

Evaluation in der Politik

Thomas Widmer
 Institut für Politikwissenschaft
 Universität Zürich

Die vorliegende Fassung ist für die Verbreitung um einige Seiten gekürzt.

Tagung „Evaluation und politische Gestaltungs- und Entscheidungsprozesse“
 Forschungsinstitut gesellschaftliche Weiterentwicklung FGW & AK Soziale Dienstleistungen DeGEval
 Düsseldorf, 15./16. März 2016

1. Zwei Vorbemerkungen

Struktur

1. Zwei Vorbemerkungen
2. Wandel von Staatlichkeit – Wandel staatlicher Steuerung
3. Fünf Thesen zur Performanz-Messung
4. Fünf Thesen zu Genese und Nutzung von Evaluation in der Politik

Fürst Sapieha

„Was ist die Mehrheit? Mehrheit ist der Unsinn;
 Verstand ist stets bei Wen'gen nur gewesen.
 [...]
 Der Staat muss untergehen, früh oder spät,
 Wo Mehrheit siegt und Unverstand entscheidet.“

Friedrich Schiller (1805): Demetrius.

Schweiz

5

Thomas Widmer, Institut für Politikwissenschaft, Universität Zürich

Wandel von Staatlichkeit

- Rückzug des Staates im Zusammenhang mit Deregulierung, Liberalisierung und Privatisierung („Leitbild ‚Schlanker Staat‘“; Werner Jann)
- Verlust an Monopolherrschaft durch Delegation an Dritte („Monopolist a. D.“; Bernhard Zangl)
- Auflösung der Grenzen zwischen Staat und Privat („Zerfaserung“; Michael Zürn/Stephan Leibfried)
- Kooperative Steuerung substituiert hierarchische Steuerung (Kooperation mit Zivilgesellschaft resp. Wirtschaft; Netzwerke; „Public-Private-Partnership“; „government“ – „governance“) („kooperative Staat“; Rüdiger Voigt, Arthur Benz).

7

Thomas Widmer, Institut für Politikwissenschaft, Universität Zürich

2. Wandel von Staatlichkeit – Wandel staatlicher Steuerung

6

Thomas Widmer, Institut für Politikwissenschaft, Universität Zürich

Wandel von Staatlichkeit – Fortsetzung

- „Vertikale Entnationalisierung“ durch Transfer der Steuerungskompetenzen von der nationalen zur internationalen oder subnationalen Ebene
- „Horizontale Segmentierung“ durch steigende Komplexität in Mehrebenenkonstruktionen (namentlich im Rahmen der Europäischen Integration, aber auch in föderalistisch-strukturierten politischen Systemen).

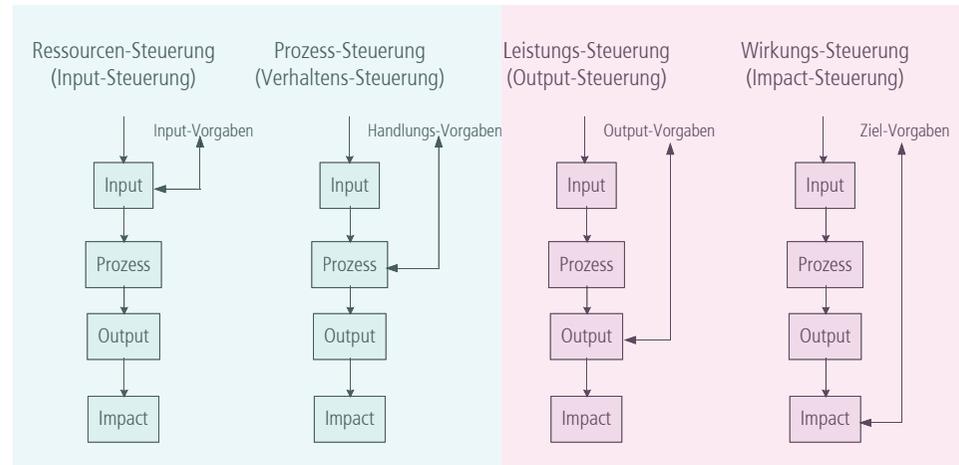
Veränderte Staatlichkeit

→ Veränderte Formen staatlicher Steuerung

8

Thomas Widmer, Institut für Politikwissenschaft, Universität Zürich

Formen der staatlichen Steuerung



Quelle: Widmer, Thomas/Rüegg, Erwin (2005): Konsequenzen von Staatsreformen für die demokratische Steuerungsfähigkeit. Vergleichende Analyse zu vier Schweizer Kantonen. *Politische Vierteljahresschrift*, 46(1): 100.

Ziel meines Beitrags

Aufzeigen, warum beide Alternativen keine Königswege darstellen.

Output- und Impact-Steuerung (kurz auch Performanz-Steuerung) überfordert herkömmliche Steuerungsinstrumente, wie Audit, Controlling, Rechnungsprüfung etc.

Zwei Alternativen:

1. Erweiterung der Reichweite der Steuerungsinstrumente: Performance Audit, Performance Measurement, Wirkungscontrolling etc. (nachfolgend Performanz-Messung)
2. Evaluation soll Steuerungsinformationen für Output- und besonders Impact-Steuerung bereitstellen.

3. Thesen zur Performanz-Messung

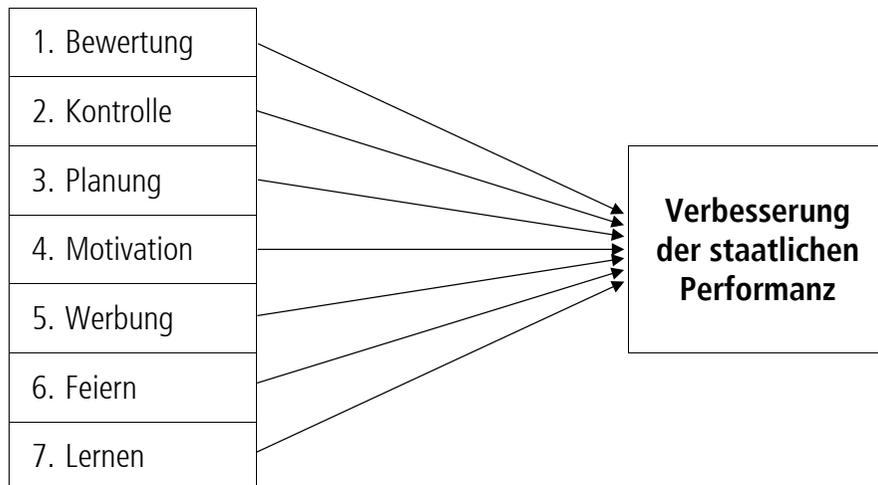
Leistung – ‚Performance‘

Drei Verständnisse des Begriffs ‚performance‘ / Performanz

- performance 1 – Leistungsfähigkeit
Output, aber auch Impact
- performance 2 – Inszenierungskraft
Show, Inszenierung, Aufführung und Darstellung
- performance 3 – Wirksamkeit
wirklichkeitsgestaltende Funktion von Zeichensystemen

Nullmeier, Frank (2004): Zwischen Performance und Performanz – Funktionen und Konzepte der Evaluierung in öffentlichen Verwaltungen. In: Kuhlmann/Bogumil/Wollmann (Hrsg.): Leistungsmessung und –vergleich in Politik und Verwaltung. Wiesbaden: VS: 47-60.

Ziele der Performanz-Messung nach Behn



Angepasst basierend auf Behn 2003

Performanz-Messung

“Why measure performance? Because measuring performance is good. But how do we *know* it is good? Because business firms all measure their performance, and everyone knows that the private sector is managed better than the public sector.”

Robert D. Behn (2003): Why Measure Performance? Different Purposes Require Different Measures. *Public Administration Review*, 63(5): 586-7. (Übersetzung TW)

These 1

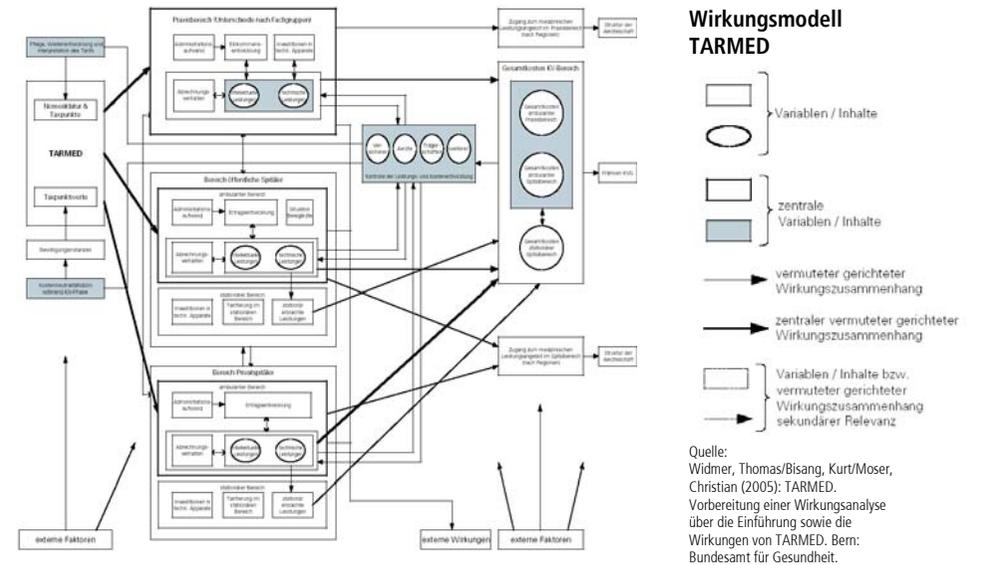
Staatliche Tätigkeiten und deren Wirkungszusammenhänge sind zu komplex, um in einem Performance-Measurement-System adäquat abgebildet zu werden.

Wirkungskette öffentlicher Politiken

Input ⇒ Prozess ⇒ Output ⇒ Outcome ⇒ Impact

These 2

Performanz-Messungen liefern im besseren Fall ungenaue und im schlechteren Fall falsche Informationen.



„Bis zum Morgenrauen stand er [Alexander von Humboldt] neben dem Kapitän und beobachtete ihn beim Navigieren. Dann holte er seinen eigenen Sextanten hervor. Gegen Mittag begann er den Kopf zu schütteln. Nachmittags um vier legte er sein Gerät beiseite und fragte den Kapitän, wieso er so unexakt arbeite.

Er mache dies seit dreissig Jahren, sagte der Kapitän.

Bei allem Respekt, sagte Humboldt, das erstaune ihn.

Man tue dies doch nicht für die Mathematik, sagte der Kapitän, man wolle übers Meer. Man fahre so ungefähr den Breitengrad entlang, und irgendwann sei man da.“

Daniel Kehlmann (2005): Die Vermessung der Welt. Reinbek: Rowohlt, S. 45.

These 3

Anstatt zu motivieren und Anlass zum Feiern zu bieten, wird die Messung der Performanz durch die Betroffenen eher als Belastung empfunden und zerstört deren intrinsische Motivation zu guten Leistungen.

„... perché gli uomini sempre ti riusciranno tristi, se da una necessità non sono fatti buoni.“

„... denn die Menschen werden sich dir gegenüber immer als böse erweisen, wenn sie nicht gezwungen werden, gut zu sein.“

Niccolò Machiavelli (1986 [1532]): Il Principe/Der Fürst. Stuttgart: Philipp Reclam, S. 186-7.

These 4

Performanz-Messungen bilden oft keine geeignete Grundlage zur Steuerung der zukünftigen staatlichen Tätigkeit.

Beispiel:

Konsolidierter Entwicklungs- und Finanzplan (KEF) des Kantons Zürich

Mittelschulen										Nr. 7301		
										Funktionale Gliederung: 0		
Finanzierung												
Erfolgrechnung (in Mio. Fr.)	R 06	B 07	Δ(P 08)	P 08	Δ(P 09)	P 09	Δ(P 10)	P 10	P 11	Δ%	06-11	
Ertrag	44.5	42.3	0.5	47.1	10.2	54.0	-1.7	48.5	45.6	2.8		
Aufwand	-342.4	-357.6	-4.0	-363.0	-3.2	-368.9	-1.4	-375.2	-386.4	12.8		
- Personalaufwand	-277.2	-282.8	-2.8	-287.7	-2.2	-293.6	-3.8	-298.7	-303.5			
Saldo	-298.0	-315.2	-3.5	-315.9	2.0	-314.9	-3.1	-326.8	-340.8			
Investitionen (in Mio. Fr.)										0 (06-11)		
Einnahmen	1.1					0.4	0.4	0.6	0.4			
Ausgaben	-26.2	-32.5	10.2	-31.6	0.1	-49.6	-10.0	-79.1	-81.9	-60.2		
Nettoinvestitionen	-25.1	-32.5	10.2	-31.6	0.1	-49.6	-18.6	-78.7	-81.3	-49.8		
Personal (Beschäftigungsumfang)	1,687.0	1,782.0	3.0	1,770.0	32.0	1,796.0	45.0	1,800.0	1,785.0			
Aufgaben												
A1 Ausbildung nach Maturitätsprofilen gemäss Maturitätsanerkennungsreglement. A2 Ausbildung an der Fachmittelschule FMS oder Diplommittelschule DMS. A3 Ausbildung an der 'Handelsmittelschule plus' oder an der Informatikmittelschule. A4 Durchführung von Vorbereitungskursen für das Aufnahmeverfahren an die Pädagogische Hochschule Zürich. A5 Durchführung des einjährigen Vorbereitungskurses und der Prüfung für die 'Passerelle'.												
Entwicklungsschwerpunkte										Direktionsziel Nr.		
E1	Erführung der zweisprachigen Maturitätsausbildung (Deutsch-Englisch, Projekt), verlängerte Projektdauer.										2008	1
E2	Zentrale Aufnahmepflichten an Mittelschulen (Projekt ZAP).										2008	5
E3	Selbstorganisiertes Lernen an gymnasialen Mittelschulen, neue Lehr- und Lernformen (Projekt SOL)										2011	3
Indikatoren												
Wirken	Art	R 06	B 07	P 08	P 09	P 10	P 11					
W1	Hohe Ausbildungsqualität für den Studienbeginn oder Berufseinstieg. Genereller Zufriedenheitsgrad zwei Jahre nach Schulabschluss. Prozentualer Anteil Antworten mit mindestens 'sehr zufrieden'	min.	n.q.	85%	n.q.	n.q.	85%	n.q.				
W2	Hohe Ausbildungsqualität für den Studienbeginn oder Berufseinstieg. Zufriedenheitsgrad betreffend Vorbereitung auf eine Hochschule. Prozentualer Anteil Antworten mit mindestens 'sehr gut'	min.	n.q.	75%	n.q.	n.q.	75%	n.q.				
W3	Maturitätsquote im 10. Schuljahr in %	20.2	20.8	20.6	20.7	20.8	20.9					
Leistungen												
L1	Einhaltung des Maturitätsreglementes: Lektionenfaktor	1.95	1.95	1.95	1.95	1.95	1.95					
L2	Anzahl Schülerinnen oder Schüler, welche in den Kontrakt finanziert werden	14'829	14'980	15'090	15'320	15'335	15'225					
L3	Anzahl Lernende in Mittelschulen	14'467	14'770	14'936	15'110	15'120	15'006					
L4	Davon Lernende in den Informatikmittelschul-Klassen	127	180	180	180	180	180					
L5	Anzahl Lernende mit Zeugnissen zweisprachige Matur	122	165	200	220	240	240					
Wirtschaftlichkeit												
B1	Nettoaufwand pro kontraktfinanzierte Schülerin oder Schüler in Fr.	20'093	21'044	20'935	20'656	21'300	22'381					

Konsolidierter Entwicklungs- und Finanzplan (KEF) des Kantons Zürich

Zufriedenheit mit der Mittelschule aufgrund einer Befragung der Maturandinnen und Maturanden zwei Jahre nach der Matur: (kumulierter Anteil „eher zufrieden“, „zufrieden“ und „sehr zufrieden“)

Jahr	2000	2001	2004	2007	2010	2013
Soll	85%	85%	85%	85%	85%	85%
Ist	90%	87%	89%	90%	89%	93%

Quelle: KEF, div. Jahrgänge (Leistungsgruppe Mittelschulen)

„Toute pierre lancée au hasard se dirige, avec une étonnante précision, vers l'endroit qu'elle finira par atteindre“.

„Jeder planlos geworfene Stein bewegt sich mit erstaunlicher Präzision auf die Stelle zu, an der er schliesslich landet.“

Jean-Claude Silbermann (exakter Fundort unbekannt)

(12) Situations of the social and economic development (Full mark:100)
Actual mark: $(\Sigma \div 4) \times \text{value} (10) =$

Evaluation item	Mark value	Evaluation content and requirement	Evaluation and grading standard	Evaluation method
12-1. Average gross national product per capita	10	Content: average gross national product per capita Requirement: The average gross national product per capita has been increasing rapidly in the administrative area of the upper level.	For counties in an area, the highest and lowest marks in respect of average gross national product per capita are 10 and 0 respectively, between which there are 10 equidistant grades (with 1 mark of difference between every two adjacent ones) The higher the average gross national product per capita, the higher the mark.	Obtain data of increase of average gross national product per capita in the past two years from the upper economic statistics departments of the unit under test.
12-2. Increase rate of GNP	10	Content: average increase rate of GNP in the past two years Requirement: same as above	For counties in an area, the highest and lowest marks in respect of increasing speed are 10 and 0 respectively, between which there are 10 equidistant grades (with 1 mark of difference between every two adjacent ones). The bigger the increase rate, the higher the mark.	Same as above
12-3. Percentage of 17-year-old population having completed junior high school education	10	Content: schooling of children of the right ages for 9-year compulsory education Requirement: The percentage of children of right ages at school should reaches 100%	In urban areas, the percentage exceeds 95%. (mark:10) 1 is deducted from the mark for decrease of every 0.5% until the mark comes to zero. In rural areas, the percentage exceeds 90% (mark:10) 1 is deducted from the mark for decrease of every 0.5% until the mark comes to zero.	Obtain related statistics from education departments in the area under test
12-4. Rate of criminal cases	10	Content: Number of criminal cases among every 10-thousand people Requirement: Rate of criminal offences should be 0.	The rate of criminal cases is 0 (mark:10). 1 is deducted from the mark for increase of every 0.5/10000 in the rate until the mark comes to zero.	Obtain related materials from the public security bureaus.
Total (Σ)	40			

Source: Fengyan, Liu/Faqian, Li/Chuhui, Wang/Zuoyong, Song (1995): Evaluating Government's Incompact Conditions with Social Indicators. Paper presented at the International Evaluation Conference, Vancouver, Beijing.

„Je planmässiger die Menschen vorgehen, desto wirksamer vermag sie der Zufall zu treffen.“

Friedrich Dürrenmatt (1998 [1962]): Die Physiker. Zürich: Diogenes, S. 91.

These 5

Performanz-Messungen rufen (unerwünschte) adaptive Verhaltensweisen hervor, die auch Lernprozesse zur Verbesserung der staatlichen Performanz behindern.

Beispiele für dysfunktionale Erscheinungen

Effekte von Bildungsstandards

Lernziele können zu "teaching to the test"-Phänomenen führen.

Erreichen der spezifischen Lernziele ohne die umfassenden Bildungsziele zu berücksichtigen.

Zielsetzungen: Reichweite und Definition

Reichweite

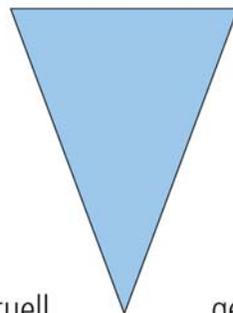
umfassend

Definition

grob

punktuell

genau



Publikationstätigkeit in der Wissenschaft

"The result is a well-documented tendency to produce large numbers of articles based on trivial research results that are easily published."

Christopher Tucci (2005): Why Europe will never have accountability in research. *Foresight Europe 2*: 28.

4. Fünf Thesen zu Genese und Nutzung von Evaluation in der Politik

33

Thomas Widmer, Institut für Politikwissenschaft, Universität Zürich

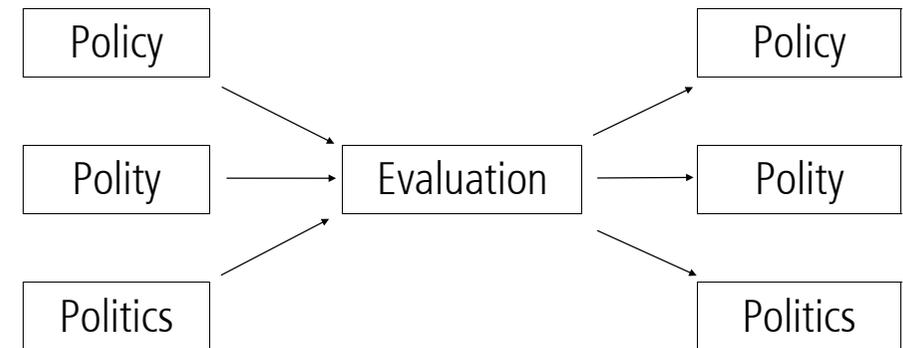
Vereinfacht betrachtet zwei Bezugspunkte zwischen Politik und Evaluation

1. Genese von Evaluation
2. Nutzung von Evaluation

35

Thomas Widmer, Institut für Politikwissenschaft, Universität Zürich

Evaluation in der Politik



34

Thomas Widmer, Institut für Politikwissenschaft, Universität Zürich

These 1: Evaluationen sollen zielgerichtet und selektiv mit klarem funktionalem Fokus eingesetzt werden.

- umfassende Evaluationsverpflichtungen führen zu unproduktiver Routinisierung
- gelingende Evaluationskonzeption erfordert funktionale Differenzierung
- fundierte Evaluationen benötigen erhebliche Ressourcen

36

Thomas Widmer, Institut für Politikwissenschaft, Universität Zürich

These 2: Evaluationen werden von der Politik oft bestellt, aber selten „abgeholt“.

- Politik fordert öfters Evaluationen, ohne deren Nutzung in Betracht zu ziehen.
- Die Evaluation ist zu langsam für die schnell voranschreitenden Aufmerksamkeitsfenster der Politik.

These 4: Politik nutzt Evaluationen und Evaluationsbefunde opportunistisch.

- strategischer Einsatz von Evaluationen
- Evaluationsbefunde dienen nicht der Bildung einer politischen Haltung, sondern zu deren Stärkung (Selbstvergewisserung, Überzeugung Dritter, Legitimation).

These 3: Evaluationsbefunde stehen bei der politischen Entscheidungsfindung in Konkurrenz zu entscheidungsmächtigeren Faktoren.

Carol H. Weiss: Four I's:

- Ideologies
- Interests
- Institutional norms and practices
- (pre-existing) information

These 5: Evaluation ist kein Steuerungsinstrument.

- Die Evaluation dient der Bewertung, nicht der Steuerung.
- Evaluation eignet sich nicht für den flächendeckenden und permanenten Einsatz.
- Evaluation erfordert Handlungsfreiheit und fügt sich schlecht in Strukturen hierarchischer Steuerung ein.

Besten Dank für Ihre Aufmerksamkeit!

Zusammenfassung der Thesen zur Performanz-Messung

1. Staatliche Tätigkeiten und deren Wirkungszusammenhänge sind zu komplex, um in einem Performance-Measurement-System adäquat abgebildet zu werden.
2. Performanz-Messungen liefern im besseren Fall ungenaue und im schlechteren Fall falsche Informationen.
3. Anstatt zu motivieren und Anlass zum Feiern zu bieten, wird die Messung der Performanz durch die Betroffenen eher als Belastung empfunden und zerstört deren intrinsische Motivation zu guten Leistungen.
4. Performanz-Messungen bilden oft keine geeignete Grundlage zur Steuerung der zukünftigen staatlichen Tätigkeit.
5. Performanz-Messungen rufen (unerwünschte) adaptive Verhaltensweisen hervor, die auch Lernprozesse zur Verbesserung der staatlichen Performanz behindern.

Zusammenfassung der Thesen zur Evaluation in der Politik

1. Evaluationen sollen zielgerichtet und selektiv mit klarem funktionalem Fokus eingesetzt werden.
2. Evaluationen werden von der Politik oft bestellt, aber selten „abgeholt“.
3. Evaluationsbefunde stehen bei der politischen Entscheidungsfindung in Konkurrenz zu entscheidungsmächtigeren Faktoren.
4. Politik nutzt Evaluationen und Evaluationsbefunde opportunistisch.
5. Evaluation ist kein Steuerungsinstrument.