

Alexandra Caspari

Chancen der Wirkungsorientierung für die entwicklungspolitische Bildungsarbeit

Zusammenfassung:

Ausgelöst durch die internationale Diskussion über die Wirksamkeit der Entwicklungszusammenarbeit (EZ) wird aktuell auch im Bereich der entwicklungspolitischen Inlandsarbeit intensiv die ‚Wirkungsorientierung‘ der eigenen Projekt- und Bildungsarbeit diskutiert. Allerdings scheinen hierbei – wie auch in der Diskussion der EZ selbst – nicht alle Begriffe einheitlich definiert, was die Diskussion bzw. Argumentation erschwert. Begriffe wie Wirkungsorientierung, Wirkungsbeobachtung, Wirkungsmessung scheinen teils synonym verwendet zu werden. Auch werden teilweise Diskussionen um Untersuchungsdesigns mit Erhebungsmethoden vermengt. Im Beitrag wird eine Klärung der verschiedenen Begrifflichkeiten aufgeführt. Insbesondere wird aufgezeigt, dass die aktuell diskutierten ‚Designs‘ für Praktikerinnen und Praktiker der entwicklungspolitischen Bildungsarbeit eine besondere Chance bieten und nicht wie teilweise angeführt umfangreiche „spezifische Kompetenzen in der empirischen Sozialforschung“ (Jungk 2010, S. 21) benötigen.

Schlüsselworte: *Begriffsdefinitionen, Wirkungsorientierung, Wirkungsbeobachtung, Wirkungsmessung, Untersuchungsdesigns und Erhebungsmethoden, empirische Sozialforschung*

Abstract:

As a result of the international discussion on the impact of development cooperation also in the field of development education ‘impact orientation’ is currently being discussed intensively. However, it seems that there are no coherent definitions of the concepts, which hampers argumentation and discussion. Terms like impact orientation, impact monitoring, impact measurement appear to be used synonymously. Discussions on survey design and data collection methods are also mixed up. The article gives a clarification on the different concepts. It shows especially that the actually discussed designs provide an opportunity for practitioners involved in development education and do not need comprehensive “specific competence in empirical research” (Jungk 2010, p. 21) as it is sometimes stated.

Keywords: *terminology, impact orientation, impact monitoring, impact measurement, survey design and data collection methods, empirical research*

Ausgelöst durch die Aid Effectiveness Debatte und verstärkt durch die Pariser Erklärung sowie den Accra Aktionsplan wird in der internationalen Entwicklungszusammenarbeit (EZ) seit einigen Jahren intensiv über die Wirksamkeit von EZ-Maßnahmen diskutiert. Zwischenzeitlich hat diese Diskussion auch die deutschen Medien erreicht, die unter Titeln wie „Mutige Experimente gegen die Armut“ (Handelsblatt 216/2009), „Was wirklich hilft“ (Die Zeit 36/2011) oder „Helfer im Härtesten“ (Der Spiegel 42/2011) von ‚Randomistas‘ berichten, die ‚wissenschaftliche Experimente‘ durchführen. Auch in der entwicklungspolitischen Inlandsarbeit wird seit geraumer Zeit die ‚Wirkungsorientierung‘ der eigenen Projekt- und Bildungsarbeit verstärkt diskutiert, wobei die besondere Bedeutung der Qualität von entwicklungspolitischen Bildungsprozessen und deren Überprüfung keinesfalls neu ist (vgl. ZEP – Zeitschrift für internationale Bildungsforschung und Entwicklungspädagogik Heft 1/2001 und 2/2008). Bereits im aktualisierten BMZ Konzept zur entwicklungspolitischen Informations- und Bildungsarbeit (vgl. BMZ 2008), welches das alte Dokument ‚BMZ-Konzepte Nr. 119‘ von 2002 ablöste, wird in einem eigenen Kapitel zur Qualitätssicherung und Erfolgskontrolle aufgeführt, dass sich „auch in der entwicklungspolitischen Informations- und Bildungsarbeit (...) zunehmend die Erkenntnis durch[setzt], dass transparente und glaubwürdige Erfolgskontrollmechanismen einen wichtigen Beitrag zur Legitimation und Qualitätssicherung leisten“ (BMZ 2008, S. 9), verstärkte Aufmerksamkeit soll daher u.a. dem Aufbau und der Weiterentwicklung von glaubwürdigen Erfolgskontrollmechanismen gewidmet werden (vgl. ebd.).

Wirkungsorientierung und Wirkungsbeobachtung oder: Wirkungsorientiertes Monitoring

Was aber wird genau unter Wirkungsorientierung verstanden? Im Vorwort der Broschüre ‚Wirkt so‘ plädiert Massing dafür, Wirkungsorientierung eher als einen „Grundanspruch, eine Zielrichtung“ zu verstehen, „nicht als ein starres Konzept“ (Massing 2010, S. 6). Wirkungsorientierung bedeutet zunächst einmal, dass bereits bei der Planung von Programmen und Projekten der Fokus auf die angestrebten Wirkungen einer Maßnahme gelegt wird und nicht mehr wie früher auf die Erfüllung von Aktivitäten oder erbrachte Inhalte und Leistungen (Inputs). Eine wirkungsorientierte Perspektive beinhaltet die de-

zidierte Formulierung der beabsichtigten Wirkungen sowie der zur Erreichung notwendigen Prozesse, meist in Form von Wirkungsketten, die die geplante Reichweite einer Maßnahme sowie den Weg dorthin in einem kausalen Zusammenhang auf-führt, d.h. die einer Maßnahme zugrundeliegenden Hypothesen über Ursache-Wirkungs-Zusammenhänge werden detailliert ausgearbeitet: Diese Programmtheorie zeigt auf, welche Annahmen über kausale Verknüpfungen zwischen der geplanten Maßnahme und den intendierten Wirkungen der Maßnahme zugrunde liegen, wobei die Logik des Wirkungsgefüges zwischen Ressourcen, Aktivitäten, Leistungen (Output), Nutzung (Outcome I), direkte Wirkungen (Outcome II) und indirekte Wirkungen (Impact) unterscheidet (vgl. z.B. Masing/Rosen/Struck 2010, S. 29). Eine Programmtheorie beschreibt demnach explizit, was eine Maßnahme wie (mit welchen Ressourcen und Aktivitäten), für wen und wozu (mit welchem übergeordneten Ziel) erreichen soll. Für jede Wirkungsebene wird hierbei eine eigene Hypothese über den vermuteten Ursache-Wirkungszusammenhang erstellt (vgl. Caspari 2009, S. 205 ff.; Caspari/Barbu 2008, S. 17 ff.). Des Weiteren werden für jede Ebene der Wirkungskette Indikatoren entwickelt, wobei im Kontext der Wirkungsorientierung neben Prozessindikatoren auf Leistungs- und Nutzungsebene auch Wirkungsindikatoren insbesondere auf Ebene der direkten Wirkungen formuliert werden. Wirkungsketten sind somit Grundlage eines strategischen Planungs-, Management- und Steuerungskonzepts nach dem Prinzip des ‚results-based Management‘, d.h. Basis eines umfassenden Monitoring-Systems. Wirkungsorientierung bedeutet somit letztendlich die Implementierung eines wirkungsorientierten Monitorings, d.h. einem Monitoring-System, das auf Basis zuvor festgelegter Indikatoren nicht nur überprüft, was getan wurde, sondern auch, welche Veränderungen bewirkt wurden, das also auf die kontinuierliche Wirkungsbeobachtung ausgerichtet ist.

Wirkungskontrolle oder: Wirkungsevaluation

Häufig werden im Kontext der Wirkungsorientierung auch Begriffe wie Wirkungskontrolle, Wirkungsanalyse oder Wirkungsüberprüfung in die Diskussion eingeworfen. Diese Begriffe sind entgegen dem zuvor dargestellten wirkungsorientierten Monitoring nun eher der Wirkungsevaluation im Sinne von erfolgsbewertender (summativer) Evaluation mit dem Ziel der Rechenschaftslegung zuzuordnen.

Ausgehend von der Definition zu Wirkungen (Impacts) des OECD/DAC, demnach Wirkungen „positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended“ (OECD/DAC 2002, S. 24) sind, wird mit Betonung auf ‚effects produced by‘ der Begriff Wirkungsevaluation definiert als: „studies that measure the impact directly attributable to a specific program or policy, as distinct from other potential explanatory factors“ (CGD 2006, S. 10; Hervorheb. AC). Zentrale Schwierigkeit bei der ‚Attribution‘, d.h. der kausalen Zuordnung von Wirkungen zu einer Maßnahme, ist die Frage, ob festgestellte (positive) Veränderungen bei Teilnehmer/inne/n einer Maßnahme tatsächlich Folge dieser Maßnahme sind oder ob sich der Erfolg auch ohne diese Maßnahme eingestellt hätte. Daher wird zwischen Brutto- und Nettowirkungen unter-

schieden (vgl. Rossi/Freeman/Hofmann 1988): Bruttowirkungen werden dabei verstanden als die nach Beendigung einer Maßnahme bei Teilnehmer/inne/n insgesamt beobachtbaren Veränderungen, die sowohl auf die Maßnahme als auch eine beliebige Anzahl andere Einflüsse zurückzuführen sind (vgl. Abb. 1, Differenz zwischen a und b). Nettowirkungen dagegen umfassen nur den isolierten Anteil an den insgesamt auftretenden Veränderungen bei den Teilnehmer/inne/n, der allein auf die durchgeführte Maßnahme zurückzuführen ist, d.h. der nicht eingetreten wäre, wenn die Maßnahme nicht durchgeführt worden wäre (vgl. Abb. 1, Differenz zwischen b und c).

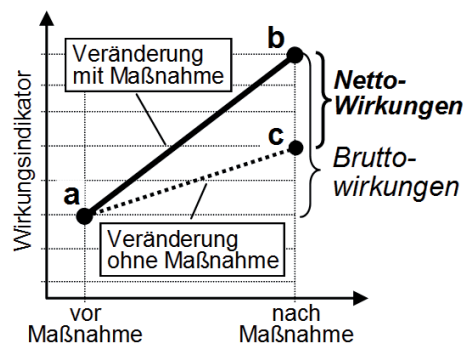


Abb. 1: Brutto- und Nettowirkungen einer Maßnahme.
Quelle: eigene Darstellung

Diese Nettowirkungen sind jedoch nicht direkt beobachtbar, da Personen entweder Teilnehmer/-innen oder Nicht-Teilnehmer/-innen einer Maßnahme sind, niemals beides gleichzeitig. D.h. die in Abb. 1 aufgezeigte Veränderung b-c kann für Teilnehmer/-innen einer Maßnahme nicht direkt gemessen werden. Daher werden Wirkungen dadurch ermittelt, dass die bei Teilnehmer/inne/n tatsächlich festgestellten Veränderungen mit Veränderungen bei einer Gruppe von Nicht-Teilnehmer/inne/n verglichen werden – einer Vergleichsgruppe. Die Differenz der Veränderungen der beiden Gruppen kann dann als kausaler Nettoeffekt interpretiert werden, der allein der evaluierten Maßnahme zuzuschreiben ist. Wirkungsevaluationen fokussieren folglich auf die Bestimmung der Nettowirkungen einer Maßnahme.

Untersuchungsdesigns und Erhebungsmethoden

Um Wirkungen kausal einer Maßnahme zuordnen zu können, müssen also Kontroll- bzw. Vergleichsgruppen gebildet werden. Zentrales Moment in Wirkungsevaluationen ist daher die Wahl des geeigneten Untersuchungsdesigns, denn das Design ist entscheidend für den Grad an Gewissheit, mit dem ein Ursache-Wirkungs-Zusammenhang aufgezeigt werden kann (Schnell/Hill/Esser 1999, S. 203).

Häufig wurden bzw. werden im Rahmen von Evaluationen vorexperimentelle Designs angelegt, d.h. die (vermeintlichen) Wirkungen einer Maßnahme werden mittels eines Vorher-Nachher-Vergleichs bei der Zielgruppe (ZG) bzw. den Teilnehmer/inne/n einer Maßnahme erhoben (vgl. Abb. 2: (b)): Baselinedaten (erhoben vor der Maßnahme) werden mit Daten aus einer Erhebung nach der Maßnahme verglichen, wobei die aufgefundenen Veränderungen der Maßnahme zugeschrieben werden.

Ein solcher Vorher-Nachher-Vergleich zeigt jedoch lediglich die Entwicklung der Zielgruppe über die Zeit hinweg

DESIGN	Vorher-Daten t_1 Baseline	Projekt/ Maßnahme X	Nachher-Daten t_2 Survey
(a)	Ein-Gruppen-Nachtest-Design	X	ZG _{t2}
(b)	Ein-Gruppen-Vortest-Nachtest-Design	ZG _{t1}	X ZG _{t2}

ZG: Zielgruppe, t: Zeitpunkt (erste, zweite Datenerhebung), X: Projekt/Maßnahme

Abb. 2: Vorexperimentelle/Nicht-experimentelle Versuchsanordnung
Quelle: eigene Darstellung

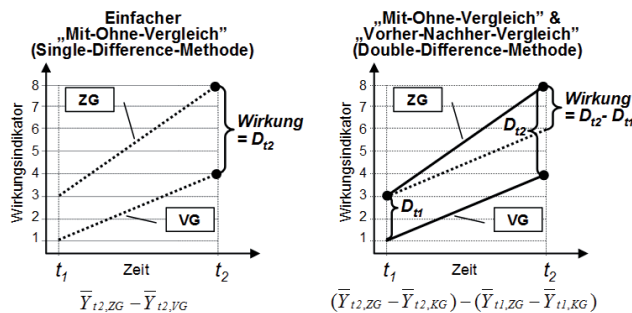
auf und kann demnach keine zuverlässige Antwort auf die Frage nach den Nettowirkungen einer Maßnahme geben – gemessen werden lediglich die Bruttowirkungen (vgl. auch Abb. 1 a-b oben). Die hierbei beobachteten Veränderungen sind selten allein sondern nur teilweise auf die Maßnahme zurückzuführen. Andere, externe Faktoren können die Wirkung der Maßnahme beeinflussen, schwächen oder auch verstärken. Gleichwohl könnten sich die Einstellungen oder Kompetenzen der Zielgruppe auch ohne Maßnahme verändert haben. Solche Faktoren, die teilweise oder ganz für die beobachtete Veränderung verantwortlich sein können, bleiben bei einem reinen Vorher-Nachher-Vergleich der Teilnehmer/-innen einer Maßnahme unberücksichtigt.

Dieses Problem kann durch das Bilden einer Kontrollgruppe (KG) bzw. Vergleichsgruppe (VG) gelöst werden, d.h. Personen, die nicht an der Maßnahme teilgenommen haben, ansonsten aber in allen anderen Aspekten identisch bzw. möglichst identisch mit der Gruppe der Teilnehmer/-innen ist. Hierdurch wird ein Mit-Ohne-Vergleich möglich (Single-Difference-Methode SD), d.h. die Wirkungen werden über einen einfachen Vergleich zwischen den Teilnehmer/inne/n und Nicht-Teilnehmer/inne/n einer Maßnahme erhoben. Zentrale Grundannahme hierbei ist, dass die Ausgangssituation der ZG und der VG vor der Maßnahme identisch ist. Diese Annahme ist in der Realität der internationalen EZ nicht immer gegeben, da Maßnahmen häufig gezielt für Personen aufgelegt werden, die entweder besondere Defizite aufweisen oder aber besondere Voraussetzungen erfüllen. Dadurch werden bei diesem Mit-Ohne-Vergleich die berechneten Wirkungen der Maßnahme je nach Situation über- oder unterbewertet. Diesem Problem kann begegnet werden, in dem eine Vorher-Messung sowohl bei der ZG als auch der VG durchgeführt wird, d.h. der Mit-Ohne-Vergleich mit dem Vorher-Nachher-Vergleich kombiniert wird (Double-Difference-Methode DD). Die Wirkung einer Maßnahme ergibt sich aus dem Unterschied zwischen ZG und VG nach der Maßnahme minus dem Unterschied zwischen ZG und VG vor der Maßnahme. Für Wirkungsevaluationen relevante Designs sind daher (echte) experimentelle Designs oder quasi-experimentelle Designs.

Experimentelle Designs/Randomized Controlled Trials (RCT)

Beim experimentellen Design werden Personen, Schulklassen oder Schulen vor der Implementation einer Maßnahme nach dem Zufallsprinzip (Randomisierung) zwei Gruppen zugeordnet. Zum einen eine Gruppe, die an der geplanten Maßnahme teilnimmt also die Zielgruppe (ZG) sein wird (Teilnehmer/-innen), und zum anderen eine Kontrollgruppe (KG), die nicht an der Maßnahme teilnehmen wird (Nicht-Teilnehmer/-innen).

Experimentelle Designs gelten als die rigoroseste Evaluationsmethodologie schlechthin, denn durch den zufallsgesteuerten Auswahlprozess werden systematische Gruppenun-



• Datenerhebung, ZG: Zielgruppe, VG: Vergleichsgruppe.
t: Zeit (erste, zweite Datenerhebung), D: Differenz

Abb. 3: Geeignete Designs zur Wirkungsmessung
Quelle: eigene Darstellung

terschiede eliminiert, d.h. systematische Auswahlverzerrungen liegen nicht vor (vgl. Schnell/Hill/Esser 1999, S. 213 f.; ADB 2006, S. 5 f.; Bloom 2006, S. 1; Baker 2000, S. 2). Da folglich die Ausgangssituation der ZG und KG vor der Maßnahme übereinstimmen, können alle nach der Maßnahme gemessenen Unterschiede in den Wirkungen zwischen ZG und der KG auch der Maßnahme zugeschrieben werden. Allerdings trifft dies nur in sogenannten Laborexperimenten zu, in sogenannten Feldexperimenten, d.h. die Anwendung einer experimentellen Versuchsanordnung in für die Beteiligten authentischer Alltagsumgebung, ist es annähernd unmöglich, alle Eigenschaften der Personen vorab zu berücksichtigen, so dass Gruppenunterschiede zwischen ZG und KG nach wie vor gegeben sind. Während der einfache Mit-Ohne-Vergleich solche systematischen Gruppenunterschiede nicht berücksichtigt, können diese anhand der Kombination eines Mit-Ohne-Vergleichs und eines Vorher-Nachher-Vergleichs erfasst werden.

DESIGN	Vorher-Daten t_1 Baseline	Projekt/ Maßnahme X	Nachher-Daten t_2 Survey
(1)	ZG _{t1} KG _{t1}	X -	ZG _{t2} KG _{t2}

ZG: Zielgruppe, KG: Kontrollgruppe (randomisiert), t: Zeitpunkt (erste, zweite Datenerhebung), X: Projekt/Maßnahme

Abb. 4: Experimentelle Designs/Randomized Controlled Trials
Quelle: eigene Darstellung

Quasi-experimentelle Designs¹

Insbesondere im Rahmen einer Wirkungsevaluation sind experimentelle Designs jedoch häufig nicht möglich, da die Evaluator/inn/en meist nicht von Beginn an einer Maßnahme in den Prozess eingebunden sind, oder die Maßnahme bereits abgeschlossen wurde – eine randomisierte Zuteilung in Zielgruppe und Kontrollgruppe folglich nicht mehr realisierbar ist. Daher ist das quasi-experimentelle Design in der Evaluationspraxis das sicherlich am häufigsten angewandte: Das Quasi-Experiment orientiert sich an der Experimentallogik, allerdings wird hierbei keine Kontrollgruppe nach dem Zufallsprinzip gebildet, sondern es wird eine Vergleichsgruppe (VG) (re-)konstruiert.² Im Unterschied zum experimentellen Design erfolgt die Bestimmung der Vergleichsgruppen daher beim quasi-experimentellen Design im Nachhinein, d.h. entweder im Laufe der Maßnahme oder gar erst nach Abschluss der Maßnahme im Rahmen einer Evaluation. Die Abbildung zeigt verschiedene quasi-experimentelle Szenarien auf.

DESIGN	Vorher-Daten t_1 Baseline	Projekt/ Maßnahme X	Nachher-Daten t_2 Survey
(2) Vortest-Nachtest mit Vergleichsgruppen-Design	ZG _{t1} VG _{t1}	X –	ZG _{t2} VG _{t2}
(3) Vortest-Nachtest mit Nachtest Vergleichsgruppen-Design	ZG _{t1}	X –	ZG _{t2} VG _{t2}
(4) Survey-Design		X –	ZG _{t2} VG _{t2}

ZG: Zielgruppe, VG: Vergleichsgruppe (nicht randomisiert),
t: Zeitpunkt (erste, zweite Datenerhebung), X: Projekt/Maßnahme

Abb. 5: Quasi-experimentelle Designs
Quelle: eigene Darstellung

Liegen Daten sowohl für die Zielgruppe als auch einer Vergleichsgruppe für den Zeitpunkt vor Implementierung einer Maßnahme vor (vgl. Abb. 5: (2)) kann die Double-Difference Methode angewandt werden, was – bei guter Datenlage und optimaler Umsetzung – zu ähnlich robusten Ergebnissen über die Wirkungen einer Maßnahme führt wie ein experimentelles Design (vgl. Abb. 4: (1) oben). Liegen Baselinedaten lediglich für die ZG vor, ist es dennoch sinnvoll, eine Vergleichsgruppe für einen Nachher-Vergleich heranzuziehen (vgl. Abb. 4: (3)) und zumindest die Single-Difference-Methode anzuwenden und dem Vorher-Nachher-Vergleich der ZG gegenüber zu stellen. Die Ergebnisse sind hierbei allerdings wesentlich weniger robust als bei den zuvor beschriebenen Designs. Die in der Evaluationspraxis häufig genutzte und fast immer umsetzbare Variante zeigt Szenario (4) in Abbildung 4: beim sogenannten ‚Survey-Design‘, d.h. Datenerhebungen nur nach Beendigung der Maßnahme, liegen keinerlei Basis-Daten vor. Der gefundene Unterschied zwischen der im Nachhinein konstruierten VG und der ZG wird hierbei als Wirkung der Maßnahme interpretiert. Da mögliche Gruppenunterschiede zum Zeitpunkt vor der Maßnahme hierbei völlig unbeachtet bleiben, ist eine äußerst bedachte Auswahl der VG nötig.

Die dargestellten experimentellen bzw. quasi-experimentellen Designs, die eine Vorher-Nachher-Messung sowie eine Vergleichsgruppe umfassen und somit eine Double-Difference-Berechnung zulassen, sind zuverlässige Methoden, um Wirkungen eindeutig einer Maßnahme zuzuschreiben und auch deren Umfang darzustellen. Allerdings wird hierdurch lediglich die Frage beantwortet „Hat die Maßnahme Wirkungen entfaltet (oder nicht)?“ nicht aber die Frage „Warum hat eine Maßnahme Wirkungen entfaltet (oder nicht)?“ bzw. „Wie hat die Maßnahme gewirkt; unter welchen Bedingungen?“. Evaluationen, die Ergebnisse über die Wirkungen geben, aber keine Angaben machen, warum eine Maßnahme die erwarteten Wirkungen gezeigt hat oder nicht, werden auch ‚black box‘ Evaluationen genannt, denn die fehlende Information über die vermittelnden Prozesse und Wirkmechanismen hinterlässt eine leere ‚black box‘ (vgl. Leeuw/Vaessen 2009, S. 20; Bloom 2006, S. 18 f.; White 2006, S. 9; Ravallion 2005, S. 1). Man könnte aber auch argumentieren, dass derartige ‚Wirkungsevaluationen‘ keine wirklichen Evaluationen sind sondern lediglich Wirkungsstudien, durchgeführt von ‚Randomistas‘.

Erhebungsmethoden

Im Rahmen der aktuellen Wirkungsdiskussion kommen immer wieder Debatten über die Frage der Datenerhebungsmethoden auf, was teilweise an den Paradigmenstreit zwischen quantitativer und qualitativer Sozialforschung erinnert (vgl. Caspari 2008, S. 139 f.; Caspari/Barbu 2008, S. 20 f.; Caspari

2004, S. 92 ff.; Caspari 2006). Hierbei werden jedoch zentrale Punkte der Diskussion um Wirkungsnachweise von entwicklungspolitischen Maßnahmen nicht berücksichtigt:

(1) Im Kontext der Diskussion um adäquate Untersuchungsdesigns ist die Frage der Datenerhebungsmethode nicht explizit enthalten. Problem ist, dass meist unterstellt wird bei Anwendung von (quasi-)experimentellen Designs könnten die notwendigen Daten ausschließlich standardisiert erhoben werden. Dies mag in der Forschungspraxis durchaus zutreffen, ist jedoch nicht zwingend notwendig: Wirkungsevaluationen setzen immer auf der Ebene der Zielgruppen an – hat diese einen ‚überschaubaren‘ geringen Umfang, können die interessierenden Informationen über die Wirkungen einer Maßnahme, die Wirkungsindikatoren, auch anhand qualitativer Methoden erhoben werden. Problem ist, dass in der Realität die Zielgruppe durchaus eine erhebliche Anzahl an Einheiten (Personen) umfassen kann, sodass es nicht immer möglich ist, qualitative Datenerhebungsmethoden anzuwenden. Hierbei ist in einem ersten Schritt die Frage der Repräsentativität zu stellen, d.h. wie groß ist die Zielgruppe und wieviele Einheiten muss ich auswählen, um repräsentative Ergebnisse zu erhalten. Je nach Umfang der notwendigen Stichprobe ist es wiederum durchaus möglich, qualitative Methoden zu nutzen. Erst wenn auch die Stichprobe einen erheblichen Umfang einnimmt, scheinen quantitative Methoden aufgrund finanzieller und zeitlicher Gründe angemessener: da quantitative im Gegensatz zu qualitativen Methoden eher in die Breite denn in die Tiefe gehen, sind sie weitaus einfacher und schneller umzusetzen, beinhalten natürlich aber dadurch einen Informationsverlust.

(2) Die Begriffe qualitative sowie quantitative Methode werden häufig sehr pauschal genutzt, wobei meist die Zuordnung zu einem der Paradigmen der Sozialforschung impliziert wird. Allerdings gibt es verschiedenen Ebenen, die zu betrachten sind: so gibt es die Frage der (qualitativen oder quantitativen) Datenerhebungsmethode – aber ebenso auch die Frage, ob qualitative oder quantitative Indikatoren besser geeignet scheinen, und ob die Datenauswertungsmethode qualitativ oder quantitativ sein soll. So werden häufig qualitativ erhobene Daten anschließend quantifiziert und statistischen Analysen zugeführt. Auch werden im Rahmen von quantitativen Datenerhebungsmethoden durchaus qualitative Größen erfasst (Geschlecht, Einschätzungen, Bewertungen, Meinungen), die allerdings häufig z.B. über Skalen quantifiziert und somit ‚messbar‘ gemacht werden. Daraus folgt, dass im Kontext der Designdiskussion in erster Linie Datenauswertungsmethoden impliziert sind, denn zur Berechnung z.B. der Double-Difference ist eine Quantifizierung der erhobenen Informationen nötig. Die Frage der Datenerhebungsmethode wird dagegen nicht dezidiert vorgegeben.

(3) Im Rahmen einer (echten) Wirkungsevaluation ist auch die der Maßnahme zugrundeliegende Programmtheorie zu überprüfen. Hier scheint die Nutzung quantitativer Datenerhebungsmethoden nicht wirklich sinnvoll. Insbesondere nicht-reaktive Verfahren wie die Analyse von Programmdokumenten und -berichten sowie wissenschaftlicher Literatur aber auch sonstige Unterlagen wie z.B. Fotos oder Karten etc. sind relevant. Für die Überprüfung der hypothetisierten Ursache-Wirkungs-Zusammenhänge sind insbesondere aber Intensiv-, Experten- und/oder Fokusgruppeninterviews notwendig. Auch

viele für die EZ-Praxis erarbeitete vorwiegend partizipative Tools wie z.B. der Problem-Baum oder das Venn-Diagramm eignen sich für diese Aufgabe sehr gut.

(4) Wirkungsevaluationen haben gemäß Definition auch nichtintendierte Wirkungen zu erfassen. Werden vor der Überprüfung der Ursache-Wirkungs-Hypothesen nicht auch mögliche nichtintendierte und negative Effekte einer Maßnahme intensiv mit den Beteiligten und auch insbesondere Vertreter/innen der Zielgruppe diskutiert, ist die Wahrscheinlichkeit groß, dass diese unberücksichtigt bleiben. Von daher sind auch Intensiv- und/oder Fokusgruppeninterviews auf der Zielgruppenebene notwendig.

(5) Insbesondere aber für die Interpretation der gefundenen Wirkungen sind Informationen, die mit Hilfe nicht-reaktiver sowie qualitativer Datenerhebungsmethoden erfasst wurden, wichtig: Ergebnisse der quantitativen Double-Difference Berechnung zur Wirkungsmessung sollten anhand qualitativer Informationen erläutert und interpretiert werden, insbesondere, wenn die Ergebnis unerwartet sind.

Es zeigt sich also, dass für angemessene Wirkungsevaluationen die Nutzung des Methodenrepertoires beider Methodologien i.S. einer Triangulation/Methodenmix notwendig ist, denn "both research traditions and the research methods they are most closely linked to (quantitative vs qualitative), are suited to answering very different types of research question" (Prowse 2007, S. 3). Die Erfordernis der Integration von qualitativen und quantitativen Ansätzen bei Wirkungsevaluationen ist auch in Deutschland zwischenzeitlich unumstritten (vgl. insb. Bamberger 2000; Chung 2000; Guijt 2000; Apleton/Booth 2001; Ezemenari u.a. 1999; Kassam 1998): „A good evaluation necessitates the combining of both qualitative and quantitative approaches to data collection and analysis. Combining both approaches provides quantified results of program impacts as well as explanations of the processes and intervening factors that yielded these outcomes. Combined approaches enrich interpretation or explanation (causality) of outcomes measured by the evaluation“ (Ezemenari/Rudqvist/Subbarao 1999, S. 19).

Chancen und Herausforderungen für die entwicklungspolitische Bildungsarbeit

Stellt sich nun die Frage, ob dies alles auch von Praktikerinnen und Praktikern der entwicklungspolitischen Bildungsarbeit umgesetzt werden kann, oder ob hierfür tatsächlich wie von Jungk angeführt umfangreiche „spezifische Kompetenzen in der empirischen Sozialforschung“ (Jungk 2010, S. 21) benötigt werden. Erreichen Maßnahmen der entwicklungspolitischen Bildungsarbeit aufgrund der Wirkungsorientierung tatsächlich „ein Komplexitätsniveau, das sich teils nur noch mit Hauptamtlichen bewältigen“ lässt (Massing 2010, S. 6)?

Wirkungsevaluationen sind – korrekt umgesetzt – in der Tat nicht ohne fundierte Kenntnisse möglich. Wie bei jeder anderen Evaluation sind auch hier Kompetenzen im Bereich der Evaluationstheorie und -geschichte sowie der Methoden der empirischen Sozialforschung nötig (vgl. DeGEval 2004; Caspari 2004, S. 83 ff.). Wirkungsorientierung i.S.v. Wirkungsbeobachtung, d.h. die Implementierung eines wirkungsorientierten Monitoring-Systems zu Beginn einer Maßnahme, kann und muss dagegen auch von Praktiker/innen in jedem Projekt oder Programm der entwicklungspolitischen Bildungs-

arbeit umgesetzt werden. Denn, wie Rosen treffend formuliert: „Wenn Organisationen bei der Projektplanung das Gerüst von Problem-Zielgruppe-Zielstellung-Strategie-Aktivitäten-Erfolgskontrolle [besser: Erfolgs-/Wirkungsbeobachtung oder Monitoring] nicht haben, wie überprüfen sie im Rahmen der Auswertung dann, ob sie adäquat handeln? Wie lernen sie als Organisationen? Wie entwickeln sie methodisch und didaktisch ihre Bildungsarbeit weiter? Wann und wo hinterfragen sie die eingesetzten Instrumente und Maßnahmen ihrer entwicklungspolitischen Bildungs- und Inlandsarbeit?“ (Rosen 2010, S. 11). Mehr noch erscheint es in Zeiten begrenzter Ressourcen nicht mehr ausreichend, als Ziel schlicht die Durchführung von x Bildungsveranstaltungen zu formulieren. Es ist notwendig, dass sich alle Beteiligten einer Maßnahme im Rahmen der Planung dezidiert darüber austauschen, welches konkrete Problem sie mit welcher Maßnahme positiv verändern wollen und welche Wirkungen hierdurch erwartet werden.

Infolgedessen würde evtl. auch deutlich, dass teilweise die erwarteten Wirkungen zu hoch angesetzt sind. So erscheinen die von Seitz aufgeführten Ziele der entwicklungspolitischen Bildungsarbeit ausgesprochen ambitioniert: „Die entwicklungspolitische Bildungsarbeit umfasst Aktivitäten, die darauf zielen, Menschen zu befähigen, an der Gestaltung einer gerechten Weltgesellschaft aktiv teilzuhaben und sich im eigenen Lebensumfeld gemeinsam mit anderen für eine zukunftsfähige und solidarische Entwicklung zu engagieren. Sie baut individuelle und kollektive Gestaltungskompetenzen für eine zukunftsfähige Entwicklung auf“ (Seitz 2010, S. 14). Dagegen werden im BMZ-Konzept zur entwicklungspolitischen Informations- und Bildungsarbeit als Ziele ‚lediglich‘ aufgeführt: ‚Interesse wecken‘, ‚Auseinandersetzung fördern‘, ‚zur aktiven Beteiligung motivieren‘, ‚entwicklungspolitisches Engagement unterstützen‘ (vgl. BMZ 2008, S. 4 f.).

Wird also im Kontext der Projektplanung intensiv über die letztendlich angestrebten direkten Wirkungen nachgedacht, so werden diese im Ergebnis auch realistischer i.S.v. ‚machbar‘ angesetzt. Darüber hinaus fällt es somit auch leichter, passende Indikatoren zu formulieren. Es kann an dieser Stelle nicht verschwiegen werden, dass die Formulierung adäquater SMARTer³ Indikatoren insbesondere auf Outcome-Ebene keine einfache Aufgabe ist. Auch in der EZ wird dies immer wieder deutlich (vgl. Caspari 2011, S. 43 f.). Es kann hier auch für die entwicklungspolitische Bildungsarbeit nur empfohlen werden, eine für alle in diesem Bereich Tätigen zugängliche Indikatoren-Datei anzulegen, in der bisher genutzte Indikatoren aufgelistet werden, sodass nicht immer wieder eigene Indikatoren neu erarbeitet werden müssen, sondern als gut erachtete Indikatoren lediglich modifiziert und evtl. weiter spezifiziert werden können. Konkrete Vorschläge für Indikatoren finden sich bereits in der ‚Handreichung zur Evaluation entwicklungbezogener Bildungsarbeit‘ (Scheunpflug/Bergmüller/Schröck 2010) sowie der ‚Handreichung zur Wirkungsorientierung und Antragstellung in der entwicklungspolitischen Inlandsarbeit‘ (Massing/Rosen/Struck 2010). Auch scheint der aktuelle pädagogische Paradigmenwechsel von der reinen Vermittlung von Wissen hin zur Vermittlung von Kompetenzen neue Möglichkeiten zu eröffnen, denn „anders als klassische Lernziele erscheinen Kompetenzen als messbar“ (Massing/Rosen/Struck 2010, S. 30).

Wird eine entwicklungspolitische Bildungsmaßnahme von Beginn an wirkungsorientiert ausgerichtet, ergeben sich hieraus konkrete Vorteile: es können mögliche Designs eruiert und implementiert werden. Experimentelle Designs werden teilweise aufgrund ethischer Vorbehalte abgelehnt. Manche Maßnahmen – auch der entwicklungspolitischen Bildungsarbeit – werden jedoch derart konzipiert, dass ein experimentelles Design ohne jegliche ethische Bedenken angewandt werden kann: Wenn im Rahmen eines größeren Programms eine geplante Bildungsmaßnahme in mehreren Phasen für einzelne Personen (bzw. Schulklassen, Schulen oder in einzelnen Städten etc.) zeitversetzt implementiert wird und es keine bewusste Entscheidung darüber gibt, warum eine Person an der ersten Phase teilnehmen sollte (andere dagegen an der zweiten oder späteren Phase), dann können die Personen, die erst in einer späteren Phase an der Maßnahme teilnehmen werden, als Kontrollgruppe für die Teilnehmer/-innen der ersten Phase herangezogen werden (vgl. ADB 2006, S. 7 ff.; Bamberger 2006, S. 11; Ravallion 2005, S. 30; Baker 2000, S. 73). D.h. wird die Entscheidung, wer an welcher Phase der Maßnahme teilnimmt per Zufall getroffen, wird bereits bei der Konzeption einer Maßnahme ein experimentelles Design angelegt. Später durchgeführte Wirkungsanalysen können somit auf adäquate Kontrollgruppen zurückgreifen. Diese Vorgehensweise nach dem sogenannten Pipeline Verfahren ist nicht nur sehr ökonomisch, sondern liefert recht zuverlässige Daten, da keine größeren systematischen Unterschiede in den Charakteristika von Zielgruppe und Kontrollgruppe vorliegen.

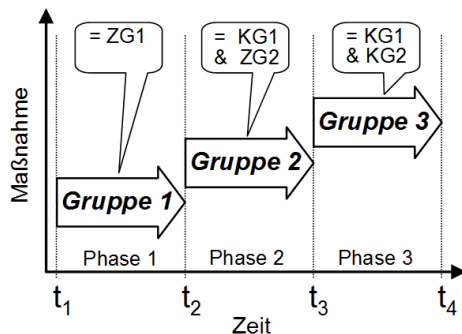


Abb. 6: Umsetzung experimenteller Designs: Das Pipeline Verfahren.
Quelle: eigene Darstellung

Aber auch wenn ein solches experimentelles Design nicht möglich ist, weil die geplante Bildungsmaßnahme lediglich an einer Schule zu einem Zeitpunkt umgesetzt wird, so führt Wirkungsorientierung dazu, dass alternative Designs berücksichtigt werden können: So kann einerseits eine Vorher-Messung der Kompetenzen bei den Teilnehmer/inne/n sowie eine Kompetenzüberprüfung bei Nicht-Teilnehmer/inne/n überlegt werden. Sicherlich mag es bei kleinen Projekten Argumente geben, die gegen solche Designs sprechen. Wirkungsorientierung bietet jedoch die Möglichkeit, diese Entscheidung zu Beginn eines Projektes bewusst zu treffen. Denn viel zu häufig werden Chancen vertan, weil erst zu spät über Wirkungsbeobachtung nachgedacht wird und somit zu spät erkannt wird, dass bei dem konkreten Projekt ein experimentelles oder zumindest quasi-experimentelles Design leicht umsetzbar gewesen wäre.

Die Herausforderung im Kontext von Wirkungsorientierung in der entwicklungspolitischen Bildungsarbeit besteht also in der Formulierung adäquater Wirkungsindikatoren. Das Design legt lediglich fest, zu welchem Zeitpunkt bei welchen Personen diese (meist ein oder zwei) Indikatoren erhoben werden, wobei die Erhebung selbst kein größeres Problem ist, insbesondere wenn auf bereits erprobte Indikatoren inkl. Anleitung zur Erhebung zurückgegriffen wird. Die Entscheidung für ein bestimmtes Design legt jedoch fest, ob letztendlich Wirkungen eindeutig der durchgeführten Bildungsmaßnahme zugeschrieben werden können (oder nicht) – eine Chance, die zukünftig hoffentlich stärker wahrgenommen wird.

Anmerkungen:

- 1 Quasi-experimentelle Designs werden insbesondere im Englischen teilweise ‚non-experimental‘ Designs genannt – dies ist jedoch verwirrend, da nicht gleichzusetzen mit vorexperimentellen Designs.
- 2 Kontrollgruppen (KG) sind demnach immer randomisiert, Vergleichsgruppen (VG) dagegen sind nicht-randomisiert.
- 3 specific, measurable, achievable/attainable, realistic/relevant, timed/time-bound.

Literatur:

- Appleton, S./Booth, D. (2001):** Combining Participatory and Survey-based Approaches to Poverty Monitoring and Analysis (Background Paper for the Workshop to be held in Entebbe, Uganda, 30 May-1 June 2001). (Online im Internet: http://www.odi.org.uk/pppg/publications/papers_reports/gov/ug_ws01/bp.pdf [30.04.12]).
- Asian Development Bank (ADB) (2006):** Impact Evaluation. Methodological and Operational Issues. Manila.
- Baker, J.L. (2000):** Evaluating the Impact of Development Projects on Poverty: a Handbook for Practitioners. Washington D.C.
- Bamberger, M. (Hg.) (2000):** Integrating Quantitative and Qualitative Research in Development Projects. Washington, D.C.
- Bamberger, M. (2006):** Conducting Quality Impact Evaluation Under Budget, Time and Data Constraints. Washington D.C.
- Bloom, H.S. (2006):** The Core Analytics of Randomized Experiments for Social Research. MDRC Working Papers on Research Methodology.
- Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung (BMZ) (2008):** Entwicklungspolitische Informations- und Bildungsarbeit (BMZ-Konzept 159). Bonn, Berlin.
- Caspari, A. (2004):** Evaluationen der Nachhaltigkeit von Entwicklungszusammenarbeit. Zur Notwendigkeit angemessener Konzepte und Methoden. Wiesbaden.
- Caspari, A. (2006):** Partizipative Evaluationsmethoden – Zur Entmystifizierung eines Begriffs in der Entwicklungszusammenarbeit. In: Flick, U. (Hg.): Qualitative Evaluationsforschung. Reinbek. S. 365–384.
- Caspari, A. (2008):** (Rigorous) Impact Evaluations – Eine nicht nur für die Entwicklungszusammenarbeit relevante internationale Diskussion. In: Zeitschrift für Evaluation, Jg. 7 H. 1, S. 137–141.
- Caspari, A. (2011):** Meta-Evaluation & Evaluationssynthese 2011 – Kurzbericht. Querschnittsauswertung dezentraler Evaluationen der Human Capacity Development Programme des Jahres 2010. Bonn.
- Caspari, A./Barbu, R. (2008):** Wirkungsevaluierungen/(Rigorous) Impact Evaluations. Zum aktuellen Stand der Diskussion und dessen Relevanz für deutsche EZ-Evaluierungen (BMZ Evaluation Working Papers), Bonn.
- Centre for Global Development (CGD) (2006):** When Will We Ever Learn? Improving Lives through Impact Evaluation. Washington.
- Chung, K. (2000):** Qualitative data collection techniques. In: Grosh, M./Glewwe, P. (Hg.): Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 Years of Living Standards Measurement Study (Bd. 2). Washington, D.C., S. 337–363.
- DeGEval – Gesellschaft für Evaluation (2004):** Empfehlungen für die Aus- und Weiterbildung in der Evaluation. Anforderungsprofile an Evaluatoreninnen und Evaluatoren. Alfter. Online unter: <http://www.alt.degeval.de/calimero/tools/proxy.php?id=24057> [30.04.12]).
- Ezemenari, K./Rudqvist, A./Subbarao, K. (1999):** Impact Evaluation: A Note on Concepts and Methods. Washington, D.C.

Guijt, I. (2000): Methodological issues in participatory monitoring and evaluation. In: Estrella, M. (Hg.): Learning from Change. Issues and Experiences in Participatory Monitoring and Evaluation. London. S. 201–228.

Jungk, S. (2010): Mühen und Chancen der Evaluation. Zur Schwierigkeit der Messung von Wirkung in der Bildungsarbeit des Globalen Lernens. In: Massing, A./Rosen, A./Struck, G. (Hg.): Wirkt so. Handreichung zur Wirkungsorientierung und Antragstellung in der entwicklungspolitischen Inlandsarbeit. Berlin. S. 18–23.

Kassam, Y. (1998): Combining participatory and survey methodologies in evaluation: the case of a rural development project in Bangladesh. In: Jackson, E.T./Kassam, Y. (Hg.): Knowledge Shared: Participatory Evaluation in Development Cooperation. West Hartford, Connecticut. S. 108–121.

Leeuw, F./Vaessen, J. (2009): Impact Evaluations and Development. NONIE Guidance on Impact Evaluation (Draft Version for Discussion at the Cairo conference March-April, 2009). (Online unter: http://www.worldbank.org/ieg/nonie/docs/Guidance_IE.pdf [20.03.12]).

Massing, A. (2010): Vorwort des Berliner Entwicklungspolitischen Ratschlags. In: Massing, A./Rosen, A./Struck, G. (Hg.): Wirkt so. Handreichung zur Wirkungsorientierung und Antragstellung in der entwicklungspolitischen Inlandsarbeit. Berlin. S. 6.

Massing, A./Rosen, A./Struck, G. (Hg.) (2010): Wirkt so. Handreichung zur Wirkungsorientierung und Antragstellung in der entwicklungspolitischen Inlandsarbeit. Berlin.

OECD/DAC (2002): Glossary of Key Terms in Evaluation and Results Based Management (Evaluation and Aid Effectiveness, No. 6). Paris.

Prowse, M. (2007): Aid effectiveness: the role of qualitative research in impact evaluation. Background Note December. ODI.

Ravallion, M. (2005): Evaluating Anti-Poverty Programs. Washington, D.C.

Rosen, A. (2010): Keine Wirkung, kein Problem? In: Massing, A./Rosen, A./Struck, G. (Hg.): Wirkt so. Handreichung zur Wirkungsorientierung und Antragstellung in der entwicklungspolitischen Inlandsarbeit. Berlin. S. 10–13.

Rossi, P.H./Freeman, H.E./Hofmann, G. (1988): Programm Evaluation. Einführung in die Methoden angewandter Sozialforschung. Stuttgart.

Scheunpflug, A./Bergmüller, C./Schröck, N. (2010): Evaluation entwicklungspolitischer Bildungsarbeit. Eine Handreichung. Münster.

Schnell, R./Hill, P./Esser, E. (1999): Methoden der empirischen Sozialforschung. München. Wien.

Seitz, K. (2010): Aus Erfahrung lernen. Wirkungsanalyse der entwicklungspolitischen Inlandsarbeit bei „Brot für die Welt“. In: Massing, A./Rosen, A./Struck, G. (Hg.): Wirkt so. Handreichung zur Wirkungsorientierung und Antragstellung in der entwicklungspolitischen Inlandsarbeit. Berlin. S. 14–17.

White, H. (2006): Impact Evaluation. The Experience of the Independent Evaluation Group of the World Bank. Washington, D.C.

Prof. Dr. Alexandra Caspari

geboren 1969, ist Diplom-Soziologin und derzeit Professorin für Evaluationsforschung, Methoden der empirischen Sozialforschung und Statistik an der Fachhochschule Frankfurt am Main. Ihr Arbeits- und Forschungsschwerpunkt liegt seit nunmehr 15 Jahren im Bereich ‚Evaluation‘, insbesondere ‚Evaluationsmethoden‘. Sie verfügt über einschlägige Erfahrungen in der Durchführung von Evaluationen und ist seit vielen Jahren für diverse Organisationen als Evaluations- und (Methoden-) Beraterin tätig. Seit 2007 beschäftigt sie sich intensiv mit Wirkungsevaluationen. Im Rahmen einer vom BMZ beauftragten Studie hat sie die internationale Methodendiskussion zum Thema ‚Rigorous Impact Evaluation‘ aufbereitet und seither auf zahlreichen Tagungen vorgetragen. Sie ist Mitherausgeberin der ‚Zeitschrift für Evaluation‘ und war von 2003 bis 2007 Vorstandsmitglied der DeGEval – Gesellschaft für Evaluation.

Die Zeitschrift für Evaluation (ZfEv) will eine Plattform sein für:

- den fachlichen Dialog zwischen Wissenschaft und Praxis
- die interdisziplinäre Bündelung sektoralen Fachwissens
- den internationalen Austausch im deutschsprachigen Raum
- die Vermittlung und Diskussion neuer Entwicklungen
- die Verbreitung von Standards in der Evaluation
- Hinweise auf Ausschreibungen, Literatur, Veranstaltungen etc.

Die ZfEv publiziert sowohl wissenschaftliche Beiträge als auch praxisorientierte Erfahrungsberichte zu(r):

- theoretischen Grundlagen von Evaluationsstudien
- methodischen Konzeptionen für Evaluationen
- Anwendungsbeispielen zur Planung und Durchführung von Evaluationen
- Präsentation von Evaluationsinstitutionen und -angeboten
- Möglichkeiten und Grenzen der Nutzung von Evaluationsergebnissen
- aktuellen Informationen über verschiedene Themenfelder
- Entwicklungen zur Qualitätssicherung von Evaluationen



Die ZfEv erscheint zwei Mal im Jahr und kann direkt vom Verlag bezogen werden. Das Jahresabonnement kostet 32,- €, das Einzelheft kostet 19,- €, jeweils zzgl. Versandkosten. Freier Bezug für DeGEval-Mitglieder, reduzierter Betrag für SEVAL-Mitglieder.



WAXMANN
Münster · New York · München · Berlin