

„Dialogtag Wirkungen“

Bericht zur Tagung am 25. November 2011
in der Saarländischen Landesvertretung in Berlin

Vera Siber,¹ Sonja Wolf²

Seit einigen Jahren beschäftigen sich staatliche wie nicht staatliche Entwicklungsorganisationen verstärkt mit der Frage, welche gewünschten nachhaltigen Veränderungen sie bewirken, und fokussieren sich stark auf die mittel- und langfristigen Wirkungen ihrer Maßnahmen. Um diese zu messen, daraus zu lernen, und sie der Öffentlichkeit zu präsentieren, haben staatliche und nicht staatliche Entwicklungsorganisationen sehr verschiedene Vorgehensweisen zur Wirkungsmessung entwickelt. Wie man Wirkungen angemessen prüft und bewertet, wird zwischen den Entwicklungsorganisationen aktuell heftig diskutiert.

Um in gegenseitigen Dialog zu treten, haben das *Bundesministerium für wirtschaftliche Entwicklung und Zusammenarbeit (BMZ)*, das *Centrum für Evaluation (CEval)* der Universität des Saarlandes und der *Verband Entwicklungspolitik deutscher Nichtregierungsorganisationen e.V. (VENRO)* am 25. November 2011 in der Saarländischen Landesvertretung Berlin eine Tagung mit dem Titel „Dialogtag Wirkungen“ veranstaltet. Die Idee dazu entstand im Rahmen des fachlichen Austauschs zwischen VENRO, BMZ und CEval.

Die Konferenz bot eine Reihe von Grundsatzstatements, Vorträgen zu Wirkungsmonitoring und -evaluation und ein abschließendes Panel, auf dem Vertreter/-innen der staatlichen und nicht staatlichen Entwicklungszusammenarbeit (EZ) sowie aus der Wissenschaft über verschiedene Ansätze von Wirkungsmonitoring und -evaluation und deren Bedeutung für die Steuerung von Projekten und Programmen in der EZ diskutierten. Ziel der Veranstaltung war der offene Austausch über unterschiedliche Sichtweisen und deren Begründungen, um die Steuerung mithilfe von Wirkungsinformationen zu verbessern. Folgender Beitrag beschreibt die diskutierten Inhalte, fasst wesentliche Ergebnisse des Dialogs zusammen und gibt einen Ausblick, wie der Austausch zukünftig fortgeführt werden könnte. Wer

-
- 1 Studierende des Studiengangs Master of Evaluation, Universität des Saarlandes (UdS), Saarbrücken; Mitglied der Arbeitsgruppe (AG) Wirkungsbeobachtung des Verbands für Entwicklungspolitik Deutscher Nichtregierungsorganisationen (VENRO) von 2008-2010, Bonn; Referentin für Humanitäre Hilfe, AWO International, e.V., Berlin
 - 2 Koordinatorin des Arbeitsbereichs Entwicklungszusammenarbeit am Centrum für Evaluation (CEval) und Dozentin im Studiengang Master of Evaluation, Universität des Saarlandes (UdS), Saarbrücken

sich für die Veranstaltungsbeiträge im Detail interessiert, dem stehen diese auf der eigens für den Dialogtag Wirkungen eingerichteten Webseite zur Verfügung.³

1. Grundsatzstatements

1.1 Nachhaltige Wirksamkeit als Maßstab in der Entwicklungszusammenarbeit

Angesichts der Millenniumserklärung und der Deklarationen von Accra, Paris bis hin nach Busan,⁴ kristallisiert sich die Fragestellung nach den erzielten Wirkungen zur Kernfrage aller beteiligten Akteure/Akteurinnen in der EZ sowohl auf staatlicher als auch auf zivilgesellschaftlicher Seite heraus.

Mit diesem Rück- und Ausblick beschäftigte sich der Eröffnungsvortrag von *Kordula Mehlhart* (BMZ). Sie führte aus, dass alle beteiligten Akteure/Akteurinnen im In- und Ausland an einer nachhaltigen Wirkungsbeobachtung interessiert seien. Dabei hätten die einen mehr die Rechenschaftslegung und die anderen mehr den Lerneffekt der Wirkungsbeobachtung im Auge.

Zu den grundsätzlichen Fragen, in welcher Form und mit welchen Instrumenten und Methoden Wirkungsbeobachtung realisiert werden kann, informierte der Beitrag über das im Aufbau befindliche unabhängige Evaluationsinstitut und der gerade neu geschaffenen Servicestelle für bürgerschaftliches und kommunales Engagement. Mit der Servicestelle plant das BMZ, sich mit programmübergreifenden Fragen zu Monitoring und Evaluation zu befassen und entsprechende Qualifizierungsmaßnahmen anzubieten. Das Evaluationsinstitut soll mit seinem externen Blick das gesamte Spektrum der vom BMZ geförderten EZ unabhängig evaluieren können, wobei die Referentin davon ausgeht, dass die Zivilgesellschaftlichen Organisationen (ZGO) von den Evaluationen und der Methodenkompetenz profitieren können. Lediglich hinsichtlich der Auswahl der zu evaluierenden Maßnahmen würde das BMZ steuernd eingreifen. Evaluationen sollten eine glaubwürdige, verlässliche und ungeschönte Bewertungsanalyse als Basis für die Initiierung von Lernprozessen bieten. Dabei spiele die Orientierung an international gültigen Standards, Kriterien und Begriffsdefinitionen für die Konzeption und Durchführung von Evaluationen eine wesentliche Rolle und mit den Erkenntnissen aus der Wissenschaft könne eine Evaluation hilfreiche und fundierte Unterstützung erhalten. Weiterhin erwähnte die Referentin das vom BMZ seit 2004 geförderte Vorhaben „NGO [Non Governmental Organisation] Impact on Development, Empowerment and Actions“ (NGO-IDEAs) und die Notwendigkeit, dessen Auswirkungen auf Evaluationen zu diskutieren.

Neben der eher institutionellen und theoretischen Fragestellung seien das BMZ und VENRO an einer konkreten Umsetzung von Wirkungsbeobachtung und Eva-

3 Siehe: <http://www.ceval-consult.com/dialogtag-wirkungen.de/vortraege.html> [Februar 2012].

4 Zeitgleich zum Dialogtag wurde in Busan von 29.11. bis 01.12.2011 über die Wirksamkeitsagenda Bilanz gezogen (vgl. BMZ 2011).

luation im Kontext der Förderung zivilgesellschaftlicher Programme interessiert. Als konkretes Ergebnis läge inzwischen die überarbeitete Handreichung zur Gestaltung unabhängiger Evaluationen als Anlage zu den Richtlinien für die Förderung entwicklungswichtiger Vorhaben privater deutscher Träger vor, würde bereits angewendet und das BMZ fördere mittlerweile auch die Durchführung der entsprechenden Evaluationen.

Die Referentin erwähnte, dass dem Dialogtag eine spannende Diskussion über die Ziele und Anforderungen von Wirkungsbeobachtung im zivilgesellschaftlichen Programmkontext zwischen CEval und VENRO vorausgegangen sei, die durch das VENRO Positionspapier „Qualität statt Beweis“ (VENRO 2010)⁵ ausgelöst wurde.

1.2 Veränderung wahrnehmen und gestalten – die Wirkungsorientierung der NRO

Heike Spielmanns (VENRO) wies zu Beginn ihres Statements darauf hin, dass zum Thema Wirkungen Mitarbeiter/-innen von Nichtregierungsorganisationen (NRO) mit der Aussage „bei den NRO ist ja immer alles anders“ konfrontiert seien. Im Bewusstsein, dass die Veranstaltung ein Ausdruck der Gemeinsamkeiten sei, wies sie auf folgende vier Besonderheiten der NRO-Arbeit hin: 1. NRO arbeiten zu vernachlässigten Themen und mit vernachlässigten Zielgruppen. 2. NRO arbeiten an der Basis und auf der Mikro- und Mesoebene. 3. NRO leisten einen Beitrag zur Demokratieentwicklung. 4. NRO in Nord und Süd arbeiten partnerschaftlich zusammen.

Anschließend bemerkte Heike Spielmanns, dass Wirkungsbeobachtung eine unerlässliche Aufgabe und ein zentrales Anliegen der Arbeit von VENRO und den NRO sei. Dabei würde neben punktuellen und vertieften Prüfungen durch Evaluation, die kontinuierliche Wirkungsbeobachtung in Form des Monitoring verstanden. Wichtig war der Hinweis, dass VENRO die Begriffe Wirkungsbeobachtung und Evaluation nicht synonym verwende und hierin wohl eine Ursache der Missverständnisse der Debatte zwischen CEval und VENRO läge. Folgende Ziele sollen durch Wirkungsbeobachtung im Rahmen des Monitorings von VENRO-Mitgliedsorganisationen erreicht werden:

- die Handlungsfähigkeit der Partnerorganisationen und Zielgruppen stärken und damit zum Empowerment der Akteure/Akteurinnen beitragen;
- aus Erfahrungen lernen und die Erkenntnisse für die weitere Steuerung von Projekten und Programmen nutzen;
- Ergebnisse nutzen, um Rechenschaft gegenüber Gebern/Geberinnen und Öffentlichkeit abzulegen.

Dabei kämen je nach Eignung, in Abhängigkeit der Organisation und Projektmaßnahme, unterschiedliche Konzepte und Methoden zur Anwendung.

⁵ Zu dieser Diskussion siehe auch die Beiträge von Meyer et al. 2011, Mack et al. 2011 und Krapp/Vahlhaus/Gajo 2011 in den beiden vorangegangenen Ausgaben der ZfEv (Debatte).

Im dritten Teil ging die Referentin auf Kritikpunkte des CEval an dem VENRO Papier „Qualität statt Beweis“ im Rahmen der kontrovers geführten Debatte⁶ zwischen CEval und VENRO ein. Auch wenn es bei der von VENRO angestoßenen Debatte teilweise um „Missverständnisse“ ginge, so die Referentin, bestünden auch echte „Differenzen“, die eine weitere Auseinandersetzung mit den jeweiligen Positionen notwendig machen. Als umstrittene Aspekte nannte Heike Spielmanns:

1. Für VENRO erfüllt Empowerment eine wesentliche Funktion der Wirkungsbeobachtung. Die Menschen sollen sich ermutigt und befähigt fühlen, ihre Angelegenheiten selbst in die Hand zu nehmen. Von daher sei es sinnvoll, sie auch am Planungs-, Monitoring- und Evaluationsprozess zu beteiligen, was eines der wirksamsten Instrumente des Empowerments darstelle.
2. Mit zunehmendem Legitimationsdruck innerhalb der letzten Jahre sei die EZ mehr denn je gefordert, die Wirksamkeit ihrer Arbeit nachzuweisen. Dessen ungeachtet, solle der Nachweis von Wirkungen im Rahmen der Legitimationsdebatte nicht überbetont werden. VENRO sieht eine Gefahr für eine vertrauensvolle und offene Zusammenarbeit mit allen Beteiligten, wenn die Wirkungsbeobachtung in erster Linie zur Rechenschaftslegung genutzt und dies von den Beteiligten als Kontrolle erfahren würde. Heike Spielmanns bemerkt zum Abschluss ihrer Ausführungen zum Legitimationsaspekt, dass die NRO mit etwas mehr Mut der Öffentlichkeit und den Spenderinnen und Spendern vermitteln könnten, dass der Erfolg von Entwicklung nur zum Teil von den Entwicklungsmaßnahmen abhängt.
3. Muss die Wirkungsbeobachtung der NRO „wissenschaftlichen Standards“ genügen? Die VENRO AG Wirkungsbeobachtung tauscht sich über Methoden-, Design- und andere evaluationsrelevante Fragen aus, die in unterschiedlichen Kontexten praktikabel und nützlich sein können. Heike Spielmanns erläuterte, dass es allerdings im Alltagsgeschäft der NRO-Akteure/Akteurinnen bei der Wirkungsbeobachtung weniger um die Durchführung externer Evaluationen ginge als vielmehr um das begleitende Projektmonitoring unter der Einbindung der Zielgruppen. Dabei sei es, unter Berücksichtigung der allseits begrenzten Ressourcen, in Verbindung mit den gestiegenen administrativen Projektanforderungen, besonders wichtig, die originären Ziele der Arbeit der NRO nicht aus den Augen zu verlieren und methodische Ansprüche und den Aufwand an die jeweils vorhandenen finanziellen und personellen Gegebenheiten einer NRO und ihrer Partnerorganisationen anzupassen.

1.3 Ziele der Veranstaltung

Prof. Dr. Reinhard Stockmann (CEval) begann seinen Kurzvortrag, indem er einen Kontrast aufzeigte: Zum einen nehmen deutsche wie internationale Organisationen in der EZ eine Ausrichtung ihrer Vorhaben auf Wirkungen hin sehr ernst; dies belegen nicht nur die Ergebnisse aus der Systemprüfung (vgl. Borrmann/Stockmann 2009), sondern auch die Gründung einer Reihe von internationalen Institutionen wie

6 Siehe Fußnote 5.

z.B. Network of Networks for Impact Evaluation (NONIE) und International Initiative for Impact Evaluations (3IE). Zum anderen jedoch gibt es nach wie vor wenig exakte Kenntnisse darüber, ob und wie die Lebensverhältnisse der Menschen in den Partnerländern nachhaltig durch entwicklungspolitische Maßnahmen verbessert worden sind. Um medienwirksamen Publikationen wie „Tödliche Hilfe“ (Erler 1985) oder „Dead Aid“ (Moyo 2009) zukünftig etwas entgegensetzen zu können, müssen nach Stockmann die Instrumente Monitoring und Evaluation mehrfach genutzt werden; einerseits um die Projekte und Programme wirksamer zu gestalten und damit ein Lernen zu ermöglichen und andererseits um deren Wirksamkeit zu überprüfen und damit auch zu legitimieren. Um die sogenannte „Evaluation Gap“ zu schließen, d.h. dem quantitativen wie qualitativen Mangel an Wirkungsevaluationen zu begegnen, fordert Stockmann, dass in Evaluationen nicht nur eine dem jeweiligen Kontext angepasste Methodik, sondern auch eine Programmtheorie Anwendung finden. Letztere stellt die Basis für eine kausale Ursache-Wirkungsüberprüfung dar.

Nach diesem Statement wurden die Ziele und der Aufbau des Dialogtags vorgestellt. Nach Stockmann würde es vor allem darum gehen, die Kluft zwischen „Randomistas“ und „Realistas“ zu reduzieren, denen unterschiedliche Wissenschaftsverständnisse zugrunde lägen. Daher sollte diese Veranstaltung genutzt werden, sich offen über unterschiedliche Sichtweisen und ihre Begründungen auszutauschen. Er betonte dabei, dass alle methodischen Verfahren Vor- und Nachteile mit sich bringen und auch „Randomized Control Trials“ (RCT) keineswegs einen Königsweg in der Wirkungsmessung darstellen.

2. Vorträge zum Thema „Wirkungsmonitoring“

2.1 NGO-IDEAs: Monitoring der eigenen Wirksamkeit durch Zielgruppen und NRO

Bernward Causemann (NGO-IDEAs) beschrieb in seinem Vortrag das Konzept von NGO-IDEAs, die Instrumente der Impact Toolbox sowie Herausforderungen und Nutzen des Ansatzes für Zielgruppen und NRO. Zu Beginn zeigte er zwei Dilemmata auf: Zum einen verliefen Veränderungen nicht kontinuierlich und damit seien für die Wirkungsmessung einmalige Querschnittsanalysen nicht ausreichend. Eine kontinuierliche Begleitung durch externe Expertinnen und Experten sei jedoch zu teuer und nur wenige Organisationen seien in der Lage, Veränderungen und ihre Ursachen eigens systematisch zu dokumentieren. Zum anderen ist es Causemann zufolge sehr schwierig, verlässliche und gleichzeitig repräsentative Daten z.B. über standardisierte Befragungen zu erheben. Daher braucht es laut Causemann eine Reflexion durch die Beteiligten. Mit NGO-IDEAs sollten Instrumente für eine wirkungsorientierte Steuerung entwickelt werden. Im Vordergrund von NGO-IDEAs stehen Bewusstseinsbildung, Selbstmonitoring und Lernen an der Basis als Grundlage für wirkungsorientiertes Handeln. Mit diesem Ansatz sollen NRO und ihre

Zielgruppen in die Lage versetzt werden, Wirkungen selbst zu beobachten. Bislang wurde NGO-IDEAs in sechs verschiedenen Sektoren von 14 deutschen NRO und 40 NRO aus Südasien, Ostafrika und den Philippinen angewandt und damit von mehreren hundert Mitarbeitern und Mitarbeiterinnen von NRO und ca. 860 Gruppen genutzt. Causemann stellte fünf NGO-IDEAs Produkte vor, wozu auch die Impact Toolbox gehört. Dabei wird betont, dass NGO-IDEAs für alle Methoden offen sei. Im Rahmen von PIAR (Participatory Impact Assessment and Reflection), einem der vier Instrumente der Impact Toolbox, findet laut Causemann auch eine Ursache-Wirkungsanalyse statt, indem Gruppendaten von NRO zusammengeführt und hinsichtlich der Projektwirkungen auf die Gruppen und ihre Mitglieder analysiert werden. Der Vortrag zeigte eine Reihe von Praxisbeispielen, bei denen Instrumente der Impact Toolbox angewandt wurden. Im Vortrag konnte der Prozess einer Ursachen-Wirkungszuschreibung jedoch nicht praktisch veranschaulicht werden. Als Herausforderungen des Ansatzes wurden beispielsweise genannt, dass die Einführung der Instrumente und damit auch die Dokumentation und Berichterstattung sehr viel Zeit und Geduld erfordern, dass oftmals nicht ausreichende Partizipations- und Moderationsfähigkeiten vorhanden sind und die Verarbeitung und Analyse von quantitativen Daten sehr schwierig ist. Als zentraler Nutzen von NGO-IDEAs für die Zielgruppen wurden Lernen, Bewusstseinsbildung und Empowerment genannt. Den NRO hingegen nütze NGO-IDEAs dahingehend, dass sie sich stärker an den Bedürfnissen der Zielgruppen orientieren und damit auch zielgerichteter arbeiten können. Außerdem verbessern sich, laut Causemann, durch den NGO-IDEAs-Ansatz die Planung, das Monitoring und die Berichterstattung. Als Desiderat wurde im Vortrag letztlich aufgezeigt, dass sich mit diesem Ansatz die Breitenwirksamkeit von Projekten nicht systematisch erfassen lässt. Außerdem gestaltet sich die Ausbreitung der Methoden noch als schwierig. Weiterhin wurden als derzeitige Herausforderungen beschrieben, ein angemessenes Niveau an Partizipation und ausreichend Qualität in der Analyse sicherzustellen. Darüber hinaus wird bei NGO-IDEAs bislang weder mit Evaluationsforschern und -forscherinnen kooperiert, noch wird das Konzept bei staatlichen Projekten oder im Rahmen von externen Evaluationen genutzt.

2.2 Monitoring-System für das Programm „Wettbewerbsfähigkeit und Umwelt“ Costa Rica

Dr. Stefanie Krapp (GIZ)⁷ und *Prof. Dr. Reinhard Stockmann* (CEval) leiteten ihren Vortrag zum Aufbau eines wirkungsorientierten Monitoring- und Evaluationssystems (M&E-System) ein, indem sie das Programm zur Wettbewerbsfähigkeit und Umwelt in Costa Rica vorstellten. Das auf sieben Jahre angelegte GIZ-Vorhaben besteht aus vier Komponenten (1. Kommunikation und Dialog, 2. Strategie, Pläne und rechtlicher Rahmen, 3. Abfallwirtschaft auf kommunaler Ebene und 4. Wettbewerbsfähigkeit und umweltverträgliches Verhalten in der Industrie) und hat zum

7 Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH.

Ziel, die Abfallwirtschaft in Kommunen und Industrie in ausgewählten Regionen Costa Ricas zu verbessern. Um die internen wie externen Wirkungen des Programms zu messen, wurde für das Programmmanagement ein integratives M&E-System entwickelt, das auf Daten aus den jeweiligen Monitoringsystemen der Gemeinden aufbaut. Die kontinuierliche Datensammlung im Rahmen des Monitorings erfolgt zu Indikatoren auf allen Wirkungsebenen in acht Pilotgemeinden des Programms (Zielgruppe) und in drei weiteren Gemeinden, die nicht durch das Programm gefördert werden (Vergleichsgruppe). Zur Erfassung der Abfallsituation in den Gemeinden kommen eine telefonische Befragung der Haushalte und ein Randomwalk mit Fotodokumentation zum Einsatz. Als Evaluierungsdesign für das M&E-System wurde ein Panel gewählt. Mit dem Panel als Längsschnittuntersuchung werden dieselben Variablen mit derselben Operationalisierung bei denselben Personen zu fünf verschiedenen Zeitpunkten (2006 als Baseline, 2007, 2008, 2010 und 2011) gemessen. Realisiert wurde diese qualitative Befragung von Hauptakteuren/-akteurinnen, Stakeholdern sowie Expertinnen und Experten zum Thema jeweils mittels leitfadengestützter Face-to-Face- und Gruppeninterviews. Im Verlauf des Vortrags wurden die Datenauswertung sowie die tabellarische und graphische Aufbereitung der Ergebnisse zu internen und externen Wirkungen im Zeitverlauf dargestellt und anhand ausgewählter Beispiele verdeutlicht. Als Lessons Learned wurde letztlich festgehalten, dass GIZ-Wirkungsketten zur Entwicklung von Indikatoren auf allen Wirkungsebenen genutzt werden können. Weiterhin empfahlen die Vortragenden ein für Anpassungen flexibles M&E-System, ein im Programmmanagement institutionalisiertes Monitoring, die Erhebung von Baselinedaten und eine Kombination aus internem Monitoring und externen Evaluierungen. Um eine systemische Betrachtung zu gewährleisten, plädierten sie für ein Längsschnittdesign auf Basis eines Wirkungsmodells, das interne und externe Wirkungsfelder berücksichtigt.

3. Vorträge zum Thema „Wirkungsevaluation“

3.1 Evaluierung in der nicht staatlichen EZ – ein Beispiel

Nach der Mittagspause präsentierte *Jens Koy* (Evangelischer Entwicklungsdienst, EED/VENRO) eine extern in Auftrag gegebene Wirkungsevaluation eines vom EED geförderten Projekts in Indien. Der EED kann auf eine 20-jährige Evaluationspraxis zurückblicken. Im konkreten Projekt sollte innerhalb von drei Jahren für etwa 60.000 Familien eine nachhaltige Einkommenssicherung geschaffen werden.

Jens Koy erläuterte das Evaluationsdesign, eine ausgewogene triangulierte Methodenkombination mit drei gleichwertigen Modulen: 1. ein Grundmodul mit Dokumentenanalyse, Feldbesuchen, Gesprächen mit den Partner-NRO, Experten/Expertinnen-, Einzelinterviews und Fokusgruppengesprächen, 2. ein partizipatives Wirkungsmodul, der Method for Impact Assessment of Programmes and Projects (MAPP[®]) und 3. ein quantitatives Modul, eine annähernd repräsentative standardi-

sierte Fragebogenerhebung. Die Module wurden von verschiedenen Gutachterinnen und Gutachtern durchgeführt.

MAPP setze, so Koy, acht definierte Instrumente (1. Lebenslinie, 2. Trendanalyse, 3. Cross-Checking, 4. Maßnahmenliste, 5. Einflussmatrix, 6. Entwicklungs- und Wirkungsprofil, 7. Millennium Development Goals (MDG)-Tabelle und 8. partizipative Entwicklungsplanung) in logischer Abfolge ein und könne in Abhängigkeit des Evaluationsgegenstands bereits mit der Minimalvariante von vier Instrumenten (1, 2, 6, 8) und einer gemeinsamen Diskussionsrunde ausreichende Ergebnisse erbringen.⁸ Die standardisierte Fragebogenerhebung, die dritte Säule des bei der Evaluation angewendeten Multi-Methoden-Ansatzes, stützt sich auf ein quasi-experimentelles Forschungsdesign unter Berücksichtigung des Kontrafaktischen,⁹ erklärte Jens Koy. Gleichzeitig wurde erwähnt, dass ein kombiniertes Auswahlverfahren von Ziel- und Vergleichsgruppen angewendet wurde.

Welche Schlussfolgerungen von dieser Projektevaluation mit dem Fokus auf Wirkungen und einem Evaluationsdesign mit Methodenmix gezogen werden können, stellte Jens Koy für jedes Modul einzeln dar, um zu erläutern welchen Mehrwert dieser Multi-Methoden-Ansatz erbracht hatte. Das Grundmodul hätte es ermöglicht, sowohl die Projektaktivitäten als auch die Situation vor Ort in den nationalen und regionalen Kontext einzuordnen. Es konnten auch Wirkungen identifiziert werden, die nicht unbedingt aus der Projektdokumentation hervorgingen. Was die Anwendung des MAPP-Moduls betrifft, sei ein Vergleich der Wirkungen von verschiedenen Programmen möglich geworden. Die externen Einflüsse wie z.B. Missernten konnten erfasst werden. Ebenfalls konnte ein systematischer Überblick der eingetretenen Veränderungen bezogen auf Lebensbedingungen und auf strukturierte Informationen auf der Outcome- und Impact-Ebene gewonnen werden. Das Modul der standardisierten Fragebogenerhebung habe die Ausweitung der Analyse auf Nachbardörfer ermöglicht. Gleichzeitig habe es Anregung zur Überprüfung einzelner Annahmen gegeben, die aus Betriebsblindheit nicht zustande gekommen wären. Koy schlussfolgerte, dass es zwar zusätzliche Informationen durch die Triangulation¹⁰ gegeben, die Fragebogenerhebung aber letztendlich keine neuen Erkenntnisse erbracht habe.

Anhand von Werkstatt-Thesen gab Jens Koy Empfehlungen, wann welcher Evaluationsansatz für den EED sinnvoll sei. Welches Design auch immer angewen-

8 Informationen zu MAPP (ehemaliges System der Wirkungserfassung armutsorientierter Projekte (SWAP), das auf dem Einsatz von Participatory Rural Appraisal (PRA)-Techniken beruht) finden Sie bei der MAPP-Autorin Susanne Neuber beim Deutschen Institut für Entwicklungspolitik (DIE). Siehe http://www.die-gdi.de/CMS-Homepage/openwebcms3.nsf/%28ynDK_contentByKey%29/MSIN-7JJHP8?Open&nav=expand%3AForschung%20und%20Beratung\Projekte%3Bactive%3AForschung%20und%20Beratung\Projekte\MSIN-7JJHP8 [19.02.2012].

9 Kontrafaktisch bedeutet, dass die tatsächlichen Wirkungen mit hypothetischen Veränderungen oder Wirkungen verglichen werden, die auch ohne die Intervention eingetreten wären (vgl. Caspari/Barbu 2008: 6).

10 Die parallele Anwendung qualitativer und quantitativer Datenerhebungs- und -auswertungsverfahren mittels Methoden-Triangulation kann zu bestätigenden oder konvergierenden Erkenntnissen führen, muss demnach nicht unbedingt ‚neue‘ oder ‚überraschende‘ Ergebnisse hervorbringen (vgl. Caspari 2004: 97).

det würde, die Ziele sowie die Adressaten und Adressatinnen einer Evaluation sollten vor ihrer Durchführung mit den Beteiligten geklärt sein, da gerade bei den Zielgruppen schnell Missverständnisse oder falsche Erwartungen entstehen könnten.

3.2 Ein „alltagstaugliches“ Konzept zur robusten Wirkungsmessung

Sonja Wolf (CEval) und Dr. Stefanie Krapp (GIZ) stellten in ihrem Vortrag zum Thema Wirkungsevaluation ein aus ihrer Sicht „alltagstaugliches“ Konzept zur robusten Wirkungsmessung vor. Einleitend machte Stefanie Krapp auf die gestiegene Wirkungsorientierung innerhalb der GIZ aufmerksam, stellte die verschiedenen Evaluierungsinstrumente der GIZ vor und benannte neun Wirkungsevaluierungen, die bislang im Auftrag der Stabsstelle Evaluierung realisiert wurden. Im Anschluss schilderte sie zentrale Herausforderungen, wenn es darum geht, Wirkungen einer Entwicklungsmaßnahme zuzuschreiben. So existieren z.B. in der Regel keine Baselinedaten zur Ausgangssituation von Ziel- und Vergleichsgruppe, eine Identifizierung von Vergleichs- bzw. Kontrollgruppe ist nicht immer möglich, wirkungsorientierte Monitoringsysteme sind nur ansatzweise eingerichtet und die zur Verfügung stehenden Ressourcen für Evaluierungen sind finanziell wie zeitlich begrenzt. Daher hat die GIZ das CEval damit beauftragt, ein methodisches Konzept zu erarbeiten, das eben genannte Prämissen berücksichtigt und dennoch valide und reliable Daten zur Wirksamkeit von Entwicklungsmaßnahmen liefert. Dieser Ansatz (vgl. Reade 2008) soll internationale Standards an Wirkungsevaluierungen einhalten und gleichzeitig der GIZ als Standardinstrument dienen. Dieses „alltagstaugliche“ Konzept zur Wirkungsevaluation beinhaltet nach Sonja Wolf folgende sechs Komponenten: 1. eine hypothesengeleitete Ursache-Wirkungsanalyse auf Grundlage der GIZ-Wirkungskette; 2. ein quasi-experimentelles Evaluationsdesign unter Anwendung der *Double-* bzw. *Single-Difference-Methode*,¹¹ 3. die Triangulation von internationalen wie lokalen Evaluatoren/Evaluatorinnen, von qualitativen und quantitativen Datenerhebungsmethoden sowie verschiedener Ergebnisse und Datenquellen; 4. die Anwendung von deskriptiven und inferenzstatistischen Verfahren, zu denen auch Hypothesentests gehören; 5. die Berücksichtigung der DAC-Kriterien¹² und der DeGEval-Standards (vgl. DeGEval 2008) und 6. eine zweimalige Datenerhebungsphase, wobei die erste Phase einen explorativen Charakter besitzt und u.a. dafür genutzt wird, den Mehrwert einer Wirkungsevaluation zu überprüfen und eine effiziente Durchführung der zweiten – stärker analytischen – Phase vorzubereiten, in der sowohl qualitative als auch quantitative Daten erhoben

11 Bei der *Single-Difference-Methode* werden Wirkungen über einen einfachen Vergleich zu einem Messzeitpunkt zwischen Faktischem (Zielgruppe) und Kontrafaktischem (Vergleichsgruppe) gemessen. Die *Double-Difference-Methode* kombiniert diesen „Mit-ohne-Vergleich“ darüber hinaus mit einer Vorher-Nachher-Messung; d.h. die Differenz zwischen Zielgruppe und Vergleichsgruppe wird zu zwei Messzeitpunkten erhoben.

12 Nähere Informationen zu den fünf DAC-Kriterien Relevanz, Effektivität, Effizient, Wirkungen und Nachhaltigkeit zur Evaluation der Entwicklungszusammenarbeit finden sich auf der Homepage von OECD/DAC unter http://www.oecd.org/document/22/0,3746,en_2649_34435_2086550_1_1_1_1,00.html [19.02.2012].

werden sollen. Die praktische Anwendung dieses Ansatzes wurde anhand eines Berufsbildungsbeispiels auf den Philippinen verdeutlicht. Im Anschluss wurden ausgewählte Ergebnisse der Studie präsentiert, die insbesondere aus methodischer Perspektive interessant erschienen. So konnte z.B. gezeigt werden, dass *Propensity Score Matching* nicht automatisch zu robusteren Ergebnissen führt, selbst wenn dadurch Fehlervariablen kontrolliert werden und qualitative Daten eine Verzerrung der Stichprobe enthüllen und Erklärungen dafür liefern können. Als Lessons Learned aus methodischer Sicht wurde festgehalten, dass die Reliabilität und Validität der Ergebnisse durch Ergebnistriangulation erhöht wird, die *Single-Difference-Methode* in Kombination mit retrospektiven Fragen in der kausalen Wirkungszuschreibung wertvolle Informationen liefern kann, und ein Methodenmix dringend notwendig ist, gerade dann, wenn die Stichprobe relativ klein ist und zusätzliche Informationen über den theoretischen Hintergrund notwendig sind. Als Empfehlungen wurde sich für die Identifizierung von Kontroll- bzw. Vergleichsgruppen zu Programmbeginn, die Formulierung von realistischen Ursache-Wirkungshypothesen und messbaren Indikatoren, die Durchführung von Baselinestudien für Ziel- und Vergleichsgruppen und die Etablierung wirkungsorientierter M&E-Systeme ausgesprochen, welche langfristig Zeitreihenanalysen erlauben sollen. Aus institutioneller Sicht wurde zum Abschluss hervorgehoben, dass die Evaluierbarkeit von Vorhaben zukünftig gesteigert werden kann, wenn die Phase der Programmplanung mit Ex-ante-Evaluierungen verbunden wird und M&E-Kapazitäten auf lokaler wie internationaler Ebene gestärkt werden.

4. Panel: Dialog über Monitoring und Evaluation von Wirkungen der Entwicklungszusammenarbeit

Am späten Nachmittag des Dialogtags fand eine abschließende Paneldiskussion statt, die von *PD Dr. Dirk van den Boom* (Freier Gutachter) moderiert wurde. Teilnehmende auf dem Podium waren in diesem Zusammenhang *Michaela Zintl* (BMZ), *Bernward Causemann* (NGO-IDEAs), *Dorothee Mack* (MISEREOR/VENRO), *Prof. Dr. Reinhard Stockmann* (CEval) und *Prof. Dr. Alexandra Caspari* (FH Frankfurt a.M.).

Trotz unterschiedlicher Herangehensweisen und Standpunkte der staatlichen, nicht staatlichen und wissenschaftlichen Organisationen ließ sich durchaus eine Basis von Gemeinsamkeiten bei der Diskussion über das Monitoring und die Evaluation von Wirkungen erkennen. Unstrittig ist die Einsicht, dass durch Monitoring und Evaluationen Erkenntnisse gewonnen werden sollen, um durch Programmsteuerungsinstrumente die Qualität von EZ-Maßnahmen zu verbessern.

Dabei kann der Fokus auf verschiedene Aspekte gerichtet sein. Die NRO-Vertreter/-innen priorisieren letztendlich die Lernfunktion und die Einbeziehung der Zielgruppen. Einigkeit besteht auch, dass für diesen Erkenntnisprozess ‚praktikable‘, d.h. dem jeweiligen Kontext bestmöglich angepasste und angemessene Instru-

mente gebraucht werden, wobei die Praktikabilität nicht zuletzt durch den Kosten-Nutzen-Aspekt bestimmt wird.

Unterschiedliche Meinungen tauchten eher in Bezug auf die Wertigkeit und Handhabbarkeit quantitativer versus qualitativer Methoden und Instrumente auf. Für einen Moment flammte der beigelegt geglaubte Wissenschaftsstreit zwischen Positivismus und Konstruktivismus auf, wobei die Erfordernis eines ‚alltagstauglichen‘ hermeneutischen Ansatzes geäußert wurde. Der Zeitrahmen war letztendlich am späten Nachmittag zu knapp bemessen, um im Detail die Vor- und Nachteile einzelner Designs und Methoden abzuwägen, was jedoch auch dazu führte, dass (quasi-)experimentelle Evaluationsdesigns fälschlicherweise qualitativen oder partizipativen Datenerhebungsmethoden gegenübergestellt wurden. Rigorose, partizipative, theoriebasierte (*Grounded Theory*) und andere – mittel- und langfristige Wirkungen erforschende – Methoden wurden im Zusammenhang mit *Good Practices* und der Frage nach deren Durchführbarkeit, insbesondere für die heterogene nicht staatliche Organisationslandschaft in Bezug auf deren sehr unterschiedlich vorhandenen personellen und finanziellen Ressourcen, angeführt. Seitens VENRO und einiger NRO wurde festgehalten, dass quasi-experimentelle Designs hilfreich sein können und mehr qualitative Auswertungsinstrumente für die Zusammenarbeit mit den Zielgruppen benötigt werden. Vertreter/-innen der NRO merkten zudem an, dass sie sich zukünftig seitens der Wissenschaft Forschung und Entwicklung zu und von „kleineren, NRO angepassten“ Tools wünschten.

Vergleichs- oder Kontrollgruppen für anspruchsvolle Evaluationsdesigns zu identifizieren, um tatsächlich verlässliche, reliable Vergleichsdaten zu ermitteln, stellt nicht nur eine Herausforderung für die NRO-, sondern auch für die wissenschaftlichen Akteure/Akteurinnen dar. Dabei ist man sich einig, dass Quasi-Experimente nicht zwangsläufig schlechtere Ergebnisse liefern als RCT und in der Praxis wesentlich häufiger anwendbar sind. Herausgefunden zu haben, dass etwas wirkt, heißt noch nicht zu wissen, warum es gewirkt hat.

Letztendlich führte der erstrebte Dialog des Panels einschließlich der Fragen und Beiträge der Teilnehmenden zu wenig neuen Erkenntnissen, die inhaltlich über den Stand der schriftlich geführten Debatte, die durch das Positionspapier von VENRO ausgelöst worden ist, hinausgehen. Allerdings konnte in der mündlichen Debatte deutlich mehr Konsens zwischen den z.T. sehr unterschiedlichen Sichtweisen der staatlichen, nicht staatlichen und wissenschaftlichen Organisationen erzielt werden, als dies in den Papieren bislang zum Ausdruck kam.

Positiv vermerkt wurden ebenso die Motivation und das ausgesprochene Interesse aller Beteiligten, im Sinne des Dialogtags weiter zu kommunizieren. Um die gegenseitige Akzeptanz zu fördern und Lernprozesse zu initiieren, wurde die Öffnung des Dialogkreises um Vertreter/-innen aus Partnerländern und weiteren sich mit diesem Thema beschäftigenden Akteurinnen und Akteuren aus wissenschaftlichen, staatlichen, nicht staatlichen und beratenden Einrichtungen als notwendig erachtet.

Wirkungsbeobachtung – sofern sie im Sinne von Wirkungsmonitoring verstanden wird – ist eben nicht gleich Wirkungsevaluation und von daher werden ggf. unterschiedliche Instrumente für das eine und für das andere benötigt.

5. Fazit: Auf dem Weg von der Debatte zum Dialog

Abschließend kann festgehalten werden, dass etwa 120 Teilnehmende die Gelegenheit wahrnahmen, gemeinsam mit Expertinnen und Experten verschiedene Ansätze zur Erfassung und Bewertung von Wirkungen in der EZ kennenzulernen und zu diskutieren.

Einen nicht unwesentlichen Hemmschuh des Dialogs stellten die zum Teil unterschiedlichen Herangehensweisen sowie insbesondere auch die divergierende Verwendung und Interpretation der Terminologien im Bereich Wirkungsbeobachtung und Evaluation dar. Eine einheitliche Festlegung und Definition der relevanten Begrifflichkeiten könnte in Zukunft die Verständigung zu diesem Thema erleichtern.

Insgesamt war die Veranstaltung stärker von einem Diskurs als einem Dialog geprägt, wenngleich sich alle Beteiligten bemühten, Brücken zwischen der eigenen und anderen Denkweisen zu bauen. Die bereits vor dem Dialogtag begonnene Debatte, die sich an dem VENRO-Positionspapier „Qualität statt Beweis“ entzündet hatte, war inhaltlich vielleicht noch zu wenig ausdiskutiert, als dass bereits eine Basis für einen ‚echten‘ Dialog vorhanden gewesen wäre. Die Veranstalter/-innen räumten am Ende des Tages ein, dass sie sich der – auch zeitlichen – Beschränkung dieser Veranstaltung bewusst seien und nun Möglichkeiten identifiziert werden müssten, um z.B. anhand konkreter Beispiele oder im Rahmen von Workshops und Arbeitsgruppen intensiver in Dialog zu treten.

Die Rückmeldung zur Veranstaltung war überwältigend positiv. Den Ergebnissen der Online-Befragung zufolge, an welcher 58 Teilnehmende im Anschluss an die Tagung partizipierten, schätzten diese die Dialogveranstaltung hinsichtlich der behandelten Themen sowie der Möglichkeiten zum Austausch und zur Vernetzung sehr positiv ein. 78 Prozent der Befragten gaben an, dass die Veranstaltung in Bezug auf unterschiedliche Sichtweisen und ihre Begründungen neue Erkenntnisse gebracht hätte. Die Ziele der Veranstaltung seien klar formuliert und die Relevanz der behandelten Themen verdeutlicht worden. Zu der Frage, ob und wie der Dialog weitergehen könne, wünschen sich fast 90 Prozent der Befragten eine Fortsetzung des Dialogs im Rahmen eines weiteren Treffens, wobei hier die Tendenz in Richtung einer internationalen Veranstaltung geht (50%). 38 Prozent der Befragungsteilnehmer/-innen gaben an, dass weitere Partner/-innen in den Dialog mit einbezogen werden sollten. Neben dieser generellen Bewertung bieten insbesondere die Angaben zu den offenen Fragen nützliche Hinweise und Anregungen für die weitere Ausgestaltung des Dialogs.

6. Literatur

BMZ – Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung: Entwicklungspolitische Informations- und Bildungsarbeit (2011): Mehr Wirkungen erzielen – 4. Hochrangiges Forum zur Wirksamkeit der Entwicklungszusammenarbeit, Busan, 29.11.-01.12.2011. BMZ-Informationsbroschüre 9/2011 (Stand September 2011).

- Borrmann, Axel/Stockmann, Reinhard (2009): Evaluation in der deutschen Entwicklungszusammenarbeit. (Bd. 1 Systemanalysen, Bd. 2 Fallstudien). Studie im Auftrag des Bundesministeriums für Wirtschaftliche Zusammenarbeit und Entwicklung – BMZ. Bd. 8 und 9 der Reihe „Sozialwissenschaftliche Evaluationsforschung“. Münster u.a: Waxmann.
- Caspari, Alexandra (2004): Evaluation der Nachhaltigkeit von Entwicklungszusammenarbeit. Zur Notwendigkeit angemessener Konzepte und Methoden. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Caspari, Alexandra/Barbu, Ragnild (2008): Wirkungsevaluierungen: Zum Stand der internationalen Diskussion und dessen Relevanz für Evaluierungen der deutschen Entwicklungszusammenarbeit. Evaluation Working Papers. Bonn: BMZ. Verfügbar unter: http://www.ceval.de/typo3/fileadmin/user_upload/PDFs/workpaper16.pdf [19.02.2012].
- DeGEval – Gesellschaft für Evaluation e.V. (2008) (Hg.): Standards für Evaluation. Mainz: Eigenverlag (4. unveränderte Auflage). Verfügbar unter: <http://www.alt.degeval.de/calimero/tools/proxy.php?id=19074> [19.02.2012].
- Erlor, Brigitte (1985): Tödliche Hilfe. Bericht von meiner letzten Dienstreise in Sachen Entwicklungshilfe. Freiburg: Dreisam Verlag.
- Krapp, Stefanie/Vahlhaus, Martina/Gajo, Michael (2011): Gedanken zum VENRO Positionspapier zur Wirkungsbeobachtung. In: Zeitschrift für Evaluation, Jg. 10, Heft 2, 2011, S. 320-324.
- Mack, Dorothee/Causemann, Bernward/Seitz, Klaus/Spielmanns, Heike (2011): Die Semantik der Wirkungsbeobachtung oder: Wie konstruieren wir Missverständnisse? Erwiderung auf den Beitrag „Wer beweist Qualität?“ in der ZfEv 1/2011. In: Zeitschrift für Evaluation, Jg. 10, Heft 2, 2011, S. 315-319.
- Meyer, Wolfgang/Rech, Jörg/Silvestrini, Stefan/Stockmann, Reinhard/Wolf, Sonja/Gaus, Hansjörg/Müller, Christoph E./Keller, Daniela (2011): Wer beweist Qualität? Stellungnahme zum VENRO Positionspapier zur Wirkungsbeobachtung. In: Zeitschrift für Evaluation, Jg. 10, Heft 1, 2011, S. 135-140.
- Moyo, Dambisa (2009): Dead Aid: Why Aid is Not Working and how There is Another Way for Africa. London: Penguin Books Ltd.
- Reade, Nicolà (2008): Konzept für alltagstaugliche Wirkungsevaluierungen in Anlehnung an Rigorous Impact Evaluations. CEval-Arbeitspapier Nr. 14. Saarbrücken: Centrum für Evaluation. Verfügbar unter: http://www.ceval.de/typo3/fileadmin/user_upload/PDFs/workpaper14_01.pdf [19.02.2012].
- VENRO – Verband Entwicklungspolitik Deutscher Nichtregierungsorganisationen (2010): Qualität statt Beweis. VENRO Positionspapier 2/2010 zur Wirkungsbeobachtung. Bonn: VENRO. Verfügbar unter: http://www.venro.org/fileadmin/redaktion/dokumente/Dokumente_2010/Publikationen/November_2010/Positionspapier_Wirkungsbeobachtung_v09_WEB.pdf [19.02.2012].

The Future of Evaluation in Modern Societies

Tagung
14. & 15. Juni 2012
Universität des Saarlandes, Deutschland

Evaluation hat sich in der (post-)modernen Gesellschaft als Instrument zur rationalen Politikgestaltung etabliert. Bedingt durch die Pluralisierung nimmt jedoch die Komplexität von Entscheidungsprozessen bei gesellschaftlich relevanten Themen zu und verändert kontinuierlich auch die an Evaluationen gestellten Herausforderungen. Die Tagung in Saarbrücken analysiert im internationalen Vergleich zentrale Trends und wagt einen gemeinsamen Blick in die Zukunft. Im Mittelpunkt steht dabei das prekäre Spannungsfeld zwischen den Standards wissenschaftlicher Exzellenz und den pragmatischen Anforderungen an eine politische Verwertbarkeit der Befunde. Kann Evaluation dazu beitragen, die gesellschaftlichen Probleme der Zukunft sinnvoll und rational zu lösen? Und: was muss dafür heute getan werden?

- ✓ Das Programm, Informationen zur Tagung sowie das Anmeldeformular finden Sie unter: <http://futureofevaluation.ceval.de>
- ✓ Tagungsgebühr: 90 €, Studierende: 45 €
- ✓ Anmeldeschluss: 31. Mai 2012
- ✓ Kontakt: Centrum für Evaluation (CEval) – Universität der Saarlandes
Maria Albrecht
Postfach 15 11 50
D-66041 Saarbrücken
+49 (0) 681 - 302 35 61
m.albrecht@ceval.de

