



Wirkungsanalysen der entwicklungspolitischen Zusammenarbeit sind machbar

Mit der neuen „Orientierung auf Wirkung“ von Planung bis Evaluierung bei Entwicklungsvorhaben hat sich die Entwicklungszusammenarbeit zum Ziel gesetzt, wirksamer und transparenter zu werden. Entsprechend sind auch Wirkungsanalysen hoch aktuell. Neben der Rechenschaftspflicht wird mit ihnen das Ziel verfolgt, aus den Ergebnissen entwicklungspolitischer Interventionen zu lernen, d.h. wenn möglich, „best practises“ zu formulieren oder – wenn es sein muss – Fehler zu korrigieren.

Im Zuge der Programm- und Budgetorientierung der Entwicklungszusammenarbeit (EZ) findet Entwicklungspolitik tendenziell auf immer höher aggregierter Ebene statt. Mit der Ausrichtung der EZ auf die *Millennium Development Goals* (MDGs) werden jedoch gleichzeitig die erhofften Wirkungen vermehrt auf der Bevölkerungsebene d.h. Mikroebene gesucht. Aufgrund der Interventionen auf der Makro- und Wirkungen auf der Mikroebene, wird die Zuordnungslücke zwischen bestimmten Interventionen und Wirkungen noch größer.

Daher verstärkt sich tendenziell auch die Skepsis der Fachwelt gegenüber der Durchführbarkeit von Wirkungsanalysen. Mit der zunehmenden Orientierung der EZ auf Programm- und Budgethilfe wird die genaue Definition von Geberbeiträgen zu bestimmten Zielen zwar womöglich an Bedeutung verlieren, es bleibt jedoch weiterhin die Frage wichtig, welche Interventionen welche Wirksamkeit entfalten und warum sie dies tun. Daher werden Wirkungsanalysen und die hiermit verknüpften Schwierigkeiten auch dann Thema sein, wenn der Wandel der EZ in die erwartete Richtung erfolgt ist.

In diesem Beitrag wird vertreten, dass Wirkungsanalysen heute und zukünftig machbar sind. Um die konzeptionelle Debatte einen Schritt weiterzubringen, werden nach einer Einführung vier provokante Thesen aufgestellt und erläutert. Es wird zudem ein Ansatz aufgezeigt, mit dem den methodischen Schwierigkeiten begegnet werden kann.

Was sind Wirkungsanalysen?

Wirkungsanalysen untersuchen Wirkungen (*Impacts*) von Interventionen im Hinblick auf das Erreichen eines übergeordneten entwicklungspolitischen Ziels (z.B. armutsmindernde Wirkungen von Wasserreformen in einem Land). Hierin unterscheiden sich Wirkungsanalysen vom bloßen Monitoring, bei dem Wirkungszuschreibungen nicht erfolgen, sondern entweder die Entwicklungstrends selbst im Vordergrund stehen – oder die Zuordnung von vornherein eindeutig ist, weil sie als unmittelbare Folge bestimmter Interventionen angesehen werden können (bei Output- und den meisten Outcome-Daten, siehe Kasten 1).

Während Inputs, Outputs und z.T. Outcomes einer Intervention, aber auch Entwicklungstrends zumeist quantitativ gemessen werden können, ist dies bei der Wirkungsanalyse kaum möglich. Der Grund hierfür ist, dass die Entwicklungstrends in einem Land ja nicht alleine durch eine Intervention ausgelöst werden, sondern die Summe aller Einflüsse aus vielen internen und externen Variablen darstellen (daher Bruttowirkungen). Die Nettowirkung einer Intervention zu isolieren und im Wirkungsgeflecht

zahlreicher Variablen darzustellen, ist aber bis heute mit quantitativen Verfahren nicht möglich. In der Evaluierungsforschung besteht die Schwierigkeit darin, dass die Zahl der vergleichbaren Fälle (z.B. Programme) gering ist, die Zahl der einwirkenden Variablen aber groß. Dadurch können Signifikanzen beim Einsatz von Statistik kaum erreicht werden. Es muss deshalb bei Wirkungsanalysen i.d.R. ein qualitatives Verfahren benutzt werden, auch wenn in Teilen der Fachwelt die Wunschvorstellung von quantitativ definierbaren Geberbeiträgen zu Zielerreichungen aufrechterhalten wird.

Typischerweise bestehen Wirkungsanalysen aus Vorher-Nachher-Vergleichen. Da geeignete Vorher-Untersuchungen jedoch zumeist nicht verfügbar sind, basieren die meisten Analysen auf ihrer Rekonstruktion. Prinzipiell ist dies akzeptabel, solange es systematisch erfolgt. In der Praxis werden jedoch häufig Informationen aus Dokumenten und Interviews mit der eigenen Wahrnehmung vermischt, so dass die Ergebnisse nicht wirklich nachvollziehbar sind. Weil der Umgang mit kontraintuitiven Informationen bei einem solchen Vorgehen zudem meistens nicht definiert ist, können partielle Ergebnisse nicht ausgeschlossen werden. Dem-

entsprechend gering ist daher häufig die öffentliche Akzeptanz der Ergebnisse. Aber auch qualitative Verfahren können valide sein, wenn sie nämlich mit den geeigneten Validierungstechniken durchgeführt werden (z.B. Cross-checking mit Daten aus anderen Quellen).

Kasten 1: Begriffsklärungen		
Intern./ DAC	Deutsche Bezeichnungen (mit Beispiel)	Ableitung/ Analyseebene
Development-Trends	Trends in Bezug auf das übergeordnete Ziel (z.B. nationale Armutstrends)	vom Kontext her (Bevölkerung)
Impact	Wirkung der Intervention (z.B. Wasserversorgung) auf die o.g. Trends	? Verknüpfung: Kontext / Intervention
Outcome	Direkter Nutzen / unmittelbare Wirkung der Intervention	von der Intervention her
Output	Ergebnisse / Leistungen	
Input	Intervention / Maßnahmen	

Die entwicklungspolitischen Durchführungsorganisationen in Deutschland evaluieren mit festgelegten Verfahren heute bis zur Outcome-Ebene. Längerfristige Wirkungen werden z.B. in Schlussprüfungen der KfW oder Querschnittsanalysen des BMZ zwar gutachterlich eingeschätzt, aber ohne dass hierfür festgelegte Methoden oder systematische Erhebungen bei Akteuren existieren.

In der internationalen Praxis werden Wirkungsanalysen häufiger *ex ante* mit Hilfe hypothetischer Wirkungsketten durchgeführt. Dies ist zwar für eine wirkungsorientierte Planung sinnvoll, die Analyse der De-facto-Wirkungen kann sie jedoch nicht ersetzen. Um diese De-facto-Analyse geht es aber im vorliegenden Beitrag.

Es wird dafür plädiert, das bisherige Instrumentarium der EZ mit gezielten Wirkungsanalysen so zu ergänzen, dass Portfolio-Analysen möglich sind. Am besten geschieht dies anhand eines logisch aufeinander aufbauenden, standardisierten Instrumentensets, das auch eine Aggregation der Daten erlaubt und organisationübergreifend eingesetzt wird. Dies ist wichtig, da es nur solch ein Ansatz erlaubt, z.B. die Wirkung des gesamten Portfolios der deutschen EZ in einem Land zu evaluieren.

Kasten 2: Vier provokante Thesen:

These 1: Negative Nebenwirkungen von EZ-Maßnahmen sollten zwar nicht akzeptiert aber „normalisiert“ werden.

These 2: Das Vorformulieren von Wirkungsketten ist in der Planung von Entwicklungsvorhaben zwar sinnvoll, bei der Analyse von De-facto-Wirkungen ist es jedoch hinderlich.

These 3: Wirkungsanalysen sollten vom Grundsatz her partizipativ gestaltet werden, dies gilt auch für Wirkungsanalysen politischer Reformen.

These 4: Der prinzipielle methodische Rahmen für Wirkungsanalysen sollte für die Mikro- bis Makroebene und für alle Sektoren der gleiche sein.

These 1: Negative Nebenwirkungen von EZ-Maßnahmen sollten zwar nicht akzeptiert aber „normalisiert“ werden.

Keinem anderen Ressort wie dem der Entwicklungspolitik wird von außen soviel „gute Intention“ und gleichzeitig so viel „Nutzlosigkeit“ unterstellt. Dabei unterscheidet sich die Entwicklungszusammenarbeit (EZ) nicht grundlegend von anderen Politikfeldern.

Gerade bei Reformprozessen auf nationaler Ebene gibt es auch in der EZ immer Gewinner und Verlierer. Das schiefe Vorhandensein von Nachteilen stellt jedoch keineswegs automatisch die Intervention selbst in Frage, wie häufig befürchtet wird. Stattdessen sollte immer abgewogen werden, wobei das Ausblenden von Nebenwirkungen das eigentlich Negative darstellt. Ziel von Wirkungsanalysen sollte es daher sein, das gesamte Spektrum der Wirkungen einzubeziehen und diese so frühzeitig zu erkennen, dass unerwünschte Effekte noch abgefedert bzw. berücksichtigt werden können. Neben Erfolgen sollten auch Rückschläge nicht verleugnet werden, sondern dazu dienen, die gemeinsame Arbeit zu optimieren.

These 2: Das Vorformulieren von Wirkungsketten ist als Planungsinstrument zwar sinnvoll, bei der Analyse von De-facto-Wirkungen ist es jedoch hinderlich.

Bei Evaluierungsfachleuten herrscht die Vorstellung vor, Wirkungsanalysen ausgehend von der Intervention, d.h. dem Programm oder Projekt vorzunehmen und den vorformulierten Wirkungsketten nachzuspüren bzw. diese zu überprüfen. Es wird hier jedoch die Auffassung vertreten, dass diese Vorgehensweise nicht zielführend ist, da sie den Blick für das Erkennen unerwarteter Wirkungen verstellt. Für die mögliche Nutzung der Analyseergebnisse, die Fehlerkorrektur und die Formulierung von *best practises* ist die Kenntnis dieser überraschenden Wirkungen jedoch grundlegend.

Das Vorgehen entlang von Wirkungsketten führt außerdem zur systematischen Überschätzung der Wirkungskraft einzelner Maßnahmen und zur Ausblendung externer bzw. anderer Effekte. Es impliziert monokausale Zusammenhänge, obwohl unsere eigene Lebenserfahrung uns zeigt, dass Ziele und Wirkungen nicht nur auf einem einzigen Wege, sondern in vielfältiger Art erreicht, aber auch durch andere Einflüsse wieder kontakariert werden können. Die Wirklichkeit ist vom Zusammenwirken vieler Variablen gekennzeichnet, die sich gegenseitig hemmen, fördern oder neutralisieren können. Das Wirkungsgeflecht zu erkennen, setzt jedoch ein offenes Vorgehen ohne vorformulierte Wirkungsketten voraus. Dabei besteht auch ein offenes Vorgehen nicht aus der Suche nach der Nadel im Heuhaufen. Wird im Vorfeld ein Rahmenkonzept mit Schlüsselkriterien festgesetzt, die das zu evaluierende Zielsystem beschreiben, und werden die Veränderungsprozesse partizipativ mit den wesentlichen Akteuren evaluiert, kann das Vorgehen im Gegenteil gezielt und einfach sein.

Die Definition von Schlüsselkriterien geschieht am besten anhand existenter Konzepte aus der Literatur: Für ein Zielsystem wie „Armutsminderung“ können z.B. die Schlüsselkriterien „Erhöhung des Lebensstandards, Zugang zu Ressourcen, Erweiterung des Wissens und Partizipation an Rechten und Macht“ festgesetzt werden. Hier spielt auch immer das Erkenntnisinteresse eine Rolle. Die Vorgabe eines festen Kriteriensets für das jeweilige Zielsystem ermöglicht die Vergleichbarkeit und das Aggregieren der Daten. Die zuzuordnenden Unterkriterien können vorgegeben oder partizipativ definiert werden.

Der erste Schritt einer Wirkungsanalyse sollte in der Ermittlung der Entwicklungstrends (also dem sog. Kontext) in der jeweiligen Programmregion gelten, wobei der Analysezeitraum den Interventionszeitraum etwas überschreiten sollte. Erst im zweiten Schritt sollte die Zuordnung zu den Interventionen erfolgen. Bei diesem „kontextorientierten Vorgehen wird „das Pferd quasi von hinten aufgezügelt“, der Ausgangspunkt ist eben nicht die Intervention, sondern die Lebenswelt bzw. Wirklichkeit (z.B. auch in einer Institution), wie sie sich für die Akteure darstellt (siehe Kasten 3).

Kasten 3: MAPP (Method for Impact Assessment of Programmes and Projects) ist eine akteurszentrierte Methode, die am Deutschen Institut für Entwicklungspolitik (DIE) entwickelt wurde und die ein offenes Vorgehen voraussetzt. Mit MAPP können die Wirkungen mehrerer Vorhaben gleichzeitig untersucht werden und die Beiträge zu den MDGs können qualitativ direkt von den Ergebnissen abgeleitet werden.

MAPP besteht aus einem Set von sieben logisch aufeinander aufbauenden Instrumenten. Um die Zuordnungslücke zu überbrücken, erfolgt in *Stakeholder-Workshops* zunächst die Erhebung der Entwicklungstrends vor Ort anhand von Schlüsselkriterien. Hierzu werden eine **Lebenslinie** und eine **Trendanalyse** erstellt, die – wie auch alle folgenden Instrumente – Punktwerte enthalten und einen Überblick über bspw. die Armutsentwicklung im letzten Jahrzehnt geben (Bruttowirkung).

Die Zuordnung der Entwicklungen zu den Interventionen erfolgt als Momentaufnahme erst mit dem dritten und vierten Instrument: Nach einem **Cross-Checking** mit anderen Datenquellen wird eine **Interventionsliste** erstellt, die Informationen über die Maßnahmen und Träger, die Relevanz, die Begünstigten sowie die Eigenbeiträge enthält und diese in einen Finanz- und Arbeitsrahmen eingeordnet (Input / Output).

Mit dem fünften Instrument, der **Einflussmatrix** wird nun die Verbindung zwischen Entwicklungstrends und Interventionen hergestellt (*Impact / u.U. Outcome*). Die positiven und negativen Einflüsse aller Interventionen auf alle Kriterien werden mit Punkten bewertet. Es werden Passiv- und Aktivsummen gebildet, wobei die Aktivsumme die Schlüsselinterventionen und die Passivsumme die stark oder gering beeinflussten Kriterien zeigt.

Das **Entwicklungs- und Wirkungsprofil** isoliert als Auswertungsschritt u.a. die Haupteinflüsse. Je nach Einheitlichkeit der Gesamttrends zeigt es zudem die Vulnerabilität der Entwicklung an. Die **partizipative Entwicklungsplanung**, bei der die bisher nur gering beeinflussten Kriterien als Ausgangspunkt neuer Planungen dienen, kann als siebtes Instrument abgeschlossen werden.

Einflussmatrix bei MAPP – Umgang mit der Zuordnungslücke

Faktoren Kriterien	P1: Was-serprogramm	P2: De-zen-tralisie-rungspr.	P3: Anti-korrupti-onsge-setz	Passiv-summe
Lebensstandard				
Einkommen	+ 3	-2	+ 3	+ 6 / -2
Landw. Erträge	+ 5	+/- 0	+ 3	+ 8
Zugang zu Ressourcen				
Boden	+/- 0	+/- 0	+ 3	+ 3
Wasser	+ 5	+2	+ 3	+ 10
Wissen				
Schulbesuch etc.	+ 1	+2	+/- 0	+ 3
Passivsumme	14	+4 / -2	+12	

Erläuterung zur Einflussmatrix: die Programme 1 und 3 haben demnach hier vielfältige und hauptsächlich positive Wirkungen auf die Armutssituation, während das Dezentralisierungsprogramm (noch) kaum positiv wirksam wurde und negative Wirkungen auf die Einkommen wahrgenommen werden. In den Workshops wird bei solchen überraschenden Effekten jeweils geklärt, warum sie aus der Sicht der Teilnehmer bestehen und welche Relevanz sie haben.

These 3: Heute sollten Wirkungsanalysen vom Grundsatz her partizipativ gestaltet werden, dies gilt auch für Wirkungsanalysen politischer Reformen.

Die letzten Jahre haben gezeigt, dass die Einbeziehung von Akteuren bei Evaluierungen bessere und umsetzungsrelevantere Ergebnisse erbringt als Evaluierungen von Externen. Akteure oder Zielgruppen können Wirkungen am besten an ihrem Wirkort beurteilen und nur sie sind in der Lage, Wirkungen mit hoher Authentizität zu beschreiben.

Konzeptionell haben sich partizipative Wirkungsanalysen daher längst durchgesetzt, in der Praxis werden sie jedoch weiterhin nur selten durchgeführt. Befürchtet werden der hohe Zeitaufwand und die „Subjektivität“ der Aussagen. Dabei kann der Zeitaufwand begrenzt werden, wenn eine bewusste Auswahl der Stichproben und eine Strukturierung der Diskussionen durch den Einsatz der Instrumente erfolgt. Zur Auswahl können z.B. besonders erfolgreiche, typische und erfolglose Gemeinden der Programmregion für die Evaluierung ausgewählt werden. So kann die Bandbreite der Wirkungen mit möglichst geringem Aufwand erfasst werden.

Die erwähnte „Subjektivität“ der Aussagen ist bei partizipativen Erhebungen wirklich gegeben, jedoch in geringerem Umfang, als dies bei Einzelinterviews der Fall ist. In Workshops mit unterschiedlichen Interessen- und sozialen Gruppen findet durch gegenseitiges Korrigieren und Argumentieren eine „kommunikative Validierung“ der verbalen Daten statt, die dem klassischen Mittelwert in seiner Aussagekraft weit überlegen sein kann. Tatsächliche Bewertungsunterschiede können als Dissens gekennzeichnet werden. Dissens tragen häufig zur

Klärung von Problemen bei, die vorher nur diffus wahrgenommen wurden. Voraussetzung einer kommunikativen Validierung ist allerdings die Existenz einer Diskussionskultur in der jeweiligen Gesellschaft, ohne die partizipative Methoden an Grenzen stoßen.

Gerade bei Interventionen in der Demokratieförderung oder Armutsbekämpfung sollte die Einbeziehung von Akteuren heute selbstverständlich sein, zumindest aber versucht werden. Denn hier soll ja der Einfluss von Zivilgesellschaft an Entscheidungsprozessen explizit erhöht werden. Ist dies nicht von Beginn an möglich, muss auch hier in Prozessen gedacht werden. Denn die Einbeziehung von Akteuren und Zielgruppen in Evaluierungen hat wichtige Nebeneffekte: Sie fördert *Ownership* und führt zum *Empowerment* und *Capacity Building*. Dies sind genau die Gründe, warum in nicht demokratischen Ländern solche Ansätze zurückhaltend bewertet werden.

Auf der anderen Seite bleibt die Frage offen, wie weit Partizipation gehen sollte. Eine grenzenlose Einbeziehung der Akteure würde das Aufgeben von Handlungsspielräumen des Trägers oder Gebers und ihrer Normen bedeuten. Ist der Umgang mit kontroversen Sichtweisen der Akteure ungeklärt, kann es zudem zu endlosen Kommunikationsschleifen kommen. Daher muss eine Balance zwischen Partizipation und Nicht-Partizipation gefunden werden.

Idealtypisch sollten die Akteure die wesentlichen Schritte der Wirkungsanalysen nach einem definierten Instrumentarium selbst durchführen. Auch sollte die organisatorische Aufhängung im Partnerland und nicht im Geberland platziert sein. Das Evaluierungsteam hat dementsprechend zunächst eine Moderatorenrolle, und nur die Bündelung der Ergebnisse unterschiedlicher Erhebungen würde ihm zufallen und u.U. Expertenmissionen aus Geberländern erfordern, da sie ja ebenfalls *Stakeholder* sind. Bei der zusammenfassenden Bewertung sollten die Standards Allparteilichkeit und Transparenz gelten, die u.a. von der Deutschen Gesellschaft für Evaluierung (DeGEval) formuliert wurden. Diese Standards durchzusetzen, ist keine methodische, sondern eine politische Frage mit durchaus hoher Brisanz.

These 4. Der geeignete methodische Rahmen für Wirkungsanalysen sollte für die Interventionsebenen und für alle Sektoren der gleiche sein.

Die Berücksichtigung der vertikalen Differenzierung sozialer Phänomene, d.h. des Phänomens, dass jede Institution vertikal in ein Institutionensetting eingeordnet werden kann und jeweils wiederum aus Individuen besteht, zeigt die Gleichwertigkeit der Aggregatensebenen Makro, Meso, Mikro: Je höher das Aggregationsniveau ist, auf dem interveniert und evaluiert wird,

desto wichtiger ist es, das jeweilige soziale System vertikal zu differenzieren. Denn jedes soziale Makrophänomen ist in sich vertikal so weitgehend differenzierbar, dass es operationalisierbar wird: Nationalstaaten oder Regierungen (Makroebene) unterhalten Institutionen (obere Mesoebene), die die Entscheidungen der Regierung umsetzen und selbst wieder Entscheidungen treffen. Diese werden wiederum auf einem niedrigeren Level umgesetzt und dort innerhalb der Institutionen (untere Mesoebene) als auch auf der Bevölkerungsebene (Mikroebene) lokal wirksam. Um aussagekräftige Ergebnisse zu bekommen, ist es daher notwendig, „Zwischenstufen“ bei Wirkungsanalysen einzubauen. Das Augenmerk sollte zunächst auf die Umsetzung der Maßnahmen gelegt werden, bevor die Wirkungsanalyse durchgeführt wird. Auf der Makroebene können die Entscheidung und partiell ihre Umsetzung, auf der Mesoebene die Umsetzung und partiell ihre Wirkungen und auf der Mikroebene vornehmlich die Wirkungen evaluiert werden. Es werden vorrangig immer solche Wirkungen evaluiert, die von der beteiligten Akteursgruppe am besten beurteilt werden können, dementsprechend werden die Gruppen zusammengesetzt. Mit einem solchen stufenweisen Vorgehen wird ein *down-* und *upscaling* ermöglicht und damit ein Lernen über die Ebenen hinaus möglich.

Der prinzipielle methodische Ansatz der Wirkungsanalyse kann aber nicht nur für die Ebenen, sondern auch für die einzelnen Sektoren (z.B. ländliche Entwicklung, Wirtschaftsförderung, Dezentralisierung) der gleiche sein. Da die Sektoren Inhalte betreffen, haben sie zumeist auf das methodische Vorgehen keinen wesentlichen Einfluss. Auf diese Weise könnte die Evaluierungsfachwelt übersektoral ihre Kräfte viel effektiver einsetzen als es bisher der Fall ist.



Dr. Susanne Neubert
Wissenschaftliche Mitarbeiterin des DIE

Literatur

Neubert, Susanne (2004): Akteurszentrierte Wirkungsanalyse – konzeptionelle Überlegungen und *Scaling Up* für die Verwaltungszusammenarbeit, in: Klaus Simon (Hrsg.), *Verwaltungen, Experten und Bürger im Reformprozess*, Nomos, Baden-Baden