

„Wirkungsanalysen in der Praxis der Entwicklungszusammenarbeit“ – Veranstaltung des AK „Evaluation von Entwicklungspolitik“ im Rahmen der DeGEval-Jahrestagung 2005

Martin Quack

Lehrstuhl Internationale Politik und Außenpolitikanalyse, Universität zu Köln

Die Evaluation von Wirkungen stellt in der Entwicklungszusammenarbeit (EZ) traditionell ein zentrales Thema dar, da die EZ in der Regel zur nachhaltigen Veränderung von Strukturen, Prozessen und Verhaltensweisen beitragen will. In den letzten Jahren ist eine verstärkte Hinwendung zu dem Thema Wirkungen in der Planung und Durchführung von Projekten und Programmen zu beobachten. Damit einher gehen höhere Erwartungen an die Wirksamkeit der entwicklungspolitischen Tätigkeiten. Die Konzepte in der EZ werden zunehmend komplexer indem sie z.B. verschiedene Interventionsebenen miteinander verbinden, Akteursnetzwerke bilden und Programme anstelle von Projektansätzen verwenden. Dadurch steigen auch die Anforderungen an die Evaluation.

Der AK „Evaluation von Entwicklungspolitik“ in der DeGEval hat aufgrund dieser aktuellen Herausforderungen im Rahmen der DeGEval-Jahrestagung in Essen entsprechend dem Tagungsthema „Den Wirkungen auf der Spur“ eine Veranstaltung zu Wirkungsanalysen in der Praxis der EZ durchgeführt, die großen Anklang fand. Es wurden drei Referate mit Praxisbeispielen präsentiert und diskutiert, die im Folgenden zusammengefasst dargestellt werden:

1. *Dr. Eva Weidnitzer* (BMZ): Computergestütztes Management-Informationssystem zum wirkungsorientierten Monitoring
2. *Dr. Stefanie Krapp* (CEval): Wirksamkeit von Berufsbildungsprojekten in der VR China
3. *Caspar Merkle* (IPEC-ILO): Measuring Longer Term Impact on Children and Families through Tracer/Tracking Methodologies

1. Computergestütztes Management-Informationssystem zum wirkungsorientierten Monitoring

Notwendig für Wirkungsanalysen sind u.a. zuverlässige Daten und bei großen Datenmengen eine übersichtliche Datenverwaltung. Mit dieser Grundlage befasste sich der erste Vortrag, in dem *Dr. Eva Weidnitzer* (BMZ) ein computergestütztes Management-Informationssystem zum wirkungsorientierten Monitoring vorstellte. Im Rahmen des Projektes ‚Institutionelle Kapazitätsentwicklung des Ministeriums für Planung und Internationale Zusammenarbeit‘ in der Republik Jemen entwickelte die GTZ ein Management-Informationssystem. Anlass für dessen Entwicklung war die Umsetzung der nationalen Armutsbekämpfungsstrategie (Poverty Reduction Strategy – PRS).

Das System geht von Wirkungsketten aus, die über Indikatorenketten dargestellt werden können. Es umfasst Indikatoren zu Input und Output (‚intermediate indicators‘) sowie Outcome und Impact (‚final indicators‘) der nationalen Strategie. So sollen nicht nur die angestrebten Wirkungen sondern auch der Prozess der Umsetzung von Strategien beobachtbar werden. Diese Monitoring-Daten dienen der Berichterstattung, Rechenschaftslegung und als Voraussetzung für zeitnah zu treffende Politikentscheidungen. Einzelne Indikatoren können ausgewählt und sowohl thematisch als auch zeitlich spezifiziert werden. Unterschiedliche Berichtsformate können sofort angezeigt, als Diagramme präsentiert und in Powerpoint oder in Word exportiert werden.

Bei der Verfolgung von Outcome- oder Impact-Indikatoren wird der schwächste Wert kursiv angezeigt, um auf das Problem hinzuweisen. Auch in der graphischen Darstellung werden schwache Werte hervorgehoben. Um möglichen Ursachen von schwachen Outcome- oder Impact-Werten auf die Spur zu kommen, können zu einem Indikator alle Projekte angezeigt werden, die einen Beitrag zu der gewünschten Wirkung liefern sollen. Diese Projekte können wiederum im Einzelnen betrachtet werden. Darüber hinaus lassen sich Projekte z.B. danach sortieren, ob sie aus Deutschland finanziert werden, einzelne Komponenten und einzelne Indikatoren-Ketten können in einem Ist-Soll-Vergleich dargestellt werden. Auch hier sind die Spezifizierung nach Jahren und graphische Darstellungen möglich. In Absprache mit dem zuständigen Ministerium wurden Standard-Berichte im System festgelegt, so dass die gewünschten Berichts-Tabellen ohne Aufwand jederzeit dargestellt werden können.

Grundlage für diese umfangreiche Monitoring-Datenverwaltung ist die Datenerhebung und die Eingabe in das System. Dazu gehört auch die Definition von Wirkungsketten. Projekte müssen genau definieren, mit welchen Maßnahmen sie zu welchen Wirkungen beitragen wollen. Als größte Risiken nannte *Dr. Weidnitzer* Probleme bei der Datenerhebung, bei der notwendigen Aktualisierung und die Gefahr, zu viele Indikatoren zu verwenden. Das Personal des jemenitischen Ministeriums sei nach mehreren Trainings in der Lage gewesen, Wirkungsketten zu entwickeln und Indikatoren zu bestimmen. Dazu müssten verschiedene staatliche Institutionen zusammen arbeiten. Die jemenitische Regierung bewerte das Instrument als wichtigen Schritt auf dem Weg zur Umsetzung der nationalen Armutsbekämpfungsstrategie.

In der anschließenden Diskussion wurden vor allem problematische Aspekte und die Grenzen des Instruments beleuchtet. Dazu gehörten die Qualität der Daten, die Quantifizierung qualitativer Indikatoren, der Umgang mit schwachen Indikatoren-Werten, die Berücksichtigung des politischen Kontextes und die Frage, inwieweit solche Systeme bereits vorhanden waren. *Dr. Weidnitzer* betonte die Möglichkeit, das System durch andere Methoden zu ergänzen, z.B. bei der Überprüfung der identifizierten Schachstellen. Das System läuft bisher in einer Testversion, in einem Klärungs- und Konzentrationsprozess soll die Zahl der Indikatoren eingeschränkt werden. Die GTZ bemüht sich das System international bekannt zu machen. Die Qualität der Daten sei zwar problematisch, durch die Implementierung des Systems und den damit verbundenen internationalen Druck steige jedoch der Druck zur Verbesserung der Datenqualität.

2. Wirksamkeit von Berufsbildungsprojekten in der VR China

Im zweiten Vortrag stellte *Dr. Stefanie Krapp* (CEval) ein Fallbeispiel der Wirkungsanalyse staatlicher und nicht-staatlicher EZ vor. In einer Ex-post-Evaluation untersuchte das CEval Projekte der Hanns-Seidel-Stiftung (HSS) und der GTZ zur Berufsbildung in China. Die GTZ verfolgte eine doppelte Strategie, in dem sie sowohl auf der Systemebene die Regierungen des Zentralstaats und zweier Provinzen bei der Neuordnung des beruflichen Bildungssystems beriet, als auch auf der operativen Ebene der Facharbeiterausbildung tätig war. Die HSS beschränkte sich in den Projekten auf die zweite Ebene, strebte jedoch ebenfalls systemische Veränderungen an.

Der Evaluationsauftrag umfasste die Untersuchung der Wirksamkeit der doppelten GTZ-Strategie und den Vergleich von staatlicher und nichtstaatlicher Arbeit. Dazu wurden Projekte beider Träger an betrieblichen Facharbeiterschulen und an betriebsungebundenen Facharbeiterschulen sowie Systemberatungsprojekte der GTZ untersucht.

Die theoretische Konzeption der Evaluation umfasste drei Modelle: Das Lebensverlaufsmodell betont das Phasenmuster und den Prozesscharakter. Das Wirkungsmodell stellt in konzentrischen Kreisen die Wirkungen zwischen dem Projekt im Kern, den Strukturen der Träger im zweiten Kreis und den Systemen der Umwelt ganz außen dar. Das Innovations-/Diffusionsmodell dient der Betrachtung von Faktoren, die zur Verbreitung spezifischer Angebote beitragen. Die Methodische Konzeption war durch ein Analyseraster gekennzeichnet, das die Grunddaten und die Qualität der Interventionen, interne Wirkungsfelder und externe Wirkungsfelder untersuchte. Zunächst wurden die einzelnen Projektprofile untersucht, dann folgte der Vergleich zwischen GTZ und HSS und die Untersuchung der ‚Doppelstrategie‘ der GTZ.

Da kein umfangreiches Monitoring-System vorhanden war, mussten die Daten zum größten Teil im Rahmen der Evaluation erhoben werden. Dazu dienten Dokumenten- und Aktenanalyse, 172 Interviews und eine schriftliche Befragung von 931 Personen (wodurch die Untersuchung um eine ‚quasiexperimentelle Dimension‘ erweitert wurde) sowie Beobachtungen.

Dr. Krapp stellte aus den Ergebnissen der Evaluation einige Punkte heraus: Die erzielten Wirkungen unterscheiden sich zwischen GTZ und HSS nur geringfügig, wobei die HSS effizienter arbeitete. An den einzelnen Standorten wurden deutliche Erfolge festgestellt, systematische Effekte waren dagegen kaum zu erkennen. Die Systemberatungsprojekte zielten auf Veränderungen in den Curricula, bei den Multiplikatoren und die Bereitstellung von Forschungsergebnissen mit dem Ziel, ein duales Berufsbildungssystem zu etablieren. Es konnten auch Diffusionswirkungen auf die operationale Ebene festgestellt werden. Die Doppelstrategie der GTZ hatte zwar nur begrenzte Wirkungen, Veränderungen auf höchster Ebene in Form einer Gesetzesänderung wurden nicht ausreichend umgesetzt. Es wurde deutlich, dass die staatliche Zusammenarbeit im Bildungsbereich mit China nicht über eine integrierte Gesamtplanung verfügt.

In der anschließenden Diskussion wurde u.a. die Zuordnung der Gesetzesänderung zur Intervention der GTZ problematisiert. *Dr. Krapp* erläuterte die klaren Aussagen im Gesetzestext und die Bestätigung des Zusammenhangs in Interviews. Der Zusammenhang zwischen den unterschiedlichen Ebenen und die mangelnde Ausstrahlung der ‚Insellösungen‘ sowie die Messung von Effizienz wurden ebenfalls diskutiert. Die Projektträger GTZ und HSS haben inzwischen Konsequenzen aus den Ergebnissen der Evaluierung gezogen und ihre Projekte entsprechend verändert.

3. Messung langfristiger Wirkungen durch ‚Tracer‘- und ‚Tracking‘-Methoden

Caspar Merkle (International Labour Organization) erläuterte im dritten Vortrag Pilotstudien zur Wirkungsmessung im Internationalen Programm zur Beseitigung der Kinderarbeit (IPEC). Mit Hilfe einer ‚Tracer‘-Methode sollen Spuren in der Vergangenheit untersucht werden, die ‚Tracking‘-Methode bezieht sich auf eine projektbegleitende Wirkungsmessung, die auch als um Wirkungsindikatoren erweitertes Monitoring bezeichnet werden könnte.

IPEC soll Kinderarbeit zum einen durch direkte Interventionen beseitigen, die arbeitende Kinder und ihre Familien z.B. durch Mikrokredite unterstützen. Zum anderen sollen indirekte Interventionen den Kontext von Kinderarbeit verändern, z.B. durch Bewusstseinsbildung oder Gesetzgebung. Beide hier vorgestellten Instrumente untersuchen direkte Aktivitäten, die auf die Kinder, deren Familien und die direkte Umwelt wirken.

Beide Methoden untersuchen die fünf Kernbereiche Bildung (z.B. Schulstunden/Tag), Arbeit (z.B. Arbeitsstunden/Tag), Haushaltseinkommen, Gesundheit und Freizeit sowie die Perzeption von Kinderarbeit. Hinzu kommen 14 Subindikatoren.

Bei der Tracer-Studie handelt es sich um eine einmalige Erhebung 1-10 Jahre nach Projektende. Im Zentrum des Interesses stehen Veränderungen im Projektverlauf. Deshalb werden ehemalige Klienten im Rahmen einer Stichprobe zu ihren Erinnerungen an drei Zeitpunkten des Projekts befragt. Der dazu entwickelte standardisierte Fragebogen wurde in fünf Pilotländern ausprobiert. Ein langer Zeitraum zwischen Projektende und Studie erwies sich als problematisch, viele ehemalige Klienten wurden nicht mehr gefunden.

Das Tracking-System wird dagegen zu Projektbeginn entwickelt und untersucht eine Stichprobe von Kindern an drei Zeitpunkten mit detaillierten Erhebungen. Die Daten werden anders als bei der Tracer-Studie von den Partnern vor Ort selbst erhoben. Diese Methode wurde in drei Pilotländern ausprobiert.

Zu den wichtigsten Ergebnissen der Tracer-Pilotstudien zählt zum einen die Bedeutung von guten Daten (Listen von Kindern), da sonst die Kinder kaum aufgespürt werden können. Zum anderen muss die Befragung verbessert werden. Lange Fragebögen zur Abfrage dreier Zeitpunkte mit sich wiederholenden Fragen lieferten unzuverlässige Ergebnisse, obwohl zur Unterstützung der Erinnerung die Zeitpunkte an eindruckliche Ereignisse (Erdbeben, Schulbau, Besuch des Präsidenten) gekoppelt wurden.

Die Pilot-Tracking-Studien ergaben, dass die Partner vor Ort partizipieren müssen, da nur sie den Zugang zu den Klienten haben. Allerdings ist eine Überlastung und eine Überforderung der Partner zu vermeiden. Zukünftig sollen Tracer-Studien mit den in den Projekten vorhandenen Monitoring-Instrumenten integriert werden.

Insgesamt, stellte *Caspar Merkle* fest, konnten Veränderungen auf individueller Ebene über den Zeitverlauf hinweg erkannt werden. Reaktion auf positive und negative Kontext-Veränderungen konnten so identifiziert werden. Problematisch sind die Aspekte ‚Erinnerung‘ und ‚Attribution‘ sowie die Quantifizierung der Beiträge der Projekte. Als nächste Schritte stehen die Reduktion der Fragebögen, die Verkleinerung des Zeitfenster für Tracer-Studien auf 2-7 Jahre nach Projektende und die Kombination von Tracer und Tracking auf der Tagesordnung.

Auch hier schloss sich eine Diskussion an, in der *Caspar Merkle* verschiedene Gesichtspunkte erläuterte: In den fünf Kernbereichen wurden Wirkungshypothesen formuliert. Die Attribution wird dadurch erschwert, dass aus ethischen Gründen auf Kontrollgruppen verzichtet wurde. Stattdessen müssen die Ergebnisse mit anderen Methoden trianguliert werden, das wurde z.B. mit focus groups bereits getan. Die Vergleichbarkeit der Tracer-Studien ist dadurch gewährleistet, dass jeweils der gleiche Fragebogen benutzt wurde, der lediglich an den lokalen Kontext adaptiert wurde. Abschließend wurden auch problematische Aspekte angesprochen: Falls in Projekten viele Klienten migrieren muss in kürzeren Abständen gemessen werden. Die Koppelung der Erinnerungs-Zeitpunkte an einprägsame Ereignisse bringt die Gefahr mit sich, dass gerade diese Ereignisse als externe Schocks Kinderarbeit beeinflussten und so die Attribution von Projektwirkungen zusätzlich erschweren.

4. Wirkungen auf der Spur

Anschließend fasste *Theo Mutter* die Vorträge zusammen. Es handele sich um drei sehr interessante Fallbeispiele die jeweils intensiv diskutiert werden müssten. Er hob hervor, dass diese aufwändigen Pilotvorhaben nicht mit ‚normalen‘ Vorhaben vergleichbar seien und forderte dazu auf, die Konsequenzen für die praktische Arbeit zu diskutieren. In jedem Fall blieben die Probleme der Zuordnung und der Datenqualität bestehen.

Ausgehend von diesem letzten Punkte forderten mehrere Anwesende eine verstärkte Nutzung von qualitativen Methoden, um systemische Aspekte besser zu er-

fassen. Demgegenüber betonte *Dr. Alexandra Caspari* die Notwendigkeit des Monitorings. Monitoring-Daten seien zwar problematisch, aber im Zweifel besser als Daten, die erst im Rahmen einer Ex-post-Evaluation erhoben werden.

Die Wirkung der vorgestellten Evaluationen war ein weiterer Diskussionpunkt. *Dr. Weidnitzer* erläuterte, dass sowohl die GTZ als auch die jemenitische Regierung das System inzwischen für andere Zwecke anwenden. *Prof. Stockmann* (CEval) betonte, der Wert der Evaluation liege in der Nutzung. Im Falle der untersuchten Projekte in China habe die HSS begonnen, ein Monitoring-System zu entwickeln, die GTZ habe sogar noch weiter gehende Konsequenzen gezogen.

Über diesen Fall hinaus, so *Prof. Stockmann*, sollten Evaluationen zukünftig verstärkt begleitend stattfinden, möglichst mit Kontrollgruppen. Dazu müsse das Projektmanagement angepasst werden, evtl. sei auch mehr Geld notwendig. Die Partner vor Ort müssten für gute Datenerhebung und für die gemeinsame Diskussion methodischer Fragen qualifiziert werden.

Abschließend machte *Michaela Zintl* (BMZ) einen Silberstreif am Horizont bei der Datenerhebung aus, sowohl in der Qualität als auch der Quantität. Sie forderte dazu auf, den Unterschied zwischen Evaluation und Wirkungsanalyse nicht zu verwischen: Evaluation solle auch untersuchen, warum Wirkungen erzielt werden. Damit kam sie auch auf die Grenzen von Projekt-Evaluationen zu sprechen. Das BMZ habe deshalb Schwerpunktbereiche definiert, gehe von Synergieeffekten aus und werde verstärkt Cluster-Evaluationen anstreben, bis hin zu Multi-Donor-Evaluierungen. Das für Evaluation notwendige Geld müsse in den Projektanträgen enthalten sein.

Der AK Evaluation von Entwicklungspolitik der DeGEval wird sich weiterhin vertiefend mit Wirkungsanalysen in der EZ befassen.