

Verwaltungsleistung messen und steuern: Konzeption, Konsens und Kosten¹

WERNER BUSSMANN

1. Steuerung über Ziele und Messung der Leistung: der „Megatrend“ in der Verwaltung

Ein neues Konzept der Verwaltungssteuerung, die Steuerung über Ziele und durch Kontrolle der Ergebnisse, ist daran, das bisherige Konzept der Verwaltungssteuerung durch Gesetz und nachgeordnete Rechtserlasse in den Hintergrund zu verdrängen. Die Steuerung über Leistungsindikatoren ist nach meiner Ansicht der wichtigste aktuelle Trend in der Führung der Verwaltung (bzw. dessen, was davon in manchen Ländern nach Privatisierung, Dezentralisierung und „Downsizing„ übrig geblieben ist).

In der Schweiz manifestiert sich dieses neue Konzept unter vielerlei Formen. Die Ansicht, die Verwaltung werde durch Ziele gesteuert und die Ergebnisse müssten kontrolliert werden, hat sich etwa im neuen Regierungs- und Verwaltungsorganisationsgesetz des Bundes vom 21. März 1997 (RVOG; SR 172.010, vgl. namentlich Art. 3 Abs. 3 und Art. 5) durchgesetzt. Letzteres hat auch ein Projekt der Verwaltungskontrolle des Bundesrats, die Aufgabenüberprüfung², massgeblich beeinflusst. Die Legislaturplanung, die in der Anfangszeit im Wesentlichen aus einer Aufzählung der dem Parlament zu unterbreitenden Verfassungs- und Gesetzesvorlagen bestand, hat sich ebenfalls auf neue Gegebenheiten ausgerichtet und listet die von der Verwaltung zu verfolgenden prioritären Ziele auf. Die Geschäftsberichterstattung gibt über die Verwirkli-

¹ Überarbeitete Version eines Referats, gehalten am zweiten Walliser Seminar über Verwaltungskontrolle in Sion (2. - 4.7.98).

² VKB-Projekt 24: Überprüfung der Bundesaufgaben, Bericht an den Bundesrat vom 20. Januar 1995.

chung dieser Ziele Aufschluss. Auf Departements- und Amtsebene werden Ziele vorgegeben und es werden Systeme der Ergebnisüberwachung eingerichtet. Die FLAG-Ämter³ auf Bundesebene und die Versuche in sog. wirkungsorientierter Verwaltungsführung auf Kantons- und Gemeindeebene sind ebenfalls dem Konzept der Zielsteuerung und Ergebnisüberprüfung verpflichtet. Im Bereich finanzieller Transfers (zwischen Bund und Kantonen sowie zwischen Kantonen und Gemeinden) soll die bisherige Steuerung der Beiträge nach Massgabe der getätigten Aufwendungen durch eine Steuerung nach Ergebnissen ersetzt werden. Im Vordergrund stehen Steuerungsmechanismen, wie sie bei Globalbeiträgen und Pauschalbeiträgen (vgl. Bussmann 1988) bisher üblich waren. Dem Ziel einer Neugestaltung der finanziellen Transfers zwischen Bund und Kantonen ist vor allem der Neue Finanzausgleich⁴ verpflichtet. Einzig der Budgetprozess hat sich, abgesehen von Ausnahmen für die erwähnten Verwaltungsversuche (FLAG-Ämter auf Bundesebene, WIF/WOV/NEF Pilotprojekte in den Kantonen), bisher weitgehend in traditionellen Bahnen bewegt (ist aber seinerseits ein Element der Abkehr von der Steuerung des Verwaltungshandelns über Gesetze, indem seit rund 20 Jahren in zahlreichen Sachbereichen die finanziellen Leistungen gemäss den jeweiligen Erlassen von einer ausreichenden Kreditzusprechung durch das Parlament abhängig gemacht wurden).

Die Reformen in Richtung Steuerung über Ziele und Ergebniskontrolle sind in zahlreichen anderen Ländern (Neuseeland, Grossbritannien, Finnland, Deutschland „Neues Steuerungsmodell“), zumindest auf dem Papier, häufig aber auch in der Realität noch weiter gediehen (Naschold, Wholey, Mawhood).

Die laufenden Verwaltungsreformen stützen sich in der Konzeption (aber aus Gründen, die noch zu erörtern sind, nicht immer in der Wirklichkeit) auf Leistungsindikatoren ab. Ihre Einführung wird, so scheint es, durch den Ausbau und die Verfeinerung statistischer Grunddaten erleichtert.

³ FLAG=Führen mit Leistungsauftrag und Globalbudget. Die gesetzliche Grundlage findet sich in Art. 44 des Regierungs- und Verwaltungsorganisationsgesetzes des Bundes vom 21. März 1997 (RVOG; SR 172.010) sowie in Art. 38a des Bundesgesetzes vom 6. Oktober 1989 über den eidgenössischen Finanzaushalt (SR 611.0).

⁴ Im Frühling 1999 in die Vernehmlassung geschickt, Botschaft an die eidg. Räte auf 1. Jahreshälfte 2001 erwartet.

Es scheint ein Grundanliegen des modernen Menschen zu sein, Sachverhalte nicht nur verbal benennen, sondern auch in numerische Zeichen fassen zu können. In diesem Jahrhundert sind im statistischen Bereich vielerlei Fortschritte erzielt worden, so etwa die Einführung der volkswirtschaftlichen Buchhaltung (mit dem Top-Indikator „Bruttosozialprodukt/Bruttoinlandprodukt“) oder die Einführung und der Ausbau der sog. Sozialindikatoren (z.B. Bildungsindikatoren). Die Entwicklung der elektronischen Datenverarbeitung hat dem Hunger der Menschen nach Zahlen weiter Auftrieb verliehen und gleichzeitig auch der Ansicht Nahrung verschafft, auf Grund der Produktivitätsfortschritte in der Datenverarbeitung liessen sich Verbesserungen ohne grössere Kosten realisieren.

2. Leistungsindikatoren können den Markt nicht ersetzen

Es entbehrt nicht einer gewissen Ironie, dass die Versuche einer Steuerung über Ziele und Leistungsmessung in einer geschichtlichen Periode erfolgen, in welche der Hauptprotagonist einer solchen Konzeption, die Sowjetunion, untergegangen ist und deren Hauptfigur heroischer Leistungserbringung, der berühmte Stachanow, langsam aber sicher dem Gedächtnis entschwindet. In der Tat ist eine der Lektionen der zwei letzten Jahrzehnte, dass alle noch so intelligenten Versuche der Steuerung von Organisationen über bürokratische Mechanismen der Steuerung mittels Wettbewerb nichts Gleichwertiges entgegenzusetzen haben. Aus diesem Grund haben Firmen ihre Betriebseinheiten in „Profitcenters“ umgebildet oder, wie beispielsweise die ABB, weitgehend verselbstständigt. Die Erfahrung lehrt jedenfalls, dass der Markt eine Leistungsmessung erbringt, die nur teilweise durch objektive Faktoren zu erklären ist. Die heute in der Privatwirtschaft dominierenden Systeme der Leistungsmessung streben nicht an, den Markt zu ersetzen. Sie versuchen vielmehr zu einem besseren Verständnis davon zu verhelfen, welche Komponenten der Leistung massgeblich zum Markterfolg beigetragen haben.

Über den Vorteil, Marktergebnisse für interne Leistungsbewertungen verwenden zu können, verfügen Verwaltungen leider nicht, insbesondere weil sie im Nachgang zu den durchgeführten Privatisierungen, Teilprivatisierungen und Autonomisierungen für jene Bereiche reserviert werden, die sich zurzeit sachinhärent dem Wettbewerb entziehen. Sicher können

Marktergebnisse und Daten für in der Privatwirtschaft erprobte Prozesse (beispielsweise Vorgaben bezüglich der Bearbeitungszeit vergleichbarer Dokumente, Betreuungsaufwand für EDV, Aufwand für Drucksachen usw.) in der Verwaltung übernommen werden. Was die Bewertung der eigentlichen Tätigkeit angeht, muss aber „Leistung“ im öffentlichen Bereich stets von neuem „erfunden“ bzw. bewertet werden.

3. „Leistung“ ist ein durch Inhalt zu füllender Begriff

„Leistung“ in der Verwaltung hat viele Gesichter. Sie ist in keiner Weise evident. Für den staatlichen Bereich ist stets von neuem zu definieren, was „Leistung“ beinhaltet⁵. Die Leistungsmessung kann sich nämlich auf eine ganze Palette von Sachverhalten beziehen. Hervorzuheben sind die folgenden:

Leistung kann sich erstens auf unterschiedliche *Aggregierungsebenen von Politik* beziehen: auf die Leistungen des Staats als Gesamtem (z.B. in Form der von Bürgerinnen und Bürgern ausgedrückten allgemeinen Zufriedenheit mit Regierung und Verwaltung oder in Form des Gesamtumfangs an öffentlichen Ausgaben), auf die einzelnen Ebenen (Bund, Kantone, Gemeinden), dann vor allem auf die Leistungen in einzelnen Politikbereichen (z.B. Gesundheit), auf die Leistungen von spezifischen Programmen in diesen Bereichen (wie AIDS-Prävention), auf die Leistungen von einzelnen Komponenten solcher Programme (wie AIDS-Prävention für heterosexuelle Personen) bis hinunter auf noch weiter desaggregierte Ebenen (z.B. einzelne Plakatkampagnen oder einzelne Sektionen der zuständigen Abteilung des Bundesamts für Gesundheitswesen).

Leistung kann sich zweitens auf *unterschiedliche Schritte des Politikumsetzungsprozesses* beziehen. Bussmann, Klöti und Knoepfel unterschei-

⁵ Je nach Finanzlage und gewählten Beratern standen und stehen eine ganze Anzahl von Devisen (und Adjektiven) zur Auswahl, welche die zu erbringende Leistung qualifizieren: rechtsstaatlich korrekte, dienende, planende, bürgernahe, freundliche, kooperative, zielorientierte, outputorientierte, effiziente, zeitgerechte, bescheidene, schlanke, proaktive, präventive, wirkungsorientierte, qualitätsorientierte etc. Verwaltung.

den folgende sieben Schritte: Politikkonzept, Verwaltungsprogramm, Behördenarrangement für die Politikumsetzung, Aktionspläne für den Vollzug, Output, Impacts und Outcome. Die Leistungsmessung kann somit auf unterschiedlichen Orten des Wirkungspfades ansetzen, der zwischen dem Politikkonzept und den daraus in der gesellschaftlichen Wirklichkeit sowie in der natürlichen Umwelt resultierenden Veränderungen liegt.

Die wichtigste Frage ist dabei wohl die, ob die Leistungsmessung bei den *Tätigkeiten* der Verwaltung (in systemtheoretischer bzw. politikanalytischer Sicht „Output“ und in betriebswirtschaftlicher Sicht „Produkt“ genannt) oder ob sie bei den *Wirkungen* des Verwaltungshandelns (in politikanalytischer Sicht „Outcome“ genannt) ansetzen soll. Aus der Sicht der Politikanalyse ist die Antwort klar (vgl. Bussmann, Klöti und Knoepfel): Es macht keinen Sinn, Verwaltungsaktivismus, auch wenn er noch so gut motiviert ist, positiv zu bewerten, wenn nicht die damit anvisierten Probleme gelöst werden⁶. Die Verwaltung hat allerdings nur ein beschränktes Interesse, an Wirkungen (und damit an Zielen) gemessen zu werden, welche sie nicht festgelegt hat. Ob das Schwergewicht der Leistungsmessung bei den Outputs oder Outcomes anzusetzen ist, hängt davon ab, welche Kompetenzen der Verwaltung übertragen werden. Wenn die Bestimmung des Umfangs und der Art und Weise der Tätigkeiten nicht in ihrer Verantwortung liegt, macht es kaum Sinn, ihre Leistung am erzielten „Outcome“ zu messen. Letzteres dürfte nur zweckmässig sein, wenn a) mit der entsprechenden Verwaltungseinheit Ziele vereinbart werden, zu deren Erreichung sie die notwendigen Handlungsspielräume hat, und wenn b) die gesellschaftlichen und ökologischen Einflüsse auf den „Outcome“ prognostizierbar sind⁷. Hier ist anzumerken, dass die Er-

⁶ Anzumerken ist einschränkend, dass es neben der hier vorliegenden Sicht der Gesetzgebung als Problemlösung (social problem solving) noch die Sicht der Gesetzgebung als „Ordnungsarchitektur“ (social ordering) gibt (vgl. Bussmann 1998).

⁷ Der Einfluss der Umwelt auf den Outcome führt namentlich bei Leistungsvereinbarungen mit (teil-) autonomen Organisationen (wie Bahnen, städtischen Verkehrsbetrieben oder Spitälern) zu ganz heiklen Zurechnungsproblemen. Es gilt zu ermitteln, inwieweit das Erreichen/Nichterreichen von Leistungszielen (z.B. Anzahl transportierter Passagiere für Unternehmen des öffentlichen Verkehrs; Aufenthaltsdauer von Patienten öffentlicher Spitälern) auf das Verhalten der entsprechenden Institution oder auf Umweltfaktoren (z.B. geändertes Mobilitätsverhalten im öffentlichen Verkehr; geänderte Struktur der Patienten von Spitälern) zurückzuführen ist.

fassung des Outcome häufig sehr viel anspruchsvoller ist als jene des Outputs, gilt es doch analytisch aus einer gesellschaftlich-wirtschaftlichen Gemengenlage den Beitrag des Staats auf die komplexe Wirkung herauszufiltrieren. Häufig herrscht die naive Meinung vor, es genüge, aussagekräftige Zielindikatoren zur Verfügung zu haben. Zielindikatoren erlauben nur zu ermitteln, ob die Ziele erreicht wurden, nicht aber, *warum sie und namentlich ob sie auf Grund staatlichen Handelns* erreicht wurden. Letzteres kann bestenfalls auf Grund von Evaluationen der Wirksamkeit oder auf Grund dichter Batterien von Monitoring-Daten und ausreichender Kenntnisse über Wirkungszusammenhänge ermittelt werden.

4. Die Qualität einer Dienstleistung zu messen erfordert Einigung über die massgeblichen qualitativen Aspekte

In der Verwaltung liegen zahlreiche quantitative Daten vor: Behandelte Asylgesuche, ausgebildete Schüler, betreute Patienten usw. Diese Daten können für eine erste Leistungsbewertung eingesetzt werden. In der Tat gab es Fälle, in welchen auf Grund einer rohen quantitativen Bewertung Entscheide gefällt wurden (so wurden beispielsweise in der Eidgenössischen Asylrekurskommission die Mitarbeiterinnen und Mitarbeiter während einer gewissen Zeit auf Grund der Zahl der von ihnen in den jeweiligen Messperioden erledigten Rekurse beurteilt). Es muss wohl nicht ausdrücklich betont werden, dass das ausschliessliche Abstützen von Managemententscheiden auf quantitative Daten zu gravierenden Fehleinschätzungen der Leistungen führen kann.

Der an sich wünschbare Einbezug qualitativer Aspekte öffnet eine ganze Pandorabüchse von Fragen. So ist beispielsweise zu fragen, welche Komponenten der Leistung zu erfassen sind und welches Messniveau angestrebt wird (vgl. den Beitrag von Markus Zürcher, „Wider ein betriebswirtschaftlich verkürztes New Public Management - Über Produkte, Indikatoren und die Wirksamkeit“ in diesem Heft).

Wenn mehrere Qualitätsaspekte unterschieden werden können, stellt sich die Frage, ob ein Gesamturteil über die Qualität zu bilden ist oder ob die Qualitätsaspekte je einzeln ermittelt werden sollen. Für Ersteres ist eine

Gewichtung der einzelnen Qualitätsaspekte und die Skalierung bzw. Indexierung der Einzelurteile notwendig.

Weiter gilt es festzulegen, wie die Daten über Qualitätsaspekte erhoben werden sollen. Einzelne Aspekte der Qualität der Dienstleistung (beispielsweise Bearbeitungszeiten) können auf Grund von Daten aus der Verwaltungsstatistik ermittelt werden; andere Daten müssen aber möglicherweise separat erhoben werden (beispielsweise im Rahmen einer Kundenbefragung oder einer Peer review), was einen entsprechenden Aufwand zur Folge hat.

5. Leistungsmessung ohne Vergleich ist wie ein Fussballmatch mit zweiundzwanzig Bällen

Die Hauptfrage jeder Leistungsmessung lautet: Womit soll die Leistung verglichen werden? Die bisherige Praxis der sog. wirkungsorientierten Verwaltungsführung hat sich „plausibler“ Vorgaben bedient, die (wenn wundert's bei Kenntnis der „20%-80%-Regel“⁸) häufig auf einer Leistungsanforderung von 80 Prozent basieren (z.B. 80% der Kunden müssen mit der Qualität der Dienstleistungen „zufrieden“ sein). Warum nicht ein tieferer oder ein höherer Wert vorgegeben bzw. angepeilt wird, wird nicht transparent gemacht.

Leistungsmessung setzt ein nachvollziehbares Verfahren zur Bildung von Standards voraus⁹. Wenn diese Voraussetzung nicht gegeben ist, liegt der Verdacht nahe, dass das Ganze eine blosse Alibiübung ist.

Nachvollziehbar sind Standards vor allem dann, wenn sie auf Grund von Vergleichen ermittelt wurden oder wenn ihnen solide Erfahrungen aus

⁸ Danach sollen 80% der Leistung in 20% der Zeit und die übrigen 20% der Leistung in 80% der Zeit erzielt werden können. Diese Regel ist indessen weder empirisch abgestützt noch im Zeitalter des „Total quality management“ eine ernstzunehmende Devise für zahlreiche Verwaltungsprozesse.

⁹ So schreibt Picciotto (1999: 16): „The credibility and fairness of evaluation hinges on carrying out fair and rigorous assessment of program goals.“

der Vergangenheit zu Grunde liegen. Dazu können sowohl Längsschnitt- wie auch Querschnittvergleiche benutzt werden.

Bei einem *Längsschnitt-Vergleich* wird die Leistungsmessung in periodischen Abständen vorgenommen. Als Standard kann beispielsweise eine kontinuierliche Verbesserung der Leistungen im Umfang plausibler Produktivitätsfortschritte innerhalb der Verwaltung (z.B. von jährlich 3%) gesetzt werden. Umgekehrt kann im Längsschnitt-Vergleich auch ermittelt werden, ob trotz kontinuierlichem Personalabbau (z.B. 3% pro Jahr bei gleicher Annahme über die Zunahme der Arbeitsproduktivität) ein bestimmtes Leistungsniveau erhalten bleibt.

Beim *Querschnittvergleich* wird die Leistung unter Einheiten desselben Typs verglichen, z.B. unter Schulen, Spitälern, Verkehrsbetrieben, Rechtsabteilungen, Werkhöfen, Zivilschutzabteilungen, Bodenfachstellen. Als expliziter oder impliziter Standard wird die Leistung derjenigen Einheit herangezogen, die in Bezug auf die als relevant erachteten Leistungsmerkmale am besten abschneidet. Bei Bedarf können vertiefte Überlegungen angestellt werden, auf welche Art und Weise die Einheit mit der „best practice“ ihre Leistungen erbringt (z.B. durch Prozessanalysen). Die dabei gewonnenen Erkenntnisse können auch für die anderen Einheiten verwendet werden. In einem solchen Fall führt der Leistungsvergleich zu einem sog. Benchmarking.

Querschnittvergleiche können auf unterschiedlichen Politikaggregierungsebenen angestellt werden. Für eine *Politik auf nationaler Ebene* sind internationale Länderstudien das geeignete Verfahren. Namentlich internationale Organisationen wie die OECD oder die Weltbank nehmen solche Länder-Examinationen vor. Bei einem *Programm* ist die Vergleichsebene ein ähnlich gelagertes Programm. Bei einer Institution (z.B. Mittelschule, Werkhof) dreht sich der Vergleich sinnvollerweise um gleichartige Institutionen. Bei einer Abteilung innerhalb einer Institution wird mit den anderen Abteilungen verglichen. Grundsätzlich bieten sich für den Querschnittvergleich umso bessere Bedingungen, je grösser die Zahl der Vergleichspartner und je ähnlicher sich diese sind. Namentlich bei Vergleichen zwischen Staaten und zwischen den unteren Gliedstaatenebenen (z.B. schweizerische Kantone, deutsche oder österreichische Bundesländer) liegt es an der *Kunst* der mit der Durchführung des Vergleichs betrauten Personen, abzuschätzen, ob und inwieweit die

vergleichende Leistungsmessung durch unterschiedliche Kontextfaktoren verzerrt wird.

6. Ein Leistungsvergleich kann nur dann gerecht sein, wenn Konsens über das zu Vergleichende besteht

Nun ist die Frage noch nicht beantwortet, was denn in einem Querschnittvergleich überhaupt verglichen werden soll. In der Tat wird fast jede Woche in einer Zeitschrift ein Leistungsvergleich präsentiert (beispielsweise von Städten, von Universitäten, von kantonalen Finanzpolitiken). Meist werden dabei bestimmte Einzelaspekte der Leistung herausgepflückt und gemessen. Häufig ist das Gemessene für die Gesamtleistung wenig relevant.

Querschnittvergleiche erhalten dann eine hohe Legitimität, wenn sie auf einem gemeinsamen Verständnis der für die Leistung relevanten quantitativen und qualitativen Faktoren beruhen. Sie setzen also voraus, dass sich die involvierten Parteien darüber einigen, was die bedeutsamen Leistungsmerkmale sind. Welche Quantitäten sind zu erfassen? Wie kann Qualität umschrieben werden? Wie kann Qualität so gemessen werden, dass sie allen Leistungserbringern gleichermaßen gerecht wird.

Um das Beispiel von Universitäten zu nehmen: Wie kann deren Qualität ermittelt werden? Durch Befragung der Studierenden? Durch Befragung der späteren Arbeitgeber? Durch eine Peer review? Durch Messung der Zitate von Publikationen des Lehrkörpers? Kann davon ausgegangen werden, dass das Urteil von Studierenden von Herkunft, politischer Einstellung u.a.m. unbeeinflusst ist? Ist allenfalls davon auszugehen, dass es an gewissen Universitäten bzw. Fakultäten mehr „kritische“ Studenten als an anderen Universitäten gibt, was sich in deren Beurteilung der Qualität der Lehre niederschlägt? Kann davon ausgegangen werden, dass Arbeitgeber in der Privatwirtschaft ihr Urteil nach denselben Kriterien abgeben wie Arbeitgeber im öffentlichen Dienst, oder gibt es auch hier Verzerrungsfaktoren? Soll überhaupt auf Befragungen verzichtet und auf so genannte „harte“ Fakten gesetzt werden (wie z.B. die Wahrscheinlichkeit, dass die Absolventen innert einer gewissen Zeit eine Stelle finden)? Diese und ähnliche Fragen müssen sinnvoll beantwortet werden, wenn

Vergleiche allen einbezogenen Einheiten gerecht werden sollen. Letzteres ist im Übrigen Voraussetzung dafür, dass die Ergebnisse auch akzeptiert und für Prozessverbesserungen genutzt werden?

7. Die Leistung einer Verwaltungseinheit kann durch viele Einflussfaktoren verzerrt werden

Sowohl bei Längsschnitt- wie bei Querschnittvergleichen gibt es, wie am Beispiel der Hochschulen dargelegt wurde, eine ganze Menge von Fallstricken und Fussangeln, welche die Aussagekraft von Indikatoren beeinträchtigen können und welche die Vorkehrungen erfordern, um störende Einflüsse auf die Leistungsmessung „herauszufiltern“.

Beim *Längsschnitt-Vergleich* gibt es die störenden Einflüsse auf Grund saisonaler, konjunktureller oder anderer Schwankungen. So können beispielsweise psychiatrische Kliniken, Wiedereingliederungsstätten der Invalidenversicherung oder Regionale Arbeitsvermittlungsstellen (RAV) bei einer konjunkturellen Abschwächung Schwierigkeiten haben, ihre Standards für die berufliche Wiedereingliederung von Klienten einzuhalten. Im gegenteiligen Fall (positive Konjunkturerwicklung) wird ein günstiges Resultat ausgewiesen, obwohl die Leistung der Dienststellen möglicherweise nicht zugenommen hat. Solche störenden zeitlichen Einflüsse können häufig nur eliminiert werden, wenn auch Querschnittvergleichsdaten verfügbar sind. Bei privaten Betrieben werden aus diesen Gründen in Zeiten konjunktureller Veränderungen die Umsatzzahlen nur als beschränkt aussagekräftig für die Leistung erachtet. Als bedeutsamer werden Daten über Marktanteile angesehen, die auf einem Querschnittvergleich basieren (Vergleich mit den wichtigsten Konkurrenten).

Beim *Querschnittvergleich* liegt das Problem darin, dass zwischen den einbezogenen Einheiten inhärente Unterschiede bestehen, die sich auf die Leistungsmessungsdaten auswirken. In der Evaluationsforschung ist ein solches Phänomen beispielsweise als „Harvard fallacy“ bekannt geworden: Es ist ein Trugschluss zu glauben, dass die Absolventen von Harvard bei Leistungsvergleichen nur deshalb gut abschneiden, weil die Qualität der Ausbildung besonders gut ist. Die Gegenthese lautet wie folgt: Harvard-Absolventen schneiden nur deshalb so gut ab, weil die

Eingangsselektion so hart ist. An diesem Beispiel wird deutlich, dass für die Leistung einer Schule nicht die Qualität des Endprodukts, sondern der durch die Schule erzielte „Mehrwert“ entscheidend ist. Jeder Schul- oder Universitätsvergleich sollte somit bei der Leistungsmessung möglichst störende Einflüsse, z.B. in Form von unterschiedlicher schichtspezifischer Zugehörigkeit der Schüler oder Studierenden, unterschiedlicher Ausländeranteile, unterschiedlicher Geschlechtszugehörigkeit, unterschiedlichen Ein- und Austrittsalters u.v.a.m., zu eliminieren versuchen. Dasselbe gilt auch für andere Institutionen wie Spitäler, Verkehrsbetriebe, Schlachthöfe usw. Die entsprechenden statistischen Methoden, die unter gewissen Umständen eine Identifizierung von Einflüssen erlauben, die sich auf die Endleistungen störend auswirken, setzen in der Regel eine ausreichende Anzahl einzubeziehender Einheiten voraus (die Zahl hängt von der Grösse der Einflüsse und dem Umfang der residuellen Wirkung ab). Bei einer nicht ausreichenden Zahl von Vergleichseinheiten müssen die spezifischen Einflussfaktoren, die einen Leistungsvergleich verzerren, qualitativ umschrieben werden.

8. Die Ergebnisse von Leistungsindikatoren sollen in die Überlegungen einbezogen, aber nicht als Steuerungsmittel verwendet werden

Die Gefahren infolge verzerrender Einflüsse bei Längsschnitt- und Querschnittdaten mahnen dazu, bei der Erarbeitung von Leistungsindikator-Systemen und bei der Veröffentlichung der Resultate behutsam vorzugehen. Es ist sorgfältig zu prüfen, wie Leistungsdaten zu verwenden sind.

Es gibt dabei zwei Probleme: Das erste ist die häufig unsichere Aussagekraft von Leistungsindikatoren, welches im vorangehenden Abschnitt bereits behandelt wurde.

Das zweite Problem ist die Gefahr einer Manipulation von Indikatoren. Alle Daten, die für die Messung von Leistungen verwendet werden, unterliegen der Gefahr, manipuliert zu werden. Der berühmte Evaluationsforscher Campbell schreibt in seinem Aufsatz „Reforms as Experiments“: „The more any social indicator is used for social decision making, the greater the corruptive pressures on it“. Sobald Daten als Leis-

tungsindikatoren verwendet werden – beispielsweise im Hinblick auf die Personalzuteilung oder auf Budgetentscheide – besteht die Gefahr von Missbräuchen. Eine Manipulation kann in Analogie zu den in der Steuerlehre unterschiedenen Tatbeständen (bei denen es ja um die Manipulation des Indikators „Einkommen“ und „Vermögen“ geht) folgende Formen annehmen: Durch Anpassung des Verhaltens an die durch die Indikatoren erfassten Sachverhalte werden *erstens* positiv beurteilte Leistungsbestandteile ausgebaut auf Kosten derjenigen, die nicht erfasst bzw. negativ bewertet werden (Steuervermeidung). Wenn sich beispielsweise die Hochschulfinanzierung bei der Abgeltung von Forschungsleistungen der Universitäten auf den Zitationserfolg abstützt, dann ist abzusehen, dass sich in den entsprechenden Institutionen Zitierseilschaften bilden werden, damit die eigene Universität oder Fakultät finanziell besser abschneidet. Ähnliches kennen wir heute bereits anhand zahlreicher Fälle, beispielsweise dem Ausgabenverhalten von Gemeinden je nach kantonalen Finanzausgleichsregelung (z.B. Bau von Mehrzweckhallen, um eine Senkung des Steuersatzes und eine Verminderung von Finanzausgleichszahlungen zu vermeiden). *Zweitens* können Indikatoren auch verfälscht werden, indem bestimmte Sachverhalte nicht gemeldet, Doppelzählungen nicht eliminiert, bestimmte Kundenkategorien nicht in die Umfragen einbezogen werden u.a.m. (Steuerhinterziehung). Als Drittes und Letztes ist die klare Verfälschung von Indikatoren, z.B. durch manipulierte Stichproben und durch Eingabe falscher Zahlenwerte, zu erwähnen - ein aus dem in der Forschung bestehenden Wettlauf nach interessanten Daten für Publikationen nicht unbekanntes Phänomen („Steuerbetrug“).

Auf Grund der Unsicherheiten über ihre Aussagekraft und auf Grund ihrer Manipulierbarkeit sind Leistungsindikatoren behutsam zu verwenden. Leistungsindikatoren sollen aus wissenschaftlicher Sicht besser dazu genutzt werden, Leistungserstellungsprozesse und ihre wichtigsten Komponenten besser zu verstehen. Nach dem Urteil von Stewart & Walsh sind Leistungsindikatoren vor allem dann nützlich, wenn sie a) als relevant, aber nicht entscheidend betrachtet werden, wenn b) der Bedarf nach einem Fächer von Indikatoren anerkannt wird, wenn c) ein Spannungsverhältnis zwischen den Indikatoren als erwünscht betrachtet wird, da damit der Komplexität öffentlicher Dienstleistungen Rechnung getragen und der politische Kontext anerkannt wird, und wenn d) die technischen Messbarkeitsprobleme, insbesondere die Gewichtung der Indikatoren, anerkannt werden.

9. Leistungsmessung im öffentlichen Sektor wird sinnvollerweise als kooperatives Unternehmen gestaltet

Leistungsmessung im öffentlichen Sektor wird aus all den erwähnten Gründen am besten nicht als Tummelfeld für Verwaltungspioniere, sondern als kooperatives Unternehmen zwischen gleich gelagerten Körperschaften, Politikbereichsangehörigen, Programmverantwortlichen und Angehörigen von Institutionen betrachtet. Je besser die Leistungsmessung koordiniert wird, desto bessere Querschnittvergleichsmöglichkeiten bestehen und desto besser können die technischen Probleme der Bildung und Gewichtung von Indikatoren und Kennzahlen angegangen werden.

Leistungsmessungen sind nicht Endpunkt, sondern Ausgangspunkt für einen ausgedehnten Dialog. Die Leistungsdaten sollen nicht ein endgültiges Urteil ermöglichen, sondern helfen, sinnvolle Fragen zu stellen (Hailstones). Leistungsmessung wird damit zu einem fortdauernden Prozess: Bildung erster Leistungsindikatoren, Dateninterpretation, Suche nach neuen aussagekräftigen Indikatoren, neue erweiterte Dateninterpretation usw. Ein solches prozesshaftes Vorgehen ist in der Privatwirtschaft unter dem Stichwort „Controlling“ seit längerem üblich. So werden etwa in Firmenkonglomeraten wie der ABB zahlreiche Kennzahlen erhoben, in Bezug gebracht zu den jeweiligen Firmenerfolgen, mit den jeweiligen Finanzchefs diskutiert, verfeinert oder ausgebaut. Da zwischen Konzern und Einzelfirmen immer noch ein Abhängigkeitsverhältnis besteht (Einsitznahme von Konzernvertretern im Verwaltungsrat, Abhängigkeit von Finanzmitteln bei Investitionen), ist aber die Situation mit dem öffentlichen Sektor, wo faktisch Gleichstellung der Körperschaften besteht, nur bedingt vergleichbar.

10. Direkte Demokratie und objektive Leistungsdaten vertraugen sich schlecht

Bund, Kantone und Gemeinden in der Schweiz haben der Leistungsmessung im privatwirtschaftlichen Bereich bis anhin nichts Ebenbürtiges entgegenzustellen. Die Schweiz steht sicher nicht an einem Nullpunkt, verfügt sie doch über zahlreiche Output- oder Outcome-Daten. Zu er-

wähnen sind beispielsweise der Umweltschutz, die Raumplanung oder der Verkehr. Dennoch besteht in der Schweiz eine eigenartige Scheu, diese Daten so weiterzuentwickeln, dass Vergleiche zwischen Gemeinden oder zwischen Kantonen möglich würden. Verpönt ist auch, solche Leistungsdaten zum Personalaufwand oder zu den finanziellen Aufwendungen in Bezug zu setzen (sog. Effizienzdaten). So ist das Institut de hautes études en administration publique, Lausanne (IDHEAP) immer wieder mit Widerständen konfrontiert, die öffentlichen Daten der Kantone in die Datenbank BADAG zu übernehmen und damit Querschnittvergleichen zugänglich zu machen. Während die Kantone im steuerlichen Bereich einen harten Wettbewerb austragen (der im Falle individueller Steuererleichterungen allerdings auch nicht dermassen transparent ist), besteht im Bereich öffentlicher Politik eine eigenartige Zurückhaltung, sich dem Vergleich zu stellen. Das Gleiche gilt übrigens auch für den Bund im Hinblick auf Vergleiche mit anderen Staaten.

Es fragt sich, ob die Scheu vor der Transparenz und dem damit einhergehenden Leistungswettbewerb auf den jahrzehntelangen Wohlstand und eine gewisse Saturiertheit zurückzuführen sei. Es bieten sich aber möglicherweise wesentlich relevantere Erklärungsgründe an, namentlich die in der Schweiz vorherrschende Kombination von Föderalismus und direkter Demokratie. Effizienz wird in einem solchen System nicht durch hierarchische Vorgaben hergestellt. Aus diesem Grund sind landesweite Vergleiche öffentlicher Körperschaften durch nationale Institutionen verpönt. So werden etwa in Frankreich landesweite Statistiken der Schulerfolge in den Départements oder in Grossbritannien umfassende Quervergleiche der lokalen Körperschaften veröffentlicht. In der schweizerischen Spielart der Demokratie wird Effizienz nicht durch solche Vergleiche und nachfolgende Eingriffe nationaler Behörden hergestellt, sondern durch eingehende demokratische Kontrolle der öffentlichen Institutionen (beispielsweise der Schulen durch demokratisch gewählte Schulkommissionen oder generell der Gemeinden und der Kantone durch die demokratischen Mitwirkungsrechte [Gesetzes- und Finanzreferendum]). Was auf kommunaler und kantonaler Ebene dermassen eingehend und detailliert debattiert wird, bedarf nicht des Querschnittvergleichs mit anderen Einheiten. Die Demokratie schweizerischer Prägung basiert in ihrer urtümlichen Ideologie ja gerade darauf, dass jede Gemeinde und jeder Kanton seine Eigenart und Unvergleichlichkeit besitzt.

Dies würde auch erklären, warum heute jede Gemeinde und jeder Kanton ihre bzw. seine eigene Spielart moderner Verwaltungsführung erfindet. Die Gemeinden und Kantone werden deshalb auch ihre jeweils eigenen Sets von Indikatoren „erfinden“, und diese Indikatoren werden, weil der Querschnittvergleich vermieden wird, von den entsprechenden Verwaltungseinheiten (vorläufig) weitgehend nach eigenem Gutdünken vorgegeben. Das Resultat dieser Spielart demokratischer Kontrolle ist paradoxerweise eine bürokratische Selbststeuerung über Indikatoren und deren Messung.

Damit besteht zwischen der Leistungsmessung mit Indikatoren, wie sie etwa in Grossbritannien entwickelt worden ist, und der direkten Demokratie ein inhärentes Spannungsverhältnis, das nie ganz aufgelöst werden kann. Die Finanzprobleme öffentlicher Haushalte, die seit langem andauernde Thematisierung der Effizienzfrage und die wieder neu aufgeworfene Frage nach der optimalen Grösse öffentlicher Körperschaften werden jedoch zur Folge haben, dass das Thema „Leistungsmessung“ aktuell bleiben wird. Leistungsmessung - auch über Querschnittvergleiche - wird sich deshalb durch die Prädominanz demokratischer Kontrolle nicht aufhalten lassen, aber zugleich auch unter der erhöhten Politisierung leiden und einen besonders sorgfältig aufgebauten Diskurs erfordern. Im Kanton Solothurn wurde diese Anforderung erkannt und in der Person von Prof. Philippe Mastronardi (Universität St. Gallen) ein externer Begleiter zur Begleitung des Dialogs Regierung/Verwaltung – Parlament beigezogen. Die Begleitung dient namentlich der Ausgestaltung einer optimalen Kontrolle der Verwaltungsreform („Schlanker Staat“) durch das Parlament, insbesondere durch die parlamentarischen Kontrollkommissionen (Geschäftsprüfungskommission und Finanzkommission).

11. Systeme der Leistungsmessung mit Indikatoren haben erhebliche Kosten zur Folge

Aus den vorangegangenen Ausführungen dürfte klar werden, dass allfällige Hoffnungen, Leistungsmessung liesse sich praktisch gratis und spesenfrei durchführen, auf Sand gebaut sind. Dass diese Hoffnungen in der Praxis von Verwaltungsreformen vorherrschen, lässt sich daraus ablesen, dass in den Projektkosten für die Pilotprojekte in so genannter wir-

kungsorientierter Verwaltungsführung keine Kredite für die Entwicklung, Ermittlung und Weiterentwicklung von Indikatoren enthalten sind. In den USA wurden für die Einführung der Leistungsmessung auf Bundesebene (im Rahmen des sog. Government Performance and Results Act 1993) keine signifikanten Kosten vorausgesehen; mittlerweile wird mit Kosten von 50 Mio. Dollar pro Jahr während der Einführung und mit mehreren Hunderttausend Dollar pro Jahr bei einer durchgängigen Verwendung der Leistungsmessung gerechnet (Wholey). Extrapoliert man diese Schätzungen auf die Schweiz, dürften auch hier - allein auf Bundesebene - Kosten in der Höhe von möglicherweise Dutzenden von Millionen Franken anfallen. Sie werden verursacht durch den Arbeitsaufwand für die Bildung der Indikatoren (Beizug externer Experten, Aufwand für Arbeitsgruppen), für die Ermittlung der Zahlen (Verwaltungsvollzugsdaten, Erhebungen bei Privaten, Umfragen), für die Interpretation der Messergebnisse und die Weiterentwicklung der Instrumente.

Werden sich diese Aufwendungen zum Aufbau von Leistungsindikatoren lohnen? Ja, wenn sie aufgewogen werden durch eine höhere Rationalität bei der Entscheidungsfindung, Arbeitssteuerung und Ergebniskontrolle. Ob dies der Fall sein wird, lässt sich erst sagen, wenn die Versuche mit Leistungsindikatoren weiter fortgeschritten sind. Für gesicherte Aussagen ist es noch zu früh. Bereits heute werden Indikatoren für vielerlei Zwecke eingesetzt. Zahlreiche Indikatoren, wie der Finanzkraftindex, der Anschlussgrad an Kläranlagen, die Maturandenquote oder die Zahl der Studierenden, sind nicht mehr aus staatlicher Politik wegzudenken. Nichts spricht dagegen, dass sich in Zukunft noch weitere erhebliche Fortschritte erzielen lassen und dass solche Indikatoren zur Beurteilung von Politiken, Programmen und Institutionen weiter ausgebaut und verfeinert werden. Alles spricht allerdings auch dafür, dass solche Bestrebungen ohne Illusionen über den dafür erforderlichen Aufwand, über die möglichen Gefahren eines Missbrauchs von Indikatoren und deren Resultaten, über die damit einhergehenden mutmasslichen Spannungen mit den bestehenden Instrumenten demokratischer Kontrolle und über die entstehenden Kosten vorgenommen werden.

12. Bibliografie

- BUSSMANN Werner, Pauschal- und Globalbeiträge, *Schweizerisches Zentralblatt für Staats- und Gemeindeverwaltung*, Band 89, Juli 1988, S. 306-316.
- BUSSMANN Werner, Instrumente der Erfolgskontrolle, *VOP* (1995), S. 345-351.
- BUSSMANN Werner, KLÖTI Ulrich & KNOEPFEL Peter (Hgg.), *Einführung in die Politikevaluation*, Basel 1997.
- CAMPBELL Donald T., Reforms as Experiments, in: Weiss Carol H., *Evaluating Action Programs, Readings in Social Action and Education*, Boston 1972, S. 186-223.
- GERMANN Raimund E., *Öffentliche Verwaltung in der Schweiz (Band 1: Der Staatsapparat und die Regierung)*, Bern 1998.
- HAILSTONES Frank, Performance Measures - A Practitioner's Perspective, in: Buschor Ernst & Schedler Kuno, *Perspectives on Performance Measurement and Public Sector Accounting*, Bern 1994, S. 189-196.
- MAWHOOD Caroline, Performance Measurement in the United Kingdom (1985-1995), in: Chelimsky Eleanor & Shadish William R. *Evaluation for the 21st Century: A Handbook*, Thousand Oaks 1997, S. 134-144.
- NASCHOLD Frieder, *Ergebnissteuerung, Wettbewerb, Qualitätspolitik: Entwicklungspfade des öffentlichen Sektors in Europa*, Berlin 1995.
- PICCIOTTO Robert, Towards an Economics of Evaluation. *Evaluation*, 5:1, 1999, S. 7-22.
- STEWART John & WALSH Kieron, *Studies in public service management*, London 1995.
- WHOLEY Joseph, Trends in Performance Measurement: Challenges for Evaluators, in: Chelimsky Eleanor & Shadish William R. *Evaluation for the 21st Century: A Handbook*, Thousand Oaks 1997, S. 124-133.