

Evaluation und New Public Management

Ernst Buschor, Bildungsdirektion des Kantons Zürich

1. Formen des New Public Management im internationalen Kontext

Im Folgenden geht es darum das Instrument der Evaluation im New Public Management zu situieren. Ausgangspunkt bilden die relevanten Merkmale des New Public Managements. In diesem Rahmen sind die unterschiedlichen Rollen der Selbst- und Fremdevaluation zu situieren, die beide in einem Wechsel-, ja „Konkurrenzverhältnis“ zur wachsenden Bedeutung des Qualitätsmanagement und zu den eher controllingorientierten Indikatorsystemen stehen. Daraus ergibt sich m.E. eine Neupositionierung der Evaluation im New Public Management.

Die Merkmale des New Public Management fasse ich einleitend in sechs Punkten zusammen:

1. Zentraler Punkt ist die wettbewerbsorientierte Steuerung. Der Wettbewerb kann auch verwaltungsintern in Form des Benchmarkings oder der freien Wahl unter mehreren Versorgungsträgern des gleichen Gemeinwesens (z.B. Schulen oder Spitäler) durchgeführt werden. Grundsätzlich soll eine Trennung zwischen Finanzierungsträger und Leistungserbringer bestehen. Häufig wird ein Leistungsauftrag definiert, sei es im externen oder im internen Wettbewerb.
2. Es ist von einer Fokussierung auf Effektivität, Effizienz und Qualität der Aufgabenerfüllung auszugehen. Im Vordergrund steht daher häufig der Output, im normativen Modell eher der Outcome. I.d.R. wird eine Delegation bzw. eine Dezentralisierung der Kompetenzen vorgenommen, im Unterschied zur herkömmlichen Verwaltung.
3. Weiter ist die Trennung der strategischen Führung, die etwa der Regierung oder dem Parlament obliegt, von der operativen Führung, d.h. der Verwaltung oder den einzelnen Betrieben charakteristisch.
4. Diese Trennung erfordert instrumentell eine neue Ausstattung. Das herkömmliche kamerale Budget genügt nicht mehr. Man braucht deshalb Formen des Globalbudgets, in denen Leistungen und Ziele benannt und wenn möglich quantifiziert werden und in denen eine pauschale Geldsumme zur Erfüllung der Aufgaben bereit gestellt wird.

5. Zum New Public Management gehört ferner die aktive Förderung der Innovation. Diese ist ein Kernpunkt. New Public Management unterstreicht in aller Form, dass Innovation auf jeder Stufe der Verwaltungsführung notwendig ist, insbesondere auch an der Basis.
6. Und schließlich als letzter Punkt: Die Delegation soll auch in den einzelnen betrieblichen Einheiten durch Dezentralisierung gefördert werden.

Wenn man New Public Management im internationalen Kontext betrachtet, hat es eher den Charakter einer Metatheorie. New Public Management wird je nach Land unterschiedlich definiert. Zu erwähnen ist etwa England mit der Next-Step-Initiative und der Citizen Charter. Neuseeland war lange Zeit das Mekka dieser Entwicklung und baute stark auf dem Privatisierungsansatz auf; heute ist wohl eher Australien führend, das einen stärker managementorientierten Ansatz aufweist. Aber auch die USA, mit den Konzepten des National Performance Review (Bundesstufe) und des Performance Measurement ist zu erwähnen, ebenso wie Deutschland mit seiner neueren Bewegung zum modernen oder „aktivierenden“ Staat (Bund) und mit dem Neuen Steuerungsmodell (KGST). Österreich hat ebenfalls zahlreiche Reformprojekte, und natürlich auch die Schweiz mit der wirkungsorientierten Verwaltungsführung (wif); beide sind eher management- als wettbewerbsorientiert.

Diese Vielfalt der Ansätze lässt sich zwar mit den genannten sechs Kriterien zusammenfassen, aber die Schwerpunkte sind bei den einzelnen Modellen recht unterschiedlich.

Anhand eines deutschen Beispiels – dem Neuen Steuerungsmodell – möchte ich auf die damit verbundenen Chancen hinweisen, generell die Steuerungsfähigkeit der Verwaltungen zu verbessern. Die früheren linearen Kürzungen sollen durch globale Vorgaben abgelöst und durch differenzierter gestaltete Vorgehensweisen ersetzt werden, die insbesondere an der Basis mehr Handlungsspielraum gewähren. Die Effizienz wie auch die Effektivität sollen gesteigert werden. Durch die Einbindung der Stakeholder in die Verwaltung entsteht ein verändertes Zusammenspiel zwischen den Kunden und den Verwaltungseinheiten. In neuerer Zeit wird der Begriff des Kunden allerdings eher vermieden. Man spricht eher vom Bürger als Benutzer: als Person, die Dienstleistungen bezieht. Der Bürger wird aber auch als Leistender gesehen, als Teil einer Gemeinschaft, die auch für den Staat etwas leisten muss. Mit diesem Modell sollen auch Motivationsimpulse ausgelöst werden. Sie tragen den MitarbeiterInnen Rechnung, indem sie ihnen Spielraum gewähren. Der heutige Mitarbeiter oder die Mitarbeiterin legt wesentlich mehr Wert auf solchen Spielraum.

Dies erfordert eine erhöhte Transparenz, die durch das Indikatorsystem, die Kostenrechnung (Kosten-Leistungsrechnung) und teilweise durch Evaluation erreicht werden soll. Damit entsteht ein neues Gleichgewicht, aber auch ein Spannungsverhältnis zwischen Rechtsstaat, Demokratie, Leistungsstaat und Wirtschaftlichkeit. Die Aspekte der Wirtschaftlichkeit und der Wirkungsorientierung gewinnen an Bedeutung und müssen mit den Fragen des Rechtsstaates oder der Rechtsgleichheit in eine neue Balance gebracht werden. Es handelt sich hierbei um normative Kriterien. Im Alltag wird die Balance nicht immer ganz erreicht, aber es ist unerlässlich, dass diese Zielrichtungen wegleitend sind.

Der US Performance and Results Act von 1993 stellt die konsequenteste Orientierung an Leistungszielen oder Outputs dar. Outputs müssen seit 1998 in der gesamten US-Verwaltung quantifiziert und durch Indikatoren gemessen werden. Für die einzelnen Programme sind Leistungsziele auszuarbeiten, für die objektive und quantifizierte Messungen durchzuführen sind. Die Strategien hierzu sind offen zu legen: durchweg werden Indikatoren festgelegt, die jährlich im Hinblick auf Inputs, Outputs und Outcomes zu messen und mit den Vorgaben zu vergleichen sind. In der Praxis werden allerdings überwiegend die Outputs betrachtet (s.u.). Es muss auf Indikatorbasis nachgewiesen werden, ob die vorgegebenen Ziele erfüllt sind. Das Ganze ist schließlich in einem Prozess der Validierung – in neuerer Zeit eher durch Qualitätssicherungsmaßnahmen – zu erhärten. Auf diese Weise lässt sich ein Stärken- und Schwächenprofil erstellen.

Das General Accounting Office of United States hat bei der Betrachtung dieser Stärken- und Schwächenprofile festgestellt, dass die Programmleistungen i.d.R. mit Leistungsindikatoren, also mit Outputs, gemessen werden (General Accounting Office 1999). Damit werden wertvolle Zeitreihen gewonnen, die Darstellung der qualitativen oder inhaltlichen Seite ist aber nur bedingt und z.T. auch gar nicht gelungen. Es besteht die Gefahr, dass kritische Führungsgrößen und Risiken nicht aufgezeigt werden, dass Indikatoren nur einen Teil der Wirklichkeit wiedergeben, und dass insbesondere Querschnittsbezüge unter Programmen zu kurz kommen. Die Darstellung von Zusammenhängen ist mit reinen Indikatorenmodellen sehr schwierig und muss durch flankierende Instrumente – ich erwähne hierbei vor allem die Evaluation, verbessert werden.

Zweifellos hat die Glaubwürdigkeit der Daten durch rigorose Erhebung gewonnen. Es entstand aber auch die Erkenntnis, dass in vielen Fällen Indikatoren ungenügend, ja z.T. willkürlich sind. Es besteht eine deutliche Tendenz zur Messung von Leistungen, weniger zur Messung von Wirkungen, wie sie die Gesetzgebung angestrebt hat.

2. Anforderungen an Indikatoren

Im Hinblick auf die hohe Steuerungsrelevanz der Indikatoren sind an diese hohe Anforderungen zu stellen. Sie müssen *zentrale* Aspekte – insbesondere Wirkungen, aber auch Nebenwirkungen – messen. Dabei muss man von der Tatsache ausgehen, dass *ein* Indikator i.d.R. *kein* Indikator ist. Arbeitet man aber mit zu vielen Indikatoren, wird die Steuerung und die Transparenz erschwert. Das gilt insbesondere gegenüber politischen Instanzen, z.B. dem Parlament, denn eine Vielzahl von Indikatoren ergibt ein komplexes System der Messung. Einseitigkeiten der Messungen sind ebenfalls gefährlich. Wenn etwa in Globalbudgets nur noch ein möglichst hoher Citation-Index als Maßstab für die Qualität der Professoren herangezogen wird, dann sind wir sicher auf dem Holzweg; oder wenn beispielsweise das Parlament im Kanton Zürich die Effizienz der Gymnasien nur noch an der unmittelbaren Übertrittsquote an die Universität messen will, sind wesentliche Dimensionen des Abiturs nicht erfasst. Die Gefahr, dass einseitige, ja verfehlte Akzentuierungen auf politischer Ebene entstehen, sollte man nicht unterschätzen, wenngleich das Positive

darin besteht, dass nicht nur über Franken oder Euro sondern auch über Leistungen gesprochen wird.

Indikatoren müssen zweitens *valide* sein, auch hier sind die Anforderungen hoch anzusetzen. Viele Indikatoren erfüllen das kaum. Sie sollten weiter *einfach, treffend und klar* sein. Das ist leichter gesagt als getan. Bei vielen Aufgabenbereichen oder Leistungen kann man auf eine gewisse Komplexität nicht verzichten. Als Beispiel erwähne ich den Lebenshaltungskostenindex. Er ist einer der komplexesten Indizes, den wir aber praktisch täglich gebrauchen, obwohl ihn nur wenige verstehen. Ein Indikator muss *verlässlich* und *genau sein*. Er muss – das ist sehr zentral – *rechtzeitig* zur Verfügung stehen. Die Regierung muss z.B. im Monat Mai ihren Geschäfts- und Rechnungsbericht veröffentlichen. Daten, die zu dieser Zeit nicht zur Verfügung stehen, nützen nichts für die Berichterstattung. Allerdings ist diese rechtzeitige Erstellung oft sehr schwierig, insbesondere wenn sie in valider Form bereitgestellt werden soll. Weiter sollte das Messsystem *vollständig* sein. Diese Anforderung ist meistens eine Illusion, aber sie gilt als Messlatte für Indikatorsysteme.

Indikatoren müssen siebtens *Vertraulichkeitsaspekte* berücksichtigen und die Persönlichkeitsschutzvorschriften beachten. Insbesondere die Rückführbarkeit auf eine Person muss ausgeschlossen werden. In der Praxis kann das schnell zu Problemen führen. Wir hatten ein solches Problem im Kanton Zürich am Beispiel der altsprachlichen Matura (Abschluss mit Griechisch). In diesem Indikator „steckten“ drei Lehrkräfte, in allen übrigen Fächern natürlich Hunderte. Solche Risiken sind sehr sorgfältig zu betrachten und zu analysieren. Als weiteren bedeutsamen Aspekt möchte ich die *Erhebungswirtschaftlichkeit* unterstreichen. Bei vielen Indikatoren besteht ein Missverhältnis zwischen dem Beitrag zur Lösung von Führungsaufgaben und den Kosten der Erstellung. Hier muss man mit großer Sorgfalt vorgehen.

Wenn wir die *Grenzen* der Leistungsindikatoren betrachten, dann kann man feststellen, dass sie zwar wichtige und nützliche Informationen liefern, dass sie aber nicht auf allen Gebieten gleichwertig und zielführend quantifiziert werden können. Hier bestehen erhebliche Unterschiede zwischen den verschiedenen Aufgabengebieten der öffentlichen Verwaltung. Leistungsindikatoren haben ferner durchaus politische Bedeutung, wobei die Evaluatoren oder die Programmverantwortlichen häufig nicht bestimmen können, welche Indikatoren das politische System als zentrale Größen herausnimmt. Die Auswahl und die Gewichtung der einzelnen Indikatoren kann sich völlig anders entwickeln als die Programmverantwortlichen sich dies wünschen. Gerade in Zeiten der Finanzknappheit besteht die Tendenz, Indikatoren die Output-Wirtschaftlichkeit messen, in den Vordergrund zu stellen. Man stellt z.B. fest, dass das Parlament vor allem an den Kosten pro Student, an den Kosten pro Bewilligung oder etwa an den Kosten pro medizinischer Behandlung interessiert ist und weniger an den Indikatoren der damit verbundenen Qualität. Das ist durch die Behörden oder die Verwaltung nicht steuerbar. In den USA wird immer deutlicher vom General Accounting Office unterstrichen, dass Indikatoren und selbst auch Evaluationen nicht genügen, dass daneben vor allem Führung wichtig ist. CAIDEN (1998) spricht sogar von „disappointing limited value“ der Evaluation und Leistungsmessung. Das scheint mir allerdings etwas pessimistisch.

Die Leistungsmessung hat typische Gefahren. Eine wichtige Gefahr ist der sog. *Tunnelblick*: man beschränkt sich auf das Nächstliegende. Das kann bis zur inversen Motivation reichen. Wird z.B. die Optimierung des Bußgeldertrags bei der Po-

lizei als Zielgröße gesetzt, kann das dazu führen, dass die Polizei Radarkontrollen an ungefährlichen Stellen mit relativ schnellem Verkehr macht. Der eigentliche Zweck der Radarkontrolle, nämlich die Senkung der Unfallhäufigkeit, steht dann nicht mehr im Zentrum.

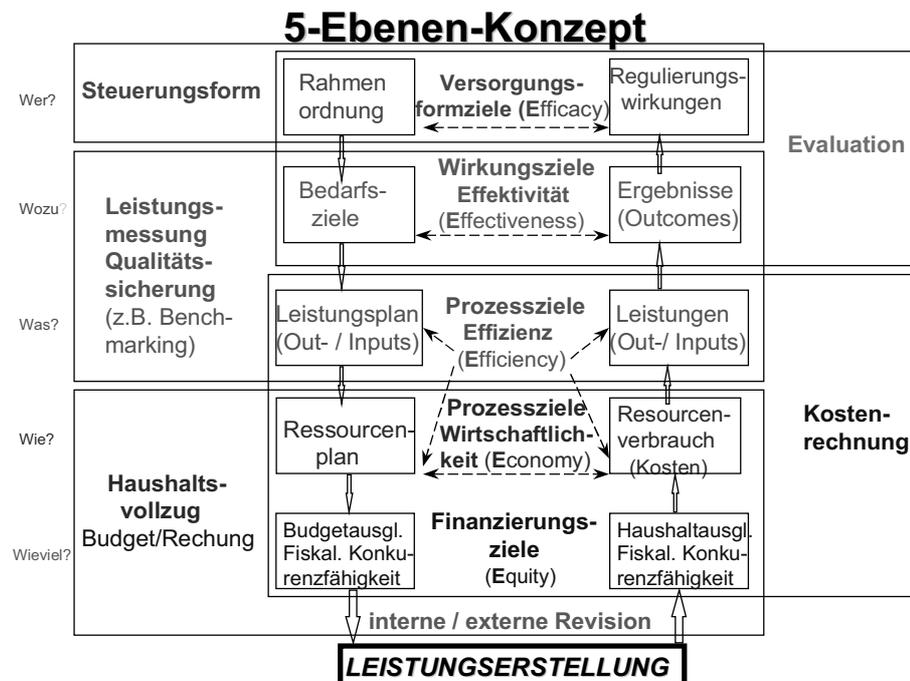
Die *Kurzsichtigkeit* ist eine weitere Gefahr, indem Indikatoren kritische Verhaltensformen auslösen können. So klagt heute das Universitätsspital darüber, dass schwierige Fälle von den übrigen Spitälern bewusst an die Universitätsklinik abgeschoben werden, um ihre Statistik der Durchschnittskosten nicht durch die Behandlung kostspieliger Fälle mit schlecht gedeckten Fallpauschalen zu schädigen. Wie weit das übertrieben ist bzw. zutrifft, müsste man zum Gegenstand einer Evaluation machen. (Ich habe andererseits festgestellt, als ich noch als Gesundheitsdirektor mit den Spitälern sprach, dass es nach Meinung der Spitalleitungen nur Spitäler mit „schwierigen Fällen“ gab.) Eine weitere Gefahr besteht im *Konvergenzdenken*. Man orientiert sich am Durchschnitt statt an den Spitzenleistungen. Das kann so weit gehen, dass Spitäler nach der Normtagezahl Patienten entlassen. Wenn man sich nur noch auf die Indikatoren konzentriert, kann Erstarrung eintreten, die Innovation vernachlässigt. Oder man betreibt Spielereien, indem man sich im System bequem einzurichten versucht und nach Indikatoren sucht, die am ehesten zu Effekten führen, die bequem erfüllt werden können. Ein solches Beispiel sind die Schülerpauschalen. Hier wurde befürchtet, dass sie dazu führen, dass die Schulen ihre Schule einfach füllen, um ein Maximum an Pauschalen auszuschöpfen. Ich befürworte aus solchen Gründen daher das prospektive Globalbudget, mit dem nicht einfach Ist-Zahlen übernommen werden, sondern vorgängig eine Schülerzahl als Grundlage der Entschädigung festgelegt wird. Schließlich besteht auch die Gefahr der bewussten *Irreführung*. Es kommt vor, dass man Indikatoren, die schlüsselrelevant sind, revidieren muss, weil mit falschen Angaben gearbeitet wird.

3. Das 5-Ebenen-Modell

Wir haben im Kanton Zürich ein Führungsmodell entwickelt, das auf fünf Ebenen arbeitet. Wenn man eine öffentliche Aufgabe erfüllt, müssen fünf Fragen beantwortet werden: Wer tut etwas? Wozu tut er/sie das? Was leistet er/sie, um das „Wozu“ zu erreichen? Wie setzt er/sie die Ressourcen ein? Und wie viel Mittel werden ihm/ihr zur Verfügung gestellt? Diese fünf Fragestellungen muss man bei jedem Programm beantworten. Wir bezeichnen sie auch als Ebene E1, E2, usw. Diese tauchen dann in den Globalbudgets und in den Entwicklungsplänen auf und führen zu entsprechenden instrumentellen Einsätzen. Der Bereich des „Wozu“, der Wirkungssteuerung oder auch der Bereich der Steuerungsformen (z.B. Aufgabenerfüllung durch die Verwaltung, selbständige Anstalt oder Leistungsvereinbarung mit privaten Trägern) erfordern eher Instrumente wie Evaluation. Der Bereich der Verwaltungs- oder Betriebssteuerung und der Bereich der Ressourcenoptimierung verlangen Instrumente wie Leistungsmessung und Kostenrechnung, die zwingend mit eingeführt werden müssen. Schließlich erfordert die Einhaltung der Budgets teilweise einen Link zu herkömmlichen Formen der Haushaltsführung.

Man kann sich das in einem Kreislauf vorstellen. Die SOLL-Seite steht links und die IST-Seite rechts (Vorgabe und erreichter Stand). Auf diese Weise lässt sich

dann auf den einzelnen Ebenen jeweils der Vergleich ziehen, der im Rahmen der Rechnungs- und Geschäftsberichte enthalten sein soll.



Die Evaluation hat eine steigende Bedeutung, insbesondere bei der Frage nach dem „Wozu“ (Wirkung) und bei den Formen der grundsätzlichen Steuerung.

4. Standards für Evaluationen

Die Entwicklung von Evaluationsstandards und die Gründung der Schweizerischen Evaluationsgesellschaft (SEVAL), an denen THOMAS WIDMER maßgeblich mitgewirkt hat¹, ist ein wichtiger Beitrag zur Professionalisierung der Evaluation in der Schweiz. Die Standards sind im Wesentlichen übernommen vom Joint Committee on Standards for Educational Evaluation der USA (1999). Sie umfassen vier Dimensionen:

- *Nützlichkeit.* Es geht um die Regelung von Stakeholderorientierung, klare Zielentscheidungen, Glaubwürdigkeit und Prozesstransparenz des ganzen Vorhabens, Vollständigkeit der Analyse und der Daten, Rechtzeitigkeit und Wirkungsorientierung. Ein zweiter Block umfasst die

1 Vgl. – www.seval.ch

- *Durchführbarkeit*. Hierbei geht es um praktikable Arbeiten, die tragfähig und kostenmäßig vernünftig sind, d.h. sie müssen ein Verhältnis von Kosten und Nutzen aufweisen, das tatsächlich vertretbar ist. Drittens ist die
- *Korrektheit* zu erwähnen. Es muss der Vertrag eingehalten werden, es muss der erforderliche Anstand gewahrt und der Persönlichkeitsschutz respektiert werden; die Daten müssen vollständig und der Prozess fair sein. D.h. die verschiedenen betroffenen Stellen sind anzuhören und Interessenkonflikte sind ggf. offen darzulegen. Schließlich die vierte Gruppe von Standards – die
- *Genauigkeit*. Das Programm muss klar und genau dokumentiert sein, der Kontext muss beschrieben werden, Informationen müssen verlässlich, valide und systematisch sein, die Schlussfolgerungen müssen begründet und unparteiisch sein.

Wir verwenden heute diese Standards sogar als Teil des Vertrags, wenn wir Evaluationsaufträge erteilen. Sie werden allenfalls bedeutsam, wenn es zu Honorarkürzungen kommt wegen mangelnder Leistungsfähigkeit der Evaluation. Insofern sind solche Standards durchaus rechtlich relevant.

5. Selbst- und Fremdevaluation

Die Erfahrungen zeigen, dass externe Evaluationen vor allem dann erfolgreich und nützlich sind, wenn sie durch interne *Selbstevaluation* unterstützt werden. Die interne Selbstevaluation ist zweifellos i.d.R. methodisch schwächer. Sie hat aber eine hohe Akzeptanz, weil sie von den Akteuren selbst getragen ist und weil die internen Strukturen und Unterlagen bekannt sind. Da die Evaluation Kenntnisse der Evaluationsmethodik voraussetzt, muss unterstrichen werden, dass eine entsprechende Ausbildung hierfür erforderlich ist. Ist diese nicht vorhanden, so leidet die Glaubwürdigkeit der Resultate nach aussen. Die Selbstevaluation hat entscheidende Vorteile aber auch entscheidende Nachteile. Bezüglich dieser Vor- und Nachteile orientiere ich mich im Folgenden am Katalog von DUGAN (1996).

Potentiale und Begabungen, die vorhanden sind, werden bei der Selbstevaluation recht gut ausgeschöpft, vor allem, wenn eine entsprechende Ausbildung vorausgesetzt werden kann. Sie erhöht auch – was mir sehr wesentlich erscheint – die Akzeptanz und Wirksamkeit von externer Evaluation, weil die Akteure mit solchen Instrumenten vertraut sind; man hat damit einen Echoeffekt, der sehr wesentlich ist. Vor allem ist es auch ein Rationalisierungsgewinn in dem Sinne, dass externe Evaluationen dort leichter durchzuführen sind, wo eine Kultur der internen Evaluation oder Selbstevaluation bereits besteht. Dadurch werden Kosten gespart. Die Grenzen zeigen sich allerdings bei sehr anspruchsvollen Programmen, es sei denn, dass entsprechend hochqualifizierte Stellen geschaffen werden. Es braucht bei der Selbstevaluation ferner eine intensive Kommunikation – mehr als bei der externen Evaluation – das kann dann allerdings auch zu sehr guten Ergebnissen führen. Die Umsetzung der Ergebnisse ist auf dieser Basis eher gewährleistet als durch externe Evaluation, zumal auch umfeldgerechtes Material zur Verfügung gestellt werden kann.

Bei den Nachteilen ist darauf hinzuweisen, dass die methodische Rigorosität häufig fehlt, wenn eine entsprechende Ausbildung und die Kenntnis der Standards

nicht vorhanden sind. Das muss als die größten Schwachpunkte der internen Evaluation angesehen werden. Schließlich ergeben sich auch bei Auswertungen Probleme, vor allem wenn Indikatoren verglichen werden müssen und wenn gehobene Ansätze erforderlich wären. Es bleibt häufig bei Leistungsindikatoren, die Wirkung wird tendenziell vernachlässigt (wie dies beim New Public Management der Tendenz nach zutrifft). Es fallen zudem nicht selten erhebliche Einführungskosten an.

Im Gegensatz dazu hat die *Fremdevaluation* den Vorteil, dass die EvaluatorInnen über gute methodische Kenntnisse und über eine hohe Unabhängigkeit verfügen. Sie hat aber den Nachteil, dass der Untersuchungsgegenstand nicht immer genügend bekannt ist, und dass infolgedessen eine Einarbeitungszeit erforderlich ist. Insofern entsteht ein größerer Zeitbedarf, bis eine Evaluation tatsächlich realisiert und politisch in der Diskussion umgesetzt werden kann.

Dazu liegt in der Schweiz auf Bundesstufe eine Untersuchung aus dem Jahr 1989 vor, welche im Rahmen der Vorarbeiten zu einem nationalen Forschungsprogramm über die „Wirksamkeit staatlicher Massnahmen“ zur Informationsbeschaffung des Bundesgesetzgebers erstellt wurde (Klöti und Schneider 1989). Sie zeigte wesentliche Schwächen auf. Diese Untersuchung ergab, dass von 20 Evaluationen auf Bundesstufe nur eine einzige bei der Entscheidung tatsächlich eine Rolle spielte: der Entscheid über die Tempolimitierung im Innerortsverkehr in der Schweiz. In allen übrigen 19 Fällen ist keine signifikante Wirkung der Evaluation auf den Entscheidungsprozess nachweisbar. Als Grund hierfür ist die schweizerische Tendenz zu vorparlamentarischen Kompromissen und zu neo-korporatistischen Gremien zu sehen. Es existieren auch Hemmnisse, die im System liegen, wie die Begrenzung auf das Machbare. Evaluationen werden i.d.R. zu spät in Auftrag gegeben bzw. vergeben. Sie müssen dann in kürzester Zeit in reduzierter Form durchgeführt werden. Das führt zu Schwierigkeiten in der Abwicklung und zu Qualitätsverlusten. Ein weiteres Problem kann auftauchen, wenn die Verwaltung versucht, die Entscheidung zu beeinflussen. Damit ist häufig die Unabhängigkeit der Evaluation in Frage gestellt.

Es ist aber auch auf z.T. *unattraktive Forschungs- und Ausbildungsbedingungen* in der Schweiz hinzuweisen. Evaluationsaufträge stellen hohe fachliche Ansprüche und erfordern daher Spezialisten, die in der Regel nicht ausgelastet werden können, vor allem, wenn wegen zu später Auftragserteilung auf einen raschen Abschluss der Evaluationen gedrängt werden muss. Einiges ist nicht optimal, aber wir haben immerhin einige Verbesserungen erreicht.

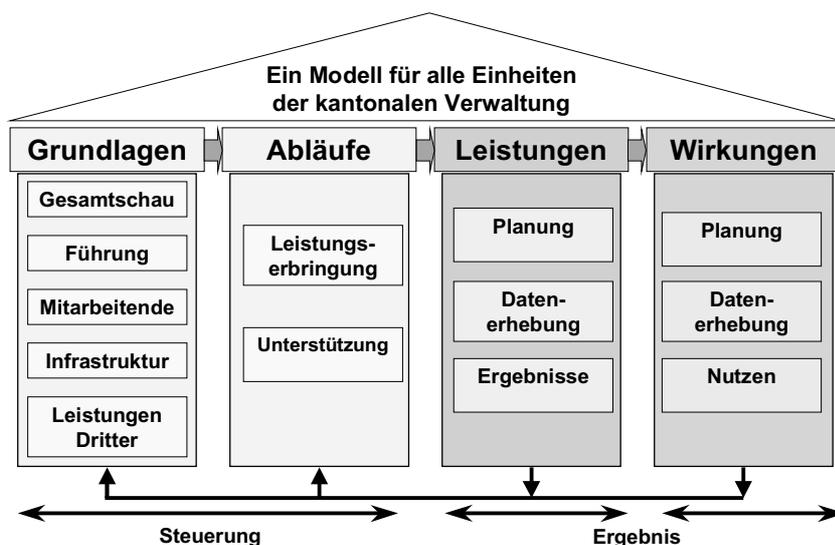
Eine weitere Wirksamkeitsbarriere besteht darin, unliebsame Ergebnisse nicht zu veröffentlichen oder nur so weit sie „liebsam“ sind. Schließlich sei auf die Tatsache hingewiesen, dass Parlamente im Prozess der Gesetzgebung im Allgemeinen wesentlich mehr von der Verwaltung beeinflusst werden als von Evaluationen – vor allem, wenn deren Ergebnisse der Verwaltung nicht genehm sind (verwaltungsgeprägte Parlamentsinformation).

6. Die Rolle des Qualitätsmanagements in New Public Management

Dies hat dazu geführt, dass die Defizite der reinen Leistungsorientierung durch Ansätze des Qualitätsmanagements ergänzt werden. Heute besteht in der Verwaltung eine klare Tendenz, diese Ansätze zu integrieren, wobei der Begriff „Qualität“ schillernd ist. Ich möchte sie hier definieren als „möglichst hoher Gebrauchs- oder Lebensnutzen“², also *fitness for use* oder *fitness for life*. Dieser Lebensnutzen wird vor allem im Bildungs- und Gesundheitswesen deutlich, denn Bildung hat keinen Gebrauchsnutzen, aber Bildung hat einen Lebensnutzen. Gleichzeitig gilt im Qualitätsmanagement, dass die Ressourcensituation zentral ist, dass also mit möglichst wenig Ressourcen möglichst viel Qualität erzeugt werden soll.

Wir haben das Züricher Qualitätsmodell angelehnt an das EFQM-Modell. Jeder Mitarbeiter, jede Mitarbeiterin und alle Ämter haben eine Kopie des Züricher Qualitätsmodells erhalten, es hängt in vielen Gängen und Büros.

Zürcher Qualitätsmodell in Anlehnung an das EFQM- Modell



Betrachten wir in der Graphik die *Grundlagen*, so stellen sich u.a. folgende Fragen: „Wie wird geführt?“, „Wie geht man um mit den MitarbeiterInnen?“, „Wie sieht die Infrastruktur aus?“ und „Was produzieren wir an Leistungen?“ Die *Abläufe* werden dann im Sinne der Prozessanalyse analysiert. Das ist die Steuerungsseite

2 Vgl. Qualitätsbild Kanton Zürich 2001. – www.wif.zh.ch

wie beim EFQM-Modell. Die Leistungsseite beinhaltet folgende Fragen: „Wie planen wir?“ „Wie beschaffen wir Daten?“ und „Wie gehen wir mit Ergebnissen bzw. Nutzen bei den Wirkungen um?“. Es ist somit ein vereinfachtes EFQM-Modell. Wir stellen eine bewusste Verbindung her zwischen Globalbudget, Planung und Qualitätsmanagement.

Instrumente der Wirkungssteuerung

	Controlling/ Qualitäts- management	Prospektive Evaluation	Retrospektive Evaluation	Revision (intern/extern)
Zielsetzung	strategische u. operative Steuerung rasche Feedbacks	umfassende Maßnahmenabschätzung	Kenntnis gewollter u. ungewollter Wirkungen bessere Controllingmaßstäbe	Prozessunabhängige Prüfung Ordnungsmäßigkeit, Wirtschaftlichkeit (u. Wirksamkeit)
Hauptformen	Berichtswesen Zeitreihenvergl. Benchmarking Qualitätssicherung	Szenarien Modelle Hypothese	Policy on/off Vergleich Zeitreihenanal. Quervergleiche (quasi-Experim.)	Rechnungsprüfung Wirtschaftlichkeitsanalysen (Kurz-Evaluationen)
Kosten	niedrig/mittel	hoch	hoch	mittel/hoch
Stärken	Kostengünstige Wirkungssteuerung rasche Feedbacks	Breiter Wirkungseinblick kausale Faktorenkenntnis	Wirkungskennntnis gewollter u. ungewollter Wirkungen besseres Controlling	Präventivwirkung Problemfrüherkennung
Schwächen	Fehlsteuerung (falsch/fehlende) Steuerungsgrößen	teuer zeitaufwendig nicht bereichsdeckende möglich	teuer zeitaufwendig nicht bereichsdeckend möglich	ressourcenorientiert Stichproben (formalistisch)

In Deutschland, im Projekt des „modernen Staates“ steht das sog. CAF-Modell im Vordergrund³. Das Common Assessment Framework arbeitet mit ähnlichen Modulen wie das Zürcher Modell, nämlich Führungsebene, Organisationspolitik und -strategie, Personalmanagement, externe Partnerschaften und Ressourcen, Prozess- und Veränderungsmanagement, Kunden- bzw. Bürgerorientierte Resultate, MitarbeiterInnenzufriedenheit, Auswirkungen auf die Gesellschaft, u.a.. Sie finden also diese Modelle in verschiedensten Formen. Ich halte deshalb die Auseinandersetzung um das beste Modell für relativ sinnlos – sei es EFQM, ISO 9000, CAF oder andere. Solche „Glaubenskriege“ sind z.T. Marketingkriege. Wichtig ist, dass Eva-

3 Common Assessment Framework, vgl. – www.staat-modern.de

luatorInnen die Fragen des Qualitätsmanagements kennen, denn die Verwaltung arbeitet zunehmend damit.

Insofern dürfen auch die Unterschiede zwischen Evaluation und Qualitätsmanagement nicht übersehen werden. Evaluation ist i.d.R. thematisch fokussierter, zeitlich begrenzt, und es ist unmöglich, eine Verwaltung flächendeckend zu evaluieren. Beim Qualitätsmanagement streben wir dagegen eine kontinuierliche, nicht eine zeitlich unterbrochene Verbesserung an; Qualitätsmanagement findet regelmäßig und grundsätzlich flächendeckend statt. Das hat auch Auswirkungen auf die Instrumente. Sie sind teils identisch, teils ähnlich, teils verschieden.

Das klassische Controlling, die Qualitätssicherung und das Qualitätsmanagement, streben vor allem eine flächendeckende repetitive Steuerung und Verbesserung des Systems zu möglichst niedrigen Kosten an. Wir müssen allerdings auch Schwächen dieser Instrumente bei der Wirkungssteuerung im Sinne von Fehlsteuerungen einräumen. Aber sie sind kostengünstiger als Evaluationsansätze, ob prospektiv oder retrospektiv. Die Evaluation hat dort entscheidende Vorteile, wo sie die Impacts analysiert – die in den übrigen Modellen kaum zum Tragen kommen – und damit wesentlich tiefergehende Informationen bereit stellt, allerdings zu einem höheren Preis. Evaluationen sind im Vergleich zu den übrigen Instrumenten teurer.

Schließlich existieren noch die Bereiche der Revision, der Rechnungshöfe und der internen Revision, die ebenfalls stark an der Wirtschaftlichkeit orientiert sind.

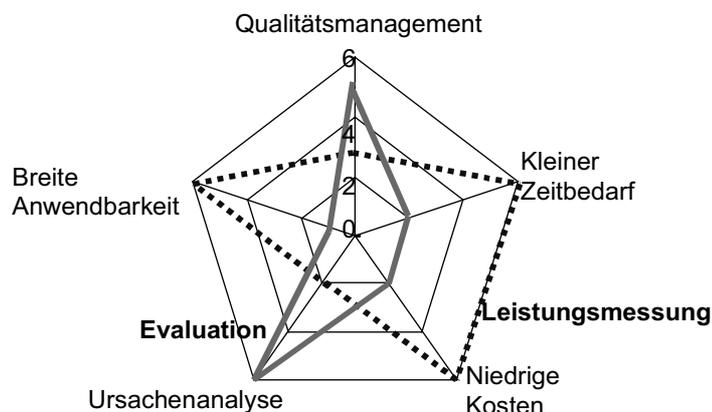
7. Evaluation, Leistungsmessung und die Bedeutung der Führung

Wesentlich ist, dass in einem Führungsmodell all die oben erwähnten Instrumente zum Zug kommen. Es ergeben sich daraus einige Regeln des Evaluationsvollzugs

1. Evaluationen dürfen nicht zu lange dauern. Sie benötigen i.d.R. aber bis gegen zwei Jahre. Das hat für das Management zur Folge, dass man zwei Jahre vor dem Entscheid der Regierung die Evaluation einleiten muss. Es macht wenig Sinn, wenn ein Jahr nach dem Regierungsentscheid die Evaluation eintrifft, wenn Entscheidungen schon gefällt sind.
2. Es braucht eine gute Begleitung einschließlich des Willens, den EvaluatorInnen Selbständigkeit einzuräumen. Eigene Interessen bzw. deren Dominanz müssen dabei abgebaut werden.
3. Es scheint wichtig, dass EvaluatorInnen Gelegenheit bekommen, mit den politischen Entscheidungsträgern ins Gespräch zu treten, damit die Kriterien, die eine politische Rolle spielen, online und life diskutiert werden können. Ich weiß, dass dies sehr schwierig zu erfüllen ist. Aber es ist entscheidend, sonst laufen Evaluationen leicht in eine Richtung, die zwar die Verwaltung will, die aber nicht den politischen Vorstellungen entsprechen.
4. Die finanziellen und zeitlichen Rahmenbedingungen müssen in Bezug auf den Auftrag eingehalten werden. Hier wird manches versäumt, indem schwammige, unpräzise Aufträge erteilt werden.
5. Die Fragestellung muss *vor* der Auftragsvergabe geklärt sein.

6. Es muss ein klares Auftragskonzept bestehen.
7. Die Zwischenergebnisse sind periodisch mit den Betroffenen zu besprechen. Die Begleitevaluation hat große Vorzüge. Meines Erachtens müssen größere Evaluationen mit Zwischenevaluationsabschlüssen durchgeführt werden, damit auch ein Lerneffekt im Prozess entsteht. Das bringt aber den Evaluator bzw. die Evaluatorin in eine schwierige Rolle, weil er/sie dann Teil des Prozesses und nicht nur Beobachter ist. Aber damit müssen EvaluatorInnen leben können.
8. Verlangt wird ferner eine frühzeitige Absprache über die Datenbeschaffung, denn dort geht die meiste Zeit verloren.
9. Es müssen die wichtigen Aspekte der Beteiligten und Betroffenen wie auch die Nebenwirkungen zum Tragen kommen.
10. Der Schlussbericht muss in verständlicher Sprache abgefasst werden. Es wäre schade, wenn Politiker den Bericht auf S. 3 beiseite legen und die ganze Evaluation wäre weitgehend umsonst. Es muss des weiteren eine Zusammenfassung angefertigt werden.
11. Die Datenquellen müssen exakt beschrieben werden.

Profil von Evaluation und Leistungsmessung



Die Abbildung zeigt, dass die Profile von Evaluation und Leistungsmessung ergänzende Merkmale haben. New Public Management benötigt zukünftig mehr Evaluation, weil man sonst weitgehend bei der Leistungsebene verbleibt und die Wirkungsebene zu wenig zum Tragen kommt.

Beides ist wiederum einzuordnen in den Führungsprozess. Ich erwähne den Appell von US-Vizepräsident GORE: „Maßgebend ist und bleibt gute Führung“. In einem klaren Konzept, wie geführt und wie gemessen wird, ist Kommunikation zentral und bedarf entsprechender Resultatverantwortungen. Indikatoren allein genügen nicht, sondern Wirkungen müssen – da haben wir nichts Besseres – immer noch qualitativ und verbal beschrieben werden.

Damit Evaluationsergebnisse umgesetzt werden, braucht man auch positive Anreize innerhalb der Verwaltung sowie einen kontinuierlichen Prozess im Sinne des Qualitätsmanagements. In einem Führungsprozess von hoher Rationalität erbringt die Evaluation (geschätzt) einen Rationalitätsgewinn von 20 – 25%. Die übrigen rund 75% liegen in der Führung, Leistungsindikatoren und in andern Instrumenten. Aber immerhin: wenn die Qualität unserer Entscheidungen dank Evaluation ein Fünftel bis ein Viertel besser wird, dann haben wir sehr viel erreicht.

Literatur

- Caiden, N. (1998): Public Service Professionalism for Performance Measurement and Evaluation. In: Public Budgeting & Finance, Vol. 18 Nr. 2: 43ff.
- Dugan, M. A. (1996): Participatory and Empowerment Evaluation: Lessons Learned in Training and Technical Assistance, in: Empowerment Evaluation, Knowledge and Tools for Self-Assessment and Accountability, ed. D. M. Fetterman/S.J.Kaftarian/A. Wandersman, Thousand Oaks/London/New Dehli (Sage), p. 296 – 300.
- General Accounting Office of United States (1999): Managing for Results, Opportunities for Continued Improvements in Agencies' Performance Plans. GAO/GGD/AIMD-99-215: 3.
- Joint Committee on Standards of Educational Evaluation. James Sander (Hg.) (1999): Handbuch der Evaluationsstandards. Aus dem Amerikanischen übersetzt von W. Beywl und Th. Widmer. Deutsche Ausgabe bearbeitet und ergänzt von W. Beywl, Th. Widmer und J. R. Sander. Leske und Budrich.
- Klöti, U. & Schneider, G.(1989): Die Informationsbeschaffung des Gesetzgebers. Eine vergleichende Evaluationsstudie zur Bedeutung von Wirkungsanalysen im Schweizer Legiferierungsprozess. Grösch: Rüeggger.