

# Evaluierung außenpolitischer Maßnahmen in fragilen Kontexten

## **Erfahrungen und Empfehlungen**

Von ANDREA BINDER, PHILIPP ROTMANN

---

**ABGESCHLOSSEN**  
**Juni 2014**

---

**VERÖFFENTLICHT**  
**Dezember 2014**

---

Außenpolitik heißt nicht mehr nur Diplomatie und Beiträge zu multilateralen Initiativen, sondern beinhaltet zunehmend den Einsatz eigener finanzieller Mittel für konkrete Projekte. Allein der Etat des Auswärtigen Amtes für „Frieden und Stabilität“ (inkl. humanitärer Hilfe) wurde in den letzten zehn Jahren verdreifacht, auf über 500 Mio. Euro. Andere Länder wie Großbritannien, die Niederlande, die Schweiz und die USA haben ähnliche Erfahrungen gemacht. Damit wird Projektarbeit ein Kernbestandteil des diplomatischen Handwerks, und dazu gehören auch Monitoring und Evaluierung (M&E) von Projekten als Lern- und Rechenschaftsinstrument. M&E außenpolitischer Maßnahmen in fragilen Kontexten steht erheblichen methodischen Herausforderungen gegenüber, deshalb können klassische Ansätze oft nicht überzeugen. Es gibt jedoch vielfältige innovative Ansätze, Kriterien und Methoden, die es für die eigenen Lern- und Rechenschaftsanforderungen anzupassen gilt.

Diese Studie wurde im Auftrag des Planungsstabs im Auswärtigen Amt durchgeführt. Die Autoren danken Susanna Krüger, Bidjan Nashat und Elias Sagmeister für hilfreiche Anregungen, Claudia Meier für hervorragende Unterstützung beim Lektorat sowie Catherine Craven, Rahel Dette und Hendrik Schopmans für Forschungsunterstützung in verschiedenen Stadien der Arbeit an der Studie. Thomas Bagger, Julian Fricke, Simon Herchen, Sven Schneider und Thomas Zahneisen auf Seiten des Planungsstabs im Auswärtigen Amt gebührt Dank für ihr Vertrauen und die angenehme Zusammenarbeit.

# Abkürzungsverzeichnis

AA	Auswärtiges Amt
AMS	Abteilung für menschliche Sicherheit, EDA (Schweiz)
BMVg	Bundesministerium der Verteidigung
BMZ	Bundesministerium für wirtschaftliche Zusammenarbeit
CHASE	Conflict, Humanitarian Assistance and Security Department, DFID
DEZA	Direktion für Entwicklung und Zusammenarbeit, EDA (Schweiz)
DPKO	Department of Peacekeeping Operations, UN-Sekretariat
EDA	Eidgenössisches Departement für Auswärtige Angelegenheiten (Schweiz)
EU	Europäische Union
FCO	Foreign and Commonwealth Office
ICAI	Independent Commission for Aid Impact, Großbritannien
INCAF	International Network on Conflict and Fragility, OECD
IOB	Inspectie Ontwikkelingssamenwerking en Beleidsevaluatie, Niederlande
KfW	Kreditanstalt für Wiederaufbau
LFA	Logical Framework Approach
M&E	Monitoring und Evaluierung
MBZ	Ministerie van Buitenlandse Zaken, Niederlande
ODA	Official Development Assistance
OECD	Organisation for Economic Cooperation and Development (Organisation für wirtschaftliche Zusammenarbeit und Entwicklung)
OECD-DAC	OECD Development Assistance Committee (Entwicklungshilfesausschuss der OECD)
OFDA	Office of US Foreign Disaster Assistance, USAID
OTI	Office of Transition Initiatives, USAID
PM&E	Planung, Monitoring und Evaluierung
PVK	Parlamentarische Verwaltungskontrolle der Schweiz
QDDR	Quadrennial Diplomacy and Development Review
RBM	Results-Based Management
RM	Bureau of Resource Management, US Department of State
SIGAR	Special Inspector-General for Afghanistan Reconstruction
SMART	specific, measurable, attainable, relevant, timebound/ spezifisch, messbar, erreichbar, relevant, zeitlich beschränkt
ToR	Terms of Reference (Pflichtenheft)
USAID	US Agency for International Development

# Inhaltsverzeichnis

<b>Einleitung</b>	9
Drei zentrale Erkenntnisse	9
Grundbegriffe	11
Vorgehen	14
Überblick über die Studie	15
<b>Großbritannien</b>	16
Zweck und Mandat	17
Akteure	18
Abläufe und Methoden	19
Einbettung in eine professionelle Projektmanagement- und Lernkultur	22
Wesentliche Erkenntnisse	23
<b>Schweiz</b>	25
Zweck und Mandat	26
Akteure	27
Abläufe und Methoden	29
Einbettung in eine professionelle Projektmanagement- und Lernkultur	31
Wesentliche Erkenntnisse	31
<b>USA</b>	33
Zweck und Mandat	33
Akteure	34
Abläufe und Methoden	35
Einbettung in eine professionelle Projektmanagement- und Lernkultur	38
Wesentliche Erkenntnisse	39
<b>Niederlande</b>	41
Zweck und Mandat	41
Akteure	43
Abläufe und Methoden	44
Einbettung in eine professionelle Projektmanagement- und Lernkultur	46
Wesentliche Erkenntnisse	47
<b>Schlussfolgerungen</b>	49
Vier Gründe für bislang oft ernüchternde Erfahrungen	50
Schwerpunktsetzung im Spannungsfeld von Lernen und Rechenschaft	53
Umgang mit Evaluierungskriterien	56
Monitoring-Praxis in Bezug auf politische Kontextanalyse stärken	56
Evaluierungsinstrumente systematisieren	57
Lernkultur	57
Aus- und Fortbildung: M&E-Kompetenz vertiefen und breiter verankern	60
Ressourcenbedarf	60
Chancen und Anforderungen für eine aktiv lernende Organisation	61
<b>Literaturverzeichnis</b>	62

# Zusammenfassung

Moderne Außenpolitik verfolgt ihre Ziele zunehmend mit eigenen finanziellen Mitteln für konkrete Projekte. Die Haushaltsmittel für „Maßnahmen und Leistungen zur Sicherung von Frieden und Stabilität einschließlich humanitärer Hilfsmaßnahmen“ betragen mittlerweile gut dreimal so viel wie noch vor zehn Jahren. Dadurch werden die Gestaltung, Vergabe, Begleitung, Monitoring und Evaluierung (M&E) von Projekten zu wichtigen Bestandteilen des diplomatischen Handwerks. Die vorliegende Studie analysiert internationale Erfahrungen mit M&E von außenpolitischen Maßnahmen in fragilen Kontexten. Sie konzentriert sich auf die Themenfelder Stabilisierung, humanitäre Hilfe, zivile Krisenprävention und demokratische Transformation sowie auf Erfahrungen aus Großbritannien, der Schweiz, den USA und den Niederlanden.

## **Besonderheiten außenpolitischer Maßnahmen in fragilen Kontexten**

Die in den meisten Ländern noch recht junge außenpolitische M&E-Praxis bietet zahlreiche Anknüpfungspunkte für eine kontinuierliche Verbesserung von außenpolitischen Maßnahmen in fragilen Kontexten. Klassische Evaluierungsansätze in der Tradition eines strikt positivistischen Wissenschaftsverständnisses können dieses Verbesserungspotential allerdings nicht realisieren. Dies liegt unter anderem daran, dass die besonderen Herausforderungen außenpolitischer Maßnahmen in fragilen Kontexten methodisch und mit Blick auf angemessene Erfolgskriterien von traditionellen Evaluierungsansätzen nicht hinreichend berücksichtigt werden können. Zu diesen Herausforderungen gehören die Kombination politisch-prozessualer (z.B. Vertrauen in staatliche Institutionen steigern) mit meist untergeordneten sektoralen Zielen (z.B. Schulbildung ausweiten) und die dynamische Veränderlichkeit fragiler Kontexte, mit der die meisten kausalanalytischen Standardansätze von Monitoring und Evaluierung nicht angemessen umgehen können. Diese Grenzen traditioneller Evaluierungsansätze sorgen für verbreitete Ernüchterung mit Evaluierungsergebnissen. In Reaktion auf diese Ernüchterung entstehen international neue, innovative M&E-Ansätze, die auf die Lern- und Rechenschaftsbedürfnisse außenpolitischer Maßnahmen zugeschnitten sind und die dafür angemessenen Kriterien und Methoden selbstbewusst gestalten.

## **Anknüpfungspunkte für eine lernorientierte außenpolitische M&E-Praxis**

Die in der Studie identifizierten Ansatzpunkte für eine kontinuierliche Verbesserung der außenpolitischen Arbeit in fragilen Kontexten sind selektiver Natur. Diese Selektivität spiegelt wider, dass die untersuchten Außenministerien noch experimentieren, wie sich M&E in außenpolitische Abläufe und Organisationstraditionen einbetten

lassen. Zu den vielversprechendsten Beispielen aus der internationalen Praxis gehören die partizipativen Evaluierungsverfahren der Niederlande, die Lern- und Rechenschaftsanforderungen effektiv zu verbinden suchen, die informelle *peer review*-Praxis bei der Programmgestaltung in der Schweiz, innovative Monitoringansätze in den USA sowie die konstruktiv-konfliktoffene Lernkultur und Forschungsorientierung in Großbritannien. Die Vielfalt akzeptierter Ansätze, Kriterien und Methoden macht deutlich, dass kein erheblicher Konformitätszwang zu einem (fremd-)bestimmten Ansatz besteht.

Die verbreiteten Evaluierungskriterien des OECD-Entwicklungshilfeausschusses (OECD-DAC-Kriterien: Relevanz, Effektivität, Effizienz, Wirkung und Nachhaltigkeit) sind nur bedingt für die Evaluierung außenpolitischer Maßnahmen in fragilen Kontexten nutzbar. Daher ist es inzwischen in den hier betrachteten Themenfeldern etablierte internationale Praxis, diese Kriterien anzupassen oder eigene Kriterien zu entwickeln. Auf der methodischen Ebene gewinnen dementsprechend neuere perzeptive oder umfrageorientierte Erhebungsmethoden an Bedeutung.

In allen betrachteten Ländern zeigte sich ein Spannungsverhältnis zwischen den zentralen Zielen von M&E: Lernen und Rechenschaftslegung. Rechenschaftsorientierte Evaluierung ist im Regelfall investigativ, oft konfrontativ gestaltet und führt deshalb zu erheblichen Reibungen. Gleichzeitig sind erfolgreiche Lerninstrumente oft methodisch und prozedural so informell, dass sie nicht zur Rechenschaftslegung taugen – weil beispielsweise keine veröffentlichungsfähigen Ergebnisse entstehen. Effektive Kombinationen aus Lern- und Rechenschaftsorientierung sind auf Grundlage der analysierten Erfahrungen aber durchaus möglich, wenn bewusst Schwerpunkte gesetzt werden.

## **Herausforderungen für die konkrete Ausgestaltung**

Die Systematisierung einer spezifischen M&E-Praxis auf Grundlage der Besonderheiten außenpolitischer Maßnahmen in fragilen Kontexten begegnet mehreren Herausforderungen. So bestehen über den Anpassungsbedarf in Kriterien, Verfahren und Methoden hinaus erhebliche Qualitätsrisiken für Evaluierungen. Diese sind zum Teil in den bestehenden Markt- und Anreizstrukturen für externe Evaluierungsdienstleister begründet, darüber hinaus liegt jedoch ein entscheidender Grund für den Erfolg oder Misserfolg jeder Evaluierung in der Qualität des Pflichtenhefts: In die Formulierung der Evaluierungsfragen, die Anwendung von Erfolgs- oder Evaluierungskriterien und die Steuerung des Evaluierungsprozesses muss erheblicher Aufwand fließen, damit externe Dienstleister zu fairen und nützlichen Ergebnissen kommen können. Deshalb ist die Funktion des Evaluierungsmanagements für eine Systematisierung der außenpolitischen M&E-Praxis zentral. Darüber hinaus profitiert eine gewinnbringende M&E-Praxis von einem professionellen Projektmanagement und von einem systematisch-konstruktiven Umgang mit Fehlern und Konflikten.

All das erfordert Zeit und Geld: Geeignete M&E-Instrumente müssen entwickelt, in Pilotversuchen getestet und behutsam eingeführt werden. Eine gewinnbringende Systematisierung von Monitoring und Evaluierung wird nur dann Aussicht auf Erfolg haben, wenn auch die Möglichkeit zur Investition zusätzlicher Personal- und Sachmittel besteht.

# Einleitung

Moderne Außenpolitik geht zunehmend über die Arbeit mit klassischen diplomatischen Instrumenten und Beiträge zu multilateralen Initiativen hinaus. Sie verfolgt ihre Ziele auch mit dem Einsatz eigener finanzieller Mittel zur Durchführung konkreter Projekte. Im Haushalt des Auswärtigen Amtes (AA) waren 2013 beispielsweise für „Maßnahmen und Leistungen zur Sicherung von Frieden und Stabilität einschließlich humanitärer Hilfsmaßnahmen“ Mittel in Höhe von € 542 Mio. vorgesehen – gut dreimal so viel wie im Haushalt 2003 (€ 175 Mio.).<sup>1</sup> Weitere Mittel werden für Auswärtige Kultur- und Bildungspolitik und anderes ausgegeben. Die Entwicklung hin zu einer Außenpolitik mit eigenen Mitteln bedeutet konkret, dass Projektarbeit – also Gestaltung, Vergabe, Begleitung und Wirkungsüberprüfung von Projekten – zunehmend Bestandteil des diplomatischen Handwerks wird.

Monitoring und Evaluierung (M&E) haben sich international als Teil eines professionellen Instrumentariums etabliert, um den wirksamen und effizienten Einsatz solcher Mittel zu unterstützen. Die vorliegende Studie analysiert internationale Erfahrungen mit Monitoring und Evaluierung von außenpolitischen Maßnahmen. Sie stellt auf dieser Grundlage Handlungsoptionen für die Systematisierung von M&E, die Auswahl geeigneter Methoden und Instrumente sowie mögliche institutionelle Weiterentwicklungen im Auswärtigen Amt zur Diskussion. Dabei konzentriert sie sich – wie vom Auftraggeber gewünscht – auf zentrale Themenfelder mit wachsender Bedeutung für das Auswärtige Amt: Stabilisierung, humanitäre Hilfe, zivile Krisenprävention und demokratische Transformation.<sup>2</sup> Den damit verbundenen Maßnahmen ist gemeinsam, dass sie in fragilen Kontexten stattfinden. Diese zeichnen sich durch ein eingeschränktes staatliches Gewaltmonopol, die unzureichende Bereitstellung sozialer Basisdienstleistungen durch den Staat und einen Mangel staatlicher Legitimität aus.<sup>3</sup>

## Drei zentrale Erkenntnisse

Die Studie gelangt zu drei zentralen Erkenntnissen. Erstens, außenpolitische Maßnahmen in fragilen Kontexten stehen spezifischen Herausforderungen gegenüber, die bei der Gestaltung geeigneter M&E-Instrumente zu berücksichtigen sind.

- 1 Angaben nach Bundesministerium der Finanzen (2003-2012): *Jahresrechnungen*; Bundesministerium der Finanzen (2013): *Bundshaushalt 2013*.
- 2 Mit dem Begriff „demokratische Transformation“ bezeichnet die Studie außenpolitische Maßnahmen zur Unterstützung politischer Übergänge, für die in der Vergangenheit jeweils länderspezifische Sonderbudgets geschaffen wurden (für Mittel- und Osteuropa nach 1989 sowie für Ägypten und Tunesien seit 2011). Nicht betrachtet wird die langfristig fortlaufende Demokratieförderung durch die politischen Stiftungen.
- 3 Auswärtiges Amt, Bundesministerium der Verteidigung und Entwicklung (2012): *Politik der Bundesregierung gegenüber fragilen Staaten*.

Außenpolitische Maßnahmen folgen multiplen Zielsetzungen: Es bestehen sowohl politisch-prozessuale (z.B. Vertrauen in staatliche Institutionen steigern) als auch sektorale Ziele (z.B. Schulbildung ausweiten), wobei die politisch-prozessualen Ziele den sektoralen zumeist übergeordnet sind. Die Erreichung der politisch-prozessualen Ziele ist häufig nicht objektiv bestimmbar, sondern beruht auf Wahrnehmungen der Bevölkerung oder politischer Eliten. Hinzu kommt, dass die positivistische Kausallogik, die den meisten M&E-Standardansätzen zu Grunde liegt, in fragilen Kontexten an ihre Grenzen stößt. Aufgrund der Komplexität und Veränderlichkeit der politischen Lage lassen sich Ursache-Wirkungs-Hypothesen nicht auf einfache Kausalketten „A führt zu B“ reduzieren.<sup>4</sup> Vielmehr müssen die jeweiligen Interventionen und damit auch M&E-Instrumente so gestaltet werden, dass sie wechselnde Ziele sowie die Notwendigkeit zur Anpassung von Zielen von Grund auf berücksichtigen.

Zweitens sind die in der entwicklungspolitischen Evaluierung verbreiteten Kriterien des OECD-DAC (Relevanz, Effektivität, Effizienz, Wirkung und Nachhaltigkeit)<sup>5</sup> nur bedingt für die Evaluierung außenpolitischer Maßnahmen in fragilen Kontexten nutzbar. Sie orientieren sich an einer linearen Kette von Aktivitäten und Investitionen (*input*) über Produkte (*outputs*) und Ergebnisse (*outcomes*) bis zu Wirkung (*impact*) von Maßnahmen. Damit sind sie der positivistischen Annahme verhaftet, dass ein spezifischer *input* kausal zu einer bestimmten Wirkung führt. Solche kausalen Annahmen sind aufgrund der politischen Volatilität fragiler Kontexte nicht tragfähig, insbesondere wenn ein Erklärungsmodell über längere Zeiträume genutzt werden soll. In den hier betrachteten Themenfeldern Stabilisierung, zivile Krisenprävention, demokratische Transformation und humanitäre Hilfe ist die Entwicklung angepasster Evaluierungskriterien deshalb zunehmend etablierte internationale Praxis.<sup>6</sup> Auch auf der Ebene von Erhebungsmethoden sind kausalanalytische Evaluierungsansätze für fragile Kontexte zunehmend überholt, während perzeptive oder umfrageorientierte Ansätze an Bedeutung gewinnen.<sup>7</sup>

Drittens besteht ein Spannungsfeld zwischen den zwei zentralen Zielen von M&E: Lernen und Rechenschaftslegung. Bei der institutionellen Zweckbestimmung von M&E für ein Ministerium wie das Auswärtige Amt sind diese Ziele zwar komplementär, doch in jedem einzelnen Evaluierungsprozess stehen sie in Konkurrenz zueinander: Die Anforderungen an effektives Lernen begrenzen den Rechenschaftswert einer Evaluierung und umgekehrt.

Aus diesen drei Kernergebnissen – den spezifischen Herausforderungen der Evaluierung von Außenpolitik in fragilen Kontexten, der Notwendigkeit angepasster Evaluierungskriterien und dem Spannungsverhältnis zwischen Lernen und Rechenschaft – folgt, dass als Teil einer eigenen Systematik außenpolitischer Evaluierung der Entwicklung eines qualitativ hochwertigen Pflichtenhefts (*terms*

4 Zu den Grenzen positivistischer Kausallogik siehe auch Rudolf und Lohmann (2013): *Außenpolitikevaluation*, 13ff.

5 OECD (1986): *Glossary of Terms*; OECD (1991): *DAC Principles for Evaluation*; OECD (2000): *Glossary of Evaluation*.

6 Beck (2006): *Evaluating humanitarian action*; OECD (2007): *Principles for Good International Engagement*.

7 Blue, Clapp-Wincek und Benner (2009): *Beyond Success Stories*; Meharg (2009): *Measuring What Matters In Peace Operations And Crisis Management*; OECD-DAC (2012): *Evaluating Peacebuilding Activities*; Ramalingam (2013): *Aid on the Edge of Chaos*.



of reference) entscheidende Bedeutung zukommt. Hier müssen Evaluierungsfragen formuliert werden, die den Spezifika außenpolitischer Maßnahmen in fragilen Kontexten Rechnung tragen; es muss der Umgang mit Evaluierungskriterien festgelegt und die Schwerpunktsetzung zwischen Lernen und Rechenschaft bestimmt werden. Nur dann können interne oder externe M&E-Expertinnen und -Experten zielführende Evaluierungen durchführen.

## Grundbegriffe

Nach Jahrzehnten der Praxis in einer Vielzahl von Politikbereichen von der Sozial- und Bildungspolitik bis zur Entwicklungs- und inzwischen auch zur Außenpolitik hat sich ein integriertes Verständnis von Monitoring & Evaluierung (M&E) entwickelt. Häufig kommt inzwischen die Planung als drittes Element hinzu (PM&E), um die Notwendigkeit einer engen Verzahnung dieser drei Funktionen im Rahmen professionellen Projektmanagements zum Ausdruck zu bringen. Das Feld hat eine eigene Fachsprache entwickelt, die teilweise schwer zugänglich ist. Daher definieren wir im Folgenden die für diese Studie zentralen Begriffe.<sup>8</sup>

**Monitoring** ist die laufende Überprüfung, ob eine Maßnahme im Kontext ihrer praktischen (z.B. Baumaßnahme) und politischen Ziele (z.B. Vertrauen in die Regierung) die gewünschten Fortschritte macht. Dazu werden vor Projektbeginn Zwischenziele und -ergebnisse formuliert, deren Einhaltung durch das Monitoring überprüft wird. Ziel ist es zu verstehen, warum bestimmte Zwischenziele nicht erreicht wurden, um die Durchführung der Maßnahme anzupassen oder, falls notwendig, ihr endgültiges Ziel zu justieren.<sup>9</sup> Damit ist die Erwartung an Monitoring eine umfassendere und kritischere als die Praxis der sogenannten „Projektfortschrittskontrollen.“ Es gilt nicht nur, einen Ausgangspunkt (*baseline*) für spätere Erfolgsüberprüfung zu etablieren und geleistete Arbeit zu dokumentieren (*performance monitoring*), sondern vor allem auch den politischen Kontext zu beobachten und fortlaufend zu analysieren (*context monitoring*).

**Evaluierung** wird traditionell als die systematische und objektive Überprüfung einer (hier: außenpolitischen) Maßnahme, ihrer Konzeption, Umsetzung und Ergebnisse anhand von messbaren Kriterien verstanden. Diese Sicht auf Evaluierung steht in der Tradition eines positivistischen Wissenschaftsverständnisses. Der Positivismus geht davon aus, dass die Sozialwissenschaften, ähnlich wie die Naturwissenschaften, beobachtete Phänomene durch kausale Zusammenhänge erklären können.<sup>10</sup> Demgegenüber hat sich in den letzten Jahren aus der Konfrontation mit den Grenzen der objektiven Messbarkeit politisch relevanter Ergebnisse und Wirkungen ein modernes Evaluierungsverständnis herausgebildet. Hier steht der Begriff der Evaluierung für eine systematische, möglichst umfassende Überprüfung von Maßnahmen anhand von perzeptionsbasierten Indikatoren und mehrheitlich

8 Wir orientieren uns dabei an Standarddefinitionen, legen aber besonderes Augenmerk darauf, die Begrifflichkeiten an die Organisationsweise und -kultur des Auswärtigen Amtes anzupassen. Daher sind nicht alle hier angeführten Definitionen oder Begriffe eins zu eins in der Evaluierungsliteratur wiederzufinden.

9 Independent Evaluation Group (2007): *Sourcebook*.

10 Siehe Rudolf und Lohmann (2013): *Außenpolitikevaluation*.

qualitativen Kriterien. Gemeinsames Ziel beider Perspektiven auf Evaluierung ist es, die Relevanz einer Maßnahme und ihrer Ergebnisse hinsichtlich der Zielerreichung und gewünschten Wirkung anhand von klar definierten Kriterien zu beurteilen.<sup>11</sup>

Im außenpolitischen Rahmen kann zwischen **Portfolio-, Programm-, und Projektevaluierung** unterschieden werden. Darüber hinausgehende Politikfeld- und Politikevaluierungen werden zwar teilweise in der Schweiz (ab S. 25) und in den Niederlanden (ab S. 41) praktiziert, werfen aber umfangreiche eigene theoretische, methodische und politische Fragen und Debatten auf, auf die nur am Rande eingegangen werden kann. Sie sind deshalb explizit kein Gegenstand dieser Studie. Hier wird M&E als Werkzeug zur Überprüfung der Zielerreichung von Maßnahmen, nicht aber der normativ-politischen Richtigkeit übergeordneter politischer Entscheidungen betrachtet. Die untenstehende Tabelle verdeutlicht, welche Evaluierungstypen Gegenstand der hier vorgenommenen Untersuchung sind:

<b>Ebene der Evaluierung</b>	<b>Beispielgegenstand</b>	<b>Beispielhafte Evaluierungsfragen</b>
Portfolioevaluierung	Transformationspartnerschaften	Haben die Transformationspartnerschaften dazu beigetragen, in den Partnerländern die Transformation hin zu Stabilität, Demokratie und Entwicklung aktiv zu unterstützen?
Programmevaluierung	Bildungsprogramme in Tunesien	Haben die in Tunesien vom AA finanzierten Bildungsprogramme dazu beigetragen, dass das Vertrauen der Zielgruppen in die demokratische Transformation des tunesischen Staates gestärkt wurde?  Haben die Bildungsprogramme zu einem erhöhten Bildungsstandard bei den definierten Zielgruppen beigetragen?
Projektevaluierung	Sprachkurse zur beruflichen Weiterqualifizierung	Hat das vom AA finanzierte Projekt dazu beigetragen, das Vertrauen in die Fähigkeit der Institutionen zu erhöhen, den Bedürfnissen der Bürger nachzukommen?  Hat das Projekt die Sprachkompetenzen der Schülerinnen und Schüler so verbessert, dass sie auf dem Arbeitsmarkt erhöhte Erfolgchancen haben?

11 Independent Evaluation Group (2007): *Sourcebook*.

Darüber hinaus lassen sich **formative** und **summative Evaluierungen** unterscheiden. Erstere begleiten die Umsetzung einer Maßnahme von Beginn an. Sie werden häufig zur Evaluierung von Pilotprojekten oder -programmen verwendet. Im humanitären Bereich hat sich eine besondere Form von formativen Evaluierungen entwickelt, sogenannte *real time evaluations* (RTE). RTEs finden ebenfalls parallel zum Projektverlauf statt. Sie sind stark auf Lernen vor Ort ausgerichtet, das heißt sie haben immer eine Feldforschungskomponente.<sup>12</sup> Summative (oder ex-post) Evaluierungen finden hingegen nach dem Abschluss eines Programmes oder Projektes statt oder werden rückblickend auf ein existierendes Portfolio durchgeführt.

**Reviews** sind eine informelle Version von Evaluierungen. Diese kritischen Begutachtungen stellen die gleichen Fragen wie Evaluierungen, verwenden aber weniger standardisierte Verfahren. So kommen zum Beispiel nicht immer externe Gutachter zum Einsatz. Damit sind *reviews* sowohl weniger zeit- und ressourcenintensiv als auch leichter durchzuführen. Ihre Ergebnisse sind in der Folge allerdings auch weniger robust als die von Evaluierungen.

Evaluierungen münden in der Regel in einen **Evaluierungsbericht**. Er beschreibt die zentralen Evaluationsfragen, das methodische Vorgehen, die wichtigsten Ergebnisse und darauf aufbauende Handlungsempfehlungen. Im Idealfall endet eine Evaluierung nicht mit dem Evaluierungsbericht. Vielmehr sollte er der Beginn der Beschäftigung mit den Ergebnissen und Empfehlungen durch die Organisation, insbesondere die jeweils verantwortlichen Entscheidungsträger sein. Auf **Nutzung** (*utilization*) ausgerichtete Evaluierungen messen deshalb dem Prozess – angefangen von der Entscheidung für eine Evaluierung, über die Gestaltung der Fragen bis hin zur Umsetzung der Empfehlungen – eine mindestens ebenso hohe, wenn nicht höhere, Bedeutung zu wie dem Evaluierungsbericht selbst.<sup>13</sup> Zu diesem Prozess gehört in der internationalen Praxis auch die sogenannte *management response* (die auch in der deutschsprachigen Fachliteratur unübersetzt bleibt). Sie ist die Antwort der politisch Verantwortlichen auf die Ergebnisse der Evaluierung. Darin begründet die Organisation, welche Ergebnisse sie akzeptiert und welche Empfehlungen in der Umsetzung priorisiert werden. Dazu gehört auch die Festlegung, wer welche Empfehlung bis wann umsetzen sollte.

Ziel von M&E ist gewöhnlich sowohl **Rechenschaft** (*accountability*) als auch **Lernen**. Bei der Rechenschaftslegung steht die Frage im Mittelpunkt, ob die verwendeten Mittel im Sinne der Vorgaben und Interessen des Parlaments, der Öffentlichkeit, der Hausleitung oder der von der Intervention betroffenen Bevölkerungsgruppen eingesetzt wurden. Um auf der politischen Ebene als Teil der Exekutive (Außenministerium) gegenüber dem Souverän (Parlament und Öffentlichkeit) das Rechenschaftsziel zu erreichen, sollten rechenschaftsorientierte Evaluierungen möglichst öffentlich und transparent sein.<sup>14</sup> Zielt die Evaluierung auf Lernen ab, ist es besonders wichtig, geschützte Räume für die an der evaluierten Maßnahme beteiligten Mitarbeiterinnen

12 Herson und Mitchell (2005): *Real-Time Evaluation*; Krüger und Sagmeister (2012): *Real-Time Evaluation of humanitarian assistance revisited*.

13 Patton (2008): *Utilization-Focused Evaluation*.

14 Rechenschaft erfordert nicht grundsätzlich Transparenz – in anderen Konstellationen kann durchaus auch mit unveröffentlichten Evaluierungen Rechenschaft abgelegt werden (z.B. auf der Ebene der Durchführungsorganisationen gegenüber dem Auftraggeber).

und Mitarbeiter zu schaffen. Nur so kann auch aus Fehlern gelernt werden. Formative Evaluierungen und Monitoring ermöglichen gleichermaßen das Lernen in Echtzeit.

## Vorgehen

Unter außenpolitischen Praktikerinnen und Praktikern ist Evaluierung durchaus umstritten – vor allem in Hinblick auf die Frage, ob Diplomatie als Kerngeschäft außenpolitischen Handelns überhaupt evaluierbar sei.<sup>15</sup> Zu dieser Diskussion leistet die vorliegende Studie keinen Beitrag. Sie beschränkt sich auf die Frage nach der Evaluierung von „Außenpolitik mit eigenen Mitteln“ in den friedens- und sicherheitspolitischen Portfolios und den damit verbundenen Projekten, die vom Auswärtigen Amt ausgewählt, finanziert und (zum Teil) politisch gesteuert werden.

Mit diesem Schwerpunkt analysiert die vorliegende Studie internationale Erfahrungen im Umgang mit den Herausforderungen von M&E in fragilen Kontexten. Der Fokus liegt dabei auf der Organisation von Monitoring und Evaluierung sowie der Gestaltung von Instrumenten und Methoden in relevanten, d.h. vergleichbaren und damit für Deutschland instruktiven Kontexten.

Die Untersuchung erfolgte in drei Schritten. Zunächst wurde die M&E-Politik von zehn staatlichen und multilateralen Akteuren ausgewertet, die in erheblichem Umfang und unter grundsätzlich vergleichbaren Rahmenbedingungen außenpolitische Maßnahmen durchführen. Dieses Mapping umfasste Dänemark, Großbritannien, Kanada, die Niederlande, Norwegen, die Schweiz, Südafrika, die Vereinigten Staaten von Amerika (USA), die Europäische Union (EU) und die Hauptabteilung für Friedenssicherungseinsätze der Vereinten Nationen (DPKO). Aufgrund mangelnder öffentlich erhältlicher Informationen über Art, Umfang und Details der M&E-Praktiken der meisten dieser Akteure war das Mapping nicht ausreichend für eine Fallauswahl der Detailstudien nach wissenschaftlichen Kriterien.

Für den zweiten Schritt, die Durchführung von vier Länderfallstudien, erfolgte die Auswahl deshalb gemeinsam mit dem Planungsstab des Auswärtigen Amtes nach dem erwarteten Nutzen und Erkenntnisgewinn für das Auswärtige Amt. Es wurden vier Länder bestimmt, die sowohl in den vier betrachteten Themenfeldern als auch im Bereich Monitoring und Evaluierung über langjährige Erfahrungen und/oder innovative Ansätze verfügen: Großbritannien, die Schweiz, die USA und die Niederlande. Die Analyse erfolgte anhand folgender Leitfragen:

- Wie wird in vergleichbaren Partnerländern außenpolitische Evaluierung verstanden und durchgeführt? Konkret: Warum wird in der Außenpolitik evaluiert? Was wird in der Außenpolitik evaluiert? Wie wird in der Außenpolitik evaluiert? Wer evaluiert Außenpolitik? Wie transparent ist außenpolitische Evaluierung?
- In welcher Beziehung steht Evaluierung zu einer Kultur der außenpolitischen Arbeit, die professionelles (Projekt-)Management, institutionelles Lernen und einen konstruktiven Umgang mit Fehlern fördert?

15 Siehe u.a. Kleistra und Van Willigen (2010): *Evaluating Diplomacy*.

Die Fallstudien wurden auf Grundlage von (zum Teil nicht-öffentlichen) Primärquellen, Literatur und semi-strukturierten, offenen Interviews mit Expertinnen und Experten in den jeweiligen Ländern erstellt. Insgesamt wurden 23 persönliche und 11 Telefoninterviews für die Studie durchgeführt. Alle empirischen Aussagen über die Evaluierungskonzepte und -praxis in den vier betrachteten Ländern, die nicht explizit mit Quellenangaben belegt sind, stammen aus diesen Interviews.

Im dritten Schritt wurden auf Grundlage der Länderanalysen, der Evaluierungserfahrung des Autorenteam sowie einer Reihe von Hintergrundgesprächen mit Expertinnen und Experten innerhalb und außerhalb des Auswärtigen Amtes Schlussfolgerungen gezogen.

## Überblick über die Studie

Im Anschluss an dieses Einführungs- und Überblickskapitel zur Evaluierung von Außenpolitik mit eigenen Mitteln stellt die Studie die Ergebnisse der Fallstudien Großbritannien (ab S. 16), Schweiz (ab S. 25), USA (ab S. 33) und Niederlande (ab S. 41) vor. Die Länderfallstudien folgen einer gemeinsamen Gliederung nach den Aspekten

- Zweck und Mandat,
- Akteure,
- Abläufe und Methoden,
- Einbettung in eine professionelle Projektmanagement- und Lernkultur sowie
- Wesentliche Erkenntnisse.

Um die Fallstudien überschaubar zu halten, sind sie nicht als möglichst vollständige Darstellung aller Facetten der oben genannten Aspekte konzipiert. Vielmehr konzentrieren sie sich auf diejenigen Erfahrungen der Studienländer, die besonders erfolgreich oder schwierig waren. Die Fallstudien berücksichtigen dabei den spezifischen politisch-bürokratischen Kontext, in dem die Länder ihre Erfahrungen gemacht haben. Auf diese Weise spannen die vier Fälle einen Horizont an Möglichkeiten für die Gestaltung von M&E in der Außenpolitik auf, vor dem sich die künftige deutsche Praxis unter Berücksichtigung ihrer eigenen Spezifika verorten kann. Kapitel 6 fasst die zentralen Erkenntnisse aus diesen Länderstudien zusammen und stellt Schlussfolgerungen für die Weiterentwicklung der M&E-Funktion im Auswärtigen Amt zur Diskussion.

# Großbritannien

---

Die britische Regierung betreibt in den Bereichen Stabilisierung, zivile Krisenprävention, demokratische Transformation und humanitäre Hilfe das umfangreichste und aufwändigste Evaluierungs- und Qualitätskontrollsystem aller untersuchten Fälle. Um zu verstehen, was das in der außenpolitischen Praxis genau bedeutet, müssen zwei Besonderheiten beachtet werden. Erstens, das *Foreign and Commonwealth Office (FCO)* und das *Department for International Development (DFID)* teilen sich die Verantwortung für den Großteil dieser außenpolitischen Portfolios in unterschiedlicher Form. Zweitens, die Bedeutung von formeller und unabhängiger Evaluierung ist in den beiden Behörden sehr unterschiedlich gelagert: Während DFID schon lange eine ausgeprägte Evaluierungs- und Qualitätskontrollpraxis pflegt, hat das FCO der formellen und unabhängigen Evaluierung lange Zeit wenig Aufmerksamkeit geschenkt. Vielmehr zog das FCO informelle, interne Feedbackmechanismen mit dem Ziel der Verbesserung des außenpolitischen Handwerks vor. Erst mit den krisenbedingten Sparmaßnahmen der Koalitionsregierung unter Premierminister David Cameron wurde der Druck auf das Außenministerium, *value for money* zu demonstrieren so groß, dass es nun auch an einem Ausbau seiner formellen Evaluierungspraxis arbeitet. So wurden 2011 und 2012 zum Beispiel unabhängige Portfolioevaluierungen der humanitären Hilfe und des *Conflict Pool* durchgeführt.<sup>16</sup> Auch wurde für die Evaluierung von *Official Development Assistance (ODA)*-finanzierten Maßnahmen aller Ministerien eine neue unabhängige Institution geschaffen, die *Independent Commission on Aid Impact (ICAI)*.<sup>17</sup> Sie nahm im Mai 2011 ihre Arbeit auf.

Die umfangreiche und aufwändige Praxis von Evaluierung- und Qualitätskontrolle in Großbritannien beruht auf einer langen Tradition des *New Public Management*, das evidenzbasierte Entscheidungsfindung in Verbindung mit „Ergebnisorientierung“ in der öffentlichen Verwaltung fördert.<sup>18</sup> Diese Praxis ist unabhängig von einzelnen Politikfeldern und politischen Mehrheitsverhältnissen tief in der britischen Politik- und Verwaltungspraxis verwurzelt. Von allen Ressorts wird der Nachweis erwartet, dass die in ihre Arbeit investierten Steuergelder dem Land entsprechenden Nutzen bringen. Dafür setzt das Finanzministerium ressortübergreifende Standards, so zum Beispiel das *Green Book: Appraisal and Evaluation in Central Government* und das *Magenta Book: Guidance for Evaluation*.<sup>19</sup>

16 Ashdown und Mountain (2011): *Humanitarian Emergency Response Review*; Independent Commission for Aid Impact (2012): *Evaluation of the Inter-Departmental Conflict Pool*; National Audit Office (2012): *Review of the Conflict Pool*.

17 The Guardian (2010): *Conservative-Liberal Democrat coalition deal*.

18 *New Public Management* ist eine Richtung der Verwaltungsreform, die mit der Einführung privatwirtschaftlicher Managementtechniken versucht, den Staatsapparat effektiver und effizienter zu gestalten. *New Public Management* ist in Großbritannien seit der Regierungszeit Margaret Thatchers ein fester Bestandteil der Verwaltungspraxis. Es hatte in den letzten Jahrzehnten aber auch in den Niederlanden und Schweden deutlichen Einfluss auf die Gestaltung von Verwaltungen. Vgl. Pollitt und Bouckaert (2004): *Public Management Reform*.

19 HM Treasury (2011a): *Green Book*; HM Treasury (2011b): *Magenta Book*.

Neben einer evaluierungsaffinen Verwaltungskultur haben FCO und DFID aber auch ganz pragmatische Gründe für ein umfangreiches Kontroll- und Lerninstrumentarium: Der Gesamtumfang der Mittel für die Portfolios Stabilisierung, zivile Krisenprävention, demokratische Transformation und humanitäre Hilfe beläuft sich auf ca. 1,1 Mrd. GBP (€ 1,3 Mrd.) im Jahr<sup>20</sup> und ist damit erheblich höher als die entsprechenden Mittel der Bundesrepublik. Daher stellt sich die Frage, wie diese Summen wirkungs- und zielorientiert ausgegeben werden können, besonders dringend.

## Zweck und Mandat

Im Spannungsfeld von Rechenschaftslegung und institutionellem Lernen liegt der Schwerpunkt der Evaluierungsinstrumente in Großbritannien daher auf der Rechenschaftsfunktion. Allerdings werden diese zunehmend mit lernorientierten Instrumenten flankiert.

Die einzelnen Ministerien bzw. Behörden haben sich auf mehr oder weniger direkten Druck des Parlaments im Laufe der Zeit eigene Evaluierungsmandate gegeben.<sup>21</sup> Das FCO hat sich selbst verpflichtet, Portfolios bzw. Programme (z.B. *Human Rights and Democracy, Counterterrorism, Arab Partnership Programme Fund*), jedes sogenannte „strategische ODA-Projekt“ ab einem jährlichen Umfang von 500 000 GBP sowie eine Auswahl von Projekten ab 100 000 GBP intern ex-post zu evaluieren. In der Praxis bedeutet dies bislang, dass die Projekte einer jährlichen *review* unterzogen werden (siehe unten). Neuerdings ist geplant, alle vier Jahre für jedes Portfolio oder Programm eine unabhängige externe Evaluierung durchzuführen.<sup>22</sup> Für humanitäre Hilfe hat DFID ein vergleichbares Evaluierungsmandat, allerdings gelten aufgrund der größeren Portfolio-, Programm- und Projektvolumina höhere Schwellenwerte.

Die *Independent Commission on Aid Impact (ICAI)* hingegen wurde vom Parlament beauftragt, Evaluierungen beliebiger Aspekte der Verwendung britischer ODA-Mittel durchzuführen. Aufgrund der enormen Vielfalt möglicher Evaluierungsthemen hat sich ICAI vier Kriterien für die Prioritätensetzung verordnet: den finanziellen Umfang einer Maßnahme, das Risiko ihres Scheiterns, die Sicherstellung einer Vielfalt von Evaluierungsschwerpunkten sowie aktuelle politische Prioritäten.

In jüngster Zeit hat das FCO zudem begonnen, sich mit Methoden zur Politik- und Politikfeld-Evaluierung zu beschäftigen. Dies geschieht auf zwei Grundlagen: zum einen der bisherigen Erfahrung mit informellen *review*- und *lessons learned*-Prozessen innerhalb des FCO und zum anderen in enger Zusammenarbeit mit DFID auf Grundlage des dort entwickelten Ansatzes zur Evaluierung von politischer Einflussnahme (*influencing evaluation*).<sup>23</sup>

20 Dazu gehören der *Conflict Pool* (etwa 220 Mio. GBP/ € 265 Mio. pro Jahr), teilselbständige Institutionen wie die *Westminster Foundation for Democracy* (etwa 5 Mio. GBP/ € 6 Mio. pro Jahr) und ad-hoc-Budgets wie der gemeinsame *Arab Partnership Programme Fund* (130 Mio. GBP/ € 158 Mio. für drei Jahre). In der humanitären Hilfe ist DFIDs Abteilung für *Conflict, Humanitarian Assistance and Security (CHASE)* mit einem Budget von etwa 723 Mio. GBP/ € 878 Mio. im Jahr zuständig. Für Menschenrechts- und Demokratieförderung verfügt das FCO im Rahmen seiner *Strategic and Bilateral Programmes* über 130 Mio. GBP/ € 158 Mio. im Jahr. Rotmann und Steinacker (2013): *Stabilisierung*. Development Initiatives (2013): *Global Humanitarian Assistance Report*.

21 Vgl. Independent Commission for Aid Impact (2011): *Effectiveness and Value for Money*; Department for International Development (2013): *International Evaluation Policy*.

22 Foreign and Commonwealth Office (2013): *Departmental Improvement Plan*.

23 Department for International Development (2013): *HTN Evaluating Influence*.

## Akteure

Innerhalb des FCO liegt die Steuerung der Portfolio-, Programm- und Projektevaluierung seit 2010 beim *Policy Programme Evaluation Board*, einer Art Evaluierungsausschuss, der von einem *Programme Office* in der Haushaltsabteilung unterstützt wird. Der Evaluierungsausschuss besteht aus hochrangigen Beamten und unabhängigen „Verwaltungsräten“ (*non-executive members* des *FCO Management Board*) sowie einem Staatsminister, wobei ein *non-executive* den Vorsitz führt.<sup>24</sup> Der Ausschuss beschließt einen mehrjährigen Evaluierungsplan, den das *Programme Office* mithilfe von akkreditierten Evaluatorinnen und Evaluatoren aus dem Kreis des FCO-Personals umsetzt. Die Akkreditierung von Regierungsbeamten für die Evaluierung wird zentral vom Finanzministerium erteilt und beinhaltet eine Weiterbildungskomponente. Das akkreditierte FCO-Personal wird für die Teilnahme an internen Evaluierungen für einige Tage oder Wochen von seiner regulären Tätigkeit freigestellt. Ab einem Projekt-, Programm- oder Portfoliobudget von 500 000 GBP dürfen diese internen Evaluatorinnen und Evaluatoren nichts mit den betreffenden Aktivitäten zu tun haben. Unterhalb dieser Schwelle ist nur die Unabhängigkeit von der einzelnen evaluierten Maßnahme sicherzustellen.<sup>25</sup> Zukünftig sollen mehr dieser internen Evaluierungen von externen Evaluatorinnen und Evaluatoren durchgeführt werden; erste Ausschreibungen wurden im Oktober 2013 vorbereitet. Zudem genießt praxisorientierte Forschung einen hohen Stellenwert: Ein *Chief Scientific Advisor* gehört als *non-executive member* der Leitungsebene des FCO an, und den Abteilungen stehen erhebliche Forschungsbudgets zur Verfügung.<sup>26</sup>

Die Forschungs- und Evaluierungsabteilung von DFID ist demgegenüber noch umfangreicher und greift auf langjährige Erfahrung zurück. Sie wird von einem Universitätsprofessor geleitet, der alle vergebenen Forschungsaufträge beaufsichtigt.<sup>27</sup> Die Forschungsabteilung hatte im Jahr 2011 ein Gesamtbudget von etwa 220 Mio. GBP (€ 267 Mio.). Davon standen ca. 775 000 GBP (€ 941 000) für Evaluierungen im Bereich *governance & security* zur Verfügung. Ein solches Budget erlaubt es, etwa sechs oder sieben umfangreiche Evaluierungen (z.B. Portfolioevaluierungen) pro Jahr umzusetzen.<sup>28</sup> Neben formellen Evaluierungen werden aus dem Forschungsbudget auch Studien im Format „*What works in stabilization/protection of civilians/etc.?*“ in Auftrag gegeben. Solche Studien leisten keinen Beitrag zur Rechenschaftslegung, helfen aber *best practices* und systematische Probleme zu identifizieren und tragen

24 Die praktische Rolle dieser Management- oder Aufsichtsstruktur ist von außen schwer zu analysieren. Öffentlich erscheint das *Management Board* unter Vorsitz des *Permanent Under Secretary*/Staatssekretärs als höchstes Gremium in grundsätzlichen Strategie- und Strukturfragen (*corporate policy*). Es umfasste im Januar 2014 die 6 höchsten FCO-Beamten, 4 professionelle Aufsichtsräte aus Wirtschaft und Gesellschaft sowie den *Chief Scientific Advisor*: <https://www.gov.uk/government/organisations/foreign-commonwealth-office#management>, letzter Zugriff am 13.01.2014; <https://www.gov.uk/government/speeches/foreign-and-commonwealth-office-spending-2013-14>, letzter Zugriff am 07.01.2014.

25 Foreign and Commonwealth Office (2013): *ODA Evaluations 2012-13*.

26 Informelle Gespräche mit britischen und internationalen Wissenschaftlerinnen und Wissenschaftlern, die in den letzten Jahren im Umfeld von Konfliktbearbeitung und Stabilisierung Studien für das FCO erstellt haben.

27 Royal Geographical Society (o.J.): *Feeding the 9 Billion*.

28 Department for International Development (2011): *Operational Plan 2011-2015*.



damit zum Lernen bei.<sup>29</sup> DFIDs gesamte Evaluierungsabteilung verfügt über etwa 40 Mitarbeiterinnen und Mitarbeiter mit Expertise in Evaluierung und Statistik sowie sogenannte *Results Advisors*, welche die Programmentwicklung unterstützen. Weitere 120 DFID-Mitarbeiterinnen und Mitarbeiter sind für die Durchführung interner Evaluierungen akkreditiert. Darüber hinaus werden in erheblichem Umfang externe Gutachter engagiert. Die bisherige Zentralisierung der Evaluierungskompetenz in einer Abteilung wird seit 2011 zunehmend aufgebrochen. Die Mitarbeiterinnen und Mitarbeiter werden nun auf die DFID-Länderbüros und die Fachabteilungen verteilt. Für die Abteilung *Conflict, Humanitarian Assistance and Security (CHASE)*, die für diese Studie relevant ist, arbeiten inzwischen nur noch zwei Evaluierungsreferentinnen in Großbritannien.<sup>30</sup> Mit dieser großzügigen personellen und finanziellen Ausstattung sowie einem hohen Qualitätsanspruch – auch im internationalen Vergleich – zählt sich DFID zur Evaluierungselite. Es entwickelt daher im Zusammenwirken mit FCO ein entsprechendes Sendungsbewusstsein bezüglich einer „richtigen“ Evaluierungspraxis.

ICAI verfolgt, wie oben bereits erwähnt, Evaluierungen vor allem mit dem Ziel der Rechenschaftslegung. Die konservativ-liberale Parlamentsmehrheit, die hinter der Gründung von ICAI steht, erwartet mittelfristig eine sichtbare Verbesserung der Wirksamkeit britischer ODA-Mittel. Ihrem Auftrag kommt die Kommission durch einen eher konfrontativen Evaluierungsprozess nach. Jede ICAI-Studie wird von einem der vier Kommissionsmitglieder selbst geleitet, die damit sowohl Datenzugang bei DFID und FCO sicherstellen als auch den Analyseprozess inhaltlich prägen. Das ist für jeden der Teilzeit-Kommissare mit erheblichem Zeitaufwand verbunden: Bis zu 20 Tage pro Studie verbringt ein Mitglied mit der Leitung von ICAI-Evaluierungen. Der Großteil der Datenerhebungs- und Analysearbeit der jeweiligen Studien wird von dem Beratungsunternehmen KPMG oder einem Unterauftragnehmer durchgeführt. Ein fünfköpfiges Sekretariat unterstützt die Kommission.<sup>31</sup> Durch ihre Unabhängigkeit von der gesamten öffentlichen Verwaltung und die politisch erzwungene Vereinfachung ihrer Urteile entlang eines Ampelsystems von grün bis rot haben viele ICAI-Berichte erhebliche politische und budgetäre Auswirkungen für die untersuchten Portfolios; vor allem die „Verlierer“ sind in Whitehall weithin bekannt – und ICAI ist deshalb zunehmend gefürchtet.<sup>32</sup>

## Abläufe und Methoden

In Großbritannien ist M&E nur ein Teil eines umfangreichen evidenzbasierten und ergebnisorientierten Steuerungsmodells für die öffentliche Verwaltung (*Results-Based Management*, RBM). Die britische Variante von RBM begünstigt formale und damit oft bürokratische Verfahren der Qualitätssicherung und der Entscheidungsfindung.

29 Vgl. Binder und Reichhold (2013): *Scoping Study*.

30 Department for International Development (2011): *Operational Plan 2011-2015*.

31 Die Kommission besteht zum Großteil aus professionellen Beraterinnen und Beratern: je ein Kommissionsmitglied hat eine Karriere bei den Beratungs- und Wirtschaftsprüfungsunternehmen PriceWaterhouseCoopers und Accenture sowie der Großkanzlei Linklaters hinter sich; das vierte Kommissionsmitglied hatte Führungspositionen bei Transparency International in Kenia und Berlin inne (sowie als Staatssekretär des kenianischen Präsidenten).

32 Gespräche mit FCO- und DFID-Mitarbeitern.

Im FCO wird auf Portfolio-, Programm- und Projektebene, wie oben beschrieben, bislang ausschließlich intern evaluiert. Bis der Rechtfertigungsdruck *value for money* zu demonstrieren 2011/12 auch das FCO erreichte, ließen sich die zusätzlichen Kosten für externe Evaluierungen schwer rechtfertigen. Darüber hinaus wurden externe Evaluierungen nicht in Betracht gezogen. Dem liegt die Annahme zugrunde, dass es externe Gutachter nicht leicht haben, über formale Zuständigkeiten hinaus die entscheidenden *stakeholder* innerhalb des Hauses zu identifizieren. Der Ablauf interner Evaluierungen bzw. *reviews* ist einfach: Das *Program Office* und der Budgetverantwortliche für den jeweiligen Haushaltstitel (*Senior Responsible Owner*, üblicherweise auf Referatsleiter- oder Beauftragtebene) einigen sich auf ein Pflichtenheft (*terms of reference*, TOR). Als Evaluatoren und Evaluatorinnen akkreditiertes Personal führen die Evaluierungen gemeinsam durch. Die Berichte dieser internen Evaluierungen werden nicht veröffentlicht. Die oben beschriebenen neu geplanten externen Evaluierungen sollen in Zukunft zumindest zusammenfassend öffentlich zugänglich sein. Darüber hinaus werden Programm- bzw. Portfolioevaluierungen eingeführt, denen jeder Haushaltstitel alle vier Jahre unterworfen werden soll.

ICAI und DFID folgen umfangreicheren, an internationalen Standards angelehnten Evaluierungsabläufen, da sie schwerpunktmäßig nicht interne *reviews*, sondern externe und unabhängige Evaluierungen durchführen. Diese Prozesse beinhalten eine öffentliche Ausschreibung, eine *inception phase*, in der das externe Evaluationsteam die Evaluierungsfragen spezifiziert, die Evaluierungsmethode vertieft, geplante Konsultationen beschreibt und einen ersten Gliederungsentwurf für den Evaluierungsbericht vorstellt.

Methodisch ist vor allem die Arbeit von DFID und ICAI interessant. Sie muss jedoch vor dem Hintergrund einer hohen Evaluierungskompetenz innerhalb der britischen Verwaltung betrachtet werden. DFID investiert seit Jahrzehnten in die Methodenentwicklung und Evaluierungspraxis. So war DFID zum Beispiel einer der Vorreiter bei der Einführung des *Logical Framework Approach (LFA)*. Der LFA ist eine vereinfachte Darstellung der Wirkungshypothese einer Maßnahme. Das heißt, er zeigt wie die geplanten *inputs* zu *outputs* und *outcomes* werden sollen, die sich dann idealerweise in die gewünschte Wirkung (*impact*) übersetzen. Aktuell gehört DFID auch zu den ersten internationalen Akteuren, die den LFA in fragilen Kontexten in Frage stellen. In diesen Kontexten stehen inzwischen sogenannte Veränderungstheorien (*Theories of Change*) im Mittelpunkt von Evaluierung und Projektmanagement. Eine Veränderungstheorie besteht dabei aus drei Elementen: einer Kontextanalyse, Hypothesen über den Beitrag einer bestimmten Maßnahme zu der gewünschten Kontextveränderung, und einer empirischen Überprüfung dieser Hypothesen.<sup>33</sup> Anders als beim LFA geht es bei der Veränderungstheorie weniger darum, vereinfachte Wirkungstheorien darzustellen. Ziel ist es vielmehr, der Kontextabhängigkeit von Programmfolgen stärker Rechnung zu tragen. Die Ergänzung klassischer *Logframes* durch Veränderungstheorien in den Planungs- und Überprüfungsprozessen von DFID und seinen Umsetzungspartnern ist kostenintensiv. Sowohl Mitarbeiterinnen und Mitarbeiter als auch Partnerorganisationen müssen darin geschult werden, Veränderungstheorien zu entwickeln und in PM&E anzuwenden. Wie der LFA bleiben

33 Vogel (2012): *Review of the use of 'Theory of Change'*.

aber auch die Veränderungstheorien in einer positivistischen Kausallogik verankert, denn das internationale Leitbild der positivistischen Wirkungsevaluierung wurde in wesentlichen Teilen von DFID geprägt. Die zugrundeliegende Weltsicht wird auch weiterhin selbstbewusst vertreten, in Aussagen wie „*We strongly believe what works is knowable.*“ Dennoch räumen auch die meisten DFID-Mitarbeiterinnen und Mitarbeiter ein, dass kausale Wirkungsmessung in fragilen Kontexten unmöglich sei. In diesen Kontexten investiert das Department seit geraumer Zeit in die Weiterentwicklung von Evaluierungsmethoden. Wegen ihrer kausallogischen Verankerung sind sowohl LFA als auch Veränderungstheorien gut mit den OECD-DAC-Kriterien zu verbinden. In der humanitären Hilfe arbeitet DFID mit den angepassten Kriterien,<sup>34</sup> ist aber in bestimmten Kontexten auch bereit, spezifische Kriterien zu entwickeln.

Im Vergleich zu DFID sind die Methoden und Prozesse, die ICAI anwendet, ohne theoretischen Überbau deutlich weniger standardisiert. Entsprechend verwendet ICAI nicht automatisch die OECD-DAC Kriterien, obwohl sie es in ihrer Arbeit ausschließlich mit ODA-Mitteln zu tun hat. Damit berücksichtigt die Kommission, dass ODA-Mittel durchaus auch für außenpolitische Maßnahmen in fragilen Kontexten verwendet werden, die andere oder zusätzliche Zielsetzungen haben als klassische ODA-Projekte. Um ihre Flexibilität zu wahren, hat sich die Kommission vier eigene *guiding criteria* gegeben, die nicht im Gegensatz zu den OECD-DAC-Kriterien stehen, aber doch deutlich andere Schwerpunkte setzen:

- realistische und angemessene Ziele,
- robuste Umsetzungsarrangements entlang der gesamten Umsetzungskette,
- transformative, positive und nachhaltige Wirkung,
- in die Projektumsetzung eingebettete Lernelemente.

Neben dem flexiblen Umgang mit Evaluierungskriterien verfügt ICAI über einen interessanten Prozess, um Evaluierungsempfehlungen nachzuverfolgen. Das Evaluierungsteam besucht die Verantwortlichen für die Umsetzung von Empfehlungen 6, 12 und 18 Monate nach der Evaluierung. Es bespricht mit den Verantwortlichen den Stand der Umsetzung, welche Probleme oder Veränderungen es gibt und ob das Evaluierungsteam bei der Umsetzung behilflich sein kann.

Durch die Arbeit von ICAI hat Evaluierung in der Außenpolitik und Entwicklungszusammenarbeit in Großbritannien nochmals deutlich an Schwung gewonnen. Gleichzeitig ist durch die Etablierung eines weiteren Akteurs neben den internen akkreditierten sowie extern individuell beauftragten Evaluatoreninnen und Evaluatoren die Kohärenz von Evaluierungen in Außen- und Entwicklungspolitik gesunken. FCO und DFID müssen nun die Ergebnisse zahlreicher, methodisch unterschiedlicher Evaluierungen in ihre Abläufe und Entscheidungsfindungen integrieren. Die Ergebnisse von ICAI Evaluierungen werden gemeinsam mit einer *management response* auf der Website der britischen Regierung veröffentlicht, was den Umsetzungsdruck erhöht.

<sup>34</sup> Die angepassten Kriterien sind *relevance/appropriateness, connectedness, coherence; coverage, efficiency, effectiveness* und *impact*, zu Deutsch: Relevanz/Angemessenheit; Verbindung zu Entwicklung, Sicherheit und lokalen Strukturen; Kohärenz; Abdeckung vorhandener Bedürfnisse; Effizienz; Effektivität und Wirkung. Vgl. Burchett, Vogler und Percival (2007): *Legends of Stonehenge*.

## Einbettung in eine professionelle Projektmanagement- und Lernkultur

Die strukturelle Vorherrschaft des *Results-Based Management (RBM)* trägt dazu bei, dass die Formulierung expliziter Ziele und messbarer Indikatoren<sup>35</sup> in der britischen Verwaltung zum Kerngeschäft gehört. Dadurch ist jedes außenpolitische Projekt, Programm oder Portfolio für Stabilisierung, Krisenprävention, humanitäre Hilfe oder demokratische Transformation Teil einer komplizierten Zielpyramide. Diese beginnt bei konkreten politischen Prioritäten (z.B. „*Agree on a sufficient and sustainable size, shape and funding for the Afghan National Security Forces*“) und endet in strategischen Visionen („*By 2015 we aim to be the best diplomatic service in the world, recognised as ambitious, confident, proud of what we do and highly rated for our policy analysis and for creative and action-oriented delivery.*“<sup>36</sup>) Gleichzeitig erzwingt RBM ein erhebliches Maß an bürokratischem Aufwand. Gerade bei komplexen, veränderlichen Politikzielen steht dieser Aufwand aus Sicht der damit konfrontierten Mittelempfänger nicht immer im Verhältnis zum Steuerungsnutzen.

Ein Beispiel hierfür sind die sogenannten *Business Cases*. Diese sind innerhalb von DFID inzwischen ein Standardinstrument zur Projektentwicklung, -steuerung und -bewertung. Ein *Business Case* beinhaltet eine Kosten-Nutzen-Kalkulation und einen *Logframe*, der später zur Ergebniskontrolle verwendet wird. Er wird gemeinsam mit der Veränderungstheorie für die Projektplanung und -überprüfung verwendet. Die Kosten-Nutzen-Kalkulation geht soweit, dass DFID seine humanitären Implementierungspartner dazu anhält „Stückkosten“ zu berechnen. Das heißt, eine Aussage darüber zu treffen, wie viele Personen von einem bestimmten Projekt profitieren und welche Kosten pro Person anfallen. Innerhalb von DFID scheint es geteilte Meinungen über die Praktikabilität der *Business Cases* zu geben. Während in unseren Interviews die Genauigkeit der Methode gelobt wurde, betonten andere DFID Mitarbeiter in informellen Gesprächen, dass sich das Department bewusst sei, dass sowohl die *Business Cases* als auch ihre Verbindung mit Veränderungstheorien und *Logframes* einen exzessiven bürokratischen Aufwand für DFIDs Partnerorganisationen nach sich zöge.

Der Formalität und Bürokratisierung des RBM steht die Bedeutung von Forschung, institutionellem Lernen und Diversifizierung analytischer Perspektiven gegenüber. Diese wissens- und lernorientierte Kultur trägt zur nationalen und internationalen Wertschätzung der Qualität und Professionalität der Arbeit von FCO und DFID bei. Diese Kultur kommt vor allem in kleinen, informellen Lernformaten, die auf unterschiedlichen Ebenen angesiedelt sind, zum Ausdruck. Auf der Ebene der Politikziele kommen bei FCO zum Beispiel *challenge sessions* zur Anwendung. Diese Runden werden von der *Policy Unit* (Planungsstab) einberufen und moderiert. Neben den Verantwortlichen für eine bestimmte Politik nehmen thematische Expertinnen und Experten und Generalisten teil. Diese haben die Aufgabe eine bestimmte Politik zu hinterfragen. Ziel ist es, die Argumentation für und Ausgestaltung von bestimmten Politiken zu stärken.<sup>37</sup> Ein weiteres Beispiel ist das Feiern von Erfolgen, das im FCO

35 Nach den SMART-Kriterien: *specific, measurable, attainable, relevant* und *time-bound*, zu Deutsch: spezifisch, messbar, erreichbar, relevant und zeitlich beschränkt.

36 Foreign and Commonwealth Office (2013): *Departmental Improvement Plan*.

37 Foreign and Commonwealth Office (2013): *Departmental Improvement Plan*.

als zunehmend wichtig erachtet wird. Auf diese Weise soll der klassische Fokus von M&E auf Defizite und Verbesserungsmöglichkeiten durch eine systematische Aufmerksamkeit für Dinge, die gut laufen, ergänzt werden. Mitarbeiterinnen und Mitarbeiter sollen sich anerkannt fühlen. Ein letztes Beispiel ist die Möglichkeit für Leserinnen und Leser, Drahtberichte im FCO-Intranet mit Sternen zu bewerten. So bekommen die Verfasserinnen und Verfasser Rückmeldungen auf ihre Berichte in Echtzeit und gleichzeitig verbessert sich durch den Wettbewerb die Qualität der Berichte, so die interne Beurteilung.

## Wesentliche Erkenntnisse

### Kurzüberblick anhand der Leitfragen (S. 14)

Warum?	Eigene Mandate aller Ministerien, Standardsetzung durch Finanzministerium
Schwerpunktsetzung	Rechenschaft und Lernen; mit sehr ausgeprägten Schwerpunkten in unterschiedlichen Instrumenten
Wie?	<ul style="list-style-type: none"> <li>• Investigative rechenschaftsorientierte Instrumente (ICAI, in geringerem Maße DFID, zukünftig auch FCO)</li> <li>• weniger systematisch: Lernerorientierte Instrumente wie interne Selbstevaluierungen/Reviews, Studien (FCO, DFID)</li> </ul>
Was?	<ul style="list-style-type: none"> <li>• FCO: Portfolios, Programme, größere Projekte werden bislang intern, in Zukunft formell extern evaluiert</li> <li>• DFID: Portfolios, Programme, größere Projekte werden extern evaluiert</li> </ul>
Wer? (spezifisch für die außenpolitischen Portfolios)	<ul style="list-style-type: none"> <li>• FCO: zentrales Program Office in der Haushaltsabteilung (DFID: de-zentral in den Abteilungen)</li> <li>• Externe Evaluierungen durch Gutachterinnen und Gutachter</li> <li>• Externe Behörde: ICAI (nur ODA-Mittel)</li> </ul>
Wie transparent?	<ul style="list-style-type: none"> <li>• ICAI, DFID: Alle formellen Evaluierungen sind öffentlich; Selbstevaluierungen bleiben intern</li> <li>• FCO: Bislang nur interne Reviews/Selbstevaluierungen</li> </ul>
Projektmanagement	M&E sind in aufwändige Projektmanagement- und Zielorientierungsverfahren eingebettet

1. M&E wird vor allem in der britischen Entwicklungszusammenarbeit und humanitären Hilfe mit erheblichem Aufwand und entsprechender Professionalität betrieben. Dies führt einerseits zu einer international als qualitativ hochwertig eingeschätzten Evaluierungspraxis. Andererseits zeigt sich – wie DFID-Mitarbeiterinnen und -Mitarbeiter bestätigen – dass der formalistisch-positivistische Ansatz vor allem in fragilen Kontexten an seine Grenzen stößt. Daher investiert DFID in die Entwicklung neuer Evaluierungsansätze, hat seine Praxis aber noch nicht entsprechend angepasst. Im FCO besteht hingegen bisher nur eine weitgehend intransparente, auf

institutionelles Lernen ausgerichtete Praxis der Selbstevaluierung, die in Zukunft durch den zunehmenden Einsatz unabhängiger Gutachterinnen und Gutachter einen stärkeren Rechenschaftscharakter gewinnen soll.

2. Das *Results-Based Management (RBM)*, das die britische Verwaltung und ihren Umgang mit M&E prägt, ist ein zweischneidiges Schwert. Einerseits hilft es dem FCO und DFID, klare Ziele zu formulieren, Prioritäten zu setzen und ein professionelles Projektmanagement zu betreiben, das M&E als integralen Bestandteil betrachtet. Gleichzeitig führt RBM zu sehr aufwändigen bürokratischen Prozessen; vor allem wenn neue Instrumente alte nicht ablösen, sondern ergänzen. Die parallele Verwendung von *Business Cases*, *Logframes* und Veränderungstheorien bei DFID steht exemplarisch für dieses Problem.
3. Die britische Außenpolitik profitiert von einer vergleichsweise stark verankerten Lernkultur. M&E wird durch praxisorientierte Forschung in erheblichem Umfang ergänzt. Dies gibt FCO und DFID die Möglichkeit, Lernprozesse von außen unterstützen zu lassen. Externe Studien sind in der Regel weniger betreuungsintensiv als externe Evaluierungen. Sie bieten die Chance, nicht nur aus eigenen Erfolgen und Fehlern zu lernen, sondern auch aus denen anderer Akteure. Auch im Kleinen praktiziert das FCO potentiell sehr wirksame Lernmaßnahmen, wie *challenge sessions* und das Feiern von Erfolgen. Dadurch werden die sonst eher defizitorientierten internen und externen Evaluierungen durch positive Lernerlebnisse ergänzt.
4. Die Gründung von ICAI hat die Rechenschaftslegung über die Verwendung der britischen ODA-Mitteln auch im Bereich des *Conflict Pool* (Stabilisierung, Krisenprävention) gestärkt. Die Kommission gilt als unabhängig, sie wird von den Betroffenen durchaus gefürchtet, aber auch als hilfreich empfunden, da sie die Evaluierungsergebnisse nicht politisiert und die Verantwortlichen bei der Umsetzung von Empfehlungen konstruktiv unterstützt. Darüber hinaus führt ICAI vor, wie man konstruktiv mit selbst gewählten Kriterien für die Erfolgsmessung von Portfolios, Programmen und Projekten eine Evaluierungsagenda prägt. Dabei orientiert sich die Kommission an internationalen Standards, setzt aber ihre eigenen Schwerpunkte.
5. Sowohl ICAI als auch das FCO arbeiten mit systematischen Verfahren, um die Umsetzung akzeptierter Empfehlungen sicherzustellen. Auf Grundlage der offiziellen *management response* besuchen Vertreter des Evaluierungsteams von ICAI in regelmäßigen Abständen die jeweils operativ Verantwortlichen, um sich über die Fortschritte zu informieren. Ähnliche Abläufe gibt es auch im FCO (und bei DFID).

Die Schweiz hat eine für die deutschen Überlegungen zur Stärkung der außenpolitischen Evaluierung sehr relevante Erfahrung gemacht: Das *Eidgenössische Departement für auswärtige Angelegenheiten (EDA)* führte 2008 eine zentral organisierte, außerhalb der operativen Strukturen angesiedelte Evaluierungsfunktion ein, um neben der Außenpolitik mit eigenen Mitteln auch Diplomatie zu evaluieren (Politikevaluierung). Diese Evaluierungsfunktion wurde 2012 unter neuer politischer Führung aufgrund mangelnder Akzeptanz im Haus reorganisiert und in Anspruch und Umfang deutlich reduziert. Politikevaluationen wurden wieder abgeschafft. Über diese interessante Erfahrung hinaus verfügt die Eidgenossenschaft außerdem über einige innovative Ansätze in der Evaluierungspraxis außenpolitischer Maßnahmen.

Wie in Großbritannien prägen auch in der Schweiz Besonderheiten des politischen Systems und Verwaltungstraditionen die Evaluierungspraxis. Hierzu gehört, dass Wirkungsmessung seit 1999 Verfassungsrang hat. Das Parlament ist damit ermächtigt, Maßnahmen des Bundes auf ihre Wirkung hin zu überprüfen. Dabei kommt der *Parlamentarischen Verwaltungskontrolle (PVK)* eine zentrale Rolle zu. Sie kann vom Parlament beauftragt werden, eine bestimmte Evaluierung durchzuführen. Im Bereich Außenpolitik wurden zwischen 2006 und 2013 vier Evaluierungen von der PVK verantwortet. Eine weitere Besonderheit ist das sogenannte Mitberichtsverfahren. Es dient der Vorbereitung von Regierungsentscheidungen und reflektiert die Natur der Schweizer Exekutive als sogenannte Kollegialbehörde. Beim Mitberichtsverfahren wird die von einem Departement erstellte Entscheidungsvorlage für den Bundesrat („Bericht“) allen anderen Departementen zur Stellungnahme vorgelegt („Mitbericht“). Bericht und Mitbericht werden separat an den Bundesrat gegeben, so dass eventuelle Meinungsverschiedenheiten zwischen den Departementen sichtbar bleiben.<sup>38</sup> Das Mitberichtsverfahren wird laut der für diese Studie interviewten Mitarbeiterinnen und Mitarbeiter des EDA in der Schweiz als Qualitätskontrollinstrument für politische Inhalte wahrgenommen.

Schließlich ist für die Evaluierung in der Außenpolitik relevant, dass es in der Eidgenossenschaft eine sogenannte Zielkaskade für das EDA gibt. Auf der höchsten Ebene dieser Kaskade befinden sich die jährlich veröffentlichten *Ziele des Bundesrates*.<sup>39</sup> Darunter folgt der *Bericht über die aussenpolitischen Schwerpunkte der Legislatur*.<sup>40</sup> Er formuliert anhand von vier Prioritäten die außenpolitische Strategie der Schweiz. Aktuell sind dies Pflege und Ausbau der Beziehungen mit den Nachbarstaaten, Anpassung und Vertiefung der Beziehung zur Europäischen Union, Fortsetzung und Anpassung des Schweizer Engagements für Stabilität in Europa und der restlichen Welt sowie Stärkung und Diversifizierung von strategischen Partnerschaften.

38 Bundesamt für Justiz (2013): *Gesetzgebungsleitfaden*.

39 Schweizerische Eidgenossenschaft (2013): *Ziele Bundesrat 2014*.

40 Eidgenössisches Departement für auswärtige Angelegenheiten (2012): *Bericht über die aussenpolitischen Schwerpunkte der Legislatur*.

Die konkreteste Ebene der Zielkaskade sind schließlich die *Botschaften an die Bundesversammlung* (einem Kabinettsbeschluss vergleichbar) für die *Abteilung für menschliche Sicherheit (AMS)* und die *Direktion für Entwicklung und Zusammenarbeit (DEZA)*. Diese formulieren strategische Grundsätze, Ziele und Indikatoren für die beiden Einheiten des EDA, die für Stabilisierung, zivile Friedensförderung, demokratische Transformation und humanitäre Hilfe (unter anderen Bezeichnungen) verantwortlich sind.<sup>41</sup>

Damit sind, ähnlich wie in Großbritannien und in den Niederlanden, außenpolitische und entwicklungspolitische Akteure gemeinsam für die hier betrachteten Themen zuständig. Als Abteilung des EDA ist DEZA allerdings stärker in dessen Strukturen integriert als DFID in das FCO. Anders als die niederländische Entwicklungspolitik hat die DEZA aber eine eigenständige Identität und eigenen Handlungsspielraum als entwicklungspolitischer Akteur.

## Zweck und Mandat

Das Mandat für Evaluierung in der Außenpolitik ist in der Schweiz auf zwei Ebenen geregelt. Zum einen wird in Artikel 170 der Bundesverfassung die Bundesversammlung angehalten, dafür Sorge zu tragen, dass „Maßnahmen des Bundes auf ihre Wirksamkeit überprüft werden.“<sup>42</sup> Dies wird damit begründet, dass sich „[s]taatliches Handeln [...] heute nicht allein durch die rechtmässige Erfüllung verfassungsmässiger und gesetzlicher Aufgaben, sondern [...] auch durch eine wirksame, effiziente und nachhaltige Lösung öffentlicher Herausforderungen [...] legitimiert.“<sup>43</sup> Zum anderen legt das *Bundesgesetz über Massnahmen zur zivilen Friedensförderung und Stärkung der Menschenrechte*<sup>44</sup> fest, dass der Bundesrat regelmässig Evaluierungen veranlasst und der Bundesversammlung für jeden Haushaltszyklus über die Verwendung der mehrjährigen Mittel zur Finanzierung außenpolitischer Massnahmen Bericht erstattet.<sup>45</sup> Für die Jahre 2012-2016 beläuft sich das Budget der AMS auf 310 Mio. CHF (ca. € 240 Mio.).<sup>46</sup>

Entsprechend geht die *Botschaft über die Weiterführung von Massnahmen zur Förderung des Friedens und der menschlichen Sicherheit* explizit auf die Funktion von Evaluierungen bei der Wahl der Schwerpunktländer, der Steigerung der Wirksamkeit des Schweizer Engagements und der Formulierung von Ausstiegsstrategien in sich

41 Eine *Botschaft des Bundesrates* ist ein Bericht der Schweizer Regierung an die Bundesversammlung. In einer Botschaft erläutert der Bundesrat seine Vorschläge für einen parlamentarischen Erlass (rechtssetzende und nicht-rechtssetzende Beschlüsse). Vgl. Schweizerischer Bundesrat (2011): *Massnahmen zur Förderung des Friedens*; Schweizerischer Bundesrat (2012): *Internationale Zusammenarbeit*.

42 Schweizerische Eidgenossenschaft (2013): *Bundesverfassung*.

43 Schweizerische Eidgenossenschaft (2003): *Medienmitteilung Bundeskanzlei*.

44 Schweizerische Eidgenossenschaft (2013): *Bundesgesetz Friedensförderung und Menschenrechte*.

45 Schweizerische Eidgenossenschaft (2013): *Bundesgesetz Friedensförderung und Menschenrechte*, Art. 5, S. 8206.

46 Davon sind 50 Mio. CHF (ca. € 41 Mio.) für ein Spezialprogramm in Nordafrika und den Mittleren Osten vorgesehen und 55% der verbleibenden 260 Mio. CHF (ca. € 213 Mio.) gehen in geografische Engagements. Der Rest ist ungebunden, um auf neue Entwicklungen eingehen zu können.



stabilisierenden Ländern ein.<sup>47</sup> Die DEZA hat demgegenüber kein gesetzlich festgelegtes Evaluierungsmandat, aber eine ausgeprägte Evaluierungskultur, die sich vor allem aus einem politischen Rechenschaftsbedarf gegenüber der Schweizer Bevölkerung und aus der Wirkungsdebatte in Fachkreisen speist.

Die offiziellen Evaluierungsmandate orientieren M&E in AMS und DEZA in Richtung Rechenschaftslegung. Formale Evaluierungen finden gegenwärtig auf Portfolio-, Programm- und Projektebene statt. Die AMS führt jährlich fünf bis zehn externe Evaluierungen sowie zehn bis zwanzig Selbstevaluierungen durch. Die Schwerpunkte liegen auf dem Nachweis von Wirkung sowie der angemessenen Berücksichtigung von Querschnittszielen (Gleichstellung von Frauen und Männern, Fokus auf besonders verwundbare Gruppen, etc.).<sup>48</sup> Jedoch gewinnen aufgrund der schlechten Erfahrungen mit der Einführung einer zentral organisierten, stark rechenschaftsorientierten Evaluierungsfunktion (siehe unten) in der AMS und der DEZA Alternativen zu formalen externen Evaluierungen an Aufmerksamkeit. Es geht zunehmend darum, die Qualität außenpolitischer Programme und Projekte in den Themenfeldern menschliche Sicherheit, Fragilität und humanitäre Hilfe durch Lernen zu erhöhen. In beiden Einheiten ist das Spannungsverhältnis zwischen lern- und rechenschaftsorientierten Evaluierungen spürbar: Während das rechtliche Mandat und die politischen Erwartungen von außen eine erhebliche Rechenschaftsorientierung vorgeben, beruht die interne Motivation der Mitarbeiterinnen und Mitarbeiter fast ausschließlich auf der Lernfunktion.

## Akteure

Die Anzahl der Akteure in der schweizerischen Außenpolitikevaluierung ist überschaubar. Politisch am bedeutsamsten ist die oben beschriebene *Parlamentarische Verwaltungskontrolle (PVK)*. Sie kann vom Parlament beauftragt werden zu überprüfen, ob die Konzeption staatlicher Erlasse sinnvoll ist und die damit verbundenen Maßnahmen des Bundes die erwünschten Resultate erzielen. Im Bereich Außenpolitik wurden zwischen 2006 und 2013 vier Evaluierungen von der PVK verantwortet. Diese betreffen die ressortübergreifende Zusammenarbeit im Bereich der Außenpolitik (2013), den Vorsitz der Schweiz im Ministerkomitee des Europarates (2011), die Kohärenz und strategische Führung der DEZA (2006) und die Aktivitäten von Pro Helvetia im Bereich der Auswärtigen Kulturpolitik (2006). Die Berichte dieser Evaluierungen sind öffentlich zugänglich und entsprechen methodisch internationalen Standards. Darüber hinaus verbinden sie Unabhängigkeit mit dem Wissen um politische und administrative Spezifika, da die PVK administrativ zwar der Geschäftsprüfungskommission des Parlaments unterstellt, inhaltlich aber unabhängig ist.<sup>49</sup>

Innerhalb des EDA ist in der AMS der Leiter Controlling für die inhaltliche Überprüfung der Arbeit zuständig, in der DEZA die Sektion Evaluierung und Controlling. Während der Leiter Controlling in der AMS als Karrierediplomat

47 Schweizerischer Bundesrat (2011): *Maßnahmen zur Förderung des Friedens*, S. 6338.

48 Schweizerischer Bundesrat (2012): *Internationale Zusammenarbeit*, S. 27.

49 Schweizer Parlament (2011): *Geschäftsprüfungskommissionen*.

nach einigen Jahren in eine andere Position rotiert, sind die Mitarbeiterinnen und Mitarbeiter der DEZA-Evaluierungseinheiten in der Regel langjährige Evaluierungsexperten. Das Generalsekretariat des EDA ist für die Wirksamkeit der Steuerungs- und Kontrollsysteme insgesamt zuständig. Diese Arbeitsteilung zwischen einer dezentral organisierten Evaluierungsfunktion und einer zentral organisierten Inspektionsfunktion gab es nicht immer. 2008 initiierte die damalige Bundesrätin für auswärtige Angelegenheiten, Micheline Calmy-Rey, eine Reorganisation des EDA. Sie wollte erreichen, dass die Schweiz nicht nur multilaterale Initiativen finanziert, sondern als einflussreicherer internationaler Akteur auftritt. Ein Eckpunkt dieser Agenda bestand darin, die Evaluierungsfunktionen der DEZA und der AMS gemeinsam mit dem Inspektorat beim Generalsekretariat (vergleichbar mit dem Leitungsbereich) in einem Evaluierungsstab zu bündeln. Dieser sollte über die bis dahin bestehende Praxis der Portfolio-, Programm- und Projektevaluierung hinaus nun auch Politikevaluierungen durchführen. Interne Kritikerinnen und Kritiker der Reform sahen die Reorganisation jedoch vor allem als Versuch, stärkere Kontrolle über das Haus zu gewinnen.

Dieser zentrale Evaluierungsstab im Generalsekretariat konnte nur wenige Evaluierungen durchführen, bevor er 2012 unter Calmy-Reys Nachfolger, Didier Burkhalter, als „zu bürokratisch“ abgeschafft wurde. Für Burkhalters liberale Partei, die FDP, ist Bürokratieabbau eine von drei politischen Prioritäten.<sup>50</sup> Neben diesem politischen Argument führte der Minister auch praktische Argumente gegen den Stab an, die innerhalb des EDA durchaus geteilt werden: Diplomatie- bzw. Politikevaluierung sei notgedrungen so abstrakt und damit methodisch so aufwändig, dass ihr Nutzen in keinem angemessenen Verhältnis zu den Kosten stehe – nicht zuletzt aufgrund des Steuerungsaufwandes für externe Evaluatorinnen und Evaluatoren. Angesichts der bestehenden Instrumente der PVK und des Mitberichtsverfahrens sei Politikevaluierung als Qualitätssicherungsmechanismus zudem redundant.<sup>51</sup> Die EDA-internen Befürworterinnen und Befürworter von Politikevaluierung begründen das Scheitern der Reform demgegenüber mit einer Ablehnungsreaktion innerhalb des Hauses gegen Qualitätskontrolle und die Einflussnahme durch die Inspektion. Den Gegnern sei es also weniger um die Ablehnung von Politikevaluierung gegangen als vielmehr um die Ablehnung von zentral organisierter Evaluierung insgesamt. Der Versuch, eine in der Inspektion aufgehängte, zentrale Evaluierungsfunktion mit Mandat für Politikevaluierungen einzuführen, gilt bei Anhängern wie Gegnern der Reform als gescheitert. 2012 kehrte man daher zu einer in der AMS und DEZA aufgehängten Evaluierungsstruktur mit Fokus auf Portfolio-, Programm- und Projektebene zurück.

50 FDP. Die Liberalen (o.J.): *Startseite* <http://www.fdp.ch/>.

51 Die PVK ist in der Bundesversammlung für Evaluierungen zuständig. Sie führt „Evaluationen über die Konzeption, Umsetzung und Wirkungen der Maßnahmen des Bundes durch.“ Schweizer Parlament (2013): *Parlamentarische Verwaltungskontrolle*. Beim Mitberichtsverfahren wird der von einem Departement erstellte Bericht an den Bundesrat den anderen Departementen und der Bundeskanzlei zur Stellungnahme vorgelegt. Bericht und Mitbericht werden separat an den Bundesrat gegeben, so dass eventuelle Meinungsunterschiede zwischen den Departementen sichtbar bleiben Schweizer Parlament (2013): *Parlamentswörterbuch*.

## Abläufe und Methoden

Die AMS und die DEZA arbeiten entlang der oben eingeführten „strategischen Kaskade“ von „Botschaft“, Strategie und Projektumsetzung. Entlang dieser Kaskade wird auch evaluiert: Greift die Strategie für eine bestimmte Aktivität die in der Botschaft formulierten Grundsätze auf? Spiegelt die Projektumsetzung die strategischen Ziele wider? Entsprechen die Projekte den Qualitätskriterien von OECD-DAC oder des *International Network on Conflict and Fragility (INCAF)*?

Grundsätze zu Anlage, Methoden und Umsetzung von Evaluierungen sind in beiden Einheiten in jeweils eigenen Evaluierungsrichtlinien festgehalten.<sup>52</sup> Die anderen Bereiche des EDA (die politische Direktion, die geografischen Einheiten usw.) geben sich zwar auch Ziele, haben aber keine vergleichbaren Finanzmittel zur Umsetzung dieser Ziele. Die Umsetzung wird auch nicht (notwendigerweise) mit anderen Abteilungen abgestimmt. Die Erfolgskontrolle erfolgt hier über Geschäftsberichte, die die Abteilungen selbst schreiben – also einer Art Selbstevaluierung. Diese Geschäftsberichte gehen an die Inspektion im Generalsekretariat.

Bezüglich M&E-Instrumenten und -Methoden lässt sich festhalten, dass sowohl DEZA als auch AMS internes Monitoring und freiwillige Selbstevaluierungen als Teil des Projektzyklus einsetzen. Die beiden Ansätze sind sehr ähnlich, da die direkt beteiligten Akteure die Wirkung, Fortschritte und Probleme ihrer Arbeit selbst einschätzen. Stärken-Schwächen-Analysen sind eine vor allem für Selbstevaluierungen bewährte Methode. Die AMS nutzt einen Fragenkatalog, anhand dessen Mitarbeiterinnen und Mitarbeiter Projekte selbst evaluieren können. Der Anspruch ist, hierbei gemäß der OECD-DAC-Standards vorzugehen.<sup>53</sup>

Was Evaluierungen betrifft, so werden formative Evaluierungen laut der Interviewten im EDA besser aufgenommen als ex-post-Wirkungsevaluierungen. Da die Wirkungsfrage vor allem in den Bereichen Stabilisierung, demokratische Transition und humanitäre Hilfe aufgrund methodischer Probleme und mangelnder Daten oft nicht abschließend oder eindeutig geklärt werden kann, fürchten Mitarbeiterinnen und Mitarbeiter, dem mehr oder weniger willkürlichen Urteil der Evaluatorinnen und Evaluatoren ausgesetzt zu sein. Diese Befürchtung untergräbt ihre Kooperationsbereitschaft. Aus dem gleichen Grund besteht zudem die verbreitete Erwartung, dass methodisch angreifbare Evaluierungsergebnisse leichter zum Spielball politischer Konflikte werden können, auch wenn es dazu bislang nicht gekommen ist.

Monitoring, Selbstevaluierung und formative Evaluierungen sind prozessbegleitend und erfordern daher eine gewisse Methodenkompetenz auf Seiten der Portfolio-, Programm- oder Projektverantwortlichen. Das heißt, M&E muss Teil der Diplomatenausbildung sein. In der Vergangenheit war dies in der Schweiz der Fall, in den letzten Jahren ist der Ausbildungsabschnitt zu M&E jedoch stark verkürzt worden.

52 Eidgenössisches Departement für auswärtige Angelegenheiten (2013): *Evaluation in der AMS*. Direktion für Entwicklung und Zusammenarbeit (2013): *Evaluationspolitik*.

53 Direktion für Entwicklung und Zusammenarbeit (2013): *Evaluationspolitik*, S.13f; Direktion für Entwicklung und Zusammenarbeit (2013): *Evaluationspolitik*, S.7; Strategisches Controlling Abteilung Menschliche Sicherheit (2013): *Selbstevaluation*.

Neben formalen externen Evaluierungen greift vor allem die DEZA zunehmend auf *peer reviews* zurück. Anders als in Großbritannien, wo *peer reviews* zur internen Bewertung von Projekt- und Programmserfolgen zum Einsatz kommen, werden *Peer-Review-Verfahren* in der Schweiz angewandt, um Länder- und Regionalstrategien weiterzuentwickeln. Dabei unterstützen Direktorinnen und Direktoren aus anderen Regionen (normalerweise Leiter von Auslandsvertretungen) in einem einwöchigen Workshop den Strategieprozess für ein bestimmtes Land bzw. Region. Die interne Debatte wurde durch dieses Verfahren sehr belebt, allerdings ist die Auswahl der *peers* heikel und es wurden hierfür eigens Kriterien entwickelt.<sup>54</sup> Dennoch kommt der Prozess bei den Beteiligten sehr gut an. Er wird durch einen externen Experten oder eine Expertin begleitet, der/die Vorbereitung, Moderation und Dokumentation übernimmt, um intern möglichst wenig Arbeitsaufwand zu erzeugen. Auch in anderen Teilen des EDA werden *peer reviews* zumindest als Möglichkeit betrachtet, um Denkprozesse anzustoßen. *Peer reviews* mit Kolleginnen und Kollegen aus anderen Ländern werden unter Schweizer Diplomatinen und Diplomaten auch geschätzt, vor allem wenn sie unter Chatham House Rules stattfinden.

Ein weiterer Ansatz, der in der Schweiz – ähnlich wie in Großbritannien – zunehmend an Unterstützung gewinnt, sind extern durchgeführte wissenschaftliche Studien mit der Frage „Was funktioniert und was funktioniert nicht?“. Dieses Format gilt als politisch weniger sensibel als externe Evaluierungen und wird von den einbezogenen Mitarbeiterinnen und Mitarbeitern als hilfreiches Lerninstrument betrachtet, um die eigene Effektivität zu erhöhen.

Mit Bezug auf fragile Kontexte wurden außerdem Konflikt- und Kontextanalysen sowie die Arbeit mit Szenarien als bedeutend hervorgehoben. In solchen Kontexten ist es häufig schwierig zu bemessen, ob die angestrebten Ziele erreicht wurden. Hinzu kommt, dass formale ex-post Evaluierungen häufig zu spät kommen, um die Qualität der Projekte während ihrer Laufzeit zu erhöhen. Hier helfen frühzeitige Kontext- und Konfliktanalysen sicherzustellen, dass das Engagement zumindest keine negativen Auswirkungen hat.

Schließlich gab es einige Befürworter von gemischten und gemeinsamen Evaluierungen. AMS und DEZA führten kürzlich gemeinsame Evaluierungen durch. Hier war ein besonderer Lerneffekt, dass die Mitarbeiterinnen und Mitarbeiter die jeweils unterschiedlichen Arbeitsweisen in der Entwicklungszusammenarbeit und Diplomatie verstehen lernten. Die DEZA-Angehörigen hätten gesehen, was es heißt, mit Personen und nicht mit Geld zu arbeiten; die AMS-Angehörigen hätten demgegenüber die Projekte besser kennengelernt und könnten die Notwendigkeit für Expertentum nun besser nachvollziehen. Die Schweizer Evaluierung zur Fragilität war im Gegensatz dazu – wie im Pflichtenheft vorgesehen – nur auf die DEZA bezogen. Andere Teile des EDA waren weder in die Evaluierung involviert noch von ihr betroffen – so konnte sie auch nicht genutzt werden, um das Thema Fragilität stärker auf der Leitungsebene zu verankern. Dies musste im Nachhinein geschehen.

54 Direktion für Entwicklung und Zusammenarbeit (2012): *Evaluation Pays*; Direktion für Entwicklung und Zusammenarbeit (2012): *Evaluation Pays*.

## Einbettung in eine professionelle Projektmanagement- und Lernkultur

Alle in der Schweiz interviewten Evaluierungsexpertinnen und -experten waren sich einig, dass M&E ohne professionelle Strategie- und Projektmanagementprozesse nicht funktionieren kann. Entsprechend stehen in AMS und DEZA Strategie-, Projektmanagement- und Evaluierungsprozesse in enger Verbindung. Hier geht es vor allem darum, Zielformulierungen und Zielerreichung miteinander vergleichen zu können.

Laut der Interviewten sind dabei mittelfristig geltende Konzepte zentral, die von der Leitung gegen tagesaktuelle Schwankungen abgeschirmt würden. Es sei schon anspruchsvoll genug, die Ziele der Maßnahmen an die Kontextveränderungen vor Ort anzupassen; wenn sie dazu auch noch kurzfristig an strategische Veränderungen angepasst werden müssen, sei eine Zielüberprüfung kaum mehr möglich. Eine weitere Voraussetzung, um PM&E gewinnbringend zu verzahnen seien außerdem wenig bürokratische Evaluierungsprozesse und professionelles Projektmanagement. Die Interviewten betonten, dass PM&E das Lernen innerhalb der Außenpolitik nur dann möglich mache, wenn die entsprechenden Instrumente auf die spezifische Art des Lernens im EDA angepasst seien. Diplomatinen und Diplomaten seien durch die Rotation im Vergleich zu DEZA-Mitarbeiterinnen und -Mitarbeitern stärker in der Gegenwart ihrer jeweiligen Aufgabe verankert und die themenspezifische Rück- und Vorschau müsse daher institutionell unterstützt werden.

## Wesentliche Erkenntnisse

1. Die Einführung einer zentralen Evaluierungsfunktion mit dem zusätzlichen Anspruch der Politikevaluierung im Außenministerium ist in der Schweiz gescheitert. Der Aufwand für institutionelle Veränderungen und zusätzliche bürokratische Prozesse war im Vergleich zur Verwendbarkeit der Ergebnisse deutlich zu hoch. Darüber hinaus betrachtet die aktuelle politische Führung des EDA die interne Politikevaluierung als redundant gegenüber dem Evaluierungsauftrag der PVK. Nicht zuletzt scheiterte die Reform aber auch am internen Widerstand gegen die Zentralisierung der Evaluierung in der Inspektion, der von Seiten vieler Mitarbeiterinnen und Mitarbeiter die Glaubwürdigkeit für eine inhaltliche Qualitätskontrolle abgesprochen wurde.
2. Die Schweizer Erfahrung mit M&E zeigt außerdem, dass zumindest im konsensorientierten schweizerischen Politikgeschäft ein Widerspruch entsteht zwischen der öffentlichen Nachfrage nach rechenschaftsorientierter Evaluierung und dem internen Bedürfnis nach lernorientierter Verbesserung der eigenen Arbeit durch Selbst- und Gemeinschaftsevaluierungen. Beide Funktionen – Lernen und Rechenschaftspflicht/Kontrolle – nicht in einer Einheit zusammen zu führen, wäre laut der befragten DEZA-Mitarbeiter daher besser gewesen. Kontrolle hätte bei der Inspektion liegen sollen und Lernen bei den fachlich zuständigen Abteilungen.

---

### Kurzüberblick anhand der Leitfragen (S. 11)

---

Warum?	Evaluierung hat Verfassungsrang; Parlamentsauftrag
Schwerpunktsetzung	Rechenschaft und Lernen; heute mit Schwerpunkt Lernen (seit die investigative rechenschaftsorientierte Evaluierung politisch und praktisch gescheitert ist)
Wie?	Rechenschaft und Lernen werden teils mit denselben, teils mit unterschiedlichen Instrumenten verfolgt – Mischung gilt als problematisch
Was?	<ul style="list-style-type: none"><li>• In der Vergangenheit: Politik – wurde im Außenministerium wieder abgeschafft, erfolgt nur noch von Parlamentsseite (Parlamentarische Verwaltungskontrolle).</li><li>• Aktuell: Portfolios, Programme, größere Projekte werden formell extern evaluiert und zusätzlich internen Selbstevaluierungen unterzogen; Peer Reviews tragen zur Strategiebildung bei.</li></ul>
Wer? (spezifisch für die außenpolitischen Portfolios)	<ul style="list-style-type: none"><li>• Abt. Menschliche Sicherheit (AMS) und Entwicklungsabteilung (DEZA, verantwortlich u.a. für humanitäre Hilfe) haben jeweils eigene Evaluationsverantwortliche</li><li>• Externe Evaluierungen durch Gutachterinnen und Gutachter</li></ul>
Wie transparent?	Alle formellen Evaluierungen sind öffentlich; Selbstevaluierungen bleiben intern
Projektmanagement	In AMS und DEZA sind M&E in Projektmanagement und Zielformulierung eingebettet

---

3. Ohne eine enge Verknüpfung mit professionellen Strategie- und Projektmanagementprozessen kann Monitoring und Evaluierung nicht greifen. Entscheidend dafür sind die Investition in die Formulierung erreichbarer und evaluierbarer Ziele, deren Abschirmung von tagesaktuellen Schwankungen sowie die systematische institutionelle Unterstützung von Lernen in einer Weise, die auf den Kontext von Außenpolitik abgestimmt ist.
4. Ein Beispiel für ein solches Lerninstrument, mit dem das Schweizer Außenministerium gute Erfahrungen gemacht hat, sind *peer reviews*. Im Schweizer Modell kommt es auf die strategische Zusammensetzung der *peer group* sowie eine zukunftsgerichtete Fragestellung an. Deshalb wird das Instrument nicht zur rückblickenden Bewertung sondern vor allem zur Weiterentwicklung von Länder- und Regionalstrategien eingesetzt.

In den USA steht die außenpolitische Evaluierung noch ganz am Anfang. Basierend auf den Erfahrungen der Entwicklungspolitik nimmt sie ihren Ausgangspunkt in der dort vorherrschenden Ernüchterung über die Qualität der Erkenntnisse und die Politikrelevanz der meisten Evaluierungen in fragilen Kontexten. Die Qualitäts-, Kriterien- und Methodendiskussion hat deshalb auch in der außenpolitischen Evaluierungsdebatte der USA einen besonderen Stellenwert. Die besonderen Herausforderungen fragiler Kontexte und außenpolitischer Maßnahmen sind damit weithin anerkannt, wie auch der Zielkonflikt zwischen Rechenschafts- und Lernorientierung. Dies hat zur Folge, dass der M&E-Werkzeugkasten derzeit in zwei Teile zerfällt: einerseits dezentral organisierte, maßgeblich lernorientierte Monitoringinstrumente und andererseits der Versuch, zentral gesteuerte Evaluierungen mit Rechenschaftsfunktion zu etablieren und dafür an den außenpolitischen Bedarf angepasste Instrumente zu schaffen.

Die Zuständigkeiten für die relevanten außenpolitischen Programmportfolios – humanitäre Hilfe, Stabilisierung, Krisenprävention, demokratische Transformation – sind zudem zersplittert: Humanitäre Hilfe liegt bei der Entwicklungsbehörde United States Agency for International Development (USAID) im *Office of U.S. Foreign Disaster Assistance (OFDA)*, in Stabilisierung und ziviler Krisenprävention tummeln sich das Pentagon, USAID mit dem *Office of Transition Initiatives (OTI)* und das *Bureau of Conflict and Stabilization Operations (CSO)* des Außenministeriums. In der demokratischen Transformation spielen sowohl bei USAID als auch im *State Department* mehrere Unterabteilungen eine Rolle.

## Zweck und Mandat

Evaluierung in der US-Außenpolitik kam 2009 auf Druck der Haushaltsbehörde des Weißen Hauses (*Office of Management and Budget*) auf den Tisch<sup>55</sup> und fand ihren Niederschlag in der ersten *Quadrennial Diplomacy and Development Review (QDDR)* 2010.<sup>56</sup> Daraufhin begann das State Department mit der Entwicklung von Evaluierungsrichtlinien, die 2012 verabschiedet wurden. Aufgrund fehlender eigener Expertise wurden zwei erfahrene Evaluierungsexperten von USAID rekrutiert, um diese zu entwickeln. In dieser *Program Evaluation Policy* beschreibt das *State Department* den Zweck von außenpolitischen Evaluierungen sowohl als Rechenschaftslegung als auch als Lerninstrument um zukünftige Aktivitäten wirksamer zu gestalten.<sup>57</sup> Ende 2013 lagen noch keine abgeschlossenen Evaluierungen nach diesen neuen Richtlinien

55 Blue, Clapp-Wincek und Benner (2009): *Beyond Success Stories*, S.15.

56 US Department of State (2010): *Diplomacy and Development Review*, S.197.

57 US Department of State (2010): *Diplomacy and Development Review*, S.2.

vor; mit ihnen ist erst für die kommenden Jahre zu rechnen (jedes der 40 *Bureaus* – etwa: Unterabteilungen – soll bis Anfang 2017 mindestens eine umfangreiche Evaluierung abschließen).<sup>58</sup> Die Verzögerung hängt auch damit zusammen, dass der anhaltende Haushaltsstreit und der damit verbundene undifferenzierte Sparzwang (*sequestration*) so verheerende Anreize gegen konstruktiv-kritische Evaluierung gesetzt hat, dass eine sinnvolle Umsetzung der Richtlinien nach Einschätzung des Rechnungshofes zumindest bis Dezember 2013 nicht möglich war.

Im Rahmen der umfassenden Verwaltungsaufsicht des *Deputy Secretary of State* wurde den fachlichen Einheiten jeweils ein dezentrales Mandat zur Evaluierung ihrer „großen“ Projekte und Programme erteilt:

„The Department of State requires that all large programs, projects, and activities be evaluated at least once in their lifetime or every five years, whichever is less. Bureaus will determine which programs, projects or activities to evaluate; upon request [the *Bureau of Resource Management, RM*] and [the *Office of U.S. Foreign Assistance Resources, F*] will assist in this effort.“<sup>59</sup>

Der Schwerpunkt der Evaluierungsbemühungen liegt dementsprechend auf einzelnen größeren Programmen und Projekten – umfassende Politikevaluierung spielt für das *State Department* keine Rolle. Es werden von Evaluierung keine bedeutenden und überraschenden Erkenntnisse für die konkrete Gestaltung von Projekten und Programmen erwartet – im Gegenteil, die langjährige Erfahrung aus der Entwicklungspolitik sei, dass Evaluierungsergebnisse selten direkt zu Veränderungen führten. Ein erfahrener Experte beschrieb ihren Einfluss als langfristig und kumulativ: Viele Evaluierungen beeinflussen die großen Paradigmen und Wirkungshypothesen, die über Jahre und Jahrzehnte entwicklungs- und zunehmend auch außenpolitische Programme prägen.

Während die tatsächliche Erfahrung mit außenpolitischer Evaluierung insofern noch auf sich warten lässt, besteht auf der Ebene einzelner mittelführender Einheiten wie dem *Bureau of Conflict and Stabilization Operations (CSO)* erhebliche Experimentierfreude auf dem Gebiet lernorientierter Monitoringinstrumente. Zusätzlich zu den geplanten Evaluierungsinstrumenten werden auch diese im Folgenden etwas näher betrachtet.

## Akteure

Die Evaluierungserfahrung der verschiedenen dezentralen Facheinheiten variiert erheblich: Die zivilen Stabilisierungsaktivitäten des Pentagon unterliegen keiner professionellen Evaluierung, im *State Department* gibt es meist keine oder minimale Kapazitäten für Evaluierung, während im entwicklungsnahe Bereich

58 Die Richtlinien selbst definieren deutlich weiter gehende Anforderungen (2-4 Evaluierungen bis Anfang 2014, Einbezug der Außenvertretungen in die Evaluierungspflicht ab 2013): US Department of State (2012): *Program Evaluation Policy*, S.2. Nach Angaben unserer Gesprächspartner scheint dies in der Praxis lockerer gehandhabt zu werden.

59 US Department of State (2012): *Program Evaluation Policy*, S.2.



– OFDA und OTI – seit Jahren eine umfangreiche Evaluierungspraxis besteht. Im Verantwortungsbereich des Außenministeriums soll die Praxis nun auf Grundlage der neuen Richtlinien von 2012 systematisiert werden. Dabei spielen zwei Arbeitseinheiten eine zentrale (und konkurrierende) Rolle.

Das *Bureau of Resource Management (RM)* und das *Office of U.S. Foreign Assistance Resources (F)* haben im Auftrag des *Deputy Secretary of State* eine Überwachungs- und fachliche Beratungsfunktion für die Durchführung von Evaluierung, ohne über realistische Sanktionsmechanismen gegenüber den Facheinheiten zu verfügen. Letztlich beruht die (zukünftige) Evaluierungspraxis also auf dem Eigeninteresse der fachlich Verantwortlichen in Verbindung mit der Autorität des *Deputy Secretary*. Diese politische Rückendeckung von ganz oben ist für den Erfolg entscheidend – und ob sie in hinreichendem Maß praktiziert werden wird, ist noch nicht abzusehen. Die Evaluierungsexperten bei F versuchen bislang, das Thema durch Weiterbildungsangebote, die Bereitstellung von Personalressourcen zur dezentralen Rekrutierung von Evaluierungsreferenten in den betroffenen Abteilungen sowie Informationsangebote im Intranet besser im Haus zu verankern und das professionelle Interesse der Kolleginnen an Evaluierung zu nähren. Diese kleinen Schritte werden wahrgenommen: Auch der ansonsten sehr kritische Rechnungshof berichtet von wachsender Aufmerksamkeit für Evaluierung im Außenministerium.

Anders als das Instrument der Evaluierung besteht weder für Monitoringwerkzeuge noch für die Lernpraxis ein Anspruch der zentralen Steuerung im US-Außenministerium. Einzelne Abteilungen schaffen sich hier ihre eigenen Instrumente und Abläufe. Eine davon ist das *Bureau of Conflict and Stabilization Operations (CSO)*. Dessen *Office of Learning and Evaluation* verfügt über etwa ein Dutzend Mitarbeiterinnen und Mitarbeiter, die zum Großteil auf die CSO-Länderteams verteilt wurden und physisch dort arbeiten. Diese Einbindung in die tägliche Arbeit der internen „Kunden“ dient dazu, die natürliche Entfremdung einer reinen Analyseeinheit von der operativen Arbeit zu verhindern. Die überwältigende Mehrheit dieser Mitarbeiterinnen und Mitarbeiter stammt aus dem Auswärtigen Dienst. Bei der Bildung des Lern- und Evaluierungsreferates konnte der erste Leiter alle Mitarbeiterinnen und Mitarbeiter nach eigenen Kriterien auswählen; das Kriterium „operativer Glaubwürdigkeit“ gegenüber Kolleginnen und Kollegen spielte dabei eine ebenso große Rolle wie das persönliche Interesse und die Fähigkeit zu analytischer Arbeit.

## Abläufe und Methoden

Über Abläufe und Methoden für Evaluierung in den relevanten Feldern liegen nach den jüngsten Veränderungen noch kaum praktische Erfahrungen vor. Ohnehin ist der Projektumfang im außenpolitischen Bereich meist zu klein für das klassische entwicklungspolitische Evaluierungsinstrumentarium, dessen Anpassung an die Bedingungen der Außenpolitik Ende 2013 erst am Anfang stand. Ausgangspunkt ist die Beobachtung „*We cannot evaluate diplomacy, but we can evaluate programs with regard to diplomatic concerns.*“<sup>60</sup>

60 Interview, Mitarbeiter des *State Department*, Washington, November 2013.

Um „Programme im Hinblick auf ihre außenpolitische Dimension“ zu evaluieren, seien zwei grundlegende methodische Hürden zu meistern. So verfolgten viele außenpolitische Maßnahmen unklare und unzureichend dokumentierte, teilweise vollkommen unrealistische Ziele (z.B. „*disrupt narcotics trade in Afghanistan*“), die nach Aktenlage zunächst nicht evaluierbar seien. Dies ist jedoch methodisch lösbar: Evaluatorinnen und Evaluatoren seien in der Lage, aufgrund von Dokumenten und Interviews die unterschiedlichen Zwecke und Prioritäten außenpolitischer Maßnahmen zu rekonstruieren. Diese Rekonstruktion erfordere allerdings möglichst vollständigen Zugang zu sensiblen, teils vertraulichen Dokumenten. Wo dies nicht möglich sei bzw. wo die Maßnahmen ziviler Akteure von viel umfangreicheren geheimdienstlichen Programmen überschattet würden, nütze es wenig, die zivilen Programme allein zu evaluieren. Die Erfassung multipler Ziele und das notwendige Kontextverständnis führen dazu, dass Außenpolitikevaluierung gemischte Teams aus organisationsexternen Evaluatorinnen und Evaluatoren einerseits und Diplomatinen und Diplomaten andererseits erfordert. Zumindest so lange es im Privatsektor keine ausdifferenzierte „außenpolitische Evaluierungsbranche“ gäbe, liefen rein externe Evaluationen sonst Gefahr, die politischen Belange außenpolitischer Projekte und die Berichts- und Entscheidungswege der Diplomatie nicht hinreichend zu erfassen.

In Bezug auf Evaluierungsmethoden besteht bei allen Beteiligten (von den Programmverantwortlichen über Evaluierungsexpertinnen und -experten bis zum Rechnungshof und Kongress) Einigkeit: Der Anspruch muss sein, systematische, unvoreingenommene Untersuchungen sicherzustellen. Dies erfordert methodische Vielfalt und Flexibilität statt der künstlichen Versteifung auf eine bestimmte positivistische Erkenntnistheorie und die damit verbundene Erwartung, kausale Wirkungen der eigenen Intervention exakt zu erfassen. In fragilen Kontexten mag es einzelne (eher langfristige) Programme geben, die auf diese Weise evaluierbar sind, doch dies ist nach dem Konsens der US-Praktiker nicht die Regel, sondern nur die Ausnahme. Gesprächspartner beim Rechnungshof und im Kongress begrüßen explizit den methodischen Spielraum, den sich Außenministerium und USAID in ihren jüngsten Evaluierungsrichtlinien gegeben haben.

Dieser methodologische Konsens entspricht allerdings nicht der bisherigen Praxis: Professionelle Evaluatorinnen und Evaluatoren sehen sich unter Reputationsdruck, mit quantitativen Methoden und dem damit verbundenen Datenerhebungsaufwand zumindest den Anschein zu erwecken, Aussagen auf robuster statistischer Basis zu treffen. Beteiligte erklären diesen Druck aus einer Kombination von Vertrautheits- und Modeeffekten und der strukturellen Macht der Arbeitsebene in der Haushaltsverwaltung, die quantifizierte Begründungen einfordert, um eine illusionäre Vergleichbarkeit über Politikfelder hinweg zu schaffen. Diese Dynamik führe zu einer erheblichen Zahl methodisch falscher und auch politisch unnötiger Evaluierungen – denn im Kongress besteht durchaus Verständnis für die Unwägbarkeiten der Projektimplementierung und Wirkungsmessung in fragilen Kontexten. Haushaltspolitisch entscheidend ist aus Sicht von Kongressmitarbeiterinnen und -mitarbeitern jedoch, dass die Methodik einer Evaluierung transparent und in ihrer Systematik nachvollziehbar sei, um den Vorwurf anekdotischer Argumentation nicht aufkommen zu lassen.

Während also die Evaluierungspraxis erst noch den besonderen Anforderungen außenpolitischer Maßnahmen in fragilen Kontexten angepasst werden soll, bieten die USA instruktive Beispiele für innovative Ansätze zum laufenden Monitoring.

In einer Reihe von Organisationseinheiten (z.B. CSO, OTI) liegt sogar ein bewusster strategischer Schwerpunkt auf der Verbesserung kontinuierlicher, möglichst schneller Lernprozesse durch Echtzeit-Monitoring, während ex-post-Evaluierung konzeptionell nur eine Nebenrolle spielt.

Bei CSO ist das zentrale Instrument der Monitoring-Plan: Eine übersichtliche Matrix von Indikatoren für wünschenswerte und schädliche Entwicklungen mit Bezug auf die Ziele einer Maßnahme, die aus einer Mischung von internen und externen unabhängigen Quellen mit wenig Aufwand und in kurzen Zeitabständen zu verfolgen sind. Zu den Datenquellen gehört die Berichterstattung der Auslandsvertretungen ebenso wie die der Geheimdienste, aber auch innovative *Big Data*-Werkzeuge zur Auswertung öffentlich zugänglicher internationaler Medien wie ICEWS (*Integrated Crisis Early Warning System*).<sup>61</sup> Dieser Monitoring-Ansatz überwacht absichtlich keine Fortschritte in der Projektumsetzung oder deren Wirkungen, sondern ausschließlich Kontextveränderungen, die eine Anpassung der Programme oder Projekte erfordern könnten.

Als Beispiel dient eine Maßnahme zur Prävention von Gräueltaten in Nigeria: Aus der Analyse wahrscheinlicher Ursachen für extreme Gewaltausbrüche wurde ein Programmziel formuliert (Förderung konstruktiver Beziehungen zwischen Staat und Gesellschaft), das einerseits mit einer Reihe konkreter Aktivitäten verfolgt wird (Medienarbeit, Unterstützung für zivilgesellschaftliche Organisationen, diplomatischer Druck auf die Regierung) und andererseits mit einer Mischung regulärer Analyseinstrumente und zusätzlicher Umfrage- und Fokusgruppensdaten zu beobachten ist. Der Beobachtungsaufwand erfüllt dabei eine Doppelfunktion: Er dient sowohl der Projektsteuerung und -kontrolle als auch der Lagebeobachtung insgesamt, um schnelle Reaktionen auf potentiell gefährliche Veränderungen zu ermöglichen. Aufgrund der Erhebung eigener Daten (die als Kapazitätsaufbaumaßnahme gemeinsam mit Meinungsforschungsinstituten und Universitäten in Nigeria erfolgt) sind die Kosten der Echtzeitbeobachtung in diesem Fall mit 8-10% des Projektumfangs relativ hoch.

Über die Echtzeitebene hinaus will CSO zukünftig mit zweierlei Werkzeugen arbeiten: Reviews und Evaluierungen. *Evaluierungen* fällt dabei die Rolle zu, den Ansprüchen an größtmögliche Präzision, Unabhängigkeit und Transparenz (volle Veröffentlichung) zu genügen. Sie sollen nur selten und gezielt eingesetzt werden. *Reviews* werden demgegenüber als organisationsinterne, vertrauliche Werkzeuge der Begutachtung und Fortschrittskontrolle verstanden, deren Ergebnisse nur in zusammenfassender Form der Öffentlichkeit zugänglich gemacht werden sollen.

61 Ursprünglich von der militärischen Forschungs- und Entwicklungsbehörde Defense Advanced Research Projects Agency (DARPA) finanziert und von Lockheed Martin entwickelt, siehe Lockheed Martin (o.J.): *Worldwide Integrated Crisis Early Warning System*. Frühe kritische Einordnung von Nutzen und Verlässlichkeit v.a. prädiktiver Anwendungen bei O'Brian (2010): *Crisis Early Warning and Decision Support*.

## Einbettung in eine professionelle Projektmanagement- und Lernkultur

Klassische, externe Evaluierungen rufen wie in der Schweiz und Großbritannien auch in den USA einen Abwehrreflex der Projektverantwortlichen und ihrer Organisationen hervor. Das Extrembeispiel hier ist das „Modell externer Überwachung“ der Sonderinspektoren für Irak und Afghanistan, mit dem weder Lern- und Evaluierungsexperten noch die Beamtinnen und Beamten des Rechnungshofs oder die Ermittlerinnen und Ermittler des *Special Inspector-General for Afghanistan Reconstruction (SIGAR)* gute Erfahrungen gemacht haben. Zwar könne man mit den notwendigen gesetzlichen Bestimmungen den Widerstand gegen die Herausgabe von Informationen brechen. Damit habe man jedoch das Klima derart vergiftet, dass die Ermittlungsergebnisse der Sondergeneralinspektoren über disziplinarische bzw. im Extremfall strafrechtliche Folgen für die Vergehen Einzelner hinaus keine Bedeutung für die Entscheidungsfindung in den betroffenen Ministerien hätten – und das trotz hoher politischer Aufmerksamkeit im Kongress. In diesem Modell sei keine kritisch-konstruktive Zusammenarbeit möglich.

Intern gesteuerte Evaluierungen werden demgegenüber nicht nur im Außenministerium und bei USAID, sondern auch von Seiten des Rechnungshofes als glaubwürdig akzeptiert. Voraussetzung dafür sei die volle Veröffentlichung nicht nur der Ergebnisse, sondern auch der Evaluierungsfragen und -methoden. Deren externe Überprüfbarkeit schaffe auch für interne, partizipativ statt investigativ arbeitende Evaluierungsakteure die volle Glaubwürdigkeit in der fachlichen und politischen Öffentlichkeit. Gleichzeitig ermöglichten partizipative Evaluierungen nützlichere Ergebnisse. Im Widerspruch zu einer verbreiteten Annahme sei ein „auskömmliches, [wenn auch] nicht zu gemütliches“ Vertrauensverhältnis zwischen Evaluatorinnen und Evaluatoren einerseits und Programmverantwortlichen andererseits nicht schädlich, sondern notwendig. Es sei ebenso wichtig für das Gelingen einer Evaluierung wie die politische und intellektuelle Unabhängigkeit der Evaluatorinnen und Evaluatoren.

Kulturelle Lernorientierung und Fehlertoleranz werden in den USA auf mehreren Wegen aktiv gefördert. Es ist eine Führungsaufgabe, Lernprozesse und -effekte aktiv nachzufragen und damit Anreize dafür zu schaffen sowie Mitarbeiter vor unangemessenen Konsequenzen identifizierter Fehler zu schützen: Es darf keine Sündenböcke geben. Zumindest auf dem Gebiet der Stabilisierung profitiert eine solche Lern- und Fehlerkultur von zunehmend realistischeren Erwartungen über den Einfluss außenpolitischer Maßnahmen. Darüber hinaus versuchen die M&E-Expertinnen und Experten, stets nachfrageorientiert zu arbeiten. Es gelte bedarfsorientierte, praktisch nützliche Angebote zu machen, die den Diplomatinen und Diplomaten bei der Maßnahmensteuerung helfen – und zwar von Anfang an. Gerade in der Startphase von Projekten, Programmen und Portfolios werden mit Zieldefinitionen die Grundlagen für eine sinnvolle Verwendung von M&E-Werkzeugen gelegt. Damit werde M&E als Teil dieses Kerngeschäfts etabliert und nicht als „Nebensache“ ständig von dringenderen, vermeintlich wichtigeren Aufgaben verdrängt.

# Wesentliche Erkenntnisse

## Kurzüberblick anhand der Leitfragen (S. 14)

Warum?	Seit 2009: zunehmender Druck aus der Finanzpolitik; eigenes Mandat des Außenministeriums
Schwerpunktsetzung	Rechenschaft und Lernen; mit unterschiedlichen Instrumenten
Wie?	<ul style="list-style-type: none"> <li>• Anpassung der Instrumente an die Besonderheiten außenpolitischer Maßnahmen in fragilen Kontexten noch nicht abgeschlossen</li> <li>• Verbreitete Rechenschaftsorientierung schafft Fehlanreize (deshalb viele methodisch unangemessene Evaluierungen), gleichzeitig zunehmend Raum für lernorientierte Instrumente</li> <li>• Praxisschwerpunkt auf Monitoring-Instrumente (CSO)</li> </ul>
Was?	<ul style="list-style-type: none"> <li>• State: Programme und größere Projekte sollen in Zukunft formell extern evaluiert werden</li> <li>• AID: Portfolios, Programme, größere Projekte werden formell extern evaluiert</li> </ul>
Wer? (spezifisch für die außenpolitischen Portfolios)	<ul style="list-style-type: none"> <li>• Evaluierung: Dezentral in den einzelnen Bureaus des State Department mit Unterstützung zentraler Kompetenzzentren (Haushaltsabteilung, Director of Foreign Assistance Resources), analog bei USAID (Stabilisierung, humanitäre Hilfe)</li> <li>• Externe Evaluierungen durch Gutachterinnen und Gutachter</li> <li>• Monitoring: Dezentral autonom</li> </ul>
Wie transparent?	<ul style="list-style-type: none"> <li>• State Department: Einzelfallentscheidung bei externen Evaluierungen</li> <li>• USAID: externe Evaluierungen werden veröffentlicht</li> </ul>
Projektmanagement	M&E sind in aufwändige Projektmanagement- und Zielorientierungsverfahren eingebettet

1. M&E in fragilen Kontexten überfordert klassische Evaluierungsmethoden, die in anderen Kontexten durchaus wissenschaftliche Anerkennung genießen. M&E-Methoden sind unter diesen speziellen Bedingungen nur dann wissenschaftlich valide und nützlich für die Praxis, wenn sie in der Lage sind, ständige Anpassungen der Maßnahmenziele an volatile politische Rahmenbedingungen nachzuvollziehen und ihrer Beurteilung zugrunde zu legen. Deshalb besteht ein erhebliches Qualitätsdefizit bei Evaluierungen in diesem Umfeld. Dieses Qualitätsdefizit wird noch verschärft durch Fehlanreize zugunsten quantitativer Evaluierungen in Bereichen, wo solche Methoden wenig belastbar sind.
2. Die Glaubwürdigkeit von Evaluierungsinstrumenten in den Augen der politischen Öffentlichkeit (Medien und Parlament) entsteht nur über den Umweg der Fachöffentlichkeit, das heißt durch Evaluierungs- und Außenpolitikexperten, die Abläufe, Methoden und Ergebnisse von Evaluierungen beurteilen können.

Hier besteht in den USA für den Bereich außenpolitische Maßnahmen in fragilen Kontexten (wie auch in der Schweiz und Großbritannien) Verständnis für notwendige Anpassungen der Kriterien und Methoden: Sowohl was politische (Teil-)Ziele als auch was fragile Kontexte betrifft, befindet sich die amerikanische Evaluierungsbranche in einer Umbruchphase hin zu perzeptionsbasierten, weniger wirkungsorientierten Ansätzen.

3. Unabhängig von den methodischen Herausforderungen verdeutlicht die Monitoring- und Lernpraxis im US-Außenministerium (CSO) die Effektivität partizipativer statt investigativer Prozesse. Dank konsequenter Nachfrageorientierung sind M&E- und Lernexperten in der Lage, einen spürbaren Mehrwert für ihre Kolleginnen und Kollegen in den operativen Arbeitseinheiten zu leisten. Dadurch wird M&E zunehmend, wenn auch mühsam, im Kerngeschäft verankert.
4. Begleitende Monitoring-Maßnahmen, die den Schwerpunkt auf den politischen Kontext (statt auf den Fortschritt der Projektimplementierung) legen, haben sowohl im Außenministerium als auch bei den USAID-Stabilisierungsmaßnahmen (OTI) einen höheren Stellenwert als ex-post-Evaluierungen. Das liegt daran, dass intensive und zielgenaue Monitoring-Instrumente kurzfristig helfen, notwendige Anpassungen bis hin zur vorzeitigen Beendigung von Projekten rechtzeitig zu erkennen und damit nicht nur einen Beitrag zum organisationsweiten Lernen leisten, sondern darüber hinaus einen direkten Nutzen für die laufende Projektsteuerung haben, der von ex-post-Evaluierung nicht zu erwarten ist.

# Niederlande

Die niederländische Evaluierungspraxis versucht, Rechenschafts- und Lernelemente so eng wie möglich zu verbinden. Durch volle Offenlegung der Berichte und die zentrale Steuerung bzw. Unterstützung aus dem „Kontrollbereich“ haben außenpolitische Evaluierungen in den Niederlanden einen starken Rechenschaftscharakter; gleichzeitig sind sie partizipativ organisiert und für die Umsetzung von Empfehlungen auf den Lernwillen der Beteiligten angewiesen. Aus der längsten Erfahrung der betrachteten Fälle hat sich zudem eine recht enge Verzahnung von M&E-Werkzeugen mit dem Projektmanagement entwickelt. Wie in der Schweiz und den USA helfen auch in den Niederlanden Pragmatismus und Experimentierfreude bei der Auswahl geeigneter Methoden zur Datenerhebung, den spezifischen Herausforderungen außenpolitischer Maßnahmen in fragilen Kontexten gerecht zu werden.

Seit der jüngsten Reorganisation des integrierten Ministeriums für Auswärtige Angelegenheiten, Außenhandel und Entwicklung (*Ministerie van Buitenlandse Zaken, MBZ*) 2012 sind alle relevanten außenpolitischen Portfolios – humanitäre Hilfe, Stabilisierung, zivile Krisenprävention und demokratische Transformation – in einem *Direktorat Stabilität und Humanitäre Hilfe* (DSH, vergleichbar einer Unterabteilung) vereint. Das Ministerium arbeitet unter zwei Ministern, deren Personal durch alle Teilbereiche rotiert. Die Ministerin für Handel und Entwicklung ist dem Außenminister zwar protokollarisch untergeordnet, handelt aber praktisch – auch aufgrund des politischen Gewichts der derzeitigen Amtsinhaberin Lillianne Ploumen (bis zur letzten Wahl die Parteivorsitzende der Arbeitspartei) – eigenständig und gleichrangig. DSH ist sowohl dem politischen als auch dem entwicklungspolitischen *Directeur-generaal* verantwortlich, praktisch jedoch enger an die entwicklungspolitische Seite des Hauses angebunden.<sup>62</sup>

## Zweck und Mandat

Evaluierung soll die Außen- und Entwicklungsminister darin unterstützen, vor dem Parlament Rechenschaft über ihre konkrete Politik und Ressourcenzuteilung abzulegen. Daneben hat Evaluierung auch eine Lernfunktion und soll zu einer effektiveren Formulierung und Umsetzung zukünftiger Politikentscheidungen beitragen.<sup>63</sup> Diese offizielle Sicht spiegelt sich auch in der Praxis: Aus der Sicht von Programmverantwortlichen kostet eine gute Evaluierung zunächst vor allem Geld und Arbeitszeit. Diese Investitionen sei eine Evaluierung aber durchaus wert, biete sie doch eine einzigartige Gelegenheit, sowohl intern (im eigenen Entscheidungsbereich) als auch extern bei Partnern und (multilateralen) Mittelempfängern notwendige

<sup>62</sup> Siehe auch Rotmann und Steinacker (2013): *Stabilisierung*.

<sup>63</sup> IOB (o.J.): *About IOB*.

Veränderungen anzustoßen, wo die Niederlande sonst als kleiner oder mittlerer Geber keinen großen Einfluss hat: „*As a bureaucrat, I have only a few moments to make changes to programs: at the very beginning of a new program, and between pre-defined phases. Evaluations give me additional opportunities for change at flexible points in time.*“<sup>64</sup> Evaluierungen schufen Gelegenheiten zum Anstoß notwendiger Veränderungen, die sonst im täglichen Fluss bürokratischer Entscheidungen nicht durchsetzbar schienen – zum Beispiel die Schließung von Länderprogrammen, die ihre Relevanz aufgrund veränderter politischer Rahmenbedingungen verloren hätten. So einen Vorschlag machten die operativ Verantwortlichen selten – dazu seien der frische Blick und die Unabhängigkeit der Evaluatorin oder des Evaluators erforderlich.

Das Evaluierungsmandat im Außenministerium beruht rechtlich auf der Kompetenz des Außenministers, ist jedoch praktisch das Ergebnis informeller Abstimmung mit dem Parlament, dem Evaluierungsberichte parallel zur Veröffentlichung direkt zugehen (ohne dass die zuständigen Minister Einfluss auf ihre Formulierung nehmen können):

“The Policy and Operations Evaluation Department [IOB] meets the need for evaluation of policy and its implementation with regard to all policy areas that are financed by the homogenous budget for international cooperation (HGIS). IOB evaluates the implementation of the Netherlands’ foreign policy, giving particular attention to its relevance, efficiency and effectiveness, the sustainability of the results achieved and the consistency and coherence of the policy.”<sup>65</sup>

Praktisch evaluiert das niederländische Außenministerium wie die Schweiz sowohl auf der Ebene von Programmen und Portfolios als auch auf der Politikebene (z.B. die niederländische Strategie zu fragilen Staaten, die niederländische EU-Ratspräsidentschaft oder die Zusammenarbeit der Benelux-Länder). Die regelmäßige Portfolio- und Programmevaluierung wird dezentral von den jeweiligen Programmverantwortlichen (Direktor DSH und unterstellte Referatsleiterinnen und -leiter) umgesetzt, folgt aber einem zentralen Evaluierungsprogramm und wird von der zentralen Evaluierungsabteilung IOB beratend unterstützt, z.B. bei der Formulierung der *terms of reference* sowie von Richtlinien und Qualitätsstandards.

Auf der Politikebene unterscheidet sich die Unabhängigkeit und Kompetenz der Evaluierungseinheit IOB erheblich zwischen dem entwicklungspolitischen und dem außenpolitischen Teil des Ministeriums. Gegenüber dem entwicklungspolitischen Teil ist IOB nach eigenen Angaben „vollständig autonom“ in der Wahl der Evaluierungsgegenstände, der Durchführung der Evaluierungen und Veröffentlichung der Ergebnisse und *issues for attention* (praktisch: Empfehlungen). Anderen Teilen des Ministeriums gegenüber ist IOB nur „begrenzt autonom“, d.h. die Auswahl eines Evaluierungsgegenstandes bedarf der Zustimmung des zuständigen *Directeur-generaal*.

64 Interview, Beamter des niederländischen Außenministeriums, Den Haag, Oktober 2013.

65 Policy and Operations Evaluation Department (2009): *Evaluation policy*, S.11.



## Akteure

Die Evaluierungsfunktion im niederländischen Außenministerium geht auf das ehemalige Handels- und Entwicklungshilfeministerium zurück, das 1977 eine Evaluierungseinheit schuf. Durch die Zusammenlegung mit dem Außenministerium 1996 erhielt die Evaluierungseinheit IOB (*Inspectie Ontwikkelingssamenwerking en Beleidsevaluatie* bzw. *Policy and Operations Evaluation Department*) die umfassende Zuständigkeit für Außen-, Außenhandels- und Entwicklungspolitik. Die Evaluierungseinheit IOB untersteht keinem der Generaldirektoren für Außen- bzw. Entwicklungspolitik, sondern dem Generalsekretär. Dieser ist als Verwaltungschef des Ministeriums den beiden Generaldirektoren formal übergeordnet, hat aber nicht die gleiche politische Bedeutung.<sup>66</sup>

Politikevaluierungen führt IOB selbst durch: IOB-Mitarbeiterinnen und -Mitarbeiter entwickeln das Pflichtenheft, engagieren externe Evaluatoren und nehmen selbst als Teil des Evaluierungsteams daran teil. Auf der Portfolio- und Programmebene berät IOB die mittelführenden Einheiten bei der Entwicklung von Pflichtenheften, bei Ausschreibungen und der Auswahl von Evaluatorinnen und Evaluatoren, bei der Steuerung der Evaluierungsprozesse sowie darüber hinaus entlang der gesamten Programmplanung, so dass professionelle Evaluierungen von vornherein in der Planung berücksichtigt und ihre Empfehlungen in spätere Entscheidungen umgesetzt werden (PM&E). Die Referatsleiterinnen und -leiter sowie der DSH-Direktor sind dadurch ständig im Gespräch mit IOB.

Im Ministerium ist IOB Teil des Rechnungsprüfungsausschusses und bildet zusammen mit der Revisionsabteilung und dem Inspektorat das sogenannte *cluster of control* des Ministeriums.<sup>67</sup> IOB selbst hat etwa 15 feste Mitarbeiterinnen und Mitarbeiter, je zu etwa einem Drittel rotierende Ministerialbeamte, dauerhaft beschäftigte Evaluierungsmanager und befristet beschäftigte Forschungsassistentinnen und -assistenten. Diese Mischung aus operativer Glaubwürdigkeit und Urteilsfähigkeit der rotierenden Beamtinnen und Beamten sowie Prozess- und Methodenexpertise der festgestellten Evaluierungsmanagerinnen und -manager leistet wichtige Beiträge zur Fähigkeit von IOB, intern wie extern akzeptierte Evaluierungen zu steuern. Deren Durchführung erfolgt durch externe Gutachterinnen und Gutachter (Budget: € 3,5 Mio. pro Jahr; zusätzliche Programm- und Portfolioevaluierungen machen weitere 1% der jeweiligen Budgets für humanitäre Hilfe, Stabilisierung etc. aus). Zudem verfügt IOB über einen Beirat professioneller Evaluatorinnen und Evaluatoren, die sowohl die Qualität der Evaluierungen als auch die Umsetzung ihrer Empfehlungen bewerten und gegebenenfalls Verbesserungsvorschläge machen sollen.<sup>68</sup> Insgesamt belaufen sich die Kosten von IOB auf 0,01% des integrierten außen- und entwicklungspolitischen Haushalts.

Entscheidend ist aus Sicht der IOB-Führung die interne Position der Evaluierungseinheit als Teil des Ministeriums: Kein organisationsfremder Akteur hätte vergleichbare Chancen, mit offener und konstruktiver Kritik wirksame Beiträge

66 Rotmann und Steinacker (2013): *Stabilisierung*, S.31.

67 Policy and Operations Evaluation Department (2009): *Evaluation policy*, S.11.

68 Ministerie van Buitenlandse Zaken (o.J.): *The Advisory Panel*.

zum internen Lernprozess zu leisten. Zwei Grundregeln sollen Interessenkonflikte vermeiden sowie die Unabhängigkeit der Einheit und ihr Vertrauensverhältnis mit den mittelführenden Referaten schützen: Erstens ist der IOB-Direktor seit 2009 erstmals kein Karrierebeamter, sondern ein extern rekrutierter Wissenschaftler, der für einige Jahre von seiner Hochschulprofessur beurlaubt wurde. Er musste auf Lebenszeit auf andere Beschäftigungsmöglichkeiten im Außenministerium verzichten. Zweitens dürfen rotierende Mitarbeiterinnen und Mitarbeiter ihre ehemaligen Arbeitsbereiche nicht direkt evaluieren.

## Abläufe und Methoden

Die Evaluierungspraxis im niederländischen Außenministerium folgt einem mehrjährigen Arbeitsprogramm. Pro Jahr fanden seit 2008 zwischen 20 und 40 interne *reviews* (Selbstevaluierungen) und externe Evaluierungen unterschiedlichen Umfangs statt, darunter auch eine Handvoll Multi-Stakeholder-Evaluierungen, die IOB mitfinanziert und -steuert. Jedes größere Programm oder Portfolio ist etwa alle 4-6 Jahre an der Reihe. So wurde die niederländische humanitäre Hilfe zuletzt 2009 evaluiert, für das laufende Jahr 2014 ist eine erneute Evaluierung geplant.<sup>69</sup> Die Themensetzung erfolgt im Zuge eines jährlich stattfindenden, intensiven Dialogs zwischen IOB und den oberen zwei Führungsebenen (*Directeurs-generaal* und *Directeurs*, entspricht etwa den oberen drei Ebenen von Staatssekretären, Abteilungsleitern und Beauftragten in Deutschland). Auch das Parlament kann Vorschläge machen, welche Programme evaluiert werden sollen.<sup>70</sup>

Der Prozess einer Evaluierung beginnt mit der Formulierung des Pflichtenhefts (*terms of reference*, TOR). Von Anfang an werden die internen *stakeholder* als Teil einer Referenzgruppe in die Formulierung der TOR einbezogen (Politikevaluierung), wenn sie nicht ohnehin federführend sind (Portfolio- und Programmevaluierung). Ein oder mehrere IOB-Referentinnen und -Referenten ergänzen im Regelfall die externen Evaluatorinnen und Evaluatoren als Teil des Evaluierungsteams. Zur Qualitätssicherung und Sicherstellung fairer Bewertungen erfolgt eine ressortinterne Kommentierung der schriftlichen Produkte durch Gutachterinnen und Gutachter außerhalb des IOB sowie eine kritische Diskussion der Berichte innerhalb des IOB. Das Evaluierungsteam gibt informelle Zwischenberichte für die Referenzgruppe heraus, um die Betroffenen am Evaluierungsprozess zu beteiligen und deren Lernen zu fördern. Sobald der Abschlussbericht vorliegt, erhält der betreffende Minister bzw. die betreffende Abteilung die Möglichkeit zur Stellungnahme, die dann mit dem Evaluierungsbericht dem Parlament übermittelt, veröffentlicht und in einer hausinternen Veranstaltung zur Diskussion gestellt wird. Diese Stellungnahme hat in der Praxis gleichzeitig die Funktion einer *management response*: Die Portfolio- oder Programmverantwortlichen verpflichten sich darin im Namen der Hausleitung auch zur Umsetzung/Reaktion auf diejenigen Empfehlungen, die sie akzeptieren.

69 Vgl. Ministerie van Buitenlandse Zaken (2010): *Evaluation Programme 2008-2015*. Anlage zum Haushalt 2011, englische veröffentlichte Version.

70 OECD-DAC (2010): *Evaluation in Development Agencies*, S.102.

Die Evaluierungspraxis des IOB versucht, volle Transparenz gegenüber Parlament und Öffentlichkeit mit maximaler Nachfrageorientierung innerhalb des Ministeriums zu verbinden. Dieser Ansatz geht allerdings auf Kosten des Zugangs zu politisch sensiblen Informationen und damit auf Kosten des Erkenntniswertes der Evaluierung. Je stärker der „diplomatisch-politische“ Anteil der evaluierten Materie, desto schwieriger wird deren Einbezug in eine Evaluierung. Kulturell tief verankerte Vertraulichkeitserfordernisse reduzieren die transparent nutzbare Datenbasis, ein volles Bild der unterschiedlichen Ziele von Aktivitäten ist schwerer rekonstruierbar und der Umgang mit Zielkonflikten zwischen kurz- und langfristigen Zielsetzungen wird schwieriger. Hier wird das Spannungsverhältnis von Rechenschafts- und Lernorientierung besonders deutlich.

Als methodische Herausforderungen werden in den Niederlanden insbesondere unklar formulierte und unzureichend dokumentierte Zielsetzungen (wie in den USA), unzugängliche und volatile Kontexte (wie in allen anderen Fallbeispielen) sowie grundsätzliche Grenzen sozialwissenschaftlicher Methoden genannt. Wirkung sei mit klassischen quantitativen Methoden schwer, wenn nicht unmöglich zu erfassen, und helfe nicht, echten politischen Wandel zu messen. So wisse man dank klassischer Evaluierung, dass niederländische Rechtsstaatsprojekte in Ruanda und Uganda auf der technischen Ebene recht erfolgreich seien, doch es gebe keine Erkenntnisse über deren Auswirkungen auf die politische Stabilität der Länder. Strikt positivistische Methoden auf Grundlage kausaler Wirkungsketten und mit dem Anspruch der eindeutigen Nachvollziehbarkeit der Wirkung sind für Maßnahmen in fragilen Kontexten nutzlos, so durchgehend die Einschätzung in Den Haag. Zudem sei die Gefahr groß, durch die Verwendung entsprechender Methoden eine unbeabsichtigte Steuerungswirkung zugunsten von messbaren statt sinnvollen Aktivitäten zu entfalten. Dementsprechend pragmatisch ist der Umgang mit den Evaluierungskriterien des OECD-DAC: Im Prinzip akzeptiert IOB diese als Grundlage, doch „Evaluierung erfordert einen maßgeschneiderten Ansatz“ mit angepassten Kriterien – insbesondere in der Außenpolitik.<sup>71</sup>

Ein niederländischer Praktiker beschreibt das methodische Problem der außenpolitischen Evaluierung in fragilen Kontexten als Zielkonflikt zwischen exaktem Wissen um Wirkung einerseits und konkretem Nutzen für die Verbesserung der Arbeit andererseits. Die Wahl falle in den Niederlanden klar auf relevante Methoden, auch wenn sie weniger exakt seien. Die Datengrundlage für Evaluierungserkenntnisse würde damit nicht willkürlich oder anekdotisch: Statt Wirkung zu „messen“, müsse jede Evaluierung diese zumindest „plausibel machen,“ so ein hochrangiger Beamter im Außenministerium. Nützlich und durchaus akzeptiert sind auch strukturierte narrative Methoden (*storytelling approaches*), die im Nachhinein den Beitrag eigener Aktivitäten an einer eingetretenen politischen Entwicklung beschreiben und damit eine Grundlage zur Beurteilung der Ausgangsannahmen und Anpassungsfähigkeit des Projekts schaffen. Es gelte grundsätzlich, Methoden und deren Anwendung im (veröffentlichten) Evaluierungsbericht detailliert zu dokumentieren, um die Erhebung der Daten nachvollziehbar zu machen – undurchsichtige qualitative Datenerhebung führe zu Recht häufig zu methodischer Kritik.

71 Policy and Operations Evaluation Department (2009): *Evaluation policy*.

Abgesehen von eigenen Evaluierungen experimentieren die Niederlande auch mit Elementen partnerschaftlicher Evaluierung als Mittel zur Versachlichung bilateraler politischer Vorhaben auf höchster Ebene. Im Rahmen eines umfangreichen Programms zur Sicherheitssektorreform in Burundi bildete die gemeinsame Bewertung von Erfolgen und Misserfolgen mit der burundischen Regierung erstmals ein Element der Vorbereitung eines Staatsbesuchs im Dezember 2013.

Über diese Evaluierungspraxis hinaus spielt auch das Monitoring laufender Projekte und Programme eine große Rolle in den Niederlanden. Statt auf formalisierten Instrumenten liegt dabei der Schwerpunkt bislang auf einer informellen Berichts- und Besuchspraxis. DSH-Mitarbeiterinnen und -Mitarbeiter besuchen jedes der zehn Maßnahmenländer mindestens einmal jährlich, in Projektentwicklungs- und Umsteuerungsphasen auch bis zu drei Mal pro Jahr für mehrere Wochen. Jeder Besuch dient explizit auch der Aktualisierung der politischen Kontextanalyse für laufende Projekte und nötigenfalls der Identifikation von Umsteuerungsbedarf gemeinsam mit der Außenvertretung vor Ort. Diese mehrwöchigen Abordnungen werden auch zur vorübergehenden personellen Verstärkung der Botschaften genutzt. Die Monitoringfunktion des Ministeriums in Den Haag soll in Zukunft durch eine Systematisierung des regelmäßigen Berichtswesens der Botschaften ausgebaut werden. Die bisherige Praxis müsse effektiver in die bestehenden Arbeitsprozesse integriert werden, damit wirklich qualitativ hochwertige Informationen gesammelt würden – auch um spätere Evaluierungen zu erleichtern.

## **Einbettung in eine professionelle Projektmanagement- und Lernkultur**

Außerhalb von Evaluierungen sind IOB-Referentinnen und -Referenten an allen programmatischen Planungsprozessen beteiligt (durch Teilnahme an Planungssitzungen und Workshops). Damit soll sichergestellt werden, dass Evaluierung von vornherein bei der Formulierung von Zielen, Indikatoren und Beobachtungsmöglichkeiten berücksichtigt wird.<sup>72</sup>

Die britische Praxis (v.a. in der Entwicklungszusammenarbeit und durch ICAI, die *Independent Commission on Aid Impact*) ist den niederländischen Gesprächspartnern sehr vertraut. Die dort üblichen, rein externen Evaluierungen sehen die niederländischen Praktiker als lernfeindlich, denn Programmverantwortliche würden damit systematisch verunsichert. Das Resultat sei wenig überraschend: Abwehrmechanismen und schlechte Moral im Inneren sowie Überreaktionen im politischen Raum. Nur partizipative Evaluierungen seien geeignet, Lernprozesse zu fördern. Politischer „Missbrauch“ von Evaluierungsergebnissen sei natürlich nie völlig auszuschließen, auch wenn keine konkreten Skandale in Erinnerung geblieben sind. Gleichzeitig – so ein ehemaliger leitender Diplomat – helfen veröffentlichte Evaluierungsergebnisse bei der gesellschaftlichen Verständigung über angemessene Erwartungen an Außenpolitik.

Was den Beitrag der eigenen Evaluierungen zum institutionellen Lernen betrifft, so besteht aus Sicht der IOB-Führung ein drastischer Unterschied zwischen dem

72 Interview, Den Haag, Oktober 2013.

entwicklungspolitischen Teil (wo dies zumindest für die wesentlichen inhaltlichen Prioritäten wie Stabilisierung, gute Regierungsführung und humanitäre Hilfe gut funktioniere) und dem außen- und europapolitischen Teil des Ministeriums (der weniger offen auf Evaluierungsergebnisse reagiere, die allerdings aufgrund der methodischen und kontextimmanenten Probleme oft auch weniger nützlich seien). Durch ein hohes Maß an informeller Interaktion mit den Beteiligten und Programmverantwortlichen – mündlich, Zwischenberichte, Referenzgruppe – betrachten die Gesprächspartner bei IOB die Wirkung ihrer Arbeit auf laufende Lernprozesse als verhältnismäßig gut organisiert. Dies wird von Seiten einer mittelführenden Arbeitseinheit im Prinzip bestätigt – ein Gesprächspartner weist allerdings darauf hin, dass es ihm praktisch frei stehe, ob und in welcher Form er Konsequenzen aus den Ergebnissen einer Evaluierung ziehe. Für erfahrene Bürokratinnen und Bürokraten sei es letztlich leicht, den Anschein von Akzeptanz und Umsetzung von Empfehlungen zu erwecken – ob er wirklich etwas ändere, hänge jedoch entscheidend davon ab, ob die Evaluatorinnen und Evaluatoren inhaltlich mit ihrer Analyse und ihren Empfehlungen überzeugen können.

Die Aufnahme von Evaluierungsergebnissen gilt auf IOB-Seite dementsprechend als nicht systematisch genug. Mehrjährige Experimente mit Aktionsplänen seien „einen langsamen bürokratischen Tod gestorben“; nun suche man gemeinsam mit dem externen Evaluierungsbeirat nach neuen Mechanismen zur Systematisierung der *management response*. Weniger erkennbar als der Lerneffekt ist aus Sicht der IOB-Experten der systemische Einfluss von Evaluierung auf eine insgesamt evidenzbasiertere Außenpolitik, z.T. weil die politische Zielformulierung im Rahmen der sogenannten *policynotes to parliament* (etwa: Regierungserklärungen in schriftlicher Form) zu unvollständig und zu vage sei. Darüber hinaus sei die Evaluierungspraxis unzureichend in die Wissensmanagementprozesse des Ministeriums integriert.

## Wesentliche Erkenntnisse

1. Die Rolle von Portfolio- und Programmevaluierung in Bezug auf die mit Stabilisierung, Krisenprävention, humanitärer Hilfe und demokratischer Transformation vergleichbaren Portfolios ist in den Niederlanden lange etabliert. Sie wird in enger Zusammenarbeit zwischen den mittelführenden Arbeitseinheiten und der zentralen Evaluierungsabteilung durchgeführt. Letztere steuert insbesondere die Erfahrung im Evaluierungsmanagement bei, d.h. bei der Formulierung robuster Pflichtenhefte, der Auswahl geeigneter Methoden und der Steuerung eines sowohl effektiven als auch inklusiven Ablaufs.
2. Die Niederlande versuchen, Rechenschafts- und Lernfunktion in einem Evaluierungsinstrument zu kombinieren. Die partizipative Form der Evaluierung erkennt an, dass Lernen immer auf Freiwilligkeit beruht; dementsprechend ist die schriftliche Selbstverpflichtung der Ministerin zur Reaktion auf Evaluierungsergebnisse (*management response*) ein weiches, aber durchaus sinnvolles Instrument. Das Ergebnis ist eine akzeptierte Praxis, die alle Beteiligten mehr oder weniger zufrieden stellt. Kritik gibt es allerdings an mangelndem Erkenntnisgewinn und begrenzter Bereitschaft zur Umsetzung von Evaluierungsergebnissen gerade in den Bereichen, wo politisch-diplomatische

Zielsetzungen nicht in voller Transparenz in die Evaluierungen eingeflossen sind. Hier stößt die Kombination von Rechenschafts- und Lernorientierung (partizipative Verfahren mit Veröffentlichungspflicht) an ihre Grenzen.

3. Die spezifischen Herausforderungen außenpolitischer Maßnahmen, vor allem multiple Zielsetzungen und fragile Kontexte, sind in den Niederlanden durchwegs anerkannt. Ihnen wird einerseits mit pragmatischer Anpassung der Standard-Evaluierungskriterien und der Methodenwahl begegnet, andererseits gilt es als Aufgabe eines professionellen und flexiblen Projektmanagements, auch außenpolitische Maßnahmen evaluierbar zu machen – angefangen mit den Planungsprozessen.
4. Die Effektivität und Akzeptanz von IOB als Kompetenzzentrum und Evaluierungsmanagement lebt einerseits von seiner internen Position, andererseits von der Mischung aus internen und externen Referentinnen und Referenten. Die rotierenden Diplomatinen und Diplomaten bringen dabei operative Glaubwürdigkeit und Urteilsfähigkeit mit, während die extern rekrutierten Evaluierungsfachleute die Prozess- und Methodenexpertise beisteuern.

---

**Kurzüberblick anhand der Leitfragen (S. 14)**

Warum?	Eigenes Mandat des Außenministeriums
Schwerpunktsetzung	Rechenschaft nach außen, Lernen nach innen
Wie?	<ul style="list-style-type: none"> <li>• Evaluierung: partizipative Verfahren mit rechenschaftsorientierten Fragestellungen</li> <li>• weniger systematisch: Lernorientierte Instrumente wie interne Reviews, Studien</li> </ul>
Was?	Politikfeld, Portfolios, Programme
Wer? (spezifisch für die außenpolitischen Portfolios)	<ul style="list-style-type: none"> <li>• Evaluierung: Politikfeldebene zentral durch Evaluierungseinheit IOB, Portfolio- und Programmevaluierungen dezentral in den jeweiligen Einheiten mit fachlicher Unterstützung durch IOB</li> <li>• Externe Evaluierungen durch Gutachterinnen und Gutachter</li> <li>• Monitoring: Dezentral autonom</li> </ul>
Wie transparent?	Alle externen Evaluierungen werden veröffentlicht
Projektmanagement	M&E sind in Strategie- und Projektmanagementverfahren eingebettet

# Schlussfolgerungen

Die vorliegende Studie hat vier Länder untersucht, in denen eine unterschiedlich ausgereifte Monitoring- und Evaluierungspraxis für außenpolitische Maßnahmen in fragilen Staaten besteht. Die Maßnahmen einer Außenpolitik mit eigenen Mitteln verfolgen in der Regel anspruchsvolle Ziele. Innerhalb relativ kurzer Zeit (meist weniger als einem Jahr) sollen komplexe, volatile und intransparente soziale und politische Prozesse in einem fremden kulturellen Kontext im Sinne der eigenen Ziele beeinflusst werden.<sup>73</sup> Der vor allem in der Entwicklungspolitik verbreitete Anspruch, mit einer Evaluierung verallgemeinerbare Aussagen über die Wirksamkeit einer Intervention analog einer medizinischen Versuchsreihe treffen zu können, hat wenig mit der Realität dieser Vorhaben zu tun.<sup>74</sup> In der politischen Öffentlichkeit steht das Schlagwort Evaluierung demgegenüber meist nur für eine vage Erwartung an die „objektive“ Überprüfung von Programm- und Projektergebnissen durch externe Expertinnen und Experten. Zwischen diesen Extremen liegt die Realität von Monitoring und Evaluierung in den jeweils für Stabilisierung, zivile Krisenprävention, demokratische Transformation und humanitäre Hilfe zuständigen außen- und entwicklungspolitischen Institutionen Großbritanniens, der Schweiz, der USA und der Niederlande.

Gerade die formale Evaluierung durch unabhängige Expertinnen und Experten anhand professioneller Evaluierungsprozesse und -methoden soll zwar in allen Fällen sowohl der Rechenschaftslegung als auch dem Lernenden dienen, doch in der Praxis dominiert meist der Rechenschaftsdruck. So ist bei DFID in Großbritannien eine Kultur extrem aufwändiger Wirkungsevaluierung entstanden, die mittlerweile zur Budgeterhaltung notwendig scheint und erhebliche Auswirkungen auf die Maßnahmengestaltung hat. Gleichzeitig zeigt diese große Schwächen im Umgang mit den Besonderheiten fragiler Kontexte. Steht der Lernzweck im Vordergrund, dann werden formale Evaluierungen häufig vermieden. Das britische FCO oder das Schweizer EDA setzen hauptsächlich auf weniger exakte und konfrontative Formen der internen *review* oder Selbstevaluierung. Nur in den Niederlanden besteht eine längere Praxis der formalen Evaluierung nach partizipativen Grundsätzen, die den politischen Rechenschaftserwartungen im Wesentlichen gerecht wird und gleichzeitig den internen Lernprozess bestmöglich zu unterstützen sucht. Formale Evaluierungsberichte werden üblicherweise veröffentlicht (so in den Niederlanden, in der Schweiz und bei DFID und USAID bezüglich ihrer humanitären und Stabilisierungsprogramme), die Ergebnisse informeller Review-Verfahren bleiben intern.

73 In der humanitären Hilfe bedeutet „eigene Ziele“ nicht notwendigerweise die Vertretung und Durchsetzung deutscher Interessen, sondern das Verfolgen prinzipienorientierter humanitärer Hilfe. Siehe Auswärtiges Amt (2012): *Strategie zur humanitären Hilfe im Ausland*.

74 Bönke, Koehler und Zürcher (2009): *Evaluation von Entwicklungszusammenarbeit zur Stabilisierung in Post-Konflikt-Zonen*; Rudolf und Lohmann (2013): *Außenpolitikevaluation*; Binder und Reichhold (im Erscheinen): *Metrics and Ethics of Protecting Civilians*.

Gegenstand von Evaluierungen oder *review*s sind meist breite, länderübergreifende Portfolios (mit Ausnahme der USA), regionale oder thematische Programme und größere Projekte. Allzu kleinteilige Evaluierungen gelten aufgrund des erheblichen Zeit- und Kostenaufwands als wenig kosteneffektiv. Während die Monitoringfunktion durchweg dezentral bei den Programmverantwortlichen und/oder Auslandsvertretungen liegt, ist die Evaluierungsfunktion in den untersuchten Außenministerien entweder bei den Haushaltsabteilungen (Großbritannien, USA) oder dezentral in den mittelführenden Arbeitseinheiten (Schweiz) angesiedelt. Die Niederlande nutzen eine effektive Mischform zwischen einem zentralen Kompetenzzentrum und dezentralisierter Verantwortung der Programmverantwortlichen.

Die Erfahrungen Großbritanniens, der Schweiz, der USA und der Niederlande zeigen, welche Vielfalt an Gestaltungsmöglichkeiten für Monitoring- und Evaluierungssysteme besteht. An einer Reihe von Beispielen sind die Chancen deutlich geworden, die sowohl zur Erfüllung von Rechenschaftsanforderungen als auch zur Förderung des Lernens bestehen. Gleichzeitig zeigen die Fallstudien auch Risiken und notwendige Entscheidungen auf: Kein M&E-Ansatz kann alle Ansprüche gleichermaßen befriedigen. Auf die konkrete Ausgestaltung kommt es deshalb an.

Auf den folgenden Seiten fassen wir zunächst vier Gründe für die verbreitete Ernüchterung über Verfahren und Ergebnisse von Evaluierungen in der außenpolitischen Praxis zusammen. Auf dieser Grundlage behandelt der Rest dieses Schlusskapitels die wesentlichen Herausforderungen für eine Weiterentwicklung der M&E-Praxis für außenpolitische Maßnahmen in fragilen Kontexten: das Spannungsfeld zwischen Rechenschafts- und Lernfunktion (S. 53), der Umgang mit Evaluierungskriterien (S. 56), die Rolle politischer Kontextanalyse im laufenden Projektmonitoring (S. 56), die systematische Ausgestaltung einzelner Evaluierungsinstrumente (S. 57) sowie damit verbundene Anforderungen an die Lernkultur der Organisation, Aus- und Fortbildung sowie Ressourcenausstattung (S. 59-60). Die Studie schließt mit einem Blick auf die Chancen und Anforderungen zur Förderung einer aktiv lernenden Organisation.

## Vier Gründe für bislang oft ernüchternde Erfahrungen

Eine wichtige Gemeinsamkeit der internationalen Praxis besteht in der verbreiteten Ernüchterung über Umsetzung und Ergebnisse von außenpolitischen Evaluierungen in den untersuchten Themenfeldern. Diese Ernüchterung speist sich aus vier Quellen.

### 1. Rechenschafts- und Lernfunktion stellen teils konkurrierende Ansprüche an die Gestaltung von Monitoring und Evaluierung.

Das Spannungsfeld zwischen der Rechenschafts- und Lernfunktion (siehe *Drei zentrale Erkenntnisse*, S. 9) ist auch in der Praxis eine große Herausforderung. Rechenschaftsorientierte Evaluierung ist im Regelfall investigativ, oft konfrontativ gestaltet und stößt deshalb auf Vorbehalte. Der offene Umgang mit Stärken und Schwächen von Programmen und Entscheidungen wird dadurch eingeschränkt; viele Mitarbeiterinnen und Mitarbeiter von Außenministerien erleben solche



M&E-Systeme als Versuch der technokratischen Kontrolle von Prozessen, die der objektiven Messbarkeit nicht zugänglich sind. Gleichzeitig sind erfolgreich eingeführte Lerninstrumente oft methodisch und prozedural so informell, dass sie nicht zur Rechenschaftslegung taugen – weil beispielsweise keine veröffentlichungsfähigen Ergebnisse entstehen.

In der Schweiz wurde das am deutlichsten. Dort ist die Einführung eines stark rechenschafts- und steuerungorientierten Evaluierungssystems gescheitert, während lernorientierte Evaluierungsinstrumente zunehmend Anklang finden. Nur in den Niederlanden knüpft M&E systematisch an die Lernfähigkeit von Außenministerien und ihren Mitarbeiterinnen und Mitarbeitern an. In Großbritannien hingegen setzen die Evaluierungspraxis von DFID und der Rechenschaftsdruck durch ICAI das Außenministerium zunehmend unter Druck, seine informellen und lernorientierten Instrumente zu formalisieren und auf Rechenschaft auszurichten.

## 2. Die spezifischen Herausforderungen für Monitoring und Evaluierung außenpolitischer Maßnahmen in fragilen Kontexten werden noch nicht angemessen berücksichtigt.

Häufig werden Datenerhebungsmethoden verwendet, die mit dem Zusammenspiel vieler politisch-prozessualer und sektoraler Ziele nicht umgehen können – insbesondere wenn die politisch-prozessualen Ziele nicht in objektiven Größen messbar sind, sondern nur im Hinblick auf Wahrnehmungen der Bevölkerung oder politischer Eliten. Die politische Volatilität fragiler Kontexte überfordert zudem die positivistische Kausallogik, die den meisten M&E-Standardansätzen zu Grunde liegt. Das führt auch dazu, dass die in der entwicklungspolitischen Evaluierung verbreiteten Kriterien des OECD-DAC nur bedingt für die Evaluierung außenpolitischer Maßnahmen in fragilen Kontexten nutzbar sind.

Hier stoßen verbreitete Erwartungen an Evaluierungsmethoden an die Grenzen des sozialwissenschaftlich Machbaren. Diese erkenntnistheoretischen Grenzen sind nicht nur eine wissenschaftstheoretische Frage, sondern ein ganz praktisches Problem: Die üblicherweise verwendeten positivistischen Methoden sind für außenpolitische Maßnahmen in fragilen Kontexten nicht hinreichend belastbar.<sup>75</sup> Der Umgang mit diesen Grenzen von evidenzfokussierter Evaluierung war in den betrachteten Ländern sehr unterschiedlich. Die britische Regierung hält am Anspruch der objektiven Messung von Ergebnissen fest und investiert stark in die Weiterentwicklung von Evaluierungsmethoden, mittlerweile zunehmend auch im Außenministerium. Die Schweizer sind noch auf der Suche nach dem richtigen Umgang mit diesem Problem, haben allerdings eine erste große Evaluierung entwicklungspolitischer Maßnahmen in fragilen Kontexten erfolgreich umgesetzt. Diese Evaluierung setzte einen starken Schwerpunkt auf Lernen, priorisierte einige wenige Fragen und nutzte eine flexible

75 Für nähere Erläuterungen siehe Donaldson et al. (2010): *The 2009 Claremont Debates*; Rudolf und Lohmann (2013): *Außenpolitikevaluation*.

Auswahl von partizipativen und wahrnehmungsorientierten Methoden.<sup>76</sup> Dies bedeutet eine Abkehr von kausalanalytischen Wirkungsevaluierungen – ein Trend, der auch in den Niederlanden und den USA deutlich wurde.

Allen untersuchten Ländern ist jedoch gemein, dass die erkenntnistheoretischen Probleme von Wirkungsevaluierungen in fragilen Kontexten kein Anlass sind, ganz auf die Rechenschafts- und Lernwirkung von Evaluierung zu verzichten. Stattdessen ist in den Themenfeldern Stabilisierung, zivile Krisenprävention, demokratische Transformation und humanitäre Hilfe die Entwicklung angepasster Evaluierungskriterien, die Verwendung von Prinzipien statt Kriterien<sup>77</sup> und innovativer Methoden wie perzeptiver oder umfrageorientierter Ansätze zunehmend etablierte internationale Praxis. Kausalanalytische Evaluierungsansätze verlieren für fragile Kontexte hingegen an Bedeutung.<sup>78</sup>

### 3. Schwächen in den Organisationsstrukturen und -abläufen erschweren eine gewinnbringende Evaluierungspraxis.

Die Strukturen und -abläufe in Außenministerien sind zur Organisation von Diplomatie entstanden. Der Einsatz erheblicher Finanzmittel zur Umsetzung von Maßnahmen durch externe Partner ist eine neue Aufgabe, die den gewohnten Organisationsabläufen zunächst fremd ist. Die Grundlagen professionellen Portfolio-, Programm- und Projektmanagements sind deshalb in Außenministerien häufig noch nicht vollständig verankert (Ausnahmen sind die Niederlande und Großbritannien). Es wurde klar, dass einzelne Referate hier nur begrenzt Abhilfe schaffen können, solange die Organisation als Ganzes das neue Instrumentarium nicht als Teil ihres Kerngeschäfts versteht und organisatorisch in ihre entsprechenden Prozesse integriert – von der Planung und Entscheidungsfindung bis zur Aus- und Fortbildung. Dazu kommt der konstruktive Umgang mit Fehlern und Konflikten, die nicht nur aus rechenschaftsorientierten M&E-Prozessen entstehen. Monitoring und Evaluierung sind immer organisationspolitische Prozesse, auch wenn sie hauptsächlich dem Lernen dienen. Sie können ihren Lernzweck nur erfüllen, wenn Kontroversen möglich sind und Entscheidungen unabhängig von Hierarchien kritisiert werden können. Dies erfordert sowohl eine gewisse Unabhängigkeit der Akteure und Prozesse (v.a. der Evaluierung) als auch wirksame politische Rückendeckung von ganz oben.

Ein maßvoller Umgang mit den notwendigen Veränderungen fällt den Akteuren in allen betrachteten Ländern schwer. Die meisten schwanken zwischen zwei Extremen: Die einen haben eine ausufernde Bürokratie des Projekt- und Qualitätsmanagements inklusive Evaluierung geschaffen, die anderen ein Maß an Freiwilligkeit und Fragmentierung zugelassen, das Lernen zwar aus Eigeninitiative ermöglicht, aber nicht sicherstellen kann. Das Schweizer EDA hat Erfahrungen mit beiden Reaktionen gemacht. Es neigte zu stark bürokratischen Prozessen beim

76 Eidgenössisches Departement für auswärtige Angelegenheiten (2012): *Evaluation of the performance of SDC instruments*.

77 Vgl. OECD (2007): *Principles for Good International Engagement*.

78 Blue, Clapp-Wincek und Benner (2009): *Beyond Success Stories*; OECD-DAC (2012): *Evaluating Peacebuilding Activities*; Ramalingam (2013): *Aid on the Edge of Chaos*.

Versuch der Zentralisierung bis 2012. Auf Widerstand der Mitarbeiterinnen und Mitarbeiter reagierte die Hausleitung mit einer Kehrtwende: Seitdem werden einzelne Lernübungen oder Evaluierungen nur noch punktuell eingesetzt. Vor allem außerhalb der von der DEZA verantworteten Themenbereiche lässt dieser Prozess bisweilen Kohärenz vermissen. Grundsätzlich scheint ein weniger bürokratischer Prozess leichter mit den Strukturen und Abläufen von Außenministerien vereinbar als die Einführung umfangreicher neuer Systeme.

#### 4. Die methodische und argumentative Qualität vieler Evaluierungen lässt erheblich zu wünschen übrig.

Für die verbreiteten Klagen über die Qualität von Evaluierungen lässt sich eine Reihe von Gründen identifizieren. Zunächst sind die Märkte für Evaluierungsspezialistinnen und -spezialisten aufgrund sprachlicher Grenzen nur teilweise internationalisiert; in kleinen, homogenen Märkten fehlt unter Umständen der Wettbewerb, der zu mehr Qualität und Innovation führen könnte. Vor allem die Schweizer Gesprächspartnerinnen und -partner hatten den Eindruck, dass der heimische Evaluierungsmarkt stark von der Entwicklungszusammenarbeit geprägt sei. Das Verständnis der Evaluatorinnen und Evaluatoren für die außenpolitische Arbeitsweise sei sehr eingeschränkt, deshalb schreibe man außenpolitische Evaluierungen in der Regel international aus.

Weitere verbreitete Gründe liegen darüber hinaus in den Anreizstrukturen des Evaluierungsgeschäfts: Evaluierung als Geschäftsmodell funktioniert am besten, wenn erprobte Standardmethoden für viele unterschiedliche Auftraggeber zur Anwendung kommen. Methodische Weiterentwicklung rechnet sich nur in geringem Maß; der Großteil der Innovation entsteht in einer Nische der empirischen Sozialwissenschaft, die sich als Grundlagenforschung versteht und deshalb meist abweichende Erkenntnisinteressen verfolgt. Grundlagenforscherinnen und -forscher sind jedoch keine Evaluatoren: Sie verfügen weder über das notwendige außenpolitische Kontextwissen noch stehen sie aufgrund ihrer Lehr- und Publikationsverpflichtungen für umfangreiche und häufig reiseintensive Evaluierungsaufträge zur Verfügung. Darüber hinaus schaffen der Zeitdruck von Kundenseite und die finanziellen Grenzen einer Evaluierung oft Rahmenbedingungen, die keine rigorose Anwendung anspruchsvoller Erhebungsmethoden erlauben.

Nicht zuletzt besteht ein erheblicher Faktor für den Erfolg oder Misserfolg von Evaluierungen in der Qualität des Pflichtenhefts: In die Auswahl und Gestaltung der Evaluierungsfragen, die Interpretation von Erfolgs- oder Evaluierungskriterien sowie die laufende Begleitung des Evaluierungsprozesses fließt beispielsweise in Großbritannien und den Niederlanden erheblicher Aufwand, damit externe Evaluatorinnen und Evaluatoren tatsächlich faire und nützliche Ergebnisse entwickeln können. Insbesondere der Rolle des Evaluierungsmanagements und der Steuerungsgruppe kommt in diesem Zusammenhang entscheidende Bedeutung zu.

## Schwerpunktsetzung im Spannungsfeld von Lernen und Rechenschaft

Systematische Überlegungen zu Monitoring und Evaluierung müssen damit beginnen, eine konkrete Antwort auf die Vielfalt der Erwartungen im Spannungsfeld von Lern- und Rechenschaftsfunktion zu geben (vgl. *Vier Gründe*, S. 53). In allen Fallstudien wurde deutlich, dass diese Spannung die praktische Anwendung und den Nutzen von Evaluierungsinstrumenten grundlegend prägt. Auf der institutionellen Ebene muss ein M&E-System immer beide Funktionen zu einem gewissen Grad erfüllen, doch bei der Gestaltung bzw. Weiterentwicklung konkreter Instrumente gilt es klare Schwerpunkte zu setzen. Davon hängt ab, an welche bestehenden Prozesse und Strukturen diese angebunden werden können, welche Anreizstrukturen geschaffen und welche organisationskulturellen Veränderungen unterstützt werden müssen.

Für die grundlegende Schwerpunktsetzung zwischen Rechenschaft und Lernen bestehen theoretisch drei Möglichkeiten: die klare Schwerpunktsetzung zugunsten der Rechenschaftslegung (d.h. mit investigativen Verfahren), der Versuch eines Mittelweges – d.h. praktisch Rechenschaftsorientierung kombiniert mit partizipativen Verfahren – sowie die klare Schwerpunktsetzung zugunsten des Lernens. In den meisten Ländern hat sich die erste Möglichkeit mit Ausnahme der britischen Praxis entweder als nicht praktikabel (Schweiz), nicht erfolgreich (USA) oder von vornherein als unangemessen (Niederlande) erwiesen und kann deshalb nicht als zielführend gelten. Die Gründe hierfür sind jedoch instruktiv für die beiden anderen Optionen.

**Rechenschaftsorientierte Evaluierungsinstrumente** priorisieren rückblickende Fragen nach den (Miss-)Erfolgen einer Maßnahme im Sinne der definierten Evaluierungskriterien und nach den Ursachen dafür. Sie werden deshalb ex-post (wenn nötig mit erheblichem Abstand zum Abschluss von Projekten und Programmen) durchgeführt. Sie erfordern ein hohes Maß an Validität der Datenerhebung, Unabhängigkeit in der Bewertung und Transparenz im Umgang mit den Ergebnissen. Deshalb sollten rechenschaftsorientierte Evaluierungen von externen Evaluatorinnen und Evaluatoren durchgeführt werden, deren professionelle Reputation von der Wahrung ihrer Unabhängigkeit gegenüber einzelnen Auftraggebern abhängt. Um in der politischen und der Fachöffentlichkeit glaubwürdig zu sein, müssen Qualitätsurteile sowohl unbeeinflusst zustande kommen als auch kritisch und ehrlich formuliert sein.

Damit entstehen Interessenkonflikte mit denjenigen, deren Informationen und Einschätzungen für die Qualität der Evaluierungsentscheidend sind, die aber gleichzeitig institutionell oder gar individuell als Verantwortliche für Erfolg oder Scheitern benannt werden sollen. Werden diese Interessenkonflikte, wie im Fall von SIGAR (USA) oder ICAI (Großbritannien) mittels investigativer Verfahren und Kompetenzen (Offenlegungszwang etc.) aufgelöst, dann entstehen erhebliche Abwehrreaktionen. Solcherart investigative Evaluierungsmechanismen können deshalb nur dann zu praktischen Veränderungen beitragen, wenn ihr konfrontativer Charakter sowohl als legitim akzeptiert als auch mit wirksamer politischer Autorität bewehrt ist. Bei ICAI ist dies der Fall, bei SIGAR nicht – hier ist der Evaluierungsanspruch deshalb gescheitert. Investigative Rechenschaftsevaluierungen werden für den deutschen Kontext nicht empfohlen.

Eine häufige Reaktion auf die Ablehnung von investigativen Verfahren zur Rechenschaftslegung ist der Versuch, **rechenschaftsorientierte Evaluierungsfragen**

**mit partizipativen Verfahren** zu kombinieren (z.B. in den Niederlanden und bei internationalen humanitären Organisationen). Das heißt, es werden Strukturen und Prozesse geschaffen, um die mit Rechenschaft verbundenen Risiken für die Beteiligten und ihre Organisationseinheiten zu minimieren, ohne den Rechenschaftsanspruch aufzugeben. Von der Formulierung des Pflichtenhefts bis zum Abschlussbericht werden die Beteiligten eingebunden, um ihre Akzeptanz der Fragestellungen und Bewertungen sicherzustellen.

Solcherart partizipative Rechenschaftsevaluierungen sind mit erheblichem Steuerungsaufwand, der Gefahr analytischer Unschärfe oder des Verzichts auf die Veröffentlichung als Zugeständnis an die beteiligten Entscheidungsträger verbunden, was ihre politische Glaubwürdigkeit reduzieren kann. Gleichzeitig bestehen durchaus politische Risiken des Missbrauchs von Evaluierungsergebnissen in Politik und Öffentlichkeit; diese sind allerdings differenziert zu betrachten. Die Skandalisierung von Evaluierungsberichten kommt vor, allerdings konnte in keiner der Fallstudien ein konkretes Beispiel identifiziert werden, in dem die Veröffentlichung einer negativen Evaluierung budgetäre oder personelle Konsequenzen gehabt hätte. Diffuse politische Nachteile sind damit jedoch nicht ausgeschlossen. Gleichzeitig unterstützt die Veröffentlichung von Evaluierungsergebnissen eine informierte Debatte in der Fachöffentlichkeit und trägt damit auch zu realistischeren Erwartungen an die Außenpolitik bei.

**Stärker lernorientierte Evaluierungen** priorisieren im Gegensatz zu rechenschaftsorientierten M&E-Instrumenten vorwärtsgewandte Fragestellungen. Grundlage für die Erkenntnisse von lernorientierten Instrumenten ist ebenfalls die Analyse bestehender Abläufe, Maßnahmen und ihrer Ergebnisse anhand von Evaluierungskriterien. Um Vertrauen zu schaffen und einen höheren Mehrwert für die Zukunft zu leisten, wird hier allerdings weniger Wert auf die Belastbarkeit von Evidenz für die hypothetischen Ursachen vergangener (Miss-)Erfolge und die Ermittlung der dafür Verantwortlichen gelegt. Im Mittelpunkt stehen die systematische Analyse der Wahrnehmung aller Beteiligten und die Konstruktion plausibler Erklärungen, nicht aber die rückwirkende Etablierung einer objektiven Wahrheit. Im Ergebnis kann ein größerer Teil des Aufwands in die Ermittlung konkreter Lehren für aktuelle Herausforderungen fließen, da die Sammlung und Triangulierung von Daten weniger Ressourcen erfordert. Auch die meisten rechenschaftsbezogenen Evaluierungen schreiben sich dieses Ziel ins Pflichtenheft, werden ihm jedoch aufgrund der enormen Schwierigkeiten bei der Erhebung rechenschaftsbezogener Daten in fragilen Kontexten nur in den seltensten Fällen gerecht.

Stärker lernorientierte Evaluierungsinstrumente sind vor allem partnerschaftlich zu gestalten. Evaluierungen müssen nicht ausschließlich von externen Gutachterinnen und Gutachtern durchgeführt und nicht zwingend vollständig veröffentlicht werden, sondern können auch (zum Teil) intern ablaufen. Über den erforderlichen Grad an Unabhängigkeit und Transparenz kann von Fall zu Fall entschieden werden. Auch die Methodenwahl steht weniger rigiden fachlichen Erwartungen gegenüber. Entscheidend ist, dass Richtlinien und (Beratungs-) Ressourcen so gestaltet sind, dass M&E in einem belastbaren Vertrauensverhältnis zwischen internen Lernmanagern, Evaluatorinnen und Evaluatoren sowie den Portfolio-, Programm- oder Projektverantwortlichen stattfinden. Beispiele dafür gibt

es in Großbritannien (FCO-interne *reviews*), in den Niederlanden und in der Schweiz, sie sind allerdings meist weniger systematisiert und transparent als die partizipativen Rechenschaftsevaluierungen in den gleichen Ländern.

Explizit lernorientierte Evaluierung bietet wenig Anknüpfungspunkte dafür, einzelne Organisationseinheiten oder gar Personen für Fehlentwicklungen zur Verantwortung zu ziehen. Damit gewinnt sie intern Akzeptanz, ist aber extern in einem Umfeld spezifisch rechenschaftsorientierter Erwartungen aus Politik und Öffentlichkeit schwerer zu vermitteln. Allerdings kommt die wissenschaftliche und praktische Evaluierungsdebatte zunehmend zu dem Schluss, dass die bewusste Lernorientierung dem Anspruch auf kontinuierliche Verbesserung der Praxis durch Evaluierung näher kommt als der häufig gescheiterte Versuch der Rechenschaftslegung – zumindest was fragile Kontexte und (außen-)politisch geprägte Aktivitäten betrifft.

## Umgang mit Evaluierungskriterien

Evaluierungskriterien sind zentral für alle M&E-Anstrengungen, unabhängig von einzelnen Instrumenten. Internationale Standards wie die OECD-DAC-Kriterien, die für entwicklungspolitische Kontexte und Maßnahmen entwickelt wurden, sind aufgrund ihrer Verankerung in einer linearen Kausallogik allerdings nur teilweise für außenpolitische Maßnahmen in fragilen Kontexten geeignet. Deshalb werden gemeinhin erhebliche Anpassungen vorgenommen oder gar eigene Kriterien definiert.

Neben der Berücksichtigung der Besonderheiten von außenpolitischen Maßnahmen in fragilen Kontexten müssen faire Evaluierungskriterien von den Ambitionen ausgehen, die tatsächlich bestehen und die in der Zielsetzung einer Maßnahme festgehalten sein sollten. Der Formulierung bzw. Anpassung von Evaluierungskriterien liegen daher drei Leitfragen zu Grunde:

- Was ist der genaue Gegenstand der Maßnahme(n)? Diese Frage ist besonders relevant für Portfolios, Programme und Projekte, die auf vage definierten Konzepten wie *state building* oder *humanitarian protection* beruhen und die multiple Ziele verfolgen.
- Wann ist die jeweilige Maßnahme „gut“ bzw. erfolgreich?
- Wie können die Güte bzw. die Resultate der Maßnahme erfasst werden?

Für die Beantwortung dieser Fragen gibt es mindestens zwei anerkannte Wege: entweder durch die Eigenentwicklung von Kriterien für den spezifischen Kontext außenpolitischer Programme und Projekte in fragilen Kontexten (wie dies beispielsweise ICAI in Großbritannien vormacht), oder die Anpassung internationaler Standardkriterien an die Spezifika der außenpolitischen Praxis (wie zum Beispiel durch IOB in den Niederlanden, und die internationale Evaluierungspraxis in der humanitären Hilfe). Evaluierungskriterien kommt eine ebenso zentrale Bedeutung zu wie Evaluierungsfragen. Mit ihnen steht und fällt, ob eine Evaluierung ihren beabsichtigten Zweck erreichen kann. Dementsprechend treiben Organisationen, die qualitativ anerkannte Evaluierungen hervorbringen – wie zum Beispiel DFID und ICAI in Großbritannien oder die DEZA in der Schweiz – hier erheblichen konzeptionellen Aufwand.

## Monitoring-Praxis in Bezug auf politische Kontextanalyse stärken

Für die tägliche Arbeit der Maßnahmensteuerung auf Projekt- und Programmebene in fragilen Kontexten ist es entscheidend, die volatilen politischen Prozesse vor Ort ständig zu verfolgen und damit schnell zielorientiert (um-)steuern zu können. Ansätze eines kontext- und zielorientierten Monitorings hat das Autorenteam vor allem in den USA und den Niederlanden gefunden. Die entsprechenden Systeme bauen teilweise auf bestehenden Erhebungs- und Berichtsverfahren auf, um Doppelstrukturen zu vermeiden.

Dabei haben sowohl die Umsetzungspartner als auch die Botschaften und Vertretungen vor Ort wichtige Rollen zu spielen, um auf Grundlage ihrer jeweiligen Informationsquellen der Analyse des politischen Kontextes und seiner Auswirkungen auf Projekt- und Programmziele ein stärkeres Gewicht in der Berichterstattung zu verleihen. Das bringt in der Regel einen höheren Personalbedarf mit sich.

## Evaluierungsinstrumente systematisieren

Für die Gestaltung und Weiterentwicklung von Evaluierungsinstrumenten spielt eine Reihe von Erwägungen eine Rolle. Einige der wichtigsten sind in den folgenden Abschnitten kurz angerissen:

### Grundfragen des Evaluierungsdesigns

- **Umfang und Ebene:** Innerhalb der linearen Kausallogik der DAC-Kriterien ist es leichter, auf der kleinteiligeren Ebene eines einzelnen Projekts belastbare Evidenz zu erhalten als für ganze Programme oder Portfolios. Daraus erklärt sich die entsprechende Schwerpunktsetzung auf kleinteiligere, allerdings aufgrund der methodischen Ansprüche an die Genauigkeit auch extrem teure Evaluierungen in Großbritannien und den USA. Finanziell günstiger sind große, zusammenfassende Evaluierungen von Programmen und Portfolios, wie dies in der Schweiz und den Niederlanden praktiziert wird. Liegen dem allerdings keine exakten Evaluierungen der einzelnen Projekte zugrunde, dann kann die Erwartung an eine solche Evaluierung nicht darin bestehen, beispielsweise die kausale Wirkung von Bildungsprojekten auf die friedliche Transformation Tunesiens in Richtung demokratischer Institutionen nachzuweisen – das ist für außenpolitische Maßnahmen in fragilen Kontexten nicht belastbar möglich. Stattdessen bieten sich nicht-positivistische Methoden wie Perzeptionsumfragen oder Narration zur Erfassung eines weniger exakten, aber oft für strategische Fragen und Lerneffekte ebenso nützlichen Gesamtbildes an. Auch solche Evaluierungen sind aufgrund des umfangreichen Erhebungsaufwandes teuer, stehen allerdings erheblich höheren Maßnahmenumfängen gegenüber.
- **Häufigkeit von Evaluierungen:** Als Teil dauerhafter und nachhaltiger Lern- und Rechenschaftsprozesse besteht üblicherweise die Erwartung, in regelmäßigen

Abständen Evaluierungen durchzuführen. In den Niederlanden ist jedes Portfolio (z.B. die humanitäre Hilfe) alle vier bis fünf Jahre Gegenstand einer Evaluierung; der genaue Zeitpunkt richtet sich auch nach politischen Rahmenbedingungen. Ähnliche Rhythmen bestehen in Großbritannien und den USA.

- **Muss jede Evaluierung von externen Profis durchgeführt werden?** Mit dem Begriff der Evaluierung ist üblicherweise die Erwartung verbunden, dass organisationsexterne Evaluatorinnen und Evaluatoren mit der Durchführung beauftragt werden. Ihre Unabhängigkeit trägt erheblich zur Glaubwürdigkeit einer Evaluierung in Fachöffentlichkeit und Politik bei. Professionelle Evaluierungsexpertinnen und -experten können zudem eine Mischung von fachlicher und Methodenkompetenz mitbringen, die auf den jeweiligen Evaluierungsgegenstand zugeschnitten ist – ein Maß an Spezialisierung, das keine Organisation intern vorhalten kann. Gleichzeitig gibt es jedoch auch eine übliche und anerkannte Praxis interner, evaluierungsähnlicher Instrumente wie *reviews* (z.B. im britischen FCO). Auf Grundlage einer gut entwickelten Lernkultur, die den Austausch konstruktiver Kritik unter Kolleginnen und Kollegen ermöglicht, können solche Instrumente leichter zum organisationalen Lernen beitragen als eine externe Evaluierung, die häufig mit einem Defizit von Vertrauen und Organisationskenntnis kämpft. Beide Ansätze haben ihre Berechtigung.
- **Schnittstellen zur Wissenschaft:** Die wenigsten Grundlagenforscher stehen selbst für die Durchführung von Evaluierungen zur Verfügung. Gleichzeitig erzeugt die Wissenschaft laufend neue methodische und thematische Erkenntnisse, die für die Gestaltung guter Evaluierungen nützlich sind. Institutionelle Schnittstellen erleichtern den Austausch.
- **Nachfragemacht und die Qualität von Evaluierungen:** Die Nachfragemacht der Auftraggeber von Evaluierungen lässt sich nutzen, um den oben ausgeführten Qualitätsproblemen (S. 53) zu begegnen. Dazu gilt es, in den Pflichtenheften praktisch relevante und wissenschaftlich bearbeitbare Evaluierungsfragen zu stellen, von externen Anbietern realistische Evaluierungsmethoden zu verlangen und die Qualität der unterschiedlichen Angebote bewerten zu können – all das stellt erhebliche Anforderungen an die Kompetenz, Erfahrung und die zeitlichen Kapazitäten des Evaluierungsmanagements. Qualitativ hochwertigen Evaluierungen würde es zudem dienen, gezielt gemischte Teams mit Evaluierungs-, Themen- und sozialwissenschaftlicher Methodenkompetenz zur Angebotsabgabe aufzufordern. M&E in fragilen Kontexten steht darüber hinaus aber auch vor grundsätzlichen methodischen Herausforderungen, die nicht durch bessere Pflichtenhefte und den richtigen Erfahrungsmix der Evaluatorinnen und Evaluatoren aufzulösen sind. Hier hilft nur die Investition in methodische Forschung, wie sie die britische Regierung beispielsweise sehr umfassend betreibt. Wissenschaftliche Studien können etwa dabei helfen, ein besseres Verständnis dafür zu entwickeln, welche Ansätze der Evaluierung für außenpolitische Maßnahmen in fragilen Kontexten (nicht) funktionieren – eine Art Forschungsfrage, die sich die Wissenschaft selbst üblicherweise nicht stellt.



## Organisation und Steuerung von Evaluierungen

- **Evaluierungsmanagement:** Die Entwicklung von Pflichtenheften, die Durchführung von Ausschreibungen, die Auswahl von Steuerungsgruppen (siehe unten) und die Nachverfolgung (*follow-up*) der aus Evaluierungen entstandenen Empfehlungen sind entscheidende Aufgaben, die über einzelne Evaluierungsprozesse hinausgehende Erfahrungen erfordern. Dieses Evaluierungsmanagement erfordert sowohl methodische und Prozessexpertise als auch ein Verständnis der Abläufe und Entscheidungsprozesse der Organisation selbst.
- **Steuerungsgruppen als zentrales Partizipationsinstrument:** Für größere Evaluierungen haben sich Steuerungsgruppen als gute Praxis erwiesen (zum Beispiel in den Niederlanden). Sie erhöhen die Nutzung von Evaluierungsergebnissen, das sie helfen, Evaluierungen für alle *stakeholder* relevant zu machen. Die Steuerungsgruppe ist an der Auswahl des Evaluierungsteams und der Formulierung des Pflichtenheftes beteiligt. Sie bringt Vertreter aller von der Evaluierung betroffenen Organisationen und Organisationseinheiten auf Arbeitsebene zusammen, um sowohl die Unabhängigkeit der Evaluierung als auch ihre Fairness zu wahren. Sie begleitet die Evaluierung von Anfang bis Ende und stellt sicher, dass Evaluierungsergebnisse in die entsprechenden Entscheidungsprozesse eingespeist werden.
- **Nutzbarmachung der Ergebnisse:** Sollen Evaluierungsergebnisse einer Organisation nützen, dann gilt es einen strukturierten Prozess zu schaffen, um die Empfehlungen der Evaluatoren an die Führungsebene zu vermitteln, über die Akzeptanz und Umsetzung auch einzelner Empfehlungen zu entscheiden (*management response*) und die Umsetzung akzeptierter Empfehlungen nachzuverfolgen.

## Lernkultur

Voraussetzung für die Nutzung von Evaluierung als Lerninstrument ist die Verankerung einer konstruktiv-kritischen Kultur der Selbstreflexion und des Lernens als Organisation. Lernbereitschaft kann durch gute Gelegenheiten und Anreize gefördert werden. Die Lernwahrscheinlichkeit ist umso höher, je mehr die Lernerfahrung dem einzelnen Menschen Erleichterung in seiner alltäglichen Arbeit bringt. Die Stärkung der Lernkultur erfordert deshalb sowohl die Vorbildfunktion von Führung als auch die Systematisierung institutioneller Prozesse und Ressourcen, die das Lernen leichter machen.<sup>79</sup> Daher muss auf allen Führungsebenen nicht nur Fehlerakzeptanz geschaffen werden, sondern es müssen auch Vorbilder gefunden werden, die sich einer

<sup>79</sup> Zu Organisationslernen in Bundesministerien ist konkret keine wissenschaftliche Literatur bekannt. Bzgl. einschlägiger internationaler Bürokratien siehe u.a. Howard (2008): *UN Peacekeeping in Civil Wars*; Benner, Mergenthaler und Rotmann (2011): *The New World of UN Peace Operations*; sowie zusammenfassend Benner, Eckhard und Rotmann (2013): *Learning in International Organizations*.

selbst-reflektierten Lernkultur verschreiben und diese bei ihren Mitarbeiterinnen und Mitarbeitern fördern.

Ein Beispiel für ein nützliches Instrument sind die *peer reviews*, die sowohl in Großbritannien als auch in der Schweiz erfolgreich eingesetzt werden: wenig formalisierte Mechanismen, um in strategischen Fragen der Programmgestaltung über den eigenen Weisungsstrang hinaus die Kompetenz und den frischen Blick gleichrangiger Kolleginnen und Kollegen zu nutzen. Während die britischen *reviews* eher dem typischen Ablauf einer Evaluierung entsprechen (nur mit internen statt externen Gutachtern), stellt die Schweizer *peer review* die Schaffung eines vertraulichen Klimas zum Austausch und Lernen in den Mittelpunkt.

## **Aus- und Fortbildung: M&E-Kompetenz vertiefen und breiter verankern**

Jedes M&E-Instrument erfordert eine erhebliche Vertrautheit mit den Zielen, Methoden und Risiken der M&E-Funktion insgesamt sowie spezifische Prozess- und Methodenkompetenz bei denjenigen, die das jeweilige Instrument einsetzen sollen. Auf beiden Ebenen ist die Vermittlung von M&E-Kompetenz von einem Verständnis für professionelles Projektmanagement im Kontext außenpolitischer Entscheidungs- und Berichtswege nicht zu trennen. Wichtig ist insbesondere, dass Monitoring- und Evaluierungsinstrumente sowohl ihre Rechenschafts- als auch ihre Lernfunktion nur erfüllen können, wenn auf Portfolio-, Programm- und Projektebene klare Ziele definiert werden. Diese mögen sich über Zeit verändern, müssen aber intern explizit dokumentiert sein. Schon deshalb sind Monitoring- und Evaluierungsprozesse vorab als Teil der Maßnahmensteuerung einzuplanen. Dementsprechend darf es bei der Entwicklung von Aus- und Fortbildungsmodulen nicht ausschließlich um die vordergründigen Instrumente für M&E gehen, sondern um den Entscheidungs- und Implementierungsprozess für außenpolitische Maßnahmen insgesamt.

## **Ressourcenbedarf**

Im Verhältnis zu den eingesetzten Mitteln muss eine erfolgreiche Monitoring- und Evaluierungspraxis nicht teuer sein, ist aber auch nicht zum Nulltarif zu haben. Für finanzielle und personelle Zielgrößen gibt es in den Bereichen humanitäre Hilfe, Stabilisierung, zivile Krisenprävention und demokratische Transformation international keinen belastbaren Vergleichsmaßstab. Wieviel Aufwand als vertretbar gilt, fällt von Land zu Land, von Behörde zu Behörde unterschiedlich aus. Quer durch alle Politikfelder schwanken wissenschaftliche Empfehlungen für die Höhe von Evaluierungsbudgets zwischen 1% und 10% der Programmkosten, während tatsächliche Budgets zwischen 0,005% und 5,5% der Programmkosten betragen.<sup>80</sup> Diese Anteile beziehen sich auf den Gesamtumfang der jeweiligen Mittel und sind ohne Bezug zu den Skaleneffekten im jeweiligen Politikbereich (Sozial-, Bildungs-,

80 Borrmann und Stockmann (2009): *Evaluation in der deutschen EZ*; OECD-DAC (2010): *Evaluation in Development Agencies*; Rieder (2011): *Kosten von Evaluationen*.

Umwelt-, Kulturpolitik) wenig aussagekräftig. Im außenpolitischen Bereich können bei einzelnen Evaluierungen schon die Kosten für externe Evaluierungsdienstleister einen Anteil von 10% des direkt evaluierten Finanzvolumens überschreiten, wie dies in mehreren Beispielen in Großbritannien und den USA der Fall war.

Die plausibelsten Schätzungen für außenpolitische Maßnahmen pendeln um 1% der Budgets, deren Maßnahmen evaluiert werden sollen. Bei Jahresbudgets von über 500 Mio. € für Stabilisierung, Krisenprävention, demokratische Transformation und humanitäre Hilfe wäre das keine unwesentliche Summe. Dafür kann eine systematische und effektive M&E-Praxis entscheidende Beiträge dazu leisten, dass die erheblichen Mittel für außenpolitische Prioritäten so sinnvoll und zielführend wie möglich eingesetzt werden. Dabei ist entscheidend, dass Ressourcen weder ausschließlich noch vorrangig zur Finanzierung externer Evaluierungsdienstleister benötigt werden, im Gegenteil: Evaluierungen können nur so gut und nützlich sein wie das zugrundeliegende Projekt- und Evaluierungsmanagement, das heißt die an sie gestellten Evaluierungsfragen und die ihnen zur Verfügung gestellten Informationen der Projektverantwortlichen.

## **Chancen und Anforderungen für eine aktiv lernende Organisation**

Es ist deutlich geworden, dass die untersuchten Außenministerien noch experimentieren, wie sich Monitoring und Evaluierung gewinnbringend in außenpolitische Abläufe und Organisationstraditionen einbetten lassen. Ein abschließendes Urteil über die Chancen und Risiken von M&E in der Außenpolitik wäre also verfrüht. Gleichzeitig fördert der Blick auf die Erfahrungen von Partnern, die sich dem Thema schon etwas länger widmen, durchaus eine Reihe nützlicher Einsichten zutage – nicht zuletzt in der selbstbewussten Auseinandersetzung mit unterschiedlichen Ansätzen und teils unrealistischen Erwartungen an M&E. Die Vielfalt akzeptierter Ansätze, Kriterien und Methoden macht deutlich, dass erheblicher Spielraum bei der Ausgestaltung einer eigenen M&E-Praxis besteht.

Diesen Gestaltungsspielraum gilt es als Chance zu ergreifen, um den eigenen Lernprozess bei der Gestaltung und Umsetzung von außenpolitischen Maßnahmen in fragilen Kontexten zu unterstützen. Das geht nicht von heute auf morgen: Geeignete M&E-Instrumente müssen entwickelt, in Pilotversuchen getestet und behutsam eingeführt werden. Klar ist allerdings, dass eine gewinnbringende Systematisierung von Monitoring und Evaluierung nur dann erfolgen kann, wenn neben dem Interesse an konstruktivem Lernen auch die Möglichkeit zur Investition zusätzlicher Personal- und Sachmittel besteht.

# Literaturverzeichnis

---

- Ashdown, P., und R. Mountain (2011): *Humanitarian Emergency Response Review*. DFID, London.
- Auswärtiges Amt (2012): *Strategie des Auswärtigen Amtes zur humanitären Hilfe im Ausland*. Auswärtiges Amt, Berlin.
- Auswärtiges Amt, Bundesministerium der Verteidigung, und B.f.w.Z.u. Entwicklung (2012): *Für eine kohärente Politik der Bundesregierung gegenüber fragilen Staaten - Ressortübergreifende Leitlinien*. Berlin.
- Beck, T. (2006): *Evaluating humanitarian action using the OECD-DAC criteria. An ALNAP guide for humanitarian agencies*. London: ALNAP.
- Benner, T., S. Eckhard, und P. Rotmann (2013): *Learning in International Organizations*. in: *Routledge Handbook of International Organization*, herausgegeben von Bob Reinalda. London, New York: Routledge, 361-373.
- Benner, T., S. Mergenthaler, und P. Rotmann (2011): *The New World of UN Peace Operations: Learning to Build Peace?* Oxford: Oxford University Press.
- Binder, A., und U. Reichhold (im Erscheinen): *The Metrics and Ethics of Protecting Civilians*. Global Public Policy Institute, Berlin.
- Binder, A., und U. Reichhold (2013): *Scoping Study - What works in protection and how do we know?* Global Public Policy Institute, Berlin.
- Blue, R., C. Clapp-Wincek, und H. Benner (2009): *Beyond Success Stories: Monitoring and Evaluation for Foreign Assistance Results*. USAID.
- Bönke, J.R., J. Koehler, und C. Zürcher (2009): *Evaluation von Entwicklungszusammenarbeit zur Stabilisierung in Post-Konflikt-Zonen: Anwendung eines Mixed-Methods-Surveys in Nordost-Afghanistan*. *Zeitschrift für Evaluation* 8: 2, 215-235.
- Borrmann, A., und R. Stockmann (2009): *Evaluation in der deutschen Entwicklungszusammenarbeit. Sozialwissenschaftliche Evaluationsforschung*. Münster/New York/Berlin/München: Waxmann Verlag.
- Bundesamt für Justiz (2013): *Gesetzgebungsleitfaden*. Bonn.
- Bundesministerium der Finanzen (2013): *Bundshaushaltsplan 2013*. Berlin.
- Bundesministerium der Finanzen (2003-2012): *Jahres- und Haushaltsrechnungen*. Berlin.
- Burchett, J., S. Vogler, und T. Percival (2007): *Legends of Stonehenge*. Oxford: Heinemann.
- Department for International Development (2011): *DFID Research and Evidence Division Operational Plan 2011-2015*. DFID, London.
- Department for International Development (2013): *Evaluating Influence. How To Note*. DFID, London.
- Department for International Development (2013): *International Evaluation Policy*. DFID, London.
- Development Initiatives (2013): *Global Humanitarian Assistance Report*. Bristol.
- Direktion für Entwicklung und Zusammenarbeit (2012): *Evaluation Pays. Principes de sélection des « Peers »*. Bern.
- Direktion für Entwicklung und Zusammenarbeit (2013): *Evaluationspolitik der Direktion für Entwicklung und Zusammenarbeit (DEZA)*. Bern.
- Donaldson, S.I., M.Q. Patton, M. Scriven, D. Fetterman, und Associates (2010): *The 2009 Claremont Debates: The Promise and Pitfalls of Utilization-Focused and Empowerment Evaluation*. *Journal of Multidisciplinary Evaluation* 6: 13, 15-19.
- Eidgenössisches Departement für auswärtige Angelegenheiten (2012): *Bericht über die ausserpolitischen Schwerpunkte der Legislatur (ausserpolitische Strategie 2012-2015), in Erfüllung der Motion 10.3212 (« Klare strategische Ausrichtung der Aussenpolitik »)*. EDA, Bern.
- Eidgenössisches Departement für auswärtige Angelegenheiten (2013): *Evaluation in der AMS. Anleitung*. EDA, Bern.

- Eidgenössisches Departement für auswärtige Angelegenheiten (2012): *Evaluation of the performance of SDC instruments in fragile and conflict-affected contexts*. EDA, Bern.
- FDP. Die Liberalen (o.J.): *Startseite* <http://www.fdp.ch/>, abgerufen am 16.12.2013.
- Foreign and Commonwealth Office (2013): *Departmental Improvement Plan*. FCO, London.
- Foreign and Commonwealth Office (2013): *Official Development Assistance Strategic Projects: FCO evaluations 2012-13*. FCO, London.
- Herson, M., und J. Mitchell (2005): *Real-Time Evaluation: where does its value lie? Humanitarian Exchange Magazine* 35.
- HM Treasury (2011a): *Green Book: Appraisal and Evaluation in Central Government* HM Treasury, [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/220541/green\\_book\\_complete.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/220541/green_book_complete.pdf), abgerufen am 08.01.2014.
- HM Treasury (2011b): *Magenta Book: Guidance for Evaluation*. HM Treasury, [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/220542/magenta\\_book\\_combined.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/220542/magenta_book_combined.pdf), abgerufen am 08.01.2014.
- Howard, L.M. (2008): *UN Peacekeeping in Civil Wars*. Cambridge: Cambridge University Press.
- Independent Commission for Aid Impact (2012): *Evaluation of the Inter-Departmental Conflict Pool*. ICAI, London.
- Independent Commission for Aid Impact (2011): *ICAI's Approach to Effectiveness and Value for Money*. ICAI, London.
- Independent Evaluation Group (2007): *Sourcebook for Evaluating Global and Regional Partnership Programs: Indicative Principles and Standards*. IEG-World Bank, Washington, D.C.
- IOB (o.J.): *About IOB*. <http://www.iob-evaluatie.nl/node/667>, abgerufen am 04.03.2014.
- Kleistra, Y., und N. Van Willigen (2010): *Evaluating Diplomacy: A Mission Impossible*. *Evaluation* 16: 2, 119-135.
- Krüger, S., und E. Sagmeister (2012): *Real-Time Evaluation of humanitarian assistance revisited: Lessons Learned and Ways forward. 10th EES Biennial Conference*. Helsinki, 03.-05-10.2012.
- Lockheed Martin (o.J.): *Worldwide Integrated Crisis Early Warning System*. [http://www.lockheedmartin.com/us/products/W-ICEWS/W-ICEWS\\_overview.html](http://www.lockheedmartin.com/us/products/W-ICEWS/W-ICEWS_overview.html), abgerufen am 03.03.2014.
- Meharg, S.J. (2009): *Measuring What Matters In Peace Operations And Crisis Management*. Kingston, ON: Queen's University Press.
- Ministerie van Buitenlandse Zaken (o.J.): *The Advisory Panel*. <http://www.iob-evaluatie.nl/node/688>, abgerufen am 04.03.2014.
- Ministerie van Buitenlandse Zaken (2010): *Evaluation Programme for the Period 2008-2015*. MBZ, Den Haag.
- National Audit Office (2012): *Review of the Conflict Pool*. Department for International Development, Foreign and Commonwealth Office.
- O'Brian, S. (2010): *Crisis Early Warning and Decision Support: Contemporary Approaches and Thoughts on Future Research*. *International Studies Review* 12: 1, 87-104.
- OECD-DAC (2012): *Evaluating Peacebuilding Activities in Settings of Conflict and Fragility. Improving Learning for Results*. OECD, Paris.
- OECD-DAC (2010): *Evaluation in Development Agencies. Better Aid*. OECD-DAC, Paris.
- OECD (1991): *DAC Principles for the Evaluation of Development Assistance*. OECD, Paris.
- (2000): *Glossary of Evaluation and Results Based Management (RBM) Terms*. OECD, Paris.
- (1986): *Glossary of Terms Used in Evaluation. Methods and Procedures in Aid Evaluation*. OECD, Paris.
- (2007): *Principles for Good International Engagement in Fragile States and Situations*. OECD, Paris.
- Patton, M.Q. (2008): *Utilization-Focused Evaluation*. 4th edition. London: SAGE Publications.
- Policy and Operations Evaluation Department, I. (2009): *Evaluation policy and guidelines for evaluation*. Ministerie van Buitenlandse Zaken, Den Haag.
- Pollitt, C., und G. Bouckaert (2004): *Public Management Reform. A Comparative Analysis*. Oxford: Oxford University Press.

- Ramalingam, B. (2013): *Aid on the Edge of Chaos: Rethinking International Cooperation in a Complex World*. Oxford University Press.
- Rieder, S. (2011): *Kosten von Evaluationen*. LeGes 22: 1, 73-88.
- Rotmann, P., und L. Steinacker (2013): *Stabilisierung: Begriffe, Strukturen und Praxis im Vergleich. Studie im Auftrag des Planungsstabs im Auswärtigen Amt*. Global Public Policy Institute, Berlin.
- Royal Geographical Society (o.J.): *Feeding the 9 Billion: Professor Tim Wheeler*. <http://www.21stcenturychallenges.org/focus/professor-tim-wheeler/>, abgerufen am 13.01.2014.
- Rudolf, P., und S. Lohmann (2013): *Außenpolitikevaluation im Aktionsfeld Krisenprävention und Friedensaufbau*. Stiftung Wissenschaft und Politik (SWP), Berlin.
- Schweizer Parlament (2011): *Geschäftsprüfungskommissionen, Berichte 2011* <http://www.parlament.ch/dokumentation/berichte/berichte-aufsichtskommissionen/geschaefspruefungskommission-GPK/berichte-2011/Seiten/default.aspx>, abgerufen am 03.03.2014.
- Schweizer Parlament (2013): *Parlamentarische Verwaltungskontrolle*. Bern.
- Schweizer Parlament (2013): *Parlamentswörterbuch*. Bern.
- Schweizerische Eidgenossenschaft (2013): *Bundesgesetz über Massnahmen zur zivilen Friedensförderung und Stärkung der Menschenrechte*. Bern.
- Schweizerische Eidgenossenschaft (2013): *Bundesverfassung der Schweizerischen Eidgenossenschaft vom 18. April 1999*. <http://www.admin.ch/opc/de/classified-compilation/19995395/201303030000/101.pdf>, abgerufen am 17.12.2013.
- Schweizerische Eidgenossenschaft (2003): *Medienmitteilung der Bundeskanzlei*. <http://www.news.admin.ch/dokumentation/00002/00015/index.html?lang=de&msg-id=19767>, abgerufen am 03.11.2003.
- Schweizerische Eidgenossenschaft (2013): *Ziele des Bundesrates 2014*. <http://www.bk.admin.ch/dokumentation/publikationen/00290/00928/>, abgerufen am 18.12.2013.
- Schweizerischer Bundesrat (2012): *Botschaft über die Internationale Zusammenarbeit 2013-2016*. Schweizerische Eidgenossenschaft, Bern.
- Schweizerischer Bundesrat (2011): *Botschaft über die Weiterführung von Massnahmen zur Förderung des Friedens und der menschlichen Sicherheit 2012-2016*. Schweizerische Eidgenossenschaft, Bern.
- Strategisches Controlling Abteilung Menschliche Sicherheit (2013): *Selbstevaluation in der Abteilung für Menschliche Sicherheit. Anleitung*. EDA, Bern.
- The Guardian (2010): *Conservative-Liberal Democrat coalition deal: full text*. <http://www.theguardian.com/politics/2010/may/12/lib-dem-tory-deal-coalition>, abgerufen am 01.03.2014.
- US Department of State (2012): *Department of State Program Evaluation Policy*. US Department of State, Washington, D.C.
- US Department of State (2010): *Quadrennial Diplomacy and Development Review*. US Department of State, Washington, D.C.
- Vogel, I. (2012): *Review of the use of 'Theory of Change' in international Development. Review report*. DFID, London.

**Global Public Policy Institute (GPPi)**  
Reinhardtstr. 7, 10117 Berlin, Germany  
Phone +49 30 275 959 75-0  
Fax +49 30 275 959 75-99  
gppi@gppi.net

[gppi.net](http://gppi.net)