werden, um zukünftige Vergleichsstudien konzeptionell, instrumentell und methodisch zu unterstützen und hierfür beratende Expertise bereitzustellen. Der Verbund bietet die Möglichkeit, professionelles Know-How über die Funktionsweise, Anwendungsvoraussetzungen und Erfolgsfaktoren von Benchmarking im Bundesstaat zu entwickeln und dieses in die Verwaltungspolitik einzubringen. Diese Forschungsziele im Blick, konzentriert sich der Projektverbund auf die folgenden Analysebereiche von Leistungsvergleich und Benchmarking, die disziplinübergreifend und teils auch international vergleichend bearbeitet werden:

- (1) Entwicklung und Steuerung von Leistungsvergleichen/Benchmarking in Mehrebenensystemen:
  - Welche Bedingungen fördern oder hemmen die Etablierung und Fortschreibung von Leistungsvergleich und Benchmarking in Politik und Verwaltung? Wie werden Vergleichsprojekte (erfolgreich und nachhaltig) initiiert, implementiert und gesteuert? Unter welchen Bedingungen entsteht (föderativer) Wettbewerb?
- (2) Rechtliche Regelung und normative Ausgestaltung von Benchmarking:
  - Welche rechtlichen Rahmenregelungen für Performanzmessung und Leistungsvergleiche gibt es, wie sind diese zu bewerten und inwieweit wirken sie begünstigend/blockierend im Hinblick auf Leistungsvergleiche?

- (3) Verfahren/Indikatoren der Performanzmessung und Methoden(-kritik):
  - Welche Ausgestaltungsvarianten, Instrumente, Techniken und Methoden von Performanzmessung und Benchmarking gibt es und wie funktionsfähig sind sie? Welche Messverfahren erweisen sich als methodisch leistungs- und tragfähig, welche als defizitär?
- (4) Anwendungsfelder/Verwaltungsbereiche des Benchmarkings:
  - Welche konkreten Anwendungsbereiche von Leistungsmessung und Benchmarking gibt es und welchen Einfluss hat die Spezifik des Aufgabenbereichs/Politikfeldes auf die Ausgestaltung und Wirksamkeit des Instruments?
- (5) Lernprozesse, Nutzung, Wirkung von Vergleichsinformationen/Evidenzbasierung:
  - Inwieweit werden Vergleichsergebnisse für Steuerungsentscheidungen in Politik und Verwaltung verwendet (*evidence-based policy making/administration*)? Wann wird aus Leistungsvergleichen gelernt und kommt es (wettbewerbsbedingt) zu Verbesserungen? Wie ist das Verhältnis von Kosten und Nutzen einzuschätzen? Welche nicht-intendierten Effekte ergeben sich und kommt es auch zu „pathologischem“ Lernen?
- (6) Handlungs- und Gestaltungsempfehlungen; Weiterentwicklung des Benchmarkings:
  - Welche Gestaltungs- und Handlungsempfehlungen sind für zukünftige Vergleichsprojekte in Bund, Ländern und Kommunen abzuleiten? Welche Lehren können aus den ausländischen Erfahrungen gezogen werden? Welche Kontextfaktoren sind für den Konzepttransfer zu beachten? Welche methodische Weiterentwicklung ist zu empfehlen?

## 2. Konturen des Forschungsfeldes

### Internationale Verbreitung von Leistungsmessung und -vergleich

Im internationalen Kontext erhielt der Gedanke der kennzahlen- und indikatorengestützten Leistungsvergleiche im öffentlichen Sektor insbesondere durch die Doktrin des New Public Management Auftrieb und Verbreitung, welche im angelsächsischen Raum in den 1980er Jahren (*Pollitt/Bouckaert* 2004) und seit den 1990er Jahren auch verstärkt in Kontinentaleuropa (*Kuhlmann* 2009; *Kuhlmann/Fedele* 2010), inzwischen teilweise ebenfalls in Mittel- und Osteuropa (vgl. *Nemec/Merickova/Ochrana* 2008; *Löffler/Vinta* 2004) aufgegriffen wurde. Der internationale Bedeutungszuwachs von Performanzmessung und Leistungsvergleich lässt sich nicht nur an einer rasanten Zunahme der wissenschaftlichen Publikationstätigkeit zum Thema des Performance Measurement und des Benchmarking in internationalen Fachjournals ablesen (vgl. *Braadbaart/Yusnandarshah* 2008), sondern auch in der Politik- und Verwaltungspraxis an der Extensivierung, Intensivierung und Externalisierung von Leistungsmessung und -vergleich im öffentlichen Sektor:

- Aufbauend auf neuen Informationstechnologien und in engem Zusammenhang mit neuen Verfahren des Finanzmanagements und der Budgetierung, werden Leistungsmessung und -vergleich heute in einer größeren Anzahl von Politikfeldern und Verwaltungsbereichen praktiziert<sup>1</sup>, worin sich die *extensivere* Nutzung entsprechender Verfahren zeigt.

- Leistungsmessung und -vergleiche gehen inzwischen über die traditionelle *monitoring*-Funktion hinaus und spielen in einer zunehmenden Zahl von öffentlichen Handlungsfeldern eine wichtige Rolle als Steuerungsinstrument<sup>2</sup>. Dies spiegelt ihre *intensivere* Nutzung wider.
- Leistungsvergleiche sind immer mehr darauf angelegt, nicht nur verwaltungsintern die Steuerungsfähigkeit zu verbessern, sondern sie werden zunehmend auch herangezogen, um politische Mandatsträger, Verwaltungsklienten und die Öffentlichkeit über Verwaltungsleistungen zu informieren<sup>3</sup>. Somit ist es im internationalen Kontext zu einer *Externalisierung* dieser Verfahren gekommen.

Wenngleich der Verbreitungsgrad, die Intensität, Methodik und Steuerungslogik sowie Wirksamkeit von Leistungsmessung und -vergleich zwischen und innerhalb der verschiedenen Länder erheblich variieren (siehe hierzu *Kuhlmann 2010*), so besteht kein Zweifel, dass sie sich im Laufe der letzten zwei Jahrzehnte zu einem international zunehmend bedeutsamen, wenn auch in Wissenschaft und Praxis nicht unumstrittenen, Ansatz der Modernisierung des Öffentlichen Sektors entwickelt haben. Eine systematische Untersuchung im Rahmen eines Projektverbunds erscheint vor diesem Hintergrund besonders lohnend, da die deutschen Erfahrungen in eine vergleichende Perspektive zu anderen OECD-Ländern gesetzt werden können. Auf diese Weise lassen sich verschiedene komparative Forschungsdesigns entwickeln.

### Aktuelle Konjunktur von Leistungsvergleichen

Zwar reicht die Praxis von Leistungsmessung und -vergleich in den öffentlichen Verwaltungen einiger Länder, etwa im Fall von Großbritannien und den USA, bereits bis ins ausgehende 19. Jahrhundert zurück<sup>4</sup>. Auch die deutsche Verwaltung kann auf Erfahrungen mit evidenz-basierter Politik und indikatoren-gestützter Leistungsmessung zurückgreifen, die mit der Etablierung der Evaluation als Analyse- und Informationsverfahren im öffentlichen Sektor seit den 1960er Jahren Verbreitung gefunden haben (*Wollmann 2003*). Allerdings ist gerade in jüngerer Zeit eine bemerkenswerte Konjunktur in diesem Reformfeld zu beobachten, die durch mehrere national und/oder international wirksame Faktoren begünstigt wurde:

- Wie oben erwähnt, hat im *internationalen* Kontext die New Public Management-Bewegung entscheidend dazu beigetragen, dass Leistungsvergleiche und Benchmarking-Ansätze als Instrumente zur Verbesserung der Effizienz der Verwaltung angesehen wurden und sich entsprechend verbreiteten. Hintergrund ist vielfach eine zunehmend prekäre öffentliche Haushaltslage und Finanzkrise in einigen Ländern, so auch in Deutschland. Zum anderen ist mit der europäischen Integration, in deren Kontext sich die horizontale Zusammenarbeit der Mitgliedsländer inzwischen als eigenständiger Modus der EU-Politik entwickelte (vgl. auch *Speer 2002*), der Ausbau von gemeinsamen Qualitätsbewertungssystemen (z.B. Common Assessment Framework - CAF etc.), Leistungsvergleichen und Benchmarking innerhalb und zwischen den europäischen öffentlichen Verwaltungen zu beobachten (*Wallace 2005, S. 85; Goetz 2006, S. 483*).
- Im *nationalen* Kontext ist zum einen auf die Grundgesetzänderung im Rahmen der Föderalismusreform II hinzuweisen (Gesetz vom 29.7.2009, BGBl I 2248), die mit

dem neu eingefügten Art 91d GG eine Rechtsgrundlage für Leistungsvergleiche/Benchmarking zur Feststellung und Förderung der Leistungsfähigkeit der öffentlichen Verwaltungen geschaffen hat. Indem die Vorschrift als „Kann-Bestimmung“ formuliert wurde, hat der Gesetzgeber am Freiwilligkeitsprinzip festgehalten. Hintergrund der Verfassungsänderung ist dabei – neben den allgemeinen (NPM-nahen) Ideen von Leistungstransparenz, Effektivitäts- und Effizienzsteigerung – vor allem das Anliegen, mittels Leistungsinformationen die parlamentarische Kontrolle zu stärken, die ebenfalls als Ziel von Verwaltungsmodernisierung begriffen wurde (BT-Drs 16/12410, 8). Daneben soll der föderative Wettbewerb um die beste Lösung angeregt werden (vgl. Ondracek Stellungnahme Sachverständigenanhörung Rechtsausschuss des BT vom 4.5.2009, S. 3; BT-Drs 16/12410, 8). Bund und Länder, die in diesem Reformfeld zusammenwirken sollen, können dabei auf vielfältige Erfahrungen und Konzeptvorschläge von Leistungsmessung und -vergleichen in zahlreichen öffentlichen Aufgabefeldern zurückgreifen, die bereits seit Anfang der 1990er Jahre im kommunalen Raum praktiziert werden. Auch diese basieren auf dem Prinzip der Freiwilligkeit, haben allerdings eine relativ geringe Transparenz nach außen und sind in ihrer Wirksamkeit durchaus nicht unumstritten (vgl. *Kuhlmann* 2005, 2007), so dass die anderen föderalen Ebenen beim Neustart von Vergleichsprojekten hieraus entsprechende „Lehren“ ziehen können.

### Besondere Eignung für Verbundforschung: Interdisziplinärer Bezug und Methodenpluralismus

Das Forschungsfeld von Leistungsvergleichen und Benchmarking ist für einen interdisziplinären Zugang besonders geeignet, da es nur aus der Perspektive mehrerer verwaltungswissenschaftlicher Teildisziplinen angemessen bearbeitet werden kann. Der Projektverbund „Leistungsvergleich und Benchmarking im Öffentlichen Sektor“ nähert sich diesem Forschungsfeld aus vier disziplinären Zugängen:

- Um die (verfassungs-)rechtlichen Implikationen und rechtspolitischen Auswirkungen von Leistungsvergleichen im Bundesstaat zu untersuchen, wird ein *rechtswissenschaftlicher* Zugang gewählt, der – speziell im deutschen Fall – besonders den Regelungsgehalt der Föderalismusreform II (Art. 91d GG) ins Auge fasst. Ferner ist geplant, mittels Rechtsvergleichung die internationale Verbreitung von entsprechenden Regelungsinhalten und eine mögliche „Vorbildfunktion“ anderer Länder zu eruieren. Dabei wird auf rechtsnormative Analysen und rechtsvergleichende Methoden auf der Basis des einschlägigen Schrifttums, Normenbestands und der Rechtsprechung sowie auf rechtstatsächliche Erhebungstechniken (Fallstatistik, Befragung, Interviews) zurückgegriffen.
- Da Leistungsvergleich und Benchmarking ihre Wurzeln in (institutionen-)ökonomischen Konzeptideen haben und ihre Verbreitung im internationalen Kontext wesentlich durch ökonomische und fiskalische Problemlagen bedingt ist, greift der Projektverbund ferner auf einen *wirtschaftswissenschaftlichen* Zugang zurück. Dieser umfasst nicht nur die theoretische Analyse und Kritik des Reforminstruments aus wirtschafts- und finanzwissenschaftlicher Perspektive, sondern aus empirischer Sicht auch die makro- und mikroökonomischen Bestimmungsfaktoren und Auswirkungen

von Leistungsvergleich und Benchmarking (Kosten-Nutzen-Analysen; Transaktionskosten etc.). Mit Blick auf den deutschen Exekutivföderalismus geht es konkret darum, die Nutzung von Benchmarks hinsichtlich der Effizienz/Effektivität der Umsetzung von Bundesrecht sowie hinsichtlich der Vollzugsprozesse auf Länder- und Kommunalebene zu untersuchen, die unterschiedliche regionale Lösungsansätze und deren Optimierungsbedarfe erkennen lassen. Hieraus werden Rückschlüsse auf Korrekturbedarfe des originären Bundesrechts gezogen. Methodisch werden dann neben makro-quantitativen Aggregatdatenanalysen und finanzstatistischen Vergleichen auch mikro-ökonomische Analyseinstrumente angewendet, insbesondere um Prozessveränderungen, Transaktionskosten und Performanzeffekte innerhalb einzelner Verwaltungsinstitutionen zu erfassen.

- Aus *politikwissenschaftlicher* Sicht wird im Projektverbund einerseits nach den politisch-institutionellen Bedingungsfaktoren und strategischen Handlungskalkülen relevanter Akteure gefragt, die es begünstigen oder auch verhindern, dass Benchmarking-Prozesse initiiert und in der öffentlichen Verwaltung implementiert werden. Andererseits ist zu ermitteln, unter welchen Umständen und befördert durch welche Anreizsysteme politisch-administrative Akteure Vergleichsinformationen und Benchmarking-Ergebnisse verwenden, in Entscheidungsprozesse einspeisen und so zu „rationaleren“ (evidenz-basierten) Entscheidungsergebnissen gelangen (*evidence-based policy making*). Aus evaluativer Sicht werden die weitergehenden Auswirkungen auf die Leistungsfähigkeit (*performance*) der Verwaltung einbezogen. Auf der Grundlage von „Large-n-studies“ wird darüber hinaus die Qualität von Instrumenten, Indikatoren und Methoden der Performanzmessung evaluiert. Zur methodischen Umsetzung wird auf die einschlägigen sozialwissenschaftlichen Erhebungsinstrumente zurückgegriffen, wobei eine Kombination qualitativer und quantitativer Techniken beabsichtigt ist (standardisierte Einstellungsforschung; Sensitivitätsanalysen von Messinstrumenten in Form von Large-n-studies sowie qualitative Fallstudien und Experteninterviews).
- Um die im Rahmen des Projektverbunds interessierende Frage zu beantworten, inwieweit „Evidenz-Basierung“ die Organisationskultur verändert und weitergehende Lern- und Veränderungsprozesse in Organisationen auslöst, kommt ein *organisationssoziologischer* Zugang zum Zuge. Dieser erlaubt es zu untersuchen, ob der Reformgedanke des Leistungsvergleichs von den handelnden Akteuren in Politik und Verwaltung internalisiert und in der Organisation „gelebt“ wird (*enacting*). Ein wesentlicher Beitrag der soziologisch orientierten Verwaltungsforschung ist eine Analyse darüber, ob Organisationen Benchmarking-Ergebnisse nutzen, um sich zu modernisieren, weiterzuentwickeln und den sich wandelnden Umweltbedingungen besser anzupassen. Von Interesse ist, wie solche reflexiven Lernprozesse verlaufen, wie nachhaltig sie sind und inwieweit möglicherweise auch pathologische Lerneffekte festzustellen sind. Neben standardisierten Umfragen werden Fallstudien über das Handeln von Verwaltungsorganisationen und weitere (qualitative und quantitative) Methoden der empirischen Sozialforschung eingesetzt.

### 3. Theoretischer Bezugsrahmen des Verbunds

Die interdisziplinäre Anlage des Projektverbunds hat zur Folge, dass den einzelnen Teilprojekten jeweils verschiedene (disziplinär verankerte) theoretische Ansätze bzw. – so

fern die Fragestellung dies nahelegt – Kombinationen dieser Ansätze zugrunde gelegt werden. *Theorienpluralismus* ist somit eine wesentliche Basis des Verbunds. In Anbetracht des Untersuchungsgegenstands und der gemeinsamen verwaltungswissenschaftlichen Forschungsorientierung bieten sich dabei bestimmte Theorieansätze für eine teilprojektübergreifende und transdisziplinäre Anwendung besonders an. Dies gilt ohne Zweifel für *institutionentheoretisch* orientierte Ansätze, da Institutionen als Forschungsgegenstand seit den späten 1970er Jahren in den verschiedenen Disziplinen – Wirtschaftswissenschaft, Soziologie, Politik- und Verwaltungswissenschaft, öffentliches Recht u.a. – unabhängig voneinander „wiederentdeckt“ worden sind (vgl. *Kaiser* 1999, S. 190; *Ede-ling* 1999, S. 7). Somit ist es sinnvoll, in der institutionell orientierten Verwaltungsforschung, wie sie im hier skizzierten Verbund vertreten ist, auf einen institutionentheoretischen Hintergrund zurückzugreifen. Die Bedeutung, die Institutionen aus verschiedenen disziplinären Blickwinkeln eingeräumt wird<sup>5</sup>, gilt gewissermaßen als einende Klammer und transdisziplinäre „Brücke“, die für die theoretische Profilierung des Projektverbunds nutzbar gemacht wird. So finden sich in der neo-institutionalistischen Theoriedebatte sowohl soziologische und politikwissenschaftliche als auch ökonomische, rechtliche und historische Konzeptideen, als deren gemeinsamer Kern die Überzeugung gelten kann, dass Institutionen eine *structural suggestion* (*Dowding* 1995, S. 44) enthalten. Bei den verschiedenen (ökonomischen, normativen, historischen, soziologischen) Varianten des Neuen Institutionalismus geht es darum, „den zeitgenössischen sozialwissenschaftlichen Theorien eine institutionalistische Perspektive beizumischen – das Mischungsverhältnis allerdings ist unterschiedlich“ (*Kaiser* 1999, S. 191). Für den Forschungsgegenstand „Leistungsvergleich und Benchmarking in der öffentlichen Verwaltung“ bietet dies die Möglichkeit, formal-rechtliche, kulturelle, akteurbezogene und ökonomische Aspekte der Verwaltungswirklichkeit jeweils unterschiedlich zu konzeptualisieren und auf ihren jeweiligen Erklärungsgehalt zu prüfen. Im Rahmen der weiteren konzeptionellen und inhaltlichen Profilierung des Projektverbunds wird der theoretischen Fundierung des gemeinsamen Forschungsgegenstandes, die hier kaum mehr als cursorisch angetippt werden konnte, besondere Beachtung geschenkt.

#### 4. Geplante Teilprojekte

Der Projektverbund setzt sich aus den o.g. vier verwaltungswissenschaftlichen Teildisziplinen (Politikwissenschaft, Rechtswissenschaft, Soziologie, Wirtschaftswissenschaft) zusammen. Die inhaltliche Zusammensetzung der Teilprojekte ist von der Überlegung geleitet, möglichst alle Ebenen des politisch-administrativen Systems und deren Interaktionen (*multi level governance*) einzubeziehen. Daher befasst sich ein Teil der Projekte eher mit Benchmarking auf der subnationalen und lokalen Ebene (*Jansen, Kuhlmann*), ein anderer Teil stärker mit Fragen der bundesstaatlichen Ordnung und dem Benchmarking auf (landes)staatlicher Ebene bzw. im Zusammenwirken von Bund/Zentralstaat und Ländern/föderativen Subjekten (*Färber, Kropp, Wieland*). Allen Projekten ist jedoch eine Mehrebenenperspektive eigen, da die jeweiligen Benchmarking-Einheiten und -Akteure nicht isoliert, sondern im Zusammenspiel mit anderen Ebenen untersucht werden. Im Einzelnen umfasst der Verbund die folgenden Themen für Teilprojekte:

- (1) „Benchmarking in der öffentlichen Verwaltung: ein europäischer Ländervergleich“ (*Kuhlmann*): Das Projekt befasst sich aus politik- und verwaltungswissenschaftlicher Perspektive mit dem Verbreitungsgrad, den Steuerungsformen und der Durchführung von Leistungsvergleichen und Benchmarking in ausgewählten europäischen Ländern („Länder-Mapping“). Der Analysefokus liegt auf der subnationalen Verwaltungsebene, d.h. Länder-/ Regional-/Kommunalebene, deren Vergleichsprojekte als – methodisch gesprochen – abhängige Variable erhoben werden sollen. Die übergeordnete (zentralstaatliche/ Bundesebene) soll – als erklärende Variable – hinsichtlich ihrer Initiativ-, Steuerungs- und Koordinationsfunktion von Vergleichsaktivitäten untersucht werden. Daneben systematisiert das Projekt weitere (akteur- und policy-bezogene, politisch-institutionelle und verwaltungskulturelle) Kontextvariablen, die den Grad der Verbreitung und die Varianten von Leistungsvergleichen erklären.
- (2) „Rechtliche Normierung und Regulierung von Leistungsvergleichen im Bundesstaat“ (*Wieland*): Ziel des Teilprojekts ist die Analyse der maßgeblichen (verfassungs-) rechtlichen Regeln und der damit verbundenen Problemlagen für Leistungsvergleiche im deutschen Bundesstaat. Hierzu werden zunächst die Maßgaben für Leistungsvergleiche unter der Geltung des Grundgesetzes, insbesondere nach der Einfügung des Art. 91d GG, herausgearbeitet und sodann aus den so gewonnenen Ergebnissen rechtliche Vorgaben, u.a. für das Verfahren und die Organisation von Leistungsvergleichen im Bundesstaat abgeleitet. Das methodische Vorgehen lässt sich demzufolge in einen Zweischritt gliedern: nach einer umfassenden verfassungsrechtlichen Analyse der Bestimmung des Art. 91d GG und der sonstigen Maßgaben für Leistungsvergleiche im Bundesstaat Deutschland werden in einem darauffolgenden zweiten Schritt Empfehlungen und Handlungsanleitungen für die weitere rechtliche Ausgestaltung von intraföderalen Leistungsvergleichen gegeben.
- (3) „Benchmarking als Instrument besserer Rechtsetzung im föderalen Mehrebenenstaat“ (*Färber*): Das Projekt untersucht den Einsatz des Benchmarking-Instruments für eine bessere Rechtsetzung im deutschen Exekutivföderalismus. Bislang verabschiedet der Bund Gesetze und Verordnungen, die von den Ländern in eigener Verantwortung vollzogen werden, ohne sich systematisch um die Effizienz und Effektivität seiner Regulierungen zu kümmern. Selbst bei Evaluationsstudien gab es höchst selten Analysen über die Wirksamkeit und Wirtschaftlichkeit unterschiedlicher Vollzugslösungen, weil die politische Fiktion des einheitlichen Verwaltungsvollzuges gewahrt werden musste. Erste Projekte unter Beteiligung der Geschäftsstelle Bürokratieabbau des Nationalen Normenkontrollrates und freiwilliger Länder und Kommunen (Elterngeld, Wohngeld, BAföG) zeigen indes, welche Differenzen hier vorliegen und in welchem Ausmaß hier knappe Ressourcen nicht nur der Normadressaten, sondern auch der vollziehenden öffentlichen Verwaltungen ineffizient verwendet werden. Das Projekt hat zum Ziel, die abgeschlossenen und die in nächster Zukunft beginnenden Projekte, die nicht mehr nur die Messung des Aufwandes für Informationspflichten, sondern den gesamten Erfüllungsaufwand zum Gegenstand haben werden, kritisch zu begleiten. Es soll diese Projekte zudem methodisch im Hinblick auf ihre Eignung als „wissensgenerierende“ Benchmarks und als Basis für den Vergleich von originären Länderaufgaben evaluieren. Kernpunkt der Analyse ist dabei immer der Interaktionsprozess zwischen Normadressaten und vollziehenden öffentlichen Verwaltungen (die z.B. einen Antrag auf Transfers bescheiden oder eine Genehmigung erteilen sollen oder auch die Einhaltung von Grenzwerten kontrollieren oder einen Verkehrsunfall

protokollieren müssen), der als Folge von (Bundes-)Recht stattfindet. Hierbei werden auch Quantifizierungsmethoden, insb. das Standardkostenmodell (SKM), ins Visier genommen, um anhand einer methodenkritischen und reflektierten Analyse und Bewertung der damit gewonnenen Messergebnisse Schlussfolgerungen im Hinblick auf eine mögliche Optimierung von Verwaltungsprozessen und eine bessere Rechtsetzung zu ziehen. Es wird zum einen von der Annahme ausgegangen, dass komplexere und methodisch reflektierte Messverfahren die Risiken des Vertrauensverlustes, die sich aus der Verwendung von simplen und unterkomplexen Leistungskennzahlen ergeben, vermeiden helfen, vor allem dann, wenn der Wissensgewinn durch Benchmarking in den Mittelpunkt gestellt wird. Zum anderen wird vermutet, dass ein solches Länder-Benchmarking die Effizienzverluste der politisch bewusst in Kauf genommenen systematischen Trennung von Gesetzgebung und Vollzug reduzieren hilft, da die vertikale Rückkoppelung zur Verbesserung der Kultur und Effizienz der Rechtsetzung selbst beitragen müsste.

- (4) „Organisationswandel und Effizienzsteigerung durch Benchmarking: das Beispiel der kommunalen Stadtwerke“ (*Jansen*): In dem Projekt wird aus soziologischer Perspektive der Frage nachgegangen, ob und inwieweit Benchmarking-Ergebnisse organisatorische Anpassungs- und Optimierungsprozesse in öffentlichen Institutionen auslösen bzw. unter welchen Bedingungen sie zu organisatorischen Lern- und Innovationsprozessen führen. Dabei wird das Problem eines möglichen pathologischen Lernens aufgegriffen, welches sich als nicht-intendierte Konsequenz aus Leistungsvergleichen ergeben kann. Das Projekt untersucht diese Fragen am Beispiel der kommunalen Stadtwerke in Deutschland, da diese bereits seit längerem einem verpflichtenden Benchmarking hinsichtlich relevanter Leistungsparameter unterworfen sind, so dass sich Effekte des (obligatorischen) Benchmarkings – im Sinne organisatorischer „Lernerfolge“ oder auch „Misserfolge“ – hier mit dem gebotenen zeitlichen Abstand untersuchen lassen. Unter anderem soll die Hypothese geprüft werden, dass Benchmarking einen Einfluss auf aktuell stattfindende Re-Kommunalisierungsprozesse hat.
- (5) „Methoden und Verfahren von Performanzmessung und Leistungsvergleich: Das Beispiel Föderalismusforschung“ (*Kropp*): Das Projekt greift die Frage nach der methodischen Eignung, Robustheit und Validität verschiedener Ansätze von Performanzmessung und -vergleich auf. Angelegt als methodenkritisches Projekt, das sich an der Methodendiskussion in der komparativen Politikwissenschaft orientiert, sollen Indikatoren, Verfahren und Vergleichskriterien von Leistungsmessung auf den Prüfstand gestellt werden. Ebenso werden die jedem Leistungsvergleich zugrunde liegenden, nicht jedoch immer explizierten theoretischen Konzepte und daran anschließende Probleme der Operationalisierung kritisch hinterfragt. Da die empirische Basis für eine solche Analyse im Bereich von Leistungsvergleichen begrenzt ist, greift das Projekt auf einschlägige Ansätze der vergleichenden Föderalismus- und Demokratieforschung zurück. Empirisch soll das Projekt mithilfe einer Kombination von Large-N-Analysen und vertiefenden Fallstudien umgesetzt werden.

Das Projekt verfolgt *drei Zielsetzungen*: Angesichts des Forschungsstands liegt es (1) nahe, sich aus inhaltlicher Sicht der Frage der Vereinbarkeit bzw. der demokratiesteigernden oder -mindernden Qualität föderaler Systeme zu widmen und dieses Untersuchungsfeld dabei (2) gleichzeitig zur methodenkritischen Analyse von Leistungsvergleichen heranzuziehen. Diese beiden Ziele werden im ersten Projektabschnitt miteinander verbun-

den. Das Projekt stellt somit einerseits einen Beitrag zur Erforschung von Problemen des Leistungsvergleichs dar, andererseits aber auch zur Forschung über föderale bzw. Mehrebenensysteme und ihre Demokratiequalität. (3) In einem zweiten Projektabschnitt sollen die Ergebnisse dieses inter- und interföderalen Vergleichs verwendet werden, um Kriterien zur Bewertung von föderalen Staaten herauszuarbeiten. Dabei sollen die unterschiedlichen Autonomiegrade subnationaler Gebietseinheiten sowie Unterschiede zwischen stärker politikverflochtenen und kooperativen bzw. trennföderalen Typen berücksichtigt werden. Ein besonderes Augenmerk soll dabei auch auf Variationen von stärker die Exekutiven, d.h. Regierungen und Verwaltungen, bzw. partizipationssteigernden Elementen von Föderalstaaten und die damit einhergehenden Wirkungen gerichtet werden.

## Anmerkungen

- 1 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), Neuseeland (umfassende Anwendung von PI-Systemen im Rahmen der Strategic Results Areas und Key Results Areas), Niederlande (enge Verknüpfung von Performance Measurement und Budgetprozess), 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, S. 87 f.).
- 2 Beispielsweise bauen nicht nur substantielle Politikentscheidungen in verschiedenen Policy-Feldern, sondern vielfach auch Entscheidungen über personalpolitische Fragen (Beförderung usw.) oder Budgetallokationen auf Leistungsmessungen anhand elaborierter PI-Systeme auf.
- 3 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* 2004, S. 89 f.).
- 4 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 und zu vergleichen (*Pollitt/Bouckaert* 2004, S. 87).
- 5 „The ‚new institutionalism‘ is not one thing but many“ (*Goodin* 1996, S. 2).

## Literatur

- Bowerman, M./Francis, G./Ball, A./Fry, J.*, 2002: The evolution of benchmarking in UK local authorities, in: *Benchmarking: An International Journal*, 9, S. 429-449.
- Bouckaert, G./Halligan, J.*, 2008: *Managing Performance, International Comparisons* New York: Routledge.
- Braadbaart, O./Yusnandarshah, B.*, 2008: Public sector benchmarking: a survey of scientific articles, 1990-2005, in: *International Review of Administrative Sciences*, 74, S. 421-433.
- Brüggemeier, M./Schauer, R./Schedler, K.*, 2007: *Controlling und Performance Management im öffentlichen Sektor*. Ein Handbuch. Bern.
- De Bruijn, H.*, 2002: *Managing Performance in the Public Sector*, London/New York.
- De Lancer, J. P.*, 2009: *Performance-Based Management Systems Effective Implementation and Maintenance*, Boca Raton: Taylor & Francis.

- Drewry, G./Greve C./Tanquerel, T.*, 2005: *Contracts, Performance Measurement and Accountability in the Public Sector*, Amsterdam.
- Goetz, K. H.*, 2006: Europäisierung der öffentlichen Verwaltung – oder europäische Verwaltung?, in: *Politische Vierteljahresschrift Sonderheft 37*. Wiesbaden, S. 472-490.
- Halachmi, A.*, 1999: *Performance and quality measurement in government. Issues and experiences*, Burke.
- Knutsson, H.*, 2003: *Conditions of inter-municipal benchmarking in the Swedish public sector*, Paper for presentation at the 9th Biennial CIGAR Conference.
- Kropp, Sabine*, 2010: *Kooperativer Föderalismus und Politikverflechtung*, Wiesbaden: VS-Verlag.
- Kropp, S./Giesbers, B./Höhmman, N./Möllers, L./Ruschke, M.*, 2008: Demokratiequalität im „Ermessen“ der Forschung. Der Tatu-Vanhanen-Index im Labor der deutschen Länder und Schweizer Kantone, in: *ZParl*, 39/4, S. 703-727.
- Kuhlmann, S.*, 2005: Selbstevaluation durch Leistungsvergleiche in deutschen Kommunen, in: *Zeitschrift für Evaluation* 2005/1, S. 7-28.
- Kuhlmann, S.*, 2007: Internationale Erfahrungen mit Performance Management: Befunde aus Großbritannien, Frankreich und Deutschland, in: *Schimanke, D.* (Hrsg.), *Qualität und Ergebnis öffentlicher Programme. Ein Werkstattbericht*. Münster, S. 166-204.
- Kuhlmann, S.*, 2009: Politik- und Verwaltungsreform in Kontinentaleuropa. Subnationaler Institutionenwandel im deutsch-französischen Vergleich, Baden-Baden.
- Kuhlmann, S.*, 2010: Performance Measurement in European local governments: a comparative analysis of reform experiences in Great Britain, France, Sweden and Germany, in: *International Review of Administrative Sciences* (i.E.).
- Kuhlmann, S./Fedele, P.*, 2010: New Public Management in Continental Europe: Local Government Modernization in Germany, France and Italy from a Comparative Perspective, in: *Wollmann, Hellmut/Marcou, Gérard* (eds.): *Production and delivery of social and public services in cross-country comparison. Between government, governance, and market*. Cheltenham/Northampton (i.E.).
- Löffler, E./Vintar, M.* 2004: *Improving the Quality of East and West European Public Services*, Bratislava.
- Moynihan, D. P.*, 2008: *The dynamics of performance management. Constructing information and reform*, Washington D.C.
- Müller, T./Pickel, S.*, 2007: Wie lässt sich Demokratie am besten messen? Zur Konzeptqualität von Demokratieindizes, in: *Politische Vierteljahresschrift*, 43, S. 511-540.
- Nemec, J./Merickova, B./Ochрана, F.*, 2008: Introducing benchmarking in the Czech Republic and Slovakia. Processes, problems and lessons, in: *Public Management Review*, 10, S. 673-684.
- Nullmeier, F.* 2005: Output-Steuerung und Performance Measurement, in: *Blanke, B./Bandemer, S. v./Nullmeier, F./Wewer, G.* (Hrsg.): *Handbuch zur Verwaltungsreform*, 3. Auflage, Wiesbaden: VS, S. 431-443.
- Pollitt, C.*, 2006: *Performance Management in Practice: A Comparative Study of Executive Agencies*, in: *Journal of Public Administration Research and Theory*, 16(1), S. 25-44.
- Pollitt, C./Bouckaert, G.*, 2004: *Public Management Reform A Comparative Analysis*, Oxford: Oxford University Press.
- Proeller, I./Siegel, J. P.*, 2009: Performance Management in der deutschen Verwaltung – eine explorative Einschätzung, in: *der moderne staat*, 2, S. 455-474.
- Reichard, C.*, 2002: Performance Management, in: *Eichhorn, P.* et al. (Hrsg.), *Verwaltungslexikon*. 3. Aufl., Baden-Baden, S. 794-795.
- Schuster, F.*, 2003: *Der interkommunale Leistungsvergleich als Wettbewerbssurrogat*, Berlin.
- Speer, B.*, 2002: Governance, Good Governance und öffentliche Verwaltung in den Transformationsländern Mittel- und Osteuropas, in: *König, K.* et al. (Hrsg.), *Governance als entwicklungs- und transformationspolitisches Konzept*, Berlin, S. 207-275.
- Vagnoni, E./Maran, L.*, 2008: Public sector benchmarking: an application to Italian health district activity plans, in: *Benchmarking: An International Journal*, 15, S. 193 - 211.
- Van Dooren, W./Van de Walle, S.* 2008: Reality is merely an illusion, albeit a persistent one: introduction to the performance measurement symposium, in: *International Review of Administrative Sciences*, 74, S. 531-34.
- Van Dooren, W./Van de Walle, S.* (Hrsg.), 2008: *Performance information in the Public Sector: How it is used*, London: Palgrave Mcmillan.

*Wallace, H.*, 2005: An institutional anatomy and five policy modes, in: (Hrsg.), Policy Making in the European Union, 5. Auflage, Oxford, S. 49-90.

*Wollmann, H.* (Hrsg.), 2003: Evaluation in Public Sector Reform. Concepts and Practice in International Perspective. Cheltenham/Northampton: Elgar.

*Anschrift:*

FÖV, Deutsches Forschungsinstitut für öffentliche Verwaltung Speyer, Freiherr-vom-Stein-Straße 2, 67346 Speyer, Prof. Dr. Sabine Kuhlmann

E-Mail: [kuhlmann@foev-speyer.de](mailto:kuhlmann@foev-speyer.de)