



Good Governance wirksamer fördern: Argumente zur Ausweitung rigoroser Wirkungsevaluierung

Zusammenfassung

Die Entwicklungsgemeinschaft ist sich weitgehend einig: Good Governance ist unverzichtbare Voraussetzung für Entwicklungshilfe und eine sozioökonomische Entwicklung. Milliarden fließen jedes Jahr in Maßnahmen zur Korruptionsbekämpfung und Demokratieförderung, zur Stärkung staatlicher Institutionen, Wahrung der Menschenrechte und Erhaltung des Friedens.

Dennoch ist über die Wirksamkeit dieser Maßnahmen nur wenig bekannt. Mit Wirkungsmonitoring und länderübergreifenden Analysen bemühen sich Geber intensiv, die Auswirkungen auf Governance aufzuzeigen. Diese Ansätze können zwar die Fortschritte im Bereich Good Governance umreißen, aber nicht zweifelsfrei belegen, dass ein wahrgenommener *Outcome* einer bestimmten Maßnahme zuzuschreiben ist.

Gebraucht wird ein belastbarer Ansatz – einer, der Wirkungsmessung statt nur Wirkungsmonitoring betreibt, der Governance *Outcomes* Maßnahmen zuschreibt, indem er andere Einflussfaktoren ausschließt.

Genau das leistet die „rigorose Wirkungsevaluierung“. Anders als in „traditionellen“ Sektoren wie Gesundheit, Bildung und Arbeit ist es ein Verfahren, das in Governance-Programmen noch nicht ausreichend genutzt wird. So zeigen Untersuchungen, dass nur eine von 165 Stabilisierungsinterventionen der USA rigoros evaluiert wurde. Schätzungen zufolge beziehen sich maximal 10 % der rund 800 abgeschlossenen und laufenden rigorosen Wirkungsevaluierungen von Entwicklungsprogrammen der letzten fünf Jahre auf Governance-Programme.

Das überrascht. Abgesehen von den Effizienzvorteilen, die sich durch Zuweisung knapper Ressourcen zu wirksamen Maßnahmen erzielen lassen, spielen rigorose Wirkungsevaluierungen im Bereich Governance eine wichtige Rolle, bedenkt man die unbeabsichtigten, potenziell schädlichen Folgen ungeeigneter Governance-Maßnahmen.

Warum also wird hier so selten rigoros evaluiert? Rigorose Governance-Evaluierung stößt auf drei große Hürden:

- *Outcomes* lassen sich schwer quantifizieren.
- Rigorose Wirkungsevaluierung kann nur eng begrenzte Aspekte einer komplexen, „systemweiten“ Governance-Förderung und kurzfristige Wirkungen erfassen.
- Es fehlen Anreize, da Evaluierungsergebnisse politische Entscheidungsfreiheit einschränken können.

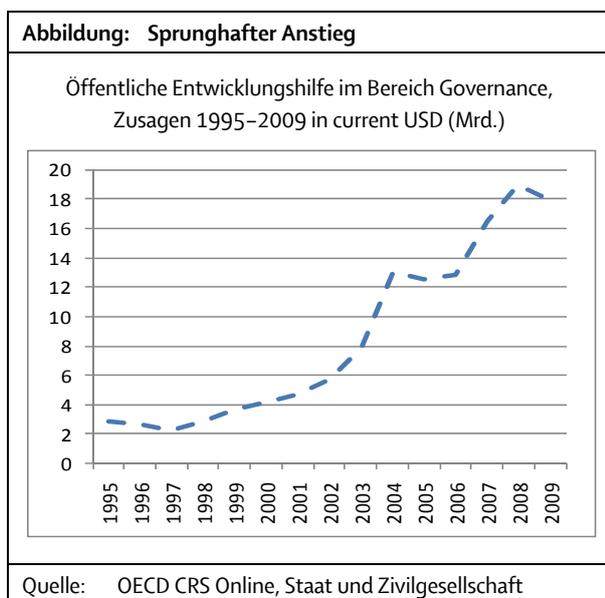
Frühere Evaluierungen geben wertvolle Hinweise auf Lösungsmöglichkeiten. *Outcomes* im Bereich Good Governance lassen sich mit kontextspezifischen Angaben oder *behavioral games* messen. Komplexe „systemweite“ Maßnahmen werden überschaubar, wenn man einzelne Komponenten des zu analysierenden Programms herausarbeitet. Verbesserte Anreize überzeugen Politiker vom Nutzen rigoroser Evaluierungen, wenn sie die Wirkungen ihrer wichtigsten Maßnahmen untersuchen und angemessene Erwartungen formulieren.

Rigorose Wirkungsevaluierung im Bereich Governance ist nicht leicht, aber machbar. Mehr denn je müssen Geber Ergebnisse vorweisen. Deshalb sollte eine rigorose Wirkungsevaluierung parallel zu verbesserter Governance-Förderung und Entwicklungswirkung arbeiten, nicht gegen sie.

Sind Governance-Maßnahmen überhaupt sinnvoll?

Good Governance ist in den letzten Jahrzehnten ein zentrales Thema internationaler Entwicklungspolitik geworden. Korruption, politische Instabilität, finanzielle Misswirtschaft, Konflikte, versagende Rechtssysteme, Menschenrechtsverletzungen und andere Faktoren können sich nachteilig auf Entwicklung auswirken und alle Bemühungen, Armut zu lindern, untergraben. Insofern ist Good Governance Voraussetzung für Entwicklungshilfe und unverzichtbares Element einer nachhaltigen Entwicklung.

Dieser Bedeutung tragen Geber Rechnung, die Governance-Maßnahmen (Politiken, Programme und Projekte) aktiv unterstützen, um in Partnerländern gut funktionierende Regierungen und ein günstiges institutionelles Umfeld zu schaffen. Das entsprechende Fördervolumen der vergangenen zehn Jahre verzeichnet deutliche Zuwächse: Wurden in 2008 rund 19 Mrd. USD in Governance-Maßnahmen investiert, waren es 1995 nur 2,8 Mrd. USD (siehe Abbildung).



Diese hohe und weiter steigende Förderung im Bereich Governance ist mit Bemühungen verbunden, ihre Ergebnisse zu messen. Ein Ansatz ist die Nutzung von Instrumenten des Wirkungsmonitorings wie Umfragen, strukturellen Indizes und qualitative Fallstudien. Diese Instrumente sind hilfreich Fortschritte und erzielte *Outcomes* zu identifizieren, haben aber eine offenkundige Schwäche: Sie können nicht zweifelsfrei ein *outcome* einer bestimmten Maßnahme zuordnen. Bestenfalls liefern sie deskriptive Erkenntnisse und anekdotische Evidenz, aber keinen echten Beweis für Wirkung.

Ein anderer Ansatz besteht in einer länderübergreifenden Analyse der Wirkung von EZ zur Förderung von Good Governance auf Entwicklung. Allerdings kranken solche Studien oft an einem gegenseitigen Feedback zwischen *Outcome* und Maßnahme. Die Förderung von Good Governance kann Entwicklung anstoßen, aber Wirtschafts-

wachstum kann auch Förderung anziehen, die dann den Aufwärtstrend verstärkt. Was aber ist Ursache, was Wirkung? Selbst wenn diese Frage geklärt ist, bleiben aggregierte länderübergreifende Statistiken zwar auf strategischer Ebene informativ, weniger jedoch auf der operativen Ebene von Projekten und Programmen.

Typische Evaluierungsberichte im Auftrag von Gebern ignorieren vielfach das Problem der Wirkungszuschreibung und bieten daher eher fragwürdige Wirkungsanalysen. So ergab ein Review von 24 USAID-Evaluierungsberichten über Demokratie- und Governance-Programme, dass man in 72 % der Fälle nicht nach alternativen Erklärungen für die beobachteten Effekte gesucht hatte.

Daher ist trotz der Bemühungen der Geber, Good Governance finanziell zu fördern und Fortschritte regelmäßig zu überwachen, nur wenig über ihre Wirkung bekannt. Es ist unabdingbar, dass Geber das „Do no harm“-Prinzip beachten. Sie sollten alles vermeiden, was das Bürger-Staat-Verhältnis belasten könnte, was Unterstützung gewährleistet, das auf falschen Annahmen über das Wesen lokaler Institutionen und Akteurskonstellationen beruht, und was existierende Anreize für lokale Reformprozesse unterminiert. Übergroßes Vertrauen in Evaluierungsberichte, die sich nicht glaubhaft mit Wirkungszuschreibung befassen, trübt unser Verständnis dessen, was wirkt, was nicht und warum.

Die Stärken rigoroser Wirkungsevaluierung

Dass sich Wirkzusammenhänge schwer präzise nachweisen lassen, weist auf ein Grundproblem von Evaluierung hin: Die Möglichkeit, dass ein *Outcome* auf andere Faktoren als vermutet zurückzuführen ist. Geber können weder die Zeit zurückdrehen und sehen, was ohne ihr Eingreifen geschehen wäre, noch können sie die geförderten und die nicht geförderten Individuen (oder Instanzen) einfach vergleichen, denn deren (fehlende) Gemeinsamkeiten müssen berücksichtigt werden. Diese Schwierigkeiten will die rigorose Wirkungsevaluierung eliminieren, indem sie mit Experimenten und anderen Mikromethoden ein plausibles Kontrafaktum findet (was wäre ohne die Maßnahme geschehen? – siehe Kasten 1).

Angetrieben durch die Erfolgsorientierung der Millenniums-Entwicklungsziele (MDG), der Erklärung von Paris und des Aktionsplans von Accra, führen inzwischen immer mehr Geber rigorose Wirkungsevaluierungen von Entwicklungsprogrammen durch. In traditionellen Sektoren wie Gesundheit, Bildung und Arbeit ist in den vergangenen Jahren durch die Messung der Wirkung einer breiten Palette von Maßnahmen eine belastbare Faktenbasis entstanden. Befürworter sind der Ansicht, dass durch Kenntnis der Wirkung einer konkreten Maßnahme (was warum funktioniert) letztendlich die schwer greifbaren makroökonomischen Wirkungen von Entwicklungszusammenarbeit benannt werden können.

Doch während die rigorose Wirkungsevaluierung in traditionellen Sektoren weit verbreitet ist, wird sie im Bereich

Kasten 1: Was ist rigorose Wirkungsevaluierung?

Das Grundproblem von Evaluierung ist die Frage: Wie lässt sich Wirkung zuschreiben, also belegen, dass ein *Outcome* das Ergebnis einer bestimmten Maßnahme und nicht eines anderen Faktors ist? In vereinfachter Algebra ist Wirkung die Differenz im *Outcome* Y ; dabei bezeichnet Y_1 den *Outcome*, wenn eine Person von der Maßnahme betroffen ist, und Y_0 den *Outcome*, wenn sie nicht betroffen ist:

$$\text{Wirkung} = Y_1 - Y_0$$

Zu einem konkreten Zeitpunkt lässt sich allein der *Outcome* für die betroffene Person beobachten, nicht jedoch der für die nicht betroffene Person. Das Dilemma besteht also darin, dass zur Berechnung der Wirkung nicht beide Zustände gleichzeitig beobachtet werden können. Da nur ein Zustand zur Zeit beobachtet werden kann (im Normalfall der nach der Maßnahme), ist es möglich, sich mit jenen quantitativen Methoden wie Experimenten und Quasi-Experimenten (die ein Experiment imitieren), dem Kontrafaktum anzunähern. Generell besagt der Begriff „rigoros“, dass die Analyse die Suche nach einem plausiblen Kontrafaktum einschließt.

Governance nur bedingt angewendet. Das steht im Widerspruch zu der allgemeinen Auffassung, dass Good Governance für Entwicklung maßgebend ist. Da auch die Effektivität anderer Reformen der Entwicklungszusammenarbeit und -politik von Good Governance abhängt, besteht ein impliziter Bedarf an rigoroser Evaluierung von Governance-Förderung.

Warum ist rigorose Wirkungsevaluierung im Bereich Governance noch immer die Ausnahme?

Immer wieder heißt es, die rigorose Evaluierung von Good Governance kranke an kleiner Stichprobengröße, Spillover-Problemen, einem Mangel an geeigneten Daten u. ä. Diese technischen Probleme sind aber nur die nächstliegenden. Drei weitere große Herausforderungen kommen hinzu:

- **Outcome-Messung.** *Outcomes* im Bereich Governance sind oft schwer zu messen, denn sie manifestieren sich vor allem in individuellem Verhalten, Einstellungen und immateriellen Konzepten. Anders als in der Landwirtschaft, wo die Wirkung von Düngemitteln auf die Erträge beobachtbar ist, oder im Bildungssektor, wo sich die bessere Versorgung mit Büchern in den Noten der Lernenden niederschlägt, ist die Messung der Wirkung von Governance-Maßnahmen auf *Outcomes* wie Korruption, Demokratie, Kriminalität und „Empowerment“ nicht so offensichtlich. Daher befürchten Kritiker, der wachsende Fokus auf Ergebnismessung werde zu einer Agenda führen, die sich einfach quantifizieren lässt.
- **Die inhärent systemweite Natur geberseitiger Governance-Förderung.** Geber, die Good Governance fördern, haben zum Ziel, das Verhältnis Bürger-Staat zu verbessern, was verschiedenste Akteure, Institutionen, Funktionen und Prozesse betrifft. So stärken Geber, die Rechenschaftspflicht und demokratische Regierungsführung fördern, politische Parteien, die Zivilgesellschaft, Parlamente, Bürger und Rechnungskontrollbehörden, etwa bei Wahlen, partizipativer Haushaltsplanung und öffentlicher Leistungserbringung. Insofern argumen-

tieren Skeptiker, rigorose Wirkungsevaluierung könne lediglich eng begrenzte Aspekte einer komplexen „systemweiten“ Governance-Förderung erfassen. Zudem bemängeln sie, diese Art der Evaluierung könne keine langfristigen Wirkungen wie Institutionenbildung und transformatorische Prozesse abbilden.

- **Anreize.** Politische Entscheidungsfreiheit kann durch Evaluierungsergebnisse eingeschränkt werden. Manche politischen Akteure und Regierungen in Partnerländern haben Vorbehalte gegen die Zusammenarbeit mit Gebern bei der rigorosen Evaluierung sensibler Governance-Fragen – wie etwa bei Menschenrechtsverstößen und Konflikten – wenn dies ihrem Ansehen schadet, ihren Raum für bequeme Beschlussfassung einengt, zukünftige Geldgeber abschreckt oder zusätzliche Konditionalität bewirkt.

Diesen Herausforderungen zum Trotz gibt es viel Raum für die Evaluierung von Governance-Maßnahmen, vielleicht sogar mehr, als man zunächst denken könnte. Aktuelle Untersuchungen, bei denen rigoros evaluiert wurde, ergaben: In Brasilien halfen Prüfungsberichte, das korrupte Verhalten von Politikern zu zügeln; in Nigeria reduzierte eine basisnahe Kampagne gegen Gewalt die Einschüchterung der Wähler und steigerte die Wahlbeteiligung und in Sierra Leone konnte ein Programm für „gemeindefaszierte Entwicklung“ keinen Wandel in den örtlichen Institutionen herbeiführen. Diese Beispiele betonen nicht allein den Erfolg oder Misserfolg von Maßnahmen. Ihre weitergehende Botschaft lautet: Rigorose Wirkungsevaluierung im Bereich Governance ist machbar.

Empfehlungen

Angesichts ihrer Dynamik und Komplexität kann die rigorose Wirkungsevaluierung im Bereich Governance eine Herausforderung sein. Daher mag sie anfangs etwas einschüchternd wirken, doch Programmverantwortliche können Erfahrungen aus früheren Evaluierungen und Anleihen bei Ansätzen anderer Sektoren nutzen:

Zwar sind Outcomes im Bereich Governance schwer zu messen, doch können behavioral games und/oder kontextspezifische Angaben hilfreich sein diese Schwäche auszugleichen. So nutzte Olken (2007) ein einfaches aber direkt anwendbares Maß zur Bemessung von Korruption beim Bau von Dorfstraßen in Indonesien. Er quantifizierte Korruption, indem er die vom Dorf genannten Kosten mit der Kostenschätzung eines unabhängigen Ingenieurs verglich. Die Differenz spiegelte den fehlenden Betrag wider. Dies ist ein geeigneter Indikator für Korruption bei Infrastrukturmaßnahmen. Werden kontextspezifische Angaben genutzt, ist es sinnvoller, einen ergebnisorientierten Indikator zu definieren als den allgemeinen Korruptionsbegriff anzuwenden. Einige Forscher führen zudem Experimente durch, um Verhalten zu messen, das anders nur schwer fassbar wäre (zum Beispiel Vertrauen und andere Formen sozialen Kapitals). Insofern deutet trotz problematischer Outcome-Messung manches darauf hin, dass rigorose Analysen machbar sind.

Mit zunehmender Konzentration auf „systemweite“ Governance-Förderung müssen einzelne Komponenten eines Programms, das rigoros analysiert werden soll, benannt werden. Tatsache ist allerdings, dass eine rigorose quantitative Analyse nicht immer möglich ist. Gleichwohl sollten Programmverantwortliche jede Gelegenheit nutzen, eine solche Analyse eines ganzen Programms oder einzelner Komponenten durchzuführen. Auf operativer Ebene bedeutet das, dass mit Beginn der Programmplanung Evaluatoren eingebunden und Mitarbeiter von Geberregierungen in rigoroser Wirkungsevaluierung ausgebildet werden sollten. Programm- und Evaluierungsverantwortliche müssen dringend zusammengebracht werden. Dies würde zu einer ausgewogeneren Nutzung verschiedener Ansätze – qualitativer wie quantitativer – führen und Programmstrukturen so verbessern, dass sich langfristige Wirkungen erfassen lassen.

Die Anreize müssen stimmen. Politiker, die wissen wollen, wie sich ihre wichtigsten Maßnahmen auswirken, lassen sich eher vom Nutzen der rigorosen Wirkungsevaluierung überzeugen. Entscheidend ist hier zu ermitteln, welche Maß-

nahmen das sind, und angemessene Erwartungen in Bezug auf die Durchführung und den Nutzen der Evaluierung zu formulieren.

Rigorose Wirkungsevaluierung erfordert Kompromisse zwischen Kosten und Nutzen. Clemens und Demombynes (2010) schlagen vor, rigorose Wirkungsevaluierung in Betracht zu ziehen, wenn sich Belege ohne großen Aufwand sammeln lassen, Ergebnisse vor einer politischen Entscheidung vorliegen können, die Umsetzung der falschen Politik besonders schädlich wäre, Raum für die Einrichtung eines Kontrollorgans besteht, politische Entscheidungsträger ernsthaft interessiert sind und der Pilotversuch genau zu der erweiterten Maßnahme passt. Diese Bedingungen sollten auch im Bereich Governance gelten. Wenn die Ergebnisse zuverlässige Daten liefern, Monitoring-Instrumente abrunden, zu evidenzbasierter politischer Entscheidungsfindung führen und letztlich die Lebensbedingungen der Menschen verbessern, wird der Nutzen der rigorosen Wirkungsevaluierung im Bereich Governance ihre Kosten bei Weitem aufwiegen.

Literatur

- Clemens, M. / G. Demombynes (2010): When does rigorous impact evaluation make a difference? The case of the Millennium Villages, Washington, DC: Center for Global Development (Working Paper 225)
- Garcia, M. (2011): Micro-methods in evaluating governance interventions, Bonn: BMZ, (Arbeitspapiere zur Evaluierung); online: http://www.bmz.de/en/zentrales_downloadarchiv/erfolg/BMZ_WP_Micro.pdf
- Olken, B. (2007): Monitoring corruption: evidence from a field experiment in Indonesia, in: *Journal of Political Economy* 115 (2), 200–249



Dr. Maria Melody Garcia

Abteilung III: Governance, Staatlichkeit, Sicherheit
Deutsches Institut für Entwicklungspolitik^{*)}

*) Aus einem vom Referat „Evaluierung der Entwicklungszusammenarbeit“ des Bundesministeriums für wirtschaftliche Zusammenarbeit und Entwicklung (BMZ) finanzierten Forschungsprojekt