

Evidenzorientierte Ansätze in der Entwicklungszusammenarbeit

Erfahrungen, Potenzial und Kernthemen

Paul Marschall



Evidenzorientierte Ansätze in der Entwicklungszusammenarbeit

Erfahrungen, Potenziale und Kernthemen

Paul Marschall

Dr. Paul Marschall ist Ökonom und war von Juli 2017 bis zum September 2021 wissenschaftlicher Mitarbeiter am Deutschen Institut für Entwicklungspolitik (DIE) im Forschungsprogramm „Inter- und transnationale Zusammenarbeit“.

E-Mail: paul.marschall@uni-bayreuth.de

Mit finanzieller Unterstützung des Bundesministeriums für wirtschaftliche Zusammenarbeit und Entwicklung (BMZ)

Dies ist eine Übersetzung von Marschall, P. (2018). *Evidence-oriented approaches in development cooperation: Experiences, potential and key issues* (Discussion Paper 8/2018). Bonn: German Development Institute / Deutsches Institut für Entwicklungspolitik (DIE).

Zitationsvorschlag:

Marschall, P. (2022). *Evidenzorientierte Ansätze in der Entwicklungszusammenarbeit: Erfahrungen, Potenziale und Kernthemen* (IDOS Discussion Paper 10/2022). Bonn: German Institute of Development and Sustainability (IDOS). <https://doi.org/10.23661/idp10.2022>

Disclaimer:

Die Publikation gibt die Meinung der Autor*innen wieder und stellt nicht unbedingt jene des German Institute of Development and Sustainability (IDOS) dar.



Sofern nicht anders angegeben, ist der Inhalt dieser Publikation unter einer Creative Commons Attribution 4.0 International License lizenziert. Es ist erlaubt, dieses Werk zu vervielfältigen, zu verbreiten und zu bearbeiten, solange das German Institute of Development and Sustainability (IDOS) gGmbH und der/die Autor*in(nen) genannt werden.

IDOS Discussion Paper / German Institute of Development and Sustainability (IDOS) gGmbH

ISSN 2751-4439 (Print)

ISSN 2751-4447 (Online)

ISBN 978-3-96021-189-1 (Print)

DOI: <https://doi.org/10.23661/idp10.2022>

© German Institute of Development and Sustainability (IDOS) gGmbH

Tulpenfeld 6, 53113 Bonn

E-Mail: publications@idos-research.de

<http://www.idos-research.de>

Gedruckt auf umweltfreundlichem, zertifiziertem Papier.



Danksagung

Ich danke Stephan Klingebiel, Sarah Holzapfel, Sven Grimm, Anna Schwachula, Julian Bergmann, Heiner Janus und Daniel Esser für ihre hilfreichen und wertvollen Hinweise.

Dieses ursprünglich 2018 auf Englisch erschienene Papier wurde im August 2022 von mir auf Wunsch des IDOS hin übersetzt und wird angesichts aktueller Relevanz evidenzorientierter Ansätze in der Entwicklungszusammenarbeit nun auch in deutscher Sprache veröffentlicht.

Bonn, August 2022

Paul Marschall

Inhalt

Danksagung	III
Abkürzungen	VI
Zusammenfassung	1
1 Einführung	6
2 Evidenz	7
2.1 Hintergrund und Kontext	7
2.2 Konzeptioneller Rahmen	9
3 Die Bereitstellung von Evidenz in der Entwicklungszusammenarbeit	19
3.1 Evidenzquellen	19
3.2 Statistik und Monitoring	20
3.3 Ergebnismanagement	21
3.4 Ergebnisbasierte Ansätze	23
3.5 Evaluierung	24
3.6 Systematische Reviews und Synthesen	28
3.7 Zwischenfazit	28
4 Die Nutzung von Evidenz in der Entwicklungszusammenarbeit	29
4.1 Wahrnehmung	29
4.2 Einbeziehung in Entscheidungsprozesse	32
4.3 Barrieren und unterstützende Faktoren	35
4.4 Zwischenergebnis	37
5 Fallstudie: Der Kopenhagen-Konsens	38
5.1 Hintergrund und Ansatz	38
5.2 Bangladesch-Prioritäten	38
5.3 Würdigung	39
6 Potenziale von Evidenz in der Entwicklungspolitik	41
6.1 Interpretation und Ergebnisdiskussion	41
6.1.1 Umfang der Evidenz	41
6.1.2 Stärke und Qualität der Evidenz	42
6.1.3 Evidenzorientierung	42
6.2 Künftiges Potenzial	44
6.3 Stärken und Schwächen	46
7 Folgerungen und Empfehlungen	46
Literatur	51
Anhang 1: Nutzung von Evidenz in verschiedenen Politikfeldern – einige Beispiele	58

Abbildungen

Abbildung 1: Konzeptionelle Wurzeln von Evidenz	11
Abbildung 2: Die Eigenschaften von Evidenz	14
Abbildung 3: Die Dimension ‚Umfang‘: Evidenztypen	15
Abbildung 4: Wahrnehmung von Evidenz durch politische Entscheidungsträger*innen	17
Abbildung 5: Evidenzorientierung: Ausmaß der Nutzung von Evidenz in der Politikgestaltung	18

Tabellen

Tabelle 1: Qualität der Evidenz im Fall der EBM	16
Tabelle 2: Qualität der Evidenz im erweiterten Rahmen in absolutem Sinn	16
Tabelle 3: Wesentliche Ansätze der Wirkungsevaluierung	27
Tabelle 4: BCURE-Rahmen zum Wert genutzter Evidenz	31
Tabelle 5: Wie Ergebnisinformationen genutzt werden	33

Abkürzungen

3ie	Internationale Initiative für Wirkungsevaluierung (<i>International Initiative for Impact Evaluation</i>)
BCURE	Kapazitätsaufbau zur Nutzung von Forschungsevidenz (<i>Building Capacity to Use Research Evidence</i>)
BMZ	Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung
DAC	Entwicklungsausschuss der OECD (<i>Development Assistance Committee</i>)
DANIDA	Dänische Internationale Entwicklungsagentur (<i>Danish International Development Agency</i>); vom dänischen Außenministerium genutzte Marke für die humanitäre Hilfe und Entwicklungszusammenarbeit
DEval	Deutsches Evaluierungsinstitut der Entwicklungszusammenarbeit gGmbH
DFID	Ministerium für internationale Entwicklung des Vereinigten Königreichs (<i>Department for International Development</i>)
EBM	Evidenzbasierte Medizin
EBPG	Evidenzbasierte Politikgestaltung (<i>evidence-based policy making – EBPM</i>)
EOF	Ergebnisorientierte Finanzierung (<i>results-based financing – RBF</i>)
EOM	Ergebnisorientiertes Management (<i>results-based management – RBM</i>)
EOZ	Ergebnisorientierte Zusammenarbeit (<i>results-based aid – RBA</i>)
EZ	Entwicklungszusammenarbeit
GIZ	Deutsche Gesellschaft für internationale Zusammenarbeit
ICAI	Unabhängige Kommission für die Wirkung von Entwicklungszusammenarbeit (<i>Independent Commission for Aid Impact</i>)
ISS	Unterstützung für Immunisierung (<i>immunisation services support</i>)
KfW	Kreditanstalt für Wiederaufbau
KK	Kopenhagen-Konsens
KNA	Kosten-Nutzen-Analyse
KWA	Kosten-Wirksamkeits-Analyse
LIC	Land mit niedrigem Einkommen (<i>low-income country</i>)
MDG	Millenniums-Entwicklungsziel (<i>Millennium Development Goal</i>)
NRO	Nichtregierungsorganisation
ODA	Öffentliche Entwicklungszusammenarbeit (<i>official development assistance</i>)
OECD	Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (<i>Organisation for Economic Co-operation and Development</i>)
EOF4MNH	Ergebnisorientierte Finanzierung für die Gesundheit von Müttern und Neugeborenen (<i>results-based financing for maternal and newborn health</i>)
RCT	Randomisierte kontrollierte Studie (<i>randomised controlled trial</i>)
SDG	Nachhaltigkeitsziel (<i>Sustainable Development Goal</i>)
USAID	Behörde der USA für die internationale Zusammenarbeit (<i>United States Agency for International Development</i>)

Zusammenfassung

Hintergrund

Die vermehrte Nutzung von Evidenz als Instrument, um die Wirkung der Entwicklungszusammenarbeit (EZ) zu verstärken, ist ein zentrales Thema in den gegenwärtigen Diskussionen darüber, wie diese Wirkung der Unterstützung effektiver gestaltet werden kann. Ausgehend von einem breiten Verständnis des Evidenz-Begriffes trägt dieses Diskussionspapier zur Beantwortung von drei Fragen bei: Erstens, wie wird Evidenz gegenwärtig in der EZ bereitgestellt? Zweitens, in welcher Weise wird Evidenz genutzt? Drittens, welches Potenzial wird die Berücksichtigung von Evidenz künftig bei der Politikgestaltung entfalten?

Evidenz

Evidenz ist ein schillernder Begriff mit verschiedenen Bedeutungen und Konnotationen. Die dieser Arbeit zugrunde liegende Definition beruht auf einem breiten Verständnis: Evidenz ist die Fähigkeit, auf Grundlage verfügbarer empirischer Daten, Informationen und/oder des Wissens darüber, dass Maßnahmen funktionieren, Schlüsse zu ziehen. Das Wissen darüber, was wirkt, ist der zentrale Hebel, um die Wirksamkeit zu erhöhen, die wiederum die geplante und die erzielte Leistung in Beziehung setzt. Bei der Auswahl der Handlungsalternativen steht dies dafür, die grundsätzliche Entscheidung für die Option mit der höchsten Wirksamkeit zu treffen, um damit die ‚richtigen Dinge zu tun‘.

Der konzeptionelle Rahmen dieser Analyse umfasst drei Dimensionen: 1) Das Ausmaß der Nutzung verschiedener Typen und Quellen von Evidenz (Umfang); 2) Die Stärke und Qualität der genutzten Evidenz (Stärke und Qualität); 3) das Ausmaß der Verwendung von Evidenz bei politischen Entscheidungen (Evidenzorientierung).

Evidenz beruht auf verschiedenen Quellen: (i) akademische Evidenz, die breiter als wissenschaftliche Evidenz ist, welche sich nur auf Natur- und Sozialwissenschaften bezieht. Sie ist idealtypisch objektiv, replizier- und verallgemeinerbar und wird als akademische Leistung bereitgestellt; (ii) Ministerien wie das ehemalige Ministerium für internationale Entwicklung des Vereinigten Königreichs (DFID), Durchführungsorganisationen wie die Deutsche Gesellschaft für internationale Zusammenarbeit (GIZ), die Verwaltung, aber auch Nichtregierungsorganisationen (NROen) – in ihrer Funktion als Durchführungsorganisationen der EZ – stellen praktisch-technische Evidenz bereit; (iii) soziale Evidenz bezieht sich auf Daten und Informationen, die direkt und meistens in aggregierter Form von Mitgliedern der Zivilgesellschaft geliefert werden, etwa durch soziale Medien oder durch soziale Bewegungen. Allgemein bezieht sich der Umfang auf die Anzahl der berücksichtigten Evidenztypen. Es gibt beispielsweise verschiedene Hierarchien, die von Akademiker*innen genutzt werden, um Stärke und Qualität der verfügbaren Evidenz in eine Rangfolge zu bringen. Es macht Sinn, derartige Schemata zu erweitern, um auch andere Quellen von Evidenz abzudecken. Schlussendlich wird angenommen, dass politische Entscheidungen auf Meinungen und/oder Evidenz beruhen. Das Ausmaß, in dem politische Entscheidungsträger*innen Evidenz berücksichtigen (Evidenzorientierung), wird durch die Niveaus meinungsbasierte, evidenzunterstützte oder evidenzbasierte Politikgestaltung ausgedrückt.

Bereitstellung von Evidenz in der Entwicklungszusammenarbeit

Ein breites Spektrum an Quellen und Informationen kann genutzt werden, um Evidenz zu generieren und verfügbare Evidenz zu verbreiten. Unstrukturierte Daten bilden das Fundament von Evidenz. Bevor Daten verarbeitet und analysiert werden, ist ihr Wert begrenzt. Besondere Formate wie akademische Zeitschriften, Berichte, andere wissenschaftliche Veröffentlichungen und Mitteilungsblätter werden verwendet, um akademische, praktisch-technische sowie soziale Evidenz zu liefern und zu kanalisieren. Es gibt Plattformen für den Wissensaustausch. Häufig

entstehen Daten und Informationen als Nebenprodukt der EZ. Regierungseinrichtungen und internationale Organisationen – sowie in zunehmendem Umfang nicht staatliche Leistungserbringer*innen der EZ – sammeln Monitoring-Daten. Auf Ebene der Partnerländer werden statistische Daten gesammelt. Eine relevante Quelle für Evidenz sind Haushaltsbefragungen. In den letzten Jahrzehnten entwickelte sich auch Monitoring zu einem wichtigen Instrument, um die Leistung von Projekten und Programmen zu messen und abzubilden. Zusätzlich ermöglichen es Erfolgsindikatoren, den Fortschritt bei der Erreichung von Entwicklungszielen zu messen, wie etwa derzeit hinsichtlich der Nachhaltigkeitsziele (Sustainable Development Goals, SDGs).

Um die Ergebnisorientierung von Projekten und Programmen zu verbessern, wurden in letzter Zeit bestehende Management-Ansätze angepasst. Ergebnisorientiertes Management (EOM) ist ein Konzept, das einen starken Fokus auf den Erfolg und die Erreichung von Outputs, *Outcomes* und Wirkungen legt. Operative Evidenz ist eine bedeutende Hauptquelle für die Implementierung von Lernprozessen. Vor wenigen Jahren wurden ergebnisorientierte Ansätze eingeführt. Sie umfassen viele Instrumente, die darauf abzielen, das Paradigma der Entwicklungszusammenarbeit zu verändern – weg von einem Fokus auf Ressourcen und Aktivitäten und hin zu einer Orientierung auf Ergebnisse bzw. Leistungen. Dies geschieht, indem die Beziehungen zwischen den Partner*innen der Entwicklungszusammenarbeit oder den Geber*innen und den Partnerländern neugestaltet werden. Deshalb müssen quantifizierbare und messbare Ergebnisse ermittelt und so eng wie möglich mit den Wirkungen der EZ verknüpft werden. So helfen beispielsweise Impfunterstützungsprogramme (ISS) von GAVI, der Impfallianz Ländern dabei, ihre routinemäßigen Schutzimpfungen auszuweiten. Sie beruhen auf leistungsabhängigen Zahlungen, die aus Impfdaten berechnet werden, die von den administrativen Berichtssystemen der Länder gewonnen werden.

In jüngerer Zeit haben verschiedene Evaluierungsmethoden jenseits der Leistungsevaluierung von Projekten und Programmen Aufmerksamkeit erregt. Wirkungsevaluierung (*impact evaluation*) zielt darauf ab, vor allem die langfristigen Ergebnisse zu bestimmen, die durch politische Entscheidungen sowie Maßnahmen, Projekte oder Programme hervorgerufen wurden. Diese Wirkungen können positiv oder negativ, beabsichtigt oder unbeabsichtigt, direkt oder indirekt sein. Im Besonderen gelten randomisierte kontrollierte Studien (RCTs, *randomised controlled trials*) – ein methodischer Ansatz, der häufig in der evidenzbasierten Medizin (EBM) und in zunehmendem Umfang in der EZ verwendet wird – als ‚Goldstandard‘, um qualitativ hochwertige Evidenz zu generieren. Dennoch hat der Einsatz von RCTs auch seine Grenzen – vor allem im sozialen Kontext, in dem jede Versuchsgruppe anders ist. Nach der Durchführung einer Studie ist die untersuchte Gruppe voreingenommen. Daher ist das Handeln dieser Personen möglicherweise verzerrt. Deshalb müssen derartige Schwachstellen in systematischen Reviews berücksichtigt werden, die mächtige Instrumente der Synthese verfügbarer Evidenz sind.

Nutzung von Evidenz in der Entwicklungszusammenarbeit

Die Verwendung von Evidenz für analytische und operative Zwecke hängt von ihrer verfügbaren Menge und Qualität ab. Bedeutsam sind darüber hinaus der zeitliche Rahmen der Herkunft sowie die Einmaligkeit der Evidenzgrundlage (beispielsweise mangelnde Reproduzierbarkeit von Studienergebnissen). Bis vor kurzem wurde Evidenz bezüglich dessen, was sowohl für sozio-ökonomische Entwicklung als auch hinsichtlich der EZ funktioniert, nicht systematisch gesammelt. Nur spezifische Evidenz zu konkreten Forschungsergebnissen hinsichtlich besonderer geographischer Gegebenheiten und einzelner Forschungsbereiche war verfügbar. Lange Zeit standen Ergebnismanagement und ergebnisorientierte Ansätze noch am Anfang ihrer Anwendung, und es gab kaum verfügbare Evidenz zur Leistung von Projekten und Programmen. Insgesamt waren die Qualität der Evidenz gering und der damit verbundene Wert für die Entscheidungsfindung sehr begrenzt. In den letzten Jahren hat sich das quantitative und qualitative Potenzial der Evidenz in der EZ als Ressource für Politikgestaltung und für die Entscheidungs-

findung wesentlich verbessert. Die Anzahl der Evaluierungen und ihre Qualität haben zugenommen. Internationale Netzwerke und Organisationen leisten nun Kapazitätsaufbau, um mehr und bessere Evidenz zu gewinnen. Es gibt Institute und Personen, die mit dem Ziel der Bewusstseinsbildung sensibilisieren. Sie sind auf eine verstärkte und verbesserte Nutzung von Evidenz auf der Ebene der Entscheidungsträger*innen sowie von Verwaltungsbehörden ausgerichtet. Es besteht dennoch weiterhin Verbesserungspotenzial – etwa mit Blick darauf, in die Politik mehr Begleitforschung einzubinden.

Evidenz spielt in der Politikgestaltung eine Rolle und wird auf zweifache Weise genutzt: 1) symbolisch, um die Glaubwürdigkeit von Entscheidungsträger*innen und ihren Entscheidungen zu erhöhen und 2) instrumentell, um bestehendes Wissen anzupassen und die Entscheidungsfindung zu verbessern. Aufgrund starker Impulse durch nationale und internationale Initiativen wurde das Bewusstsein für den Wert von Evidenz in der EZ auf markante Weise geschärft. Allerdings zeigen die Wissensübertragung und Nutzung von Evidenz in verschiedenen Kontexten ein gemischtes Bild. Die Nachfrage nach und die Nutzung von Evidenz unterscheidet sich beträchtlich. Pionierstaaten wie Australien, Kanada, Neuseeland oder das Vereinigte Königreich haben das Thema auf die Agenda gesetzt und arbeiten am evidenzorientierten Aufbau von Kapazitäten, auch in Partnerländern. Ein wichtiges Thema ist die Wirkungsevaluierung (Impact Evaluation). Sie liefert Informationen sowohl für die Verbesserung laufender Aktivitäten als auch für die Gestaltung neuer Projekte. Die Vorteile aus der Bereitstellung von Evidenz aus dem Ergebnismanagement sind bisher noch nicht voll ausgeschöpft. Grund dafür sind die Komplexität von Ergebnismatrizen sowie Messprobleme. Es gibt noch weitere Hürden, die dafür sorgen, dass Evidenz nicht genutzt wird. Politische Entscheidungsträger*innen sind oftmals politische Akteur*innen. Deshalb sind die zugrunde liegenden Werte, politischen Vorstellungen und das Zusammenwirken mit dem politischen System bedeutsam. In der Praxis ist Evidenz nur einer von mehreren Faktoren, die in Politikprozessen eine Rolle spielen. Manchmal wird Evidenz bei Entscheidungen nicht berücksichtigt, da sie zum Zeitpunkt des Bedarfs nicht verfügbar ist. Neben anderen Gründen hängt die Berücksichtigung von Evidenz durch politische Entscheider*innen in kritischer Weise davon ab, ob das bereitgestellte Wissen als Lösung für ein bestehendes Problem aufgefasst wird oder nicht.

Fallstudie: der Kopenhagen-Konsens

Aufgrund wenig geeigneter Transmissionsformate gibt es nach wie vor gewaltige Herausforderungen, um akademische Evidenz in die praktische Verwendung zu überführen. Gegenwärtig sind soziale Medien und Marketingkampagnen, wie der Kopenhagen-Konsens (KK), wichtige Instrumente, um Aufmerksamkeit zu wecken. Das *Copenhagen Consensus Centre* ist ein internationaler Thinktank, der in Kopenhagen ansässig ist und im Jahr 2002 gegründet wurde. Seit 2004 wurden verschiedene Projekte durchgeführt, die sich weitgehend auf Entwicklungsprobleme auf nationaler, regionaler und globaler Ebene fokussierten. Auf Grundlage des ökonomischen Standardkonzepts Kosten-Nutzen-Analyse (KNA), einem Bewertungsverfahren, das bei Ex-post-Evaluierungen verwendet wird, wurden optimale Lösungen für die betrachteten Entwicklungsherausforderungen ermittelt. Bei KNA wird der Nettonutzen von Investitionen berechnet, indem der erwartete Ertrag alternativer Maßnahmen (= Preis-Leistungs-Verhältnis) miteinander verglichen wird. Die Logik dahinter beruht darauf, effiziente Optionen auszuwählen. Die Analyse des KK-Ansatzes ist aufschlussreich, da sie als Paradebeispiel für akademische Evidenz und die entsprechenden Barrieren bei der Nutzung akademischer Erkenntnisse für politische Entscheidungen verstanden werden kann. Das KK-Konzept wird allerdings kontrovers diskutiert. Denn es handelt sich um ein Vorhaben, das von Ökonom*innen dominiert wird. Viele Akademiker*innen schätzen den Wert von Hintergrundberichten, die als wichtige Ressourcen für eine KNA genutzt werden. Sie lehnen jedoch einige der Modelle und Annahmen ab, die in einer KNA verwendet werden. Viele Akademiker*innen, die im Bereich der Umweltwissenschaften arbeiten, akzeptieren beispielsweise die Ergebnisse des KK aufgrund der Art und Weise nicht, wie sie methodisch implementiert wurden. Resultate des KK wurden dafür kritisiert, dass

sie sektorenübergreifende Ansätze vernachlässigen. Gestützt auf simultane Entscheidungen zwischen zahlreichen Problemen und Lösungen entstehen isolierte Silos der Finanzierung und spezifische Prioritätenlisten.

Das Potenzial der Evidenz in der Entwicklungszusammenarbeit

Wege zur erfolgreichen Evidenznutzung beruhen auf begleitenden Maßnahmen, einschließlich laufender Politikberatung. Es ist wichtig, verfügbare Evidenz kritisch zu beleuchten, da diese nicht immer konsistent und implementierbar ist. Aufgrund der fortlaufenden Verbindung zwischen den Anbieter*innen und Nachfrager*innen von Evidenz, ist geeignete Evidenz ermittelbar und wird bereits zur Verfügung gestellt. Um zu verstehen, warum bekannte Evidenz bei politischen Entscheidungen und deren Umsetzung nicht berücksichtigt wird, muss die Theorie der politischen Ökonomie beachtet werden. Trotz internationaler Vereinbarungen, die auf eine höhere Wirksamkeit der Entwicklungszusammenarbeit abzielen, und trotz ausgesprochener Lippenbekenntnisse zur ‚Bedeutung von mehr evidenzbasierten politischen Entscheidungen‘ sind politische Akteur*innen oftmals nicht daran interessiert, das vollständige Menü bereitgestellter Evidenz zu berücksichtigen. Von politischen Entscheidungsträger*innen werden oft Bewertungsschemata begrüßt, zu denen Darstellungen des Preis-Leistungs-Verhältnisses verschiedener Maßnahmen zählen, wie sie vom *Copenhagen Consensus Centre* veranschaulicht werden. Mit derartigen Listen wird jedoch häufig Rosinenpickerei betrieben. Bestehende Unsicherheiten hinsichtlich der Zukunft und ernsthafte Zweifel an Berechnungen, die von Wissenschaftler*innen und Thinktanks geliefert werden, können sehr leicht als Entschuldigung dafür genutzt werden, Entscheidungen gerade nicht auf Grundlage von Evidenz zu treffen.

Die gegenwärtige Diskussion über den Einsatz von mehr Evidenz in der EZ fokussiert wesentlich auf akademische und praktisch-technische Evidenz. Die direkte Einbindung von sozialer Evidenz, verstanden als aggregierte Erfahrungen und Wahrnehmungen von Einzelpersonen und im Besonderen von der Zivilgesellschaft, hängt vom konkreten Kontext ab. Beispielsweise ist es nachgewiesen, dass sie im Gesundheitssektor manchmal genutzt wird, weil die Zivilgesellschaft als eine wichtige beteiligte Partei verstanden wird. Auf hochrangiger Ebene wird soziale Evidenz berücksichtigt, indem die Zivilgesellschaft oder ihre Vertreter*innen bei politischen Entscheidungen einbezogen werden. Auf der operativen Ebene besteht mehr Spielraum, um soziale Evidenz zu beachten.

Die Praxis der EZ zeigt, dass der ‚Ruf nach mehr akademischer Evidenz‘ oftmals zu kurz greift. Enge und intensive Zusammenarbeit zwischen Akademiker*innen und den agierenden Einheiten von Durchführungsorganisationen sowie ein offener Dialog können eine Win-win-Situation für alle Partner*innen ermöglichen. Sie können auch ein Beweis für valide Evidenz sein, die auf der operativen Ebene eine wichtige Lernquelle ist. Im Zeitablauf hat sich die Stärke und die Qualität der verfügbaren Evidenz auf beeindruckende Weise verbessert. So sind mittlerweile viele Evaluierungen auf Grundlage rigoroser Methoden verfügbar.

Im Vergleich zur Vergangenheit hat sich das Nutzungsniveau in der EZ substanziell erhöht. Allerdings ist der Ruf politischer Entscheidungsträger*innen nach mehr Evidenz manchmal dadurch motiviert, bereits getroffene Entscheidungen zu rechtfertigen. Im fortgeschrittenen Stadium wirtschaftlicher Entwicklung besteht wahrscheinlich ein besseres Verständnis und eine gesteigerte Wertschätzung von Evidenz. In einem derartigen Umfeld ist die verfügbare Qualität der Evidenz vermutlich besser. Die Bevölkerung – vor allem die Personen mit einem akademischen Hintergrund – wachsen in einer stärkeren evidenzorientierten Kultur auf. Trotzdem bedeutet dies nicht gleichzeitig, dass politische Prozesse automatisch Evidenz absorbieren. Evidenz muss auf angemessene Weise bereitgestellt werden, etwa im Rahmen einer Ko-Produktion mit den relevanten beteiligten Parteien. Sehr wichtig sind die Anzahl der beteiligten Parteien und die Homogenität ihrer (besonderen) Interessen. Im Fall von Unternehmen mit einem vorherrschenden Fokus auf Gewinnmaximierung ist die Nutzung von Evidenz vermutlich höher als

im Fall politischer Entscheidungen in einem mehrdimensionalen politischen Raum. Politische Prioritäten spielen ebenfalls eine Rolle. Aus diesem Grund weicht die Zuteilung finanzieller Mittel im EZ-Kontext von den Preis-Leistungs-Überlegungen ab, die der Kopenhagen-Konsens vorschlägt. Zusammengefasst bedeutet dies, dass Politikgestaltung gegenwärtig als evidenzunterstützt verstanden werden kann. NROen nehmen die Bedeutung verfügbarer Evidenz bei der Akzentuierung ihrer Forderungen zunehmend wahr.

Ähnlich wie in den Aktivitätsfeldern Gesundheit und Bildung wird die Bedeutung von Evidenz in der Entwicklungszusammenarbeit künftig ansteigen. Um das konkrete Relevanzniveau von Evidenz in der Politikgestaltung zu verbessern, muss daran gearbeitet werden, Evidenz viel mehr als bisher zu generieren und bereitzustellen. Die Art und Weise, Evidenz für die praktische Anwendung zu transformieren, ist zu verbessern. Zudem sollten politische Entscheidungsträger*innen ein besseres Verständnis für die Nützlichkeit von Evidenz entwickeln. Dies ist zumindest aus Gründen der Rechenschaftspflicht erforderlich, damit die Maßnahmen erfolgreicher werden, die dazu dienen, ihre politischen Ziele zu erreichen. Für das Gelingen dieser Prozesse ist es sehr hilfreich, eine evidenzorientierte Kultur zu etablieren, die auf transformativer Forschung basiert und die alle gesellschaftlichen Akteure im Forschungsprozess durch die Mitgestaltung und eine Gemeinschaftsproduktion des Wissens einbezieht.

Folgerungen und Empfehlungen

Um die Wirkung der deutschen EZ zu steigern, ist ein stärkerer Fokus auf Evidenz notwendig. Allgemein ist es hilfreich, eine starke evidenz- und ergebnisorientierte Kultur in allen Teilen des deutschen EZ-Systems zu etablieren. Es besteht ein Mehrbedarf an systematischem Lernen. Indem die institutionellen Strukturen verbessert werden, kann ihm nachgekommen werden.

Es sollte das Ziel sein, Evidenz zeitnah und effizient in die Entscheidungsfindung mit einzubeziehen. Unter Berücksichtigung der spezifischen Facetten der Problemstellung, einschließlich des Zeithorizonts, sollten die notwendigen Evidenztypen und ihre Bestandteile identifiziert und gesammelt werden.

Es ist eine zentrale Mission der deutschen EZ und derjenigen, die von ihr unterstützt werden, immer wieder neu zu verorten, ‚was funktioniert‘. Wenn Entscheidungen anstehen oder getroffen werden, sollten die zu nutzende Art der Evidenz sowie ihre jeweilige Stärke spezifiziert werden. Es kann hilfreich sein, geeignete Kriterien für die Bewertung der Evidenzstärke, wie Validität, Zeithorizont und Zuverlässigkeit, zu bedenken und zu definieren.

Um evidenzorientierte Politikgestaltung zu unterstützen, kann es nützlich sein, spezifische Studien und Evaluierungen durchzuführen. Verfügbare Evidenz muss das Rückgrat von Politikgestaltung bilden. Wichtig und verbesserungsfähig sind zudem das Ergebnismanagement für die Rechenschaft und eine Optimierung der zu treffenden Entscheidungen.

Entscheidungsträger*innen in Partnerländern müssen dabei unterstützt werden, Evidenz in ihrer Politikgestaltung anzuwenden. Um deren Wirkung in vernachlässigten Politikfeldern zu verbessern, müssen bestehende Evidenzlücken identifiziert und adressiert werden. Um Handlungen effektiver zu gestalten, müssen Leistungsträger*innen ihre Arbeit im Netzwerk und die Bündelung verfügbarer Ergebnisse verbessern.

1 Einführung

Ein unlängst veröffentlichter Bericht der Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD) widmet sich dem Thema ‚Daten für Entwicklung‘ (OECD, 2017). Die Autor*innen weisen auf den Bedarf an besserer Evidenz hin, damit fundiertere Entscheidungen über die Prioritäten und Strategien zur Erreichung der nachhaltigen Entwicklungsziele (Sustainable Development Goals, SDGs) getroffen werden können.

‚Evidenz‘ ist ein schillernder Begriff. In der Wissenschaft wird dieses Wort bereits seit Jahrhunderten verwendet. Die lateinische Wurzel (*evidere*) erinnert uns daran, dass wir die Aufmerksamkeit auf das richten sollten, was gesehen werden kann. Der Begriff Evidenz wird inzwischen sowohl in der beruflichen Praxis als auch in der Alltagssprache verwendet.

‚Evidenz‘ ist zu einem häufig verwendeten Schlagwort mit verschiedenen Konnotationen geworden. Es ist jedoch nicht immer klar, worauf Evidenz beruht. Sie wird als Überbegriff verwendet und ist entweder austauschbar mit ‚empirischer Information‘ als solcher oder im engeren Sinn, als die Fähigkeit zu verstehen, Folgerungen zu ziehen oder zu unterstützen. Und zwar dann, wenn verfügbares Tatsachenmaterial oder Informationen vorliegen, die belegen, dass eine Überzeugung, eine Behauptung oder Theorie offensichtlich wahr oder valide ist.

Ganz konkret ist es oftmals bei Vorträgen etwa für das Publikum unklar, was der oder die Vortragende mit dem Ausdruck ‚Evidenz‘ tatsächlich meint. Manchmal erschließt sich die Bedeutung erst, wenn der konkrete Kontext berücksichtigt wird.

Die Welt der Evidenz ist paradox. Evidenz beruht auf unstrukturierten Daten, die aufbereitet und analysiert werden müssen. Selbst die besten statistischen Ämter sind nicht in der Lage, alle Unterziele der SDGs adäquat abzudecken. In einigen Entwicklungsländern bestehen immer noch gewaltige Datenlücken. Dennoch ist die Nutzung von ‚Big Data‘, also von großen Datenmengen, ein wichtiges Thema. Um Evidenz zu produzieren, müssen verfügbare (Meta-)Daten analysiert sein. Gegenwärtig gibt es eine intensive Diskussion über Datengewinnung (data mining), auch mit Blick auf den Datenschutz.

Mit Blick auf die akademische und politikberatende Literatur scheint es auf globaler Ebene eine Dynamik hin zu mehr evidenzbasierter Politikgestaltung (EBPG) zu geben (OECD, 2017), obwohl es unterschiedliche Meinungen darüber gibt, worum es sich bei evidenzinformierter Politik tatsächlich handelt oder welcher Zusammenhang zu Agenden besteht, die Politik beeinflussen (Hewman, Capillo, Famurewa, Nath, & Siyanbola, 2013). Eine Vielzahl an evidenzorientierten Ansätzen wurde in verschiedenen Politikfeldern und Disziplinen wie Psychologie, Rechtspflege, Pflegewissenschaft, Bildung, Sozialpflege, Transport sowie Bibliotheks- und Informationswissenschaften entwickelt (vgl. z. B. Davies, Nutley, & Smith, 2000; Young, 2013). Diese neuen Diskussionsforen wurden weitgehend von der ursprünglichen philosophischen Debatte über Evidenz entkoppelt (Achinstein, 2001). Wichtige Philosoph*innen wie Franz Brentano und Edmund Husserl brachten vor, dass Evidenz nicht verlagert bzw. anderweitig verwendet werden kann. Aufgrund ihrer direkten Verbindung zur Wahrheit verfüge Evidenz immer über eine absolute Bedeutung. Von zentraler Bedeutung ist im Unterschied dazu das relative Verständnis von Evidenz mit ihren verschiedenen Hierarchieniveaus und den damit in Zusammenhang stehenden Folgen für Anwendungen im Bereich der evidenzbasierten Medizin (EBM) oder für die Verwendung von Evidenz im Bereich der Entwicklungszusammenarbeit (EZ).

Die Diskussion über ‚mehr Evidenz‘ weist eine starke Verbindung zu anderen erkennbaren gesellschaftlichen Dynamiken auf. Wie etwa eine verbesserte Informiertheit der Bevölkerung oder die zunehmend allgegenwärtigen Daten, Informationen und bestimmten Fakten. Der Ruf nach Entscheidungen, die stärker evidenzorientiert sind, geht auf die Mängel bestehender Konzepte in der Entscheidungsfindung zurück und auf die damit verbundenen enttäuschenden Ergebnisse. Es besteht die Hoffnung, dass stärker evidenzorientierte Entscheidungen die Ergebnisse

bzw. Leistungen im Sinne von Outputs, *Outcomes* und Wirkungen verbessern und so knappes, verfügbares Geld eingespart werden kann. Die Anzeichen mehren sich, dass Akademiker*innen Evidenz zunehmend bereitstellen. Auch die Zahl akademischer Zeitschriften, die Evidenz liefern, wächst kontinuierlich. Zahlreiche Studien, Analysen und Evaluierungen sind auf breiter Basis im Internet frei zugänglich.

Ein Großteil der Literatur, die Evidenz verfügbar macht, legt ihren Fokus klar auf die akademische Gemeinschaft. Allerdings ist ihr praktischer Nutzen aufgrund der restriktiven Art und Weise ihrer Darstellung beschränkt. Zahlreiche Berichte, die politische Entscheider*innen direkt adressieren, erreichen die intendierte Zielgruppe nicht. Beispielsweise investiert die Weltbank mehr als ein Viertel ihres Budgets in länderbezogene Dienstleistungen in Form von Wissensprodukten. Unlängst haben Forschende ermittelt, dass nahezu ein Drittel der von der Weltbank archivierten, politikbezogenen Berichte, die sich auf die Wirkungen zahlreicher Projekte beziehen – vom Dammbau bis hin zu Mikrokrediten – niemals aus dem Internet heruntergeladen wurden (Doemeland & Trevino, 2014). Das dahinterliegende Problem beschränkt sich nicht auf die Weltbank.

Evidenzbasierte Politikgestaltung und ihre Praxis ist kein neues Thema in der EZ. Das Konzept wurde bereits in den Leitfaden der Millenniumsziele der Vereinten Nationen (Millennium Development Goals, MDGs) aufgenommen. Dort steht, dass sich evidenzbasierte Politik auf einen Politikprozess bezieht. Dieser hilft den Planenden, Entscheidungen auf einer besseren Informationsgrundlage zu treffen, indem die beste verfügbare Evidenz im Zentrum des inhaltlichen Politikprozesses steht (United Nations Development Programme, 2007). Dennoch hat erst unlängst der Ruf nach mehr Evidenz eine intensivere Diskussion in der EZ ausgelöst.

Es ist das wesentliche Ziel dieses Arbeitspapiers, Bereitstellung und Nutzung von Evidenz im Politikfeld der EZ zu analysieren.

Auf der Grundlage eines breiten Verständnisses von Evidenz trägt diese Arbeit zur Beantwortung von drei Fragen bei: Erstens, wie wird Evidenz gegenwärtig in der EZ bereitgestellt? Zweitens, in welcher Weise wird Evidenz genutzt? Drittens, welches Potenzial entwickelt die Berücksichtigung von Evidenz künftig bei der Politikgestaltung? Dies bezieht sich darauf, wie und in welchem Kontext Evidenz verwendet wird.

Deshalb wird im zweiten Kapitel ein umfassender Rahmen für die Untersuchung der verschiedenen Kategorien von Evidenz entwickelt. Dazu zählen Typen und Formen von Evidenz sowie ihre Qualität und Stärken. Weiterhin wird die Frage der Bewertung der Verwendung von Evidenz untersucht. Das dritte Kapitel gibt einen Überblick über verschiedene Formen der Evidenz in der Entwicklungszusammenarbeit. Kapitel 4 bietet einige Informationen über die Nutzung von Evidenz in der EZ. Als Fallstudie über die Bereitstellung von Evidenz für die politische Entscheidungsfindung wird der Ansatz des dänischen Thinktanks *Copenhagen Consensus Centre* in Kapitel 5 vorgestellt und diskutiert. Nachfolgend in Kapitel 6 wird die Wirkung bestehender Evidenz in der Entwicklungszusammenarbeit analysiert. Schließlich werden in Kapitel 7 einige Folgerungen bezüglich der Perspektiven von Evidenz in diesem Politikfeld gezogen.

2 Evidenz

2.1 Hintergrund und Kontext

Die Wurzeln des konzeptionellen Verständnisses von Evidenz können bis zu den frühen Schriften der Philosophie der Wissenschaft und der Epistemologie zurückverfolgt werden. Dort ist mit Evidenz eine Vorstufe der Wahrheit gemeint. Einem breiten Verständnis nach drückt der Begriff aus, dass Tatsachenmaterial oder Informationen vorliegen, die belegen, dass eine Überzeugung, eine Behauptung oder Theorie offensichtlich wahr oder valide ist.

Die dieser Arbeit zugrundeliegende Definition beruht auf diesem Verständnis: Evidenz ist die Fähigkeit, auf Grundlage verfügbarer empirischer Daten, Informationen und/oder Wissen darüber, dass Maßnahmen funktionieren, Schlüsse zu ziehen.

Das Wissen darüber, was wirkt, ist der zentrale Hebel zur Stärkung dieser Wirksamkeit, also des Verhältnisses von geplanter und erzielter Leistung. Bei der Auswahl der Handlungsalternativen bedeutet dies, sich grundsätzlich für die Option mit der höchsten Wirksamkeit zu entscheiden, um dadurch die ‚richtigen Dinge zu tun‘. Allerdings müssen dabei noch zusätzliche Faktoren wie die damit verbundenen Kosten (Kosten-Wirksamkeit) berücksichtigt werden.

Forderungen nach dem Einsatz von mehr Evidenz gibt es in vielen Handlungsfeldern und in Politikbereichen, die sich in einem dynamischen und sich verändernden sozialen, politischen, umweltbezogenen und technologischen Kontext befinden. Anhang 1 zeigt dazu einige Beispiele. Seit einigen Jahren gibt es Initiativen zur Förderung engerer und effektiverer Verbindungen zwischen Forschung und Politik (Oliver, Lorenc, & Innvær, 2014). Das Feld von Medizin und öffentlichem Gesundheitswesen ist die erfolgsversprechendste Disziplin. Auf der Grundlage erster Überlegungen im 19. Jahrhundert (Hjørland, 2011) und bahnbrechender Arbeiten zur Effektivität und Effizienz im Jahr 1972 wurde 1992 das Konzept der evidenzbasierten Medizin (EBM) eingeführt (Evidence-Based Medicine Working Group, 1992). Es wurde zu einer wichtigen Grundlage im Denken über den Wert von Evidenz in Disziplinen wie Soziologie und Politikwissenschaft. Dennoch gibt es bedeutende Unterschiede zwischen der Implementierung von EBM und der Bereitstellung von Evidenz in anderen Kontexten. So etwa bei politikbezogenen Entscheidungen in anderen Gesellschaftsbereichen (Pawson, 2006).

Im Allgemeinen gibt es zwei wesentliche Stränge evidenzorientierter Diskussionen.

- *Strang A* fokussiert die Mikro-Ebene der *Praxis*. Programme oder Praktiken sind evidenzbasiert, wenn sie effektiv die beste Forschungsevidenz mit einer Art von praktischer Erfahrung und – falls zutreffend – kultureller Kompetenz sowie mit den Werten der Personen, welche die entsprechenden Leistungen erhalten, vereinen.
- *Strang B* zielt darauf ab, Evidenz aus verschiedenen Quellen auf die politische Ebene zu übertragen. Häufig beschränkt sich die Diskussion auf die akademische Evidenz. Dennoch können andere Quellen sowie andere Typen von Evidenz bedeutsam sein. Ein Beispiel dafür ist evidenzbasierte öffentliche Gesundheitspolitik.

Zahlreiche Faktoren haben zur verstärkten Bedeutung von Evidenz in Politik und Praxis beigetragen: So hat etwa die Nachfrage nach Evidenz vonseiten der gut gebildeten und gut informierten Öffentlichkeit zugenommen. Oder die ‚Explosion‘ im Bereich der Datenverfügbarkeit aller Typen, das Wachstum des Umfangs und der Fähigkeiten der Forschungsgemeinschaft sowie eine zunehmende Betonung der Produktivität und der internationalen Wettbewerbsfähigkeit (Nutley & Webb, 2000). In der EZ wurde das anwachsende Bewusstsein bezüglich ‚funktionierender Entwicklungszusammenarbeit‘ durch die MDGs, den Monterrey Konsens und die Pariser Erklärung zur Wirksamkeit der Entwicklungszusammenarbeit vorangetrieben. Viele Nichtregierungsorganisationen (NROen) haben ebenfalls das Bedürfnis nach breiter aufgestellten Interventionen zur Erreichung von Entwicklungszielen erkannt.

Die Wahrnehmungen darüber, was Evidenz ist und wie sie bei der Ausgestaltung von Politik und in der Praxis genutzt werden sollte, sind vielfältig. Es ist allgemein anerkannt, dass es nicht nur einen einzigen Typ von Evidenz gibt, der alle Antworten geben kann. Mehrere Typen oder Dimensionen können unterschieden werden. So differenziert Tytler (2001) zwischen drei Dimensionen von Evidenz.

Formale akademische Evidenz beruht auf akademischen Ergebnissen. Häufig bilden Theorien und Modelle den Ausgangspunkt für Forschungsfragen. Manchmal werden Theorien auf der

Grundlage von Datensätzen überprüft. Es gibt eine umfangreiche Forschung dazu, kausale Zusammenhänge zwischen Variablen zu identifizieren, um darauf gestützt Folgerungen zu ziehen. EBM ist ein Beispiel dafür. Sie legt den Fokus auf konkrete Krankheiten und beruht auf zahlreichen akademischen Studienergebnissen. Empfehlungen (Richtlinien) werden basierend auf der Information gegeben, ob eine konkrete Therapie, die etwa auf die Einnahme von Arzneimittel gestützt sein kann, für eine Patientin oder einen Patienten mit bestimmten Eigenschaften vorteilhaft ist. Evidenz ist der Beweis dafür, dass eine Therapie unter bestimmten Umständen funktionieren könnte. Im Unterschied zur formalen Evidenz bezieht sich *informelle Evidenz* auf den 'gesunden Menschenverstand' oder auf persönliche Erfahrungen, die außerhalb der begrenzten Reichweite der akademischen Evidenz zu verorten sind. Ein ähnlicher Ausdruck ist die *anekdotische Evidenz*. Ein dritter Evidenztyp beinhaltet *weitere Aspekte*, die durch Evidenz beeinflusst werden, wie etwa *umweltbezogene oder gesetzliche Belange*. In der Rechtsprechung stellt Evidenz das Mittel zur Wahrheitsfindung dar, wonach – sanktioniert von Regeln und Gesetzen – in Gerichtsverfahren Tatsachen beachtet werden.

In der Literatur werden noch weitere Evidenzklassifikationen vorgeschlagen. Das wesentliche Problem dieser Typologien ist, dass die Evidenztypen, die dabei berücksichtigt werden, oft nicht klar voneinander abgegrenzt werden. So unterscheiden etwa Buetow und Kenealy (2000) zwischen wissenschaftlicher, theoretischer, praktischer, Expert*innen-bezogener, rechtlicher und ethisch basierter Evidenz. In vielerlei Hinsicht überlappen sich diese Kategorien. Es kann auch angezweifelt werden, ob detailliertere Untergliederungen wirklich hilfreich sind. Beispielsweise beinhaltet praktische und Expert*innen-bezogene Evidenz auch informelle Evidenz. Zusätzlich kann es hilfreich sein, im Sinne eines situationsbezogenen Verständnisses Praktiker*innen oder Expert*innen als ganzheitlich Personen, die in ihrer eigenen Umwelt agieren, zu betrachten und zudem die Natur sozialer Interaktionen zu berücksichtigen.

Buetow und Kenealy (2000) weisen darauf hin, dass viele Fragen unterschiedliche Facetten beinhalten. Sogar eher technische Entscheidungen können nicht in einem moralischen oder sozialen Vakuum betrachtet werden, insbesondere dann nicht, wenn sie soziale Phänomene betreffen. Dies ist auch ein wichtiger Punkt in der Entwicklungszusammenarbeit. Interventionen dürfen nicht nur durch ein Verständnis geleitet werden, das in der europäischen Kultur und in deren Werten verwurzelt ist. Sie müssen auch zur Kultur der intendierten Begünstigten passen. Mit Blick auf wissenschaftlich fundierte Entscheidungen ist es wichtig, die relevanten Evidenztypen zu identifizieren. Die Regressionsanalyse ist eine wichtige Methode zur Identifikation von Evidenz. Allerdings zählt es zu den Schwachpunkten der Regressionsanalyse, dass die korrespondierende Evidenz darauf beruht, was für den Durchschnitt einer bestimmten Population gilt, also etwa wie sich Personen oder Objekte verhalten. Dafür werden Maße zentraler Tendenz genutzt. Entscheidungen, die auf der Evidenz des Durchschnitts beruhen, können der Mehrheit der Bevölkerung zugutekommen. Es ist aber möglich, dass benachteiligte Gruppen ('vulnerable groups') dadurch ausgeschlossen werden. Dies wird durch das zentrale Motiv der SDGs thematisiert, das darauf abzielt, niemanden zurückzulassen ('Leaving No One Behind').

2.2 Konzeptioneller Rahmen

Daten, die nicht manipuliert sind und über ein bestimmtes Qualitätsniveau verfügen, werden als Fakten genutzt. Sie bieten eine Beschreibung der realen Welt. Daten sind Bausteine, die dafür verwendet werden, Wissen zu bilden. Zusätzlich können sie als Informationen erfasst werden, so dass andere Menschen zu unterschiedlichen Zeitpunkten darauf Zugriff haben können. Information und Wissen sind keine synonymen Begriffe, da verfügbare Informationen verarbeitet und absorbiert werden müssen. Wissen kann als Kapitalstock verstanden werden. Zudem resultiert Wissen auch aus der Kombination einzelner Bestandteile von Wissen. Es ist an die einzelne Person gebunden. Scott-Findlay & Pollock (2004) betonen, dass nur Informationen, Theorien, Erfahrungen und Forschung von Person zu Person übertragbar sind.

Wissen basiert auf theoretischen und/oder empirischen Einsichten. Allerdings muss Wissen nicht immer als wahr aufgefasst werden. Unser Verständnis kann auf falschen oder verzerrten Daten beruhen. Schwierigkeiten bezüglich der Auslegung von Daten kann zu einem verzerrten Wissen führen. Lernen ist ein Prozess, durch den der angesammelte Wissenskapitalstock erweitert wird. Es ist möglich, Fachwissen entweder empirisch durch (eigene) Erfahrungen, Beobachtungen oder durch theoretische Einsichten zu erwerben.

Evidenz ist nicht synonym zu Fachwissen, auch wenn – entsprechend dem Evidenztyp – die Konturen verschwimmen können. Im für diese Studie genutzten konzeptionellen Rahmen wird angenommen, dass es sich bei Evidenz um eine Teilmenge des Fachwissens handelt, in der vier klare Kriterien erfüllt sind:

1. Evidenz beinhaltet nur Fachwissen auf empirischer Grundlage.
2. Im Fall von Evidenz werden Daten und Informationen auf systematische Weise gesammelt.
3. Die vorliegenden Informationen unterliegen einer bestimmten Art der Beurteilung der Vertrauenswürdigkeit.
4. Infolgedessen beruht der Anspruch vorliegender Evidenz auf einer bestimmten Mindestqualität.

Dies bezieht sich auf die genutzten Daten, Informationen und Methoden, die zur Schaffung von Fachwissen verwendet werden. ‚Empirische Evidenz‘ ist dafür ein passendes Beispiel. Häufig werden statistische Methoden genutzt, um zu beurteilen, ob eine bestimmte Aussage durch Daten oder Erfahrungen gestützt wird. Beispielsweise wird ein gegebener Datensatz für Hypothesentests verwendet. Auf Grundlage eines gewählten Signifikanzniveaus kann berechnet werden, ob eine statistische Signifikanz auf einem bestimmten Niveau vorliegt. An diesem Beispiel wird auch klar, dass die Relevanz einer Aussage auf einen gegebenen Datensatz beschränkt ist. Im Fall eines kleinen Datensatzes ist es nicht möglich, ohne Weiteres zu folgern, ob die ermittelte Beziehung auch für einen umfassenderen Datensatz gültig ist. Daten mit schlechter Qualität oder mit einigen Ausreißern können zu einem hohen Niveau an Verzerrung führen. Das Beispiel zeigt auch, dass Fachwissen, das als Evidenz verstanden wird, sehr sorgfältig genutzt und interpretiert werden muss. Abbildung 1 zeigt die Bausteine von Evidenz.

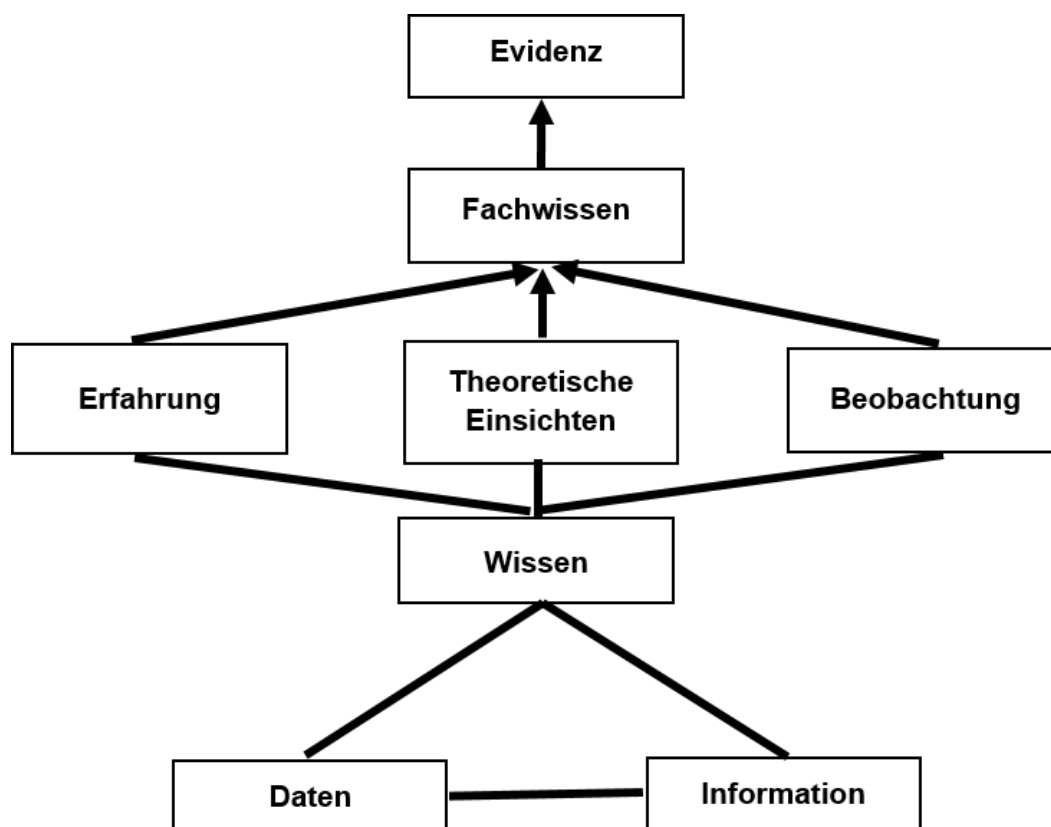
Entscheidungsträger*innen nutzen Evidenz in einem bestimmten Umfang. Die verwendete Information wird nicht auf einer homogenen Menge an Informationen beruhen, sondern eher auf verschiedenen Teilen, die nicht immer kohärent und konsistent sind und dabei gemäß den Präferenzen oder Werten der entscheidenden Person gewichtet werden. Allgemein gibt es für die beteiligten Parteien Anreize, Informationen aus der Vergangenheit als ‚Evidenz‘ zu bezeichnen, da gemäß dem allgemeinen Sprachgebrauch ‚Evidenz‘ den Beiklang hat, dass ‚eine Art Beweis‘ vorliegt. Es besteht ebenfalls die Gefahr, dass Evidenz missbraucht wird, also unterdrückt oder zur Erreichung von politischen Zielen eingesetzt wird (Witter, Kardan, Scott, Moore & Shaxson, 2017). Die entscheidende Person bestimmt, ob es erforderlich ist, einen ‚Evidenzbeweis‘ vorzulegen oder nicht. Diesbezüglich muss unterschieden werden, ob Evidenz objektiv oder eher subjektiv ist.

Der Prozess und die Verfahren, Evidenz zu ermitteln, variieren gemäß dem verwendeten relevanten Evidenzkonzept. Beispielsweise wurde die Diskussion um evidenzbasierte Politikgestaltung (EBPG) wesentlich von evidenzbasierter Medizin (EBM) inspiriert, auch wenn sich der Fokus qualitativ unterscheidet. EBPG zielt auf die gesamte Bevölkerung ab, wohingegen EBM patientenorientiert ist und die Entscheidungen, die mit EBPG in Verbindung gebracht werden, eine umfassendere öffentliche Untersuchung erfordern (Dobrow, Goel, & Upshur, 2004).

Die zugrunde liegende Logik des EBPB-Ansatzes ist für viele derjenigen Theoretiker*innen und Praktiker*innen attraktiv, die mit der Prämisse übereinstimmen, dass sich bessere Entscheidungen auf die besten verfügbaren Informationen stützen (Howlett & Craft, 2013).

Allerdings werden Entscheidungen nicht mit Blick auf die Vergangenheit, sondern auf die Zukunft getroffen. Deshalb besteht nicht nur Unsicherheit über den ‚Wert von Evidenz‘ als solcher, sondern auch hinsichtlich unterschiedlicher künftiger Zustände der Welt. Aus der Vergangenheit gewonnene Evidenz muss nicht notwendigerweise in der Zukunft valide sein. Um Wissen, das seiner Natur zufolge historisch ist, in die Zukunft zu übersetzen, ist ein angemessenes Modell erforderlich. Zudem leben wir in einer dynamischen Umwelt und in einer Epoche des schnellen Wandels und steigender Unsicherheit, die von Komplexität und Nichtlinearität geprägt ist. Evidenz, die dafür genutzt wird, Strategien zu formen und zu gestalten, muss Wissen aus der Vergangenheit auf angemessene Weise adaptieren. Deshalb ist es vermutlich nicht adäquat, Evidenz auf lineare Weise fortzuschreiben.

Abbildung 1: Konzeptionelle Wurzeln von Evidenz



Quelle: Autor

Um Entscheidungen zu treffen, kombinieren politische Entscheidungsträger*innen verschiedene Teilinformationen auf Grundlage unterschiedlicher Informationsquellen, die wiederum von uneinheitlicher Qualität sind. Sie gewichten diese, um Handlungsalternativen zu ermitteln. Politische Entscheidungsträger*innen sind häufig nicht nur an harten Fakten interessiert, die von Forscher*innen bereitgestellt werden. Die entsprechenden Ergebnisse werden häufig vor dem Hintergrund der eigenen (politischen) Werte und Meinungen und denen der wesentlichen Unterstützer*innen oder denen von interessierten und betroffenen Stakeholdern sowie von der allgemeinen Öffentlichkeit gewichtet (Lavis, Ross, & Hurley, 2002). Politische Entscheidungsträger*innen sind eine extrem heterogene Gruppe. Unter anderem zählen zu ihnen Regierungsvertreter*innen, Gesetzgeber*innen, öffentlich Bedienstete, die Justiz und die Medien. Zusätzlich sind diese Akteure in vielfältige Aktivitäten eingebunden (Oliver, Lorenc, & Innvær, 2014).

Nach dem Verständnis der politischen Ökonomie beabsichtigen Entscheidungsträger*innen, die Wahrscheinlichkeit zu maximieren, wiedergewählt zu werden bzw. in den nächsten Wahlen gewählt zu werden. Deshalb haben sie ein Interesse daran, die Art von Informationen zu nutzen, die ihre Vorstellungen stützt. Aus der Perspektive politischer Entscheidungsträger*innen bedeutet ‚beste Evidenz‘ nicht notwendigerweise ‚beste Qualität der Evidenz‘, sondern eher ‚beste Evidenz gemäß den eigenen politischen Vorstellungen‘. Es ist ihnen weniger wichtig, dass die verwendete Evidenz valide und von hoher Qualität ist, sondern dass sie vonseiten der Wähler*innen antizipiert wird. Die politische Ökonomie von Naturkatastrophen ist ein prominentes Beispiel (Cohen & Werker, 2008). Aus der Perspektive der Sozialwissenschaften bestehen Naturkatastrophen aus zwei verschiedenen Bestandteilen: der Wahrnehmung einer ‚Auslenkung der Umwelt‘ als solche und die korrespondierende Bewertung des Handlungsbedarfs. Damit eine Auslenkung als ‚Naturkatastrophe‘ verstanden wird, muss ein kritisches Schwellenniveau übertroffen werden. Die Medien können die entsprechende Evidenz liefern, wonach politikbezogene Handlungen erforderlich sein können (Miles & Morse, 2007). Auf dieser Grundlage erhalten politische Entscheidungsträger*innen die Information, dass beispielsweise im Fall des Erdbebens in Haiti, Wähler*innen die Bereitstellung von humanitärer Hilfe unterstützen und diesbezügliche Entscheidungen honorieren (Lundahl, 2013).

Für Politiker*innen, insbesondere für gewählte Amtsinhaber*innen, ist die öffentliche Wahrnehmung eigener Erklärungen und Handlungen von hoher Bedeutung, vor allem kurz vor Wahlen. Vorhandene Evidenz kann für den Nachweis nützlich sein, dass bestimmte politische Maßnahmen erfolgreich waren. Demzufolge kann Evidenz als Instrument verwendet werden, um eigene Pläne und Aktivitäten zu rechtfertigen. Deshalb besteht die Gefahr, dass ein Forschungsprojekt in Auftrag gegeben wird, um bestimmte politische Maßnahmen zu unterstützen, deren Entscheidungsprozess bereits abgeschlossen ist, da ‚Evidenz‘ über ein stark verbindendes Element verfügt (Marmot, 2004). Allerdings haben politische Entscheidungsträger*innen auch ein lebhaftes Interesse an Evidenz als Feedback-Instrument, um politische Maßnahmen zu korrigieren, die nicht erfolgreich gestaltet wurden. Da Evidenz häufig mit Unsicherheit in Verbindung gebracht wird, bedeutet es nicht automatisch, dass die Maßnahmen erfolgreich sein werden, die auf ihrer Grundlage getroffen werden. Deshalb steht es Politiker*innen frei, Evidenz auf strategische Weise zu nutzen oder Evidenz infrage zu stellen. So wie es etwa der frühere US-Präsident Donald Trump getan hat, der den menschengemachten Klimawandel leugnete. Evidenz darf nicht mit Wahrheit verwechselt werden. Aufgrund bestehender Mängel der Bereitstellung von Evidenz und Unsicherheit, kann es falsch sein, Entscheidungen auf der Grundlage von Evidenz zu treffen. Zudem ist nach Jasanoff (2006) wissenschaftliches Denken immer ein gesellschaftliches Vorhaben, das in ein besonderes Umfeld eingebettet ist; es ist kontextspezifisch, zweckbestimmt und kulturell eingebunden. Dennoch sollte die Zurückweisung von Evidenz gut begründet sein. Politiker*innen können auch Vorteile daraus ziehen, dass informelle Evidenz nicht auf rigorosen Methoden beruhen muss.

Abbildung 2 zeigt einen Analyserahmen zur Bereitstellung und Nutzung von Evidenz. Der Evidenz-Würfel beinhaltet drei Dimensionen:

- (1) Der Umfang von Evidenz
- (2) Die Stärke und die Qualität von Evidenz
- (3) Das Ausmaß der Nutzung von Evidenz in der Politikgestaltung

Alle Dimensionen werden nachfolgend erläutert.

(1) Der Umfang von Evidenz

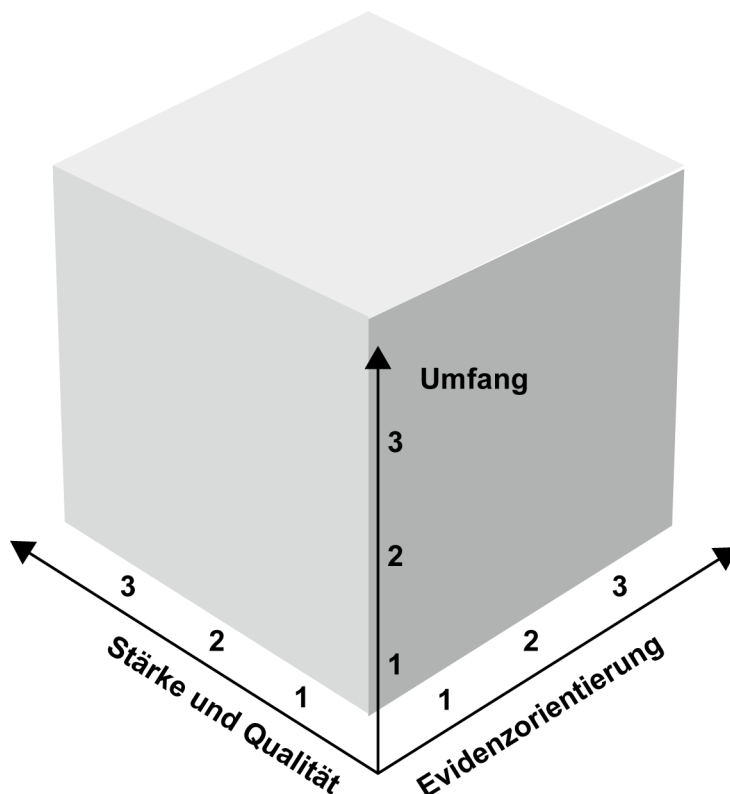
Die vertikale Achse des Evidenz-Würfels stellt den Umfang der genutzten Evidenz dar. Die Vielzahl an Evidenztypen besteht aus drei breiten Kategorien:

1. *Akademische Evidenz* greift auf empirische Erkenntnisse aller Disziplinen zurück und beinhaltet philosophische und ethische Evidenz sowie alle Politikfelder wie etwa Bildung und Gesundheit. *Idealerweise* ist sie:
 - a. objektiv, das heißt, sie ist durch Andere beobachtbar, basiert auf Fakten und ist frei von Verzerrung oder Vorurteilen, die etwa durch persönliche Meinungen hervorgerufen werden;
 - b. replizierbar, in dem Sinn, dass Andere die Ergebnisse reproduzieren können, indem sie die ursprünglich verwendeten Methoden anwenden;
 - c. entweder verallgemeinerbar, d.h. sie kann auf andere Umgebungen übertragen werden, und/oder sie ist für interne Lernprozesse nützlich, wie bei Fallstudien mit eng gefassten kontextspezifischen Umgebungen.

Diese Definition steht nicht im Widerspruch zu dem Verständnis, wonach soziale Prozesse den Umfang des wissenschaftlichen Wissens färben, das als sicher aufgefasst wird (Jasanoff, 1987).

2. *Praktisch-technische Evidenz* beinhaltet das Fachwissen von Verwaltungen und Ministerien – wie etwa dem ehemaligen Ministerium für internationale Entwicklung des Vereinigten Königreichs (DFID, Department for International Development) und dem Ministerium für auswärtige Angelegenheiten von Dänemark (DANIDA) oder von Durchführungsorganisationen wie die Französische Entwicklungsagentur (AFD), die Deutsche Gesellschaft für internationale Zusammenarbeit (GIZ) und die Kreditanstalt für Wiederaufbau (KfW) – aber auch von NROen, die im EZ-Bereich arbeiten. Häufig werden zur Bereitstellung dieses Evidenztyps Instrumente genutzt, die auf von Akademiker*innen entwickelten Methoden beruhen.
3. *Soziale Evidenz* beruht auf den aggregierten Erfahrungen und Wahrnehmungen von Individuen. Diese Personen sind etwa die Begünstigten von Projekten oder Programmen. Sie kommunizieren miteinander über ihre Vorstellungen und Überlegungen. Soziale Evidenz beinhaltet auch Beiträge der Zivilgesellschaft in Form von Instrumenten wie Petitionen und Manifesten, aber sie basiert auch auf dem gesunden Menschenverstand oder den persönlichen Erfahrungen von Mitgliedern der Zivilgesellschaft, die jenseits des beschränkten Bereichs der akademischen und praktisch-technischen Sphären zu verorten sind. Soziale Evidenz liefert zusätzliche Informationen über die Glaubwürdigkeit und die Eignung anderer Evidenztypen. Ob, wie und in welchem Umfang zivilgesellschaftliche Erkenntnisse berücksichtigt werden, hängt von der öffentlichen Wissensgenerierung ab, die sich dem nationalen und kulturellen Kontext nach unterscheiden (Jasanoff, 2005). Zu diesem Evidenztyp zählen unter anderem: die Weitergabe von Informationen durch die Zivilgesellschaft mittels Berichtskarten zur Darstellung der Wahrnehmung, wie mit öffentlichen Geldern umgegangen wird (*Open Budget Initiative*); die Meldung von Korruption mittels Smartphone-Apps; Informationen über Proteste oder die öffentliche Beratung von Sachverhalten. Soziale Netzwerke sind mittlerweile wichtige Instrumente für die Kommunikation sozialer Bedürfnisse.

Abbildung 2: Die Eigenschaften von Evidenz



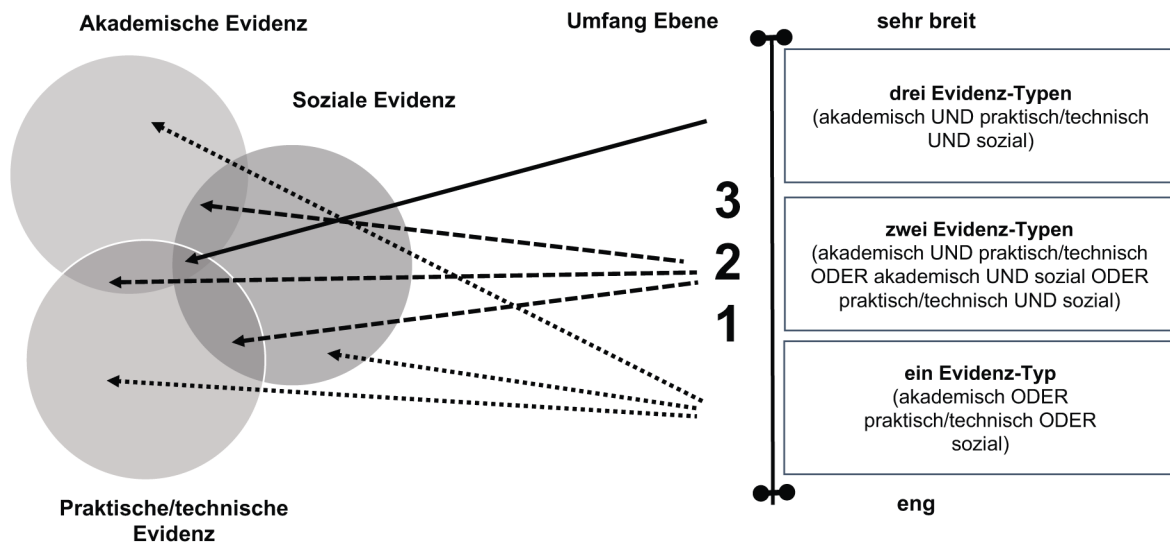
Dimension	Beschreibung	Niveaus
Umfang	Anzahl der genutzten Quellen; mit den Hauptkategorien: <ul style="list-style-type: none"> • akademische Evidenz • praktische/technische Evidenz • soziale Evidenz 	1. ein Haupttyp (= eng) 2. zwei Haupttypen 3. drei Haupttypen (= breit)
Stärke und Qualität	Qualität und Vertrauenswürdigkeit der verfügbaren Quellen	1. weich 2. mittel 3. hart
Evidenzorientierung	Ausmaß der Nutzung von Evidenz in der Politikgestaltung	1. meinungsbasiert 2. evidenzunterstützt 3. evidenzbasiert

Quelle: Autor

Der Umfang der Evidenz wird durch die Endpunkte ‚eng‘ (= Umfang-Ebene 1) und ‚sehr breit‘ (= Umfang Ebene 3) definiert – siehe Abbildung 3.

Nicht überlappende Bereiche beziehen sich auf Umgebungen, in denen Entscheidungsträger*innen nur an besonderen Evidenztypen interessiert sind. Die Schnittmengen bilden Fälle ab, bei denen mehr als ein Evidenztyp berücksichtigt wird. Möglicherweise beachten Entscheidungsträger*innen Evidenz, die sowohl von Forscher*innen als auch von Durchführungsorganisationen bereitgestellt wird. Bereitgestellte Evidenz kann auch aus einer Zusammenarbeit von Akademiker*innen und Praktiker*innen resultieren.

Abbildung 3: Die Dimension ‚Umfang‘: Evidenztypen



Quelle: Autor

Die breiten Kategorien bestehen aus Teilkategorien. Die bereitgestellte akademische Evidenz kann etwa auf den Ergebnissen nur einer Disziplin beruhen, wie der Ökonomie. Es ist auch möglich – so bei sehr wichtigen Themen wie Nachhaltigkeit –, dass Entscheidungsträger*innen an breit-basierter akademischer Evidenz aus vielen Disziplinen interessiert sind. Abhängig von der Themenstellung, dem Typ der Aktivität, der Intervention und dem Zweck, werden wahrscheinlich verschiedene Typen und Quellen der Evidenz bevorzugt (Guijt & Roche, 2014). Dies ist in den Fällen wichtig, in denen die Bewertung der Herausforderungen zwischen Disziplinen unterschiedlich ausfällt. Höhere Komplexitätsniveaus und zusätzlich benötigte Zeit, die erforderlich ist, um unterschiedliche Evidenz zu vergleichen, erschweren möglicherweise Entscheidungen und verzögern sie.¹

Der Kontext interner und externer Entscheidungsfindung bestimmt, was als Evidenz gilt und wie die Evidenz genutzt wird (Dobrow, Goel, & Upshur, 2004).



(2) Die Stärke und Qualität der Evidenz

Die horizontale Achse des Evidenz-Würfels zeigt die Stärke und Qualität von Evidenz. Im Kern bezieht sie sich auf die Qualität und Vertrauenswürdigkeit der verfügbaren Quellen, wie sie in den Ranglisten verwendet werden, die ursprünglich in der EBM entwickelt wurden (Tabelle 1). Mittlerweile werden diese aber auch in anderen Politikfeldern genutzt, so etwa im Umweltmanagement (Dicks, Walsh, & Sutherland, 2014). Befragungstechniken, Expert*innen-Meinungen ohne ausdrückliche Bewertung usw. stellen das niedrigste Niveau dar (‚weich‘). Das höchste Niveau (‚hart‘) beinhaltet systematische Reviews auf Grundlage experimenteller oder quasi-experimenteller Evidenz. Tabelle 1 zeigt, dass systematische Reviews nicht immer am besten sind. Im Fall schlechter Qualität, etwa bei nur wenigen Beobachtungen, wird das Niveau nach

1 Vollkommen rationale Entscheidungsträger*innen würden den gesamten Nutzen und sämtliche Kosten bezüglich der Berücksichtigung zusätzlicher Komponenten und Typen an Evidenz miteinander vergleichen. Der Ökonomie der Evidenz folgend, würden sie den Umfang der Evidenz, die als Ressource bei der Entscheidungsfindung berücksichtigt wird, bis zu der Menge ausweiten, bei der die Grenzkosten (d.h. die Verzögerung bei der Entscheidung, Verwaltungskosten etc.) dem Grenznutzen (d.h. Nutzen für ausgegrenzte Menschen) entsprechen. Der jeweilige Umfang an Evidenz maximiert die öffentliche Wohlfahrt.

unten korrigiert. Wenn Forschungsergebnisse mangelhaft sind, liegt das meist an unklaren Zielen, schlechtem Design, an methodischen Schwächen, an einer nicht angemessenen Darstellung der statistischen Sachverhalte oder der Analyse, an einer selektiven Verwendung von Daten sowie an Folgerungen, die von den genutzten Daten nicht unterstützt werden. Diesen Aspekt decken ebenfalls die Kategorien ‚Stärken‘ und ‚Qualität‘ ab.

Tabelle 1: Qualität der Evidenz im Fall der EBM

Niveau	Datenquelle	Niveau
1	Expert*innen-Meinung ohne kritische Bewertung	weich 
2	Fallserie (eine Kohorte mit schlechter Qualität und Fall-Kontroll-Studien)	
3a	Individuelle Fall-Kontroll-Studie	
3b	Systematische Reviews (homogen) von Fall-Kontroll-Studien	
4a	‚Outcome‘ Forschung, ökologische Studien	
4b	Individuelle Kohortenstudien (einschließlich randomisierte kontrollierte Studien – RCTs - mit geringer Qualität)	
4c	Systematische Reviews (homogen) von Kohortenstudien	
5a	Alles-oder-nichts-Studien	
5b	Einzelne RCT (mit engem Konfidenz-Intervall)	
5a	Systematische Reviews (homogen) von RCTs	hart 

Quelle: Auf Grundlage von Camfield, Duvendack & Palmer-Jones (2014) und OCEBM Levels of Evidence Working Group (2011)

Tabelle 2 zeigt eine auf Tabelle 1 gestützte Metrik, die es ermöglicht, auch Quellen jenseits der Evidenz zu berücksichtigen.

Tabelle 2: Qualität der Evidenz im erweiterten Rahmen in absolutem Sinn

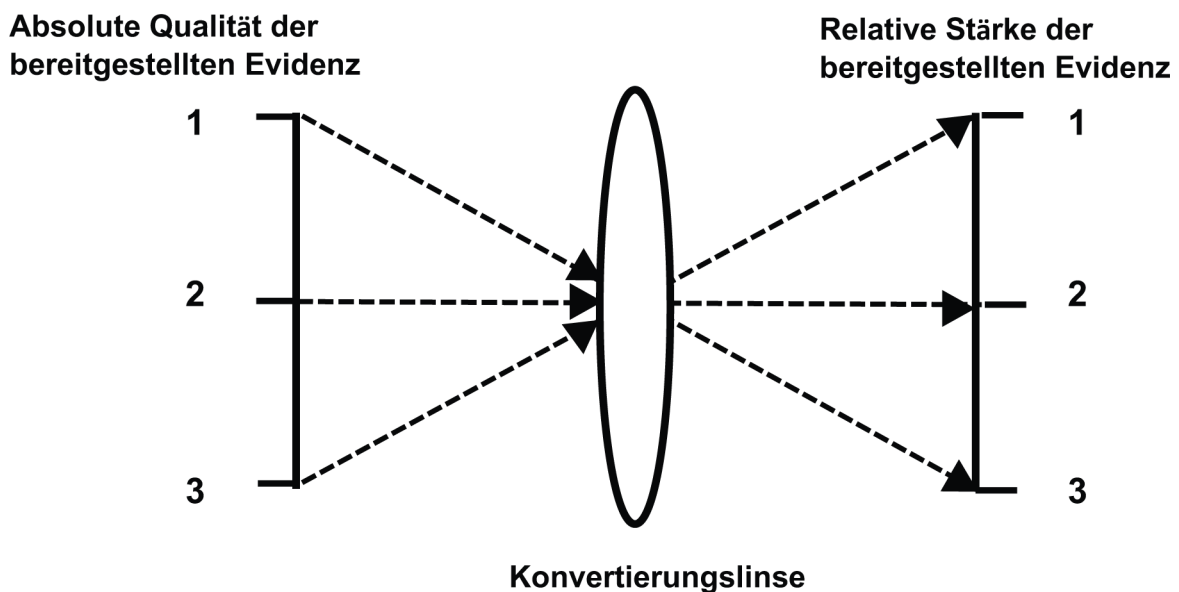
#	Datentypen (Beispiel)	Niveau
1	Expert*innen-Meinungen mit begrenzter kritischer Bewertung, Ergebnisse aus Fallstudien, Daten aus ergebnisbasiertem Management und Leistungsdaten mit geringer Qualität; einzelne und individuelle Äußerungen von Vertreter*innen von Interessengruppen	weich
2	Individuelle Forschungsergebnisse gestützt auf Methoden mit qualitativer Grundlage; ergebnisorientiertes Management und/oder Leistungsdaten mit guter Qualität; Äußerungen von Bürger*innen-Bewegungen	mittel
3	Systematische Reviews mit Forschungsergebnissen auf Grundlage von rigorosen Methoden; ergebnisorientiertes Management mit höchster Qualität und/oder Leistungsdaten; sehr breit-basierte öffentliche Petitionen und Stimmen aus der Zivilgesellschaft	hart

Quelle: Autor

Wird ein Studiendesign auf einen anderen Kontext übertragen, bedeutet dies nicht gleichzeitig, dass dieselbe Qualität auch in einer anderen Umgebung erreicht wird. Disziplinen oder Fachbereiche können auch ein unterschiedliches Verständnis von Qualität oder von deren Bedeutung haben. Von der Einsicht ausgehend, dass die Qualität wissenschaftlicher Evidenz, die in Gerichtsverfahren genutzt wird, von der in der Forschung abweicht, weist Jasanoff (2006) darauf hin, dass Erstere nie davon abhängen sollte, wie sie übergreifend verstanden wird. Dies ist jedoch nicht der Anspruch von Tabelle 2.

Die Qualität von Evidenz im absoluten Sinn darf nicht mit der Stärke von *Evidenz für politische Entscheidungsträger*innen* verwechselt werden. Absolut gesehen, sind alle drei Evidenztypen, die in der Dimension ‚Umfang‘ enthalten sind, inhärent mit verschiedenen Niveaus verbunden. Akademische Evidenz ist wahrscheinlich von höherer absoluter Qualität als soziale Evidenz (Abbildung 4 – linke Seite). Die relative Stärke der bereitgestellten Evidenz (Abbildung 4, rechte Seite), ist eher die Folge davon, wie diese Evidenz von politischen Entscheidungsträger*innen wahrgenommen wird. Weiche Evidenz, die von qualifizierten beteiligten Parteien geliefert wird, die als sehr bedeutsam gelten, wird beispielsweise in der Rangfolge hochgestuft werden. Der entsprechende Prozess wird in Abbildung 4 mit einer Konvertierungslinse dargestellt.

Abbildung 4: Wahrnehmung von Evidenz durch politische Entscheidungsträger*innen



Niveaus der absoluten Qualität und der relativen Stärke

- 1 = weich
- 2 = mittel
- 3 = hart

Quelle: Autor

(3) Das Ausmaß der Nutzung von Evidenz in der Politikgestaltung

Das Ausmaß der Evidenz, die bei der Politikgestaltung genutzt wird, hängt von vielen kritischen Faktoren ab und ist in hohem Maße vom Umfeld abhängig.

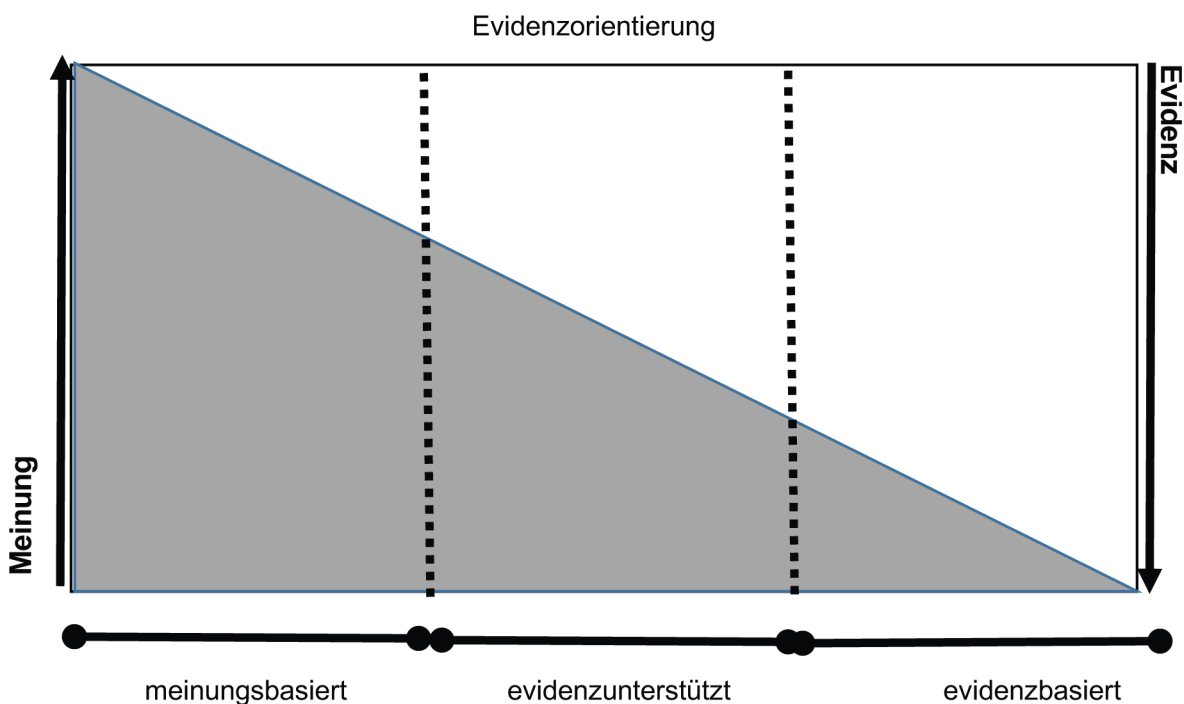
Allgemein betrachtet, gibt es drei verschiedene *prozessbezogene* Schritte für die Berücksichtigung von Evidenz:

1. Wahrnehmung: Politische Entscheidungsträger*innen werden sich verfügbarer Evidenz bewusst.
2. Ressource: Verfügbare Evidenz wird im Entscheidungsprozess theoretisch berücksichtigt, etwa durch Verpflichtungserklärungen.
3. Implementierung. Evidenz wird tatsächlich in der Praxis eingesetzt.

Abbildung 5 zeigt die Niveaus der Evidenzorientierung bei politischen Entscheidungen. Werden ‚Meinung‘ und ‚Evidenz‘ als Variablen verwendet, so lassen sich drei grundlegende Niveaus voneinander unterscheiden:

- Auf dem untersten Niveau ignorieren politische Entscheidungsträger*innen verfügbare Evidenz vollkommen. Dies ist der Fall bei *meinungsbasierter* Politikgestaltung. Wird Evidenz genutzt, dann auf sehr selektive Weise, meistens für eine nachträgliche Rechtfertigung der eigenen Weltanschauung. Dieses Niveau kann mit dem Vorherrschen ideologischer Standpunkte und Vorurteile charakterisiert werden. Meinungen dominieren auch in den Fällen, in denen keine Evidenz verfügbar ist. Wenn etwa ein Problem neu auftritt und begründete Zweifel an der bereitgestellten Evidenz bestehen.
- Das andere Extrem bildet die *evidenzbasierte* Politikgestaltung (*evidence-based policy making*, EBPM). Dies ist durch eine Situation gekennzeichnet, in der Entscheidungen vollkommen auf Evidenz beruhen. Dies ist etwa der Fall bei äußerst pragmatischen Entscheidungsträger*innen, für die ideologische Aspekte unbedeutend sind. Diese Ebene korrespondiert mit eher wohlwollenden Entscheidungsträger*innen, die ergebnisorientiert ausgerichtet sind. Die Informationen dazu, was funktioniert, können Outputs, *Outcomes* und die Wirkungen von Interventionen und Aktivitäten verbessern. EBPM bezieht sich auch auf Ansätze, die bei der Entscheidungsfindung stark auf Rationalität beruhen und beim Vergleich von Handlungsalternativen verfügbare Evidenz filtern. Die Distanz zwischen den unteren und oberen Enden des Spektrums korrespondiert mit einer Situation, in der Entscheider*innen im Vorfeld der Entscheidung viele verschiedene Komponenten bedenken, einschließlich einiger Arten verfügbarer Evidenz. Es kann viele Gründe dafür geben, dass Evidenz eingeschränkt genutzt wird. In einer Situation, die von politischen Merkmalen geprägt ist, wird verfügbare Evidenz vermutlich nur dafür genutzt, bestehende politische Prioritäten zu stützen.
- *Evidenzunterstützte* Politikgestaltung kann auch die Folge eines mangelhaften Verständnisses des Wertes von Evidenz sein oder eines beschränkten Vertrauens in dessen Verlässlichkeit. Auf dieser Ebene können Meinungen immer noch relevant sein.

Abbildung 5: Evidenzorientierung: Ausmaß der Nutzung von Evidenz in der Politikgestaltung



Quelle: Autor

Theoretisch kann eine Situation erwünscht sein, in der alle Elemente den höchsten Wert erreichen. Realistisch betrachtet, ist dieser Zustand häufig nicht erreichbar oder es könnten die Kosten zu hoch sein, die mit dem Erreichen dieses Punktes verbunden sind. Es ist zeitintensiv, verfügbare Evidenz zu sammeln (Suchkosten), was wiederum in Bezug zu anderen (Opportunitäts-)Kosten steht. Deshalb ist die Ökonomie der Evidenz bedeutsam.

Die korrespondierende Evidenz muss verarbeitet und aufbereitet werden, um als Übersicht von Wahlmöglichkeiten für die Entscheidungsfindung fungieren zu können. Beispielsweise genügt es nicht, wenn akademische Evidenz in akademischen Zeitschriften veröffentlicht wird, da politische Entscheidungsträger*innen den Beitrag vermutlich nicht lesen. Deshalb sind Thinktanks und Wissensvermittler*innen (*Knowledge Broker*) wichtig, um verfügbare Evidenz zu übersetzen, zu aggregieren und zu synthetisieren, so dass sie wahrgenommen wird. Nicht alle Quellen von Evidenz sind aber hinreichend solide, um als Grundlage für die Politikgestaltung zu fungieren. Im Fall sehr schlechter Daten und Informationen ist eine kritische Diskussion erforderlich, um zu entscheiden, ob Evidenz bei der Entscheidung ein- oder ausgeschlossen wird.

Der Beratungsprozess verläuft nicht linear. Es gibt Rückkopplungsschleifen. Die Auswahl der geeigneten Evidenz kann als Form einer Koproduktion von politischen Entscheidungsträger*innen, Berater*innen und einigen anderen Stakeholdern aufgefasst werden: In derartigen Systemen verändern sich die Formen, Wissen zu schaffen und Entscheidungen zu treffen, fortlaufend (Clark, van Kerkhoff, Lebel, & Gallopin, 2016).

3 Die Bereitstellung von Evidenz in der Entwicklungszusammenarbeit

3.1 Evidenzquellen

Die Bereitstellung von Evidenz in der EZ ist aus vielen Gründen herausfordernd. Häufig wird der Fall von EBM als Blaupause betrachtet. Infolgedessen wird das EBM-Konzept unmittelbar auf den EZ-Bereich übertragen, ohne zu berücksichtigen, dass der EZ-Kontext viel komplexer ist. Das Feld der menschlichen Gesundheit steht im Blickpunkt insbesondere einer Disziplin (Biomedizin). Es gibt verschiedene Konzepte von Gesundheit, einschließlich moderner und traditioneller Medizin. Wenn geeignete Kandidatinnen für Interventionen ausgewählt werden, wird allerdings die Evidenz, die auf Grundlage verschiedener Gesundheitskonzepte ermittelt wird, nicht synthetisiert. Bei der EZ aber handelt es sich um ein Forschungsgebiet mit viel mehr (Sub-) Disziplinen. Diese arbeiten auf der Grundlage eines unterschiedlichen Verständnisses und verschiedener Interpretationen der Realität. Deshalb können sich Evidenzkonzepte in der EZ auch unterscheiden. In einigen Kontexten ist es besonders wichtig, verfügbare Evidenz zu kombinieren. EZ bietet eine reiche Basis für akademische, praktisch-technische und soziale Evidenz.

Es kann ein breites Spektrum an Quellen und Instrumenten verwendet werden, um verfügbare Evidenz zu generieren und zu verbreiten. Bevor die Daten verarbeitet und analysiert sind, ist ihr Wert allerdings beschränkt. Als besondere Formate werden akademische Zeitschriften, Berichte, Arbeitspapiere und Newsletter genutzt, um akademisch, praktisch-technische und soziale Evidenz bereitzustellen und zu kanalisieren. Darüber hinaus arbeiten die korrespondierenden beteiligten Akteur*innen bei der Generierung von Evidenz oftmals eng zusammen, so etwa im Fall von Projekten. Es gibt auch Plattformen für den Wissensaustausch. Manchmal veröffentlichen Mitarbeiter*innen von praktisch-technischen Einrichtungen ihre erzielten Ergebnisse in akademischen Zeitschriften. Forscher*innen sammeln auch quantitative und qualitative Daten und Informationen, die wichtig dafür sind, Evidenz bereitzustellen. Die Ergebnisse werden häufig in Artikeln oder Büchern veröffentlicht. Oftmals werden Daten und Informationen in der EZ

als Nebenprodukt generiert, wie etwa in technischen Projektberichten. Regierungseinrichtungen, internationale Organisationen – und in zunehmendem Umfang nicht staatliche Bereitsteller*innen von EZ – sammeln Monitoring-Daten, um den Erfolg von Programmen und Projekten abzubilden. Auf Ebene der Partnerländer werden statistische Daten (beispielsweise Zensusdaten) gesammelt. Haushaltsbefragungen sind wichtige Quellen der Evidenz. In jüngster Zeit ist es wichtig geworden, Managementdaten, insbesondere Evaluierungsergebnisse zu sammeln. Diese Daten und Informationen liegen häufig nur eingeschränkt in der organisatorischen Domäne vor. Allerdings sind zusammenfassende Berichte besonders aus Rechenschaftsgründen verfügbar. Durchführungsorganisationen stellen auch technische Berichte bereit, die Projektergebnisse reflektieren. In der EZ ist zudem die Zivilgesellschaft von besonderer Bedeutung. Es herrscht ein gemeinsames Einvernehmen, dass die Bedürfnisse der Bevölkerung adäquat bei der Politikgestaltung berücksichtigt werden sollten. Organisationen in der Zivilgesellschaft sammeln soziale Evidenz und verbreiten sie über Massenmedien, öffentliche Veranstaltungen und/oder politische Einrichtungen. Dem jeweiligen Kontext der Länder entsprechend wird soziokulturelle Evidenz zusätzlich als Ressource verwendet, um Entscheidungen zu treffen.

Das Thema ‚Daten für Entwicklung‘ hat zuletzt viel Aufmerksamkeit erhalten. Einerseits wird die bestehende Datenlücke diskutiert – insbesondere bei der Abbildung der SDG-Unterziele, mit besonderem Fokus auf Partnerländern und einigen Kontexten. Beispielsweise mangelt es an Evidenz bei der Implementierung von Sektorprogrammen. Andererseits weitet sich die verfügbare Menge an Datenmassen (Big Data) aus. Hier stellt sich die Frage, wie die entsprechenden Informationen nutzbar gemacht werden können. Um Evidenz zu generieren, nimmt fortlaufend die Zahl der Quellen und das korrespondierende Material sehr stark zu. Die Quellen umfassen breit genutzte Technologien wie Smartphones, Satellitennavigationssysteme, soziale Medien, internetbasierte Datenbanken, Audio und automatische Anrufverteiler (*interactive voice response*) sowie digitale Videos und Kameras. Mittlerweile sind solche Informationen auch über das Internet zugänglich, weitestgehend ohne technische und finanzielle Barrieren.

Die nachfolgende Darstellung fokussiert sich auf Instrumente, die in jüngster Zeit stärker beachtet wurden.

3.2 Statistik und Monitoring

Seit vielen Jahren werden unterschiedliche Datentypen auf internationaler, nationaler, regionaler und lokaler Ebene gesammelt, und zwar durch internationale Organisationen, nationale Regierungen, Fachministerien und Verwaltungen, von Träger*innen der EZ und ihren Durchführungsorganisationen, aber auch von Forscher*innen. In der EZ sind Indikatoren sehr relevant, bei denen es sich um eine spezielle Kategorie von Daten handelt. Sie bieten wesentliche Informationen für verschiedene Stakeholder und werden für Planung, Monitoring und Evaluierung genutzt (Holzapfel, 2014).

Das Wissen darüber, was wirkt, ist der zentrale Hebel, um die Wirksamkeit zu erhöhen, also die Beziehung zwischen geplanter und erzielter Leistung. Erfolgs- und Ergebnisindikatoren sind wichtige Referenzpunkte für die Bereitstellung von operativer Evidenz, um ‚das Richtige zu tun‘. Hierzu dient das Monitoring. Dies ist ein laufender Prozess regelmäßiger, systematischer Bewertung. Diese beruht auf Partizipation, Reflexion, Feedback, Datensammlung, Analyse des tatsächlichen Erfolgs unter Nutzung von Indikatoren sowie auf der regelmäßigen Berichterstattung während der Lebensdauer eines Programms oder Projekts. Der Zweck ist es, hinsichtlich der Ergebnisse, Qualität und Zielerreichung die Projekt- und Programmfortschritte zu verfolgen. Monitoring-Ergebnisse können dabei helfen, Maßnahmen zu ergreifen, um Politiken, Programme und Praktiken zu korrigieren, die verbessert werden müssen.

Die MDGs steigerten das Interesse an der Rolle von Indikatoren, die dazu dienen, die globalen Entwicklungsfortschritte zu messen. Sie trugen dazu bei, evidenzorientierte Entscheidungsfindung zu untermauern. Diesbezüglich bilden Haushaltsbefragungen eine reichhaltige Quelle an Informationen. Allerdings sind seit der Einführung der MDGs schwere Mängel im Prozess der Sammlung und Bereitstellung von Daten deutlich geworden. In Partnerländern gibt es nach wie vor große Probleme bei der Datensammlung. In vielen Fällen sind keine wesentlichen Informationen vorhanden. Beispielsweise verfügen nicht weniger als 100 Länder über keine akkuraten Aufzeichnungen von Geburten und Todesfällen. Die SDGs betreffen auch viele Bereiche, in denen Regierungen noch niemals zuvor versucht haben, akkurate Daten zu erheben. Es bestehen auch Schwierigkeiten, adäquate Indikatoren zu einigen Unterzielen der SDGs zu ermitteln. Deshalb wurden verschiedene Initiativen gestartet, um die Datensammlung in der EZ zu stärken (Keijzer & Klingebiel, 2017; OECD, 2017).

3.3 Ergebnismangement

Der Wunsch, Ergebnisse zu erzielen, war schon immer Bestandteil der EZ-Agenda. Historisch betrachtet, kann der Fokus auf zu erzielende Ergebnisse als Teil des *New-Public-Management*-Paradigmas gesehen werden. Dieser Ansatz stammt aus Prozessen der Privatwirtschaft. Er zielt darauf ab, Unternehmenspraktiken auf den öffentlichen Sektor zu übertragen. Dabei werden Ergebnisse als Mittel zu dem Zweck betrachtet, die Gewinne von Anteilseigner*innen zu maximieren bzw. um die Kosten zu senken. Ideologische Komponenten spielen in diesem Zusammenhang keine Rolle (Eyben, 2015). In den 1980er und 1990er Jahren wurde Ergebnismangement bereits beim Management von Entwicklungsprojekten eingesetzt. Dies wurde durch Methoden wie den *Logical-Framework*-Ansatz (LogFrame) unterstützt (Vähämäki, Schmidt, & Molander, 2011). Derzeit macht die EZ eine radikale Veränderung durch, und es hat sich die Wirklichkeit der EZ-Landschaft verschoben. Probleme gibt es wegen ihrer Fragmentierung (Klingebiel, Mahn, & Negre, 2016). Es ist bekannt, dass die bestehende Evidenzbasis zu den Wirkungen entwicklungsbezogener Interventionen eher schwach ist (Sayedoff, Levine, & Birdsall, 2006) oder unklar, wie im Fall des ‚Mikro-Makro-Paradoxon‘ der EZ (McGillivray, Feeny, Hermes, & Lensink, 2005). Seit der Jahrtausendwende ist das Interesse an einer Antwort auf die Frage groß, wie EZ effektiver gestaltet werden kann (Klingebiel, 2014). Diese Aspekte haben Geber*innen und Partner*innen herausgefordert, die Ergebnisse von EZ-Aktivitäten zu verbessern. Und zwar ungeachtet der Wahrnehmung, wonach im Laufe der Zeit die EZ-Abhängigkeit von Partnerländern gesunken ist. Außerdem hat die Forderung, EZ-Ausgaben zu rechtfertigen, dazu geführt, dass die Nachfrage nach messbareren und glaubwürdigeren Ergebnissen stieg und dazu, dass die konkreteste Evidenz dargestellt wurde (Klingebiel, 2012).

Deshalb wurden ‚neue‘ ergebnisorientierte Modelle eingesetzt, die die bisherige Interventionslogik umgekehrt haben. Diese Konzepte stützten sich auf frühere Arbeiten jenseits der EZ, wie etwa den ‚*Performance and Results Act*‘ der US-Regierung von 1993. Bis dato stand es bei der Bewertung der Projektimplementierung im Vordergrund, Ressourcenverwendung, Aktivitäten und Prozesse nachzuverfolgen. Durch ergebnisorientierte Konzepte erfolgte eine Verschiebung hin zu den Ergebnissen des Monitorings. In ihnen werden *Outputs* und *Outcomes* als analytische Startpunkte genutzt, um die Ergebniskette zu erforschen. Dazu müssen numerische Indikatoren passend definiert werden. Den neuen Leitprinzipien zufolge sollten zu erreichende Meilensteine bestimmt werden, die mittels numerischer Werte klar benannt werden. In diesem Kontext wurde das Ergebnismangement auf Programme und auf die organisatorische Ebene übertragen. Wesentliche Unterschiede gegenüber früheren Ansätzen des Managements bestehen auch bei der verwendeten Terminologie.

Ergebnisorientiertes Management (EOM, results-based management) wurde von der OECD bzw. von ihrem Entwicklungsausschuss (OECD/DAC, 2002) als Managementstrategie definiert. EOM fokussiert sich auf den Erfolg und darauf, *Outputs*, *Outcomes* und *Impacts* zu erreichen.

Das vorherige Konzept wurde durch die OECD/DAC zum ‚Management für Entwicklungsergebnisse‘ (*management for development results*) weiterentwickelt, um neue Vorstellungen von Zusammenarbeit, Partnerschaften, *Ownership* der Länder, von Harmonisierung und Ausrichtung mit aufzunehmen. So wird für das Management ein höherer Standard ermöglicht, weil die beteiligten Parteien aufgefordert sind, ihren Fokus fortlaufend auf den Erfolg der Länder-Wirkungen (*Outcomes*) im Gegensatz zu kurzfristigen Ergebnissen zu richten.

In die Pariser Erklärung zur Wirksamkeit der Entwicklungszusammenarbeit (2005) wurde ‚Ergebnisorientierung‘ (*management for results*) als viertes Prinzip aufgenommen. Als wichtige Instrumente, um Ergebnisse zu erzielen, wurden ergebnisorientiertes Management (EOM, *results-based management*) sowie ergebnisorientierte Berichterstattung und Bewertungsrahmen (*results-oriented reporting and assessment frameworks*) hervorgehoben.

Die Implementierung des Ergebnismanagements unterscheidet sich aufgrund unterschiedlicher Prioritäten und Strategien von Land zu Land. Ergebnismatrizen oder Wirkungsrahmen knüpfen Verbindungen zwischen den strategischen Zielen eines Landes auf höherem *Outcome*-Niveau, den organisatorischen Strukturen eines Landes und den Beziehungen zwischen den wesentlichen beteiligten Parteien und den Entwicklungspartner*innen. An ihnen zeigt sich die Hierarchie von wesentlichen *Outcomes* für Sektor- oder für umfassende Programme, und sie können aus mehreren Ebenen bestehen. Dazu zählen die sektoralen, regionalen, individuellen oder die Agentur-bezogenen Ebenen. Sie werden verwendet, um ergebnisorientierte Indikatoren zu erfassen, damit – analog zur Bereitstellung – operative Evidenz darüber vorliegt, eng definierte Ergebnisse erreicht zu haben (Roberts & Khattri, 2012). Die Ergebnismatrix ist ein wichtiger Baustein in ergebnisorientierten Konzepten. Sie liefert durch die Ergebniskette – von Aktivitäten bis hin zu Wirkungen – eine solide Basis, um wichtige Analysen zu unterstützen.

Die deutsche EZ nutzt eine Wirkungsmatrix als Basis für die Planung und die Berichterstattung von GIZ und KfW. Sie beruht auf Indikatoren und Daten aus den Partnerländern, aber es werden auch eigene Daten gesammelt.

Die Globale Partnerschaft für wirksame Entwicklungszusammenarbeit (*Effective Development Cooperation*) ist eine Multi-Akteurs-Plattform, die die Wirksamkeit der Entwicklungsbemühungen aller Akteur*innen fördert. Ihrem Bericht aus dem Jahr 2016 zufolge, konnten nur 54,7% der Ergebnisindikatoren aus den von den Ländern getragenen Ergebnismatrizen gewonnen werden (Global Partnership for Effective Development Co-operation, 2016). Quantitative und qualitative Indikatoren werden genutzt, um zu ermitteln, wie gut ein Projekt, ein Programm oder eine Politik gegenüber den erwarteten Zielen abschneidet. Evidenz wird durch einen fortlaufenden Prozess von Rückmeldungen, Lernen und Anpassung geschaffen. Dieser Prozess wird durch interne Finanzkontrolle, Managementprüfungen, Erfolgs-Monitoring und durch die Implementierung iterativer Programme und Evaluierungen unterstützt (OECD, 2015).

Im Fall des Vereinigten Königreichs hat DFID ein Ergebnissystem entwickelt, das alle wesentlichen und notwendigen Elemente beinhaltet. Es berücksichtigt einen Politik- und Management-Rahmen, Unterziele und Indikatoren. Außerdem bietet es die Kapazität und Anreize, um Ergebnisinformationen zu sammeln, zu verarbeiten und zu analysieren. Dieser Rahmen ist ähnlich wie die Modelle ausgestaltet, die von den großen multilateralen Entwicklungsbanken implementiert wurden. So werden sowohl Monitoring und Management des Fortschritts von Entwicklungsergebnissen auf der Organisationsebene ermöglicht als auch die Möglichkeit eröffnet, über die Ergebnisse öffentlich zu berichten. Dieses System stützt sich auf allgemeine Indikatoren für die organisationale Berichterstattung; spezifischere Indikatoren werden für die Projektberichterstattung verwendet. Das System bezieht sich weitgehend auf Output-Indikatoren (OECD, 2014b) und besteht aus verschiedenen Ebenen. Ebene 1 misst den Fortschritt anhand der wesentlichen *Outcomes* der Entwicklung, zu der die DFID in Partnerländern beitragen will. Ebene 2 gilt den Indikatoren, die solche Outputs und intermediäre *Outcomes* messen, die direkt mit den Interventionen von DFID verknüpft werden können. Ebene 3 deckt die operative Leistungsfähigkeit

ab. Korrespondierende Verbesserungen können dazu führen, dass Ergebnisse und ein besseres Preis-Leistungs-Verhältnis leichter erreicht werden. Auf Ebene 4 fokussieren sich Indikatoren darauf, die organisationale Wirksamkeit zu messen. Die korrespondierenden Indikatoren zielen darauf ab, Verbesserungen in internen Geschäftsprozessen zu überwachen (OECD, 2014b).

DANIDA hat einen Rahmen für das Ergebnismanagement und für die Berichterstattung auf Länderebene eingeführt. Er besteht aus einem länderbezogenen Policy-Paper, einem länderbezogenen Programm-Dokument sowie aus dem Entwicklungsengagement in jedem Land. Der Wirkungsrahmen umfasst strategische und thematische Ziele sowie Output und *Outcome* des Engagements (OECD, 2014a).

Einigen Autor*innen zufolge, ist der Anspruch von EOM aufgrund gravierender Schwierigkeiten, die komplexe EZ auf der Ebene der Organisationen zu handhaben, gescheitert. Probleme bereiten die Unterschiede zwischen den Ländern, den Sektoren und zwischen den Kontexten, in denen die Entwicklungsorganisationen arbeiten. Darüber hinaus bestehen andauernde Schwierigkeiten, weil ein gemeinsames Verständnis dafür fehlt, was EOM ist und für welche Zwecke es eingesetzt werden sollte. Da EOM im Management-Denken verankert ist und sich auf Kontrolle und Prozesse fokussiert, erfolgt die Anpassung an Veränderungen nach wie vor schwerfällig. EOM, das von den Entwicklungsorganisationen genutzt wird, wird zudem vorgeworfen, dass es sich meistens auf die Prioritäten der Geber*innen fokussiert und nicht auf die nationalen Prioritäten der Partnerländer. Dennoch ist EOM ein wichtiges Werkzeug, um operative Evidenz bereitzustellen (Vähämäki et al., 2011).

3.4 Ergebnisbasierte Ansätze

Ergebnisbasierte Ansätze sind der Überbegriff für eine Vielzahl von Ansätzen, die in jüngster Zeit eingeführt wurden. Sie haben das Paradigma der Entwicklungszusammenarbeit von seinem Fokus auf Ressourcen und Aktivitäten hin zur Orientierung auf Ergebnisse verschoben. Und zwar, indem sie die Beziehung zwischen den Entwicklungspartner*innen oder Geber*innen mit den Partnerländern neugestaltet haben. Dafür müssen quantifizierbare und messbare Ergebnisse erzeugt werden, die so nah wie möglich mit den Wirkungen der EZ verbunden sind (Klingebiel & Janus, 2014).

Diese Ansätze können für jedes Programm gelten, bei dem der Prinzipal finanzielle oder andere Anreize für einen Agenten setzt, damit dieser zuvor bestimmte Outputs oder *Outcomes* erreicht und bei denen die Entlohnung an den Nachweis geknüpft ist, dass Letztere auch erzielt wurden (Musgrove, 2011). Normalerweise handelt im Fall der EZ eine nationale oder subnationale Regierungseinrichtung des Partnerlandes als Prinzipal. Die Vertragspartnerin oder der Vertragspartner ist gewöhnlich eine implementierende Agentur oder eine Person (Grittner, 2013).

Es gibt sowohl nachfrage- als auch angebotsseitige Ansätze. Im Fall ergebnisorientierter Zusammenarbeit (EOZ, *results-based aid*) werden finanzielle Mittel von einer Regierung an eine andere ausgezahlt. Wenn Einzelpersonen, Haushalte, öffentliche Einrichtungen, Einrichtungen ohne Gewinnerzielungsabsicht oder die Privatwirtschaft Zahlungen für selbst erbrachte Leistungen erhalten, wird der Begriff ergebnisorientierte Finanzierung (EOF, *results-based financing*) verwendet. Jenseits dieser allgemeinen Klassifikation gibt es eine Vielzahl von Ansätzen, deren Eigenschaften von den Entwicklungsorganisationen ausgestaltet wurden (Grittner, 2013). Sie variieren sehr stark und beziehen sich auf Anreize und Zielgruppen sowie auf das konkrete Niveau der Ergebniskette, auf das sie ausgerichtet sind. Nach Perakis und Savedoff (2015) sind dies Beispiele dafür:

- Impfunterstützungsprogramme (ISS) von GAVI, die Ländern helfen, ihre routinemäßigen Schutzimpfungen auszuweiten. Sie beruhen auf leistungsabhängigen Zahlungen, die aus

den Impfdaten errechnet werden, die von den administrativen Berichtssystemen der Länder gewonnen werden.

- Der Amazonasfonds zielt darauf ab, die Walderhaltung zu fördern. Öffentliche Entwicklungszusammenarbeit (*Official Development Assistance*) von Norwegen und Deutschland an Brasilien erfolgt, wenn als Folge reduzierter Abforstung der Kohlendioxid ausstoß vermindert wird und dies verifiziert ist.
- Für Schüler*innen, die Prüfungen im Rahmen ihrer Sekundärschulbildung absolvieren, leistet DFID Zahlungen.

Im Auftrag des Bundesministeriums für wirtschaftliche Zusammenarbeit und Entwicklung (BMZ) implementierte die KfW im Rahmen der deutschen finanziellen Zusammenarbeit mehrere bilaterale Projekte mit EOZ-Mechanismus:

- Ein Gutscheinprogramm in Kenia für medizinische Behandlungen auf Grundlage subventionierter Preise sowie ein Output-basiertes Hilfsprogramm im Energiesektor von Uganda (Blume, 2012).
- Ergebnisorientierte Finanzierung für die Gesundheit von Müttern und Neugeborenen (EOF4MNH) in Malawi. Ein Projekt, das von der Abteilung für reproduktive Gesundheit des Gesundheitsministeriums von Malawi geleitet wird. Durch angebotsseitige Anreize, nachfrageseitige Zahlungen und Investitionen in Ausstattung und Infrastruktur soll es die Qualität ausgewählter Einrichtungen der Geburtshilfe und der Neugeborenenpflege verbessern (White-Kaba, 2017).

Zur Bereitstellung von Evidenz werden Erfolgsindikatoren genutzt. Sie müssen gut ausgewählt werden, so dass Anreize bestehen, mit denen Partner*innen motiviert werden, Ergebnisse mit guter Qualität zu erzielen. In der Praxis kommt es zu Datenproblemen, wenn die Ergebnisse nicht als Folgen externer Effekte zugeordnet werden können. In diesem Fall könnte zusätzliche Evidenz hilfreich sein (Holzapfel & Janus, 2015). In den letzten Jahren haben zahlreiche Entwicklungsorganisationen Experimente als Pilotprojekte mit spezifischen EOZ-Modellen initiiert, die häufig im Gesundheits- und Bildungssektor angewandt werden. Im Landwirtschaftssektor allerdings gibt es nur begrenzte Erfahrung mit ihnen (Janus & Holzapfel, 2016). Deshalb ist Evidenz meist nur in konkreten Kontexten anwendbar. Ergebnisorientierte Ansätze sind in der EZ keine Universallösung. Sie sind zudem anfällig für Fehlanreize (Paul et al., 2018), aber nützlich, wenn sie gut ausgestaltet sind. Im Vergleich zu EOM liefern ergebnisbasierte Ansätze stärkere Anreize, sich auf verfügbare Evidenz zu beziehen, da die Zahlungen mit Indikatoren verbunden sind. Empirische Evidenz, die EOZ unterstützt, ist noch immer rar gesät. Eine jüngst veröffentlichte Studie von Bernal, Celhay und Martinez (2018) zeigt im Fall der *Saludad Mesoamerica Initiative* in El Salvador bessere Ergebnisse im Vergleich zu konventioneller Entwicklungszusammenarbeit. Mittlerweile gibt es mehr praktische Erfahrungen mit Pilotprojekten, die zu Verbesserungen neuer Vorhaben genutzt werden (Kemper & Sieler, 2016).

3.5 Evaluierung

Zwischen der Entwicklung der Evaluierungsfunktion und dem zunehmenden Bewusstsein für die Bedeutung von Evidenz besteht eine starke Verbindung. In den 1950er Jahren wurden Evaluierungen ursprünglich in Organisationen eingeführt, die in den USA angesiedelt waren. Sie sollten Ergebnisse messen und vergleichen. Zu den Organisationen zählten etwa die Weltbank und die US-amerikanische Agentur für internationale Entwicklung (*United States Agency for International Development*, USAID). Allerdings erfolgten jene Evaluierungen nur in Form von Schnappschüssen; Prozesse wurden überhaupt nicht berücksichtigt. Erst in den 1980er Jahren wurden Evaluierungen zu einem wichtigen Instrument für Transparenz und Rechenschaft, als

internationale Organisationen begannen, stärker auf die langfristigen Wirkungen von EZ zu achten. Mittlerweile existieren viele unterschiedliche Typen von Evaluierung, die für verschiedene Zwecke eingesetzt werden (Segone, 2007).

Obwohl sowohl Evaluierungen als auch ergebnisorientierte Ansätze auf Ergebnissen beruhen, haben sie doch unterschiedliche Funktionen. Während EOM eher breit aufgestellt ist, fokussieren sich Evaluierungen auf tiefgehende Analysen. Erstere werden routinemäßig durchgeführt, um ausgehend von einem Startpunkt, den Fortschritt etwa zu Lernzwecken zu verfolgen. Tiefgehende Analysen hingegen werden nur gelegentlich vorgenommen. EOM versucht zu beantworten, welche Ergebnisse erzielt werden. Evaluierungen haben zwei wesentliche Anwendungsbereiche. Erstens dienen sie instrumentellen Zwecken, um die Implementierung einer laufenden Maßnahme oder Politik zu verbessern. Hierzu wird Prozessevaluierung oder Erfolgsevaluierung genutzt. Zweitens, um zu bestimmen, ob eine Politik oder ein Programm funktioniert haben. In dieser Funktion trägt die durch die Evaluierung gelieferte Evidenz dazu bei, in spezifischen Politikbereichen dazuzulernen. Evaluierungen können sich auf einzelne Projekte, Programme oder auf das Organisationsniveau fokussieren. Evaluierungen sind auch im EOM integriert – als besondere Phase, um operative Evidenz bereitzustellen. Das konkrete Verständnis für den Zweck von Evaluierungen im Vergleich zum EOM unterscheidet sich von Organisation zu Organisation (Binnendijk, 2000).

Im neuen Jahrtausend haben die MDGs eine zusätzliche Nachfrage nach Evaluierungen hervorgerufen, und zwar einerseits durch Geber*innen, die nach Beratung suchten und andererseits durch Erfordernisse aufseiten der NROen (Bamberger, Vaessen, & Raimondo, 2015).

Auf der Projekt- oder Programmebene werden verschiedene Evaluierungen und Bewertungsmethoden genutzt, um Entscheidungen, besonders solche der Prioritätensetzung, zu erleichtern (Barr, Rinnert, Lloyd, Dunne, & Henttinen, 2016). Ökonomische Evaluierungen bewerten den Gesamtwert eines Projekts, eines Programms oder einer Politik und berücksichtigen die jeweiligen Kosten. Eine Kosten-Wirksamkeits-Analyse (KWA, *cost-effectiveness analysis*) wird dafür eingesetzt, die Kosten und *Outcomes* verschiedener Optionen, wie etwa Interventionen, miteinander zu vergleichen. Seit vielen Jahren werden Kosten-Nutzen-Analysen im öffentlichen Sektor verwendet. Mit einer Kosten-Nutzen-Analyse (KNA, *cost-benefit analysis*) wird versucht, die positiven und negativen Folgen eines Projekts zu erfassen. Dabei werden die *Outcomes* in Geldeinheiten ausgedrückt. Eine KNA kann Entscheidungsträger*innen auf verschiedene Art und Weise unterstützen: Erstens bei der Entscheidung, ob ein vorgeschlagenes Projekt oder Programm überhaupt durchgeführt werden sollte. Zweitens bei der Frage, ob ein bestehendes Projekt oder Programm fortgesetzt werden sollte. Drittens als Instrument, um zwischen alternativen Projekten oder Programmen auszuwählen. Das Kosten-Nutzen-Verhältnis (oder Nutzen-Kosten-Verhältnis) liefert Informationen darüber, ob Maßnahme A ein besseres Preis-Leistungs-Verhältnis hat als Maßnahme B.

In jüngster Zeit zogen Wirkungsevaluierungen (*impact evaluations*) mehr Aufmerksamkeit auf sich. Im Vergleich zu Projekt- oder Programmevaluierungen, die vor allem prüfen, ob eine Intervention allgemein als erfolgreich angesehen werden kann, fokussieren sich Wirkungsevaluierungen darauf, ob entwicklungsbezogenes Handeln bei bestimmten Zielgruppen Wirkungen erzielt hat. Dazu wird die alternative kontrafaktische Situation herangezogen. Sie betrifft die Outputs und *Outcomes* für den Fall, dass überhaupt keine Intervention stattgefunden hat. Evidenz, die sich aus Wirkungsevaluierungen ergibt, kann bei der Entscheidung hilfreich sein, ob ein bestehendes Projekt aufgestockt oder zurückgefahren werden sollte. Aus der Evidenz können Lehren gezogen werden und sie kann Verbesserungen dienen. Indem sie es ermöglicht, die Wirksamkeit verschiedener Projekte und Programme zu vergleichen, kann sie dazu beitragen, die Ausgestaltung von Entwicklungsprojekten, Programmen oder Politiken zu anzupassen. Allerdings hat der Einsatz von RCTs und von robusten Wirkungsanalysen auch seine Grenzen. Beispielsweise können sie nicht in der programm-basierten Entwicklungszusammenarbeit und beim Aufbau von Institutionen eingesetzt werden.

Es herrscht ein breiter Konsens unter Forscher*innen und Praktiker*innen in der Entwicklungszusammenarbeit darüber, dass Wirkungsevaluierungen eingesetzt werden sollten, um die Wirkung von Interventionen zu analysieren (Rudolph, von Schiller, & Strupat, 2017). Deshalb ist die Zahl der durchgeführten Evaluierungen im Entwicklungssektor stark angestiegen. Nach Angaben der *International Initiative for Impact Evaluation* (3ie) wurden vor 2004 jährlich weniger als 20 Studien veröffentlicht. Nach 2008 stieg diese Zahl sehr stark an. Cameron, Mishra und Brown (2016) fanden in einem internetbasierten Archiv 2.259 Studien, die zwischen 1981 und 2012 veröffentlicht wurden. Vor dem Jahr 2000 wurden 81,1% der Studien in Zeitschriften veröffentlicht, die zum Thema Gesundheit publizieren. Danach wurden Wirkungsevaluierungen auch in anderen Themenbereichen populärer (Cameron et al., 2016). Im Lauf der Zeit hat die Vielfalt der Typen und Ansätze zugenommen.

Die Ansätze einer Wirkungsevaluierung zeigen, ob Entwicklungsprogramme zu Ergebnissen führen. Dies bedeutet analytisch, dass die betreffende Intervention eine bestimmte Wirkung erzielt hat. Es können intendierte und nicht intendierte – durch andere Ursachen bedingte – Veränderungen stattgefunden haben. Diese können positiv oder negativ ausfallen. Selbst wenn ein Projekt oder Programm hinsichtlich seiner Zielerreichung als wirksam bezeichnet werden kann, ist es möglicherweise dennoch nicht erfolgreich, wenn es zu große negative Nebenwirkungen hat. Die Definition von Wirkung bestimmt Umfang und Inhalt der entsprechenden Studie. Die wesentlichen Ziele von Wirkungsevaluierungen sind: 1) Beratung (Wert einer Maßnahme als Information für die künftige Ausgestaltung der Politik); 2) Allokation (Unterstützung bei der Priorisierung, welche Projekte, Menschen und Institutionen finanziell gefördert werden sollten); 3) Analyse (etwas über Verbesserungen zu lernen und darüber, was funktioniert) sowie 4) Rechenschaft (wie sie der jeweiligen Gesetzgebung und den Grundsätzen bester Praxis zufolge erforderlich ist) (Rogers, Hawkins, McDonald, Macfarlan, & Milne, 2015). Verschiedene Konzepte von Kausalität (Was verursacht die Wirkung?) und unterschiedliche Ansätze einer ‚kausalen Inferenz‘ dienen diesen Zielen jeweils auf ihre Weise. Es werden diverse Evaluierungsdesigns benötigt, um eine bestimmte grundlegende Logik abzubilden (Stern et al., 2012). Jedes Design, das in Tabelle 3 aufgeführt ist, hat seine Vor- und Nachteile. Diese wiederum beeinflussen ihre Qualität und die herausgearbeitete Evidenz. Beispielsweise können statistische und ökonometrische Modelle Schwierigkeiten mit multiplen Ursachen aufweisen. Folglich können sie die Interaktionen zwischen Variablen möglicherweise mangelhaft erfassen oder sie mittels komplexer Pfade unsachgemäß darstellen. Insgesamt aber können diese Ansätze dazu beitragen, Schlüsselfragen des EBPM zu beantworten (Donaldson, Christie, & Mark, 2009).

Auch wenn RCTs oft als ‚Goldstandard‘ bei Evaluierungen gelten (Bédécarrats, Guérin, & Rouband, 2017), sind ihre Ergebnisse sowie ihre Relevanz durchaus mit Vorsicht zu interpretieren (Vaessen, Raimondo, & Bamberger, 2015). Beispielsweise beantwortet das Kontrafakt nur kontextspezifische Fragen (beispielsweise: funktioniert es hier, für diese besondere Gruppe?). Sie kann nicht verallgemeinert werden, um auch in anderen Kontexten zu passen (niedrige externe Validität). Die Ausgestaltungsanalyse stellt nur Verbindungen zwischen den Outputs der Intervention (Ursachen) und den *Outcomes* (Wirkungen) her und untersucht keine Prozesse. Also das, was zwischen den Outputs einer Intervention und den *Outcomes* geschieht: Sie erklärt nicht, wie *Outcomes* erreicht werden, oder wie und warum angenommene Ursachen zu den *Outcomes* beigetragen haben (Bamberger, 2015).

In den letzten Jahren haben sich Menge und Qualität der verfügbaren Evaluierungen stark verbessert. Das liegt an den zunehmenden Bemühungen vieler Länder, ihre Evaluierungssysteme zu reformieren. Einige neue öffentliche und private Institute wurden aufgebaut. So etwa wurde 2011 im Vereinigten Königreich die *Independent Commission for Aid Impact* (ICAI) eingerichtet. Im gleichen Jahr hat das BMZ das Deutsche Evaluierungsinstitut der Entwicklungszusammenarbeit (DEval) gegründet. Es existieren nun viele Vereinigungen, Gesellschaften und weltweite Netzwerke für Evaluierungen, die Kapazitätsbildung betreiben (OECD, 2016a). Auch NROen haben starke Evaluierungskapazitäten entwickelt und ihre eigenen Netzwerke aufgebaut. Diese tauschen untereinander ihre Erfahrungen aus und verbinden lokale, regionale und internationale

Einrichtungen (Segone & Ocampo, 2006). Im Jahr 2008 wurde 3ie gegründet, das einen sehr nützlichen Internetauftritt² hat und ein umfangreiches Paket an Ressourcen zur Verfügung stellt.

Tabelle 3: Wesentliche Ansätze der Wirkungsevaluierung

Design-Ansatz	Besondere Varianten	Basis für kausale Inferenz
Experimentell	RCTs, quasi-experimentell, natürliche Experimente	Kontrafakte, die Gleichzeitigkeit von Ursachen und Wirkungen
Statistisch	Statistische Modellierung, Langzeitstudien, Ökonometrie	Korrelation zwischen Ursache und Wirkung oder zwischen Variablen, Einflüsse von (für gewöhnlich) zahlreichen isolierbaren Ursachen einer einzelnen Wirkung, Kontrollgröße für Störgrößen
Theoriebasiert	<i>Kausale Prozess-Designs</i> : Theorie des Wandels, Beitragsanalyse, Wirkungspfade <i>Kausale Mechanismus-Designs</i> : Wirklichkeitsevaluierung, Kongruenzanalyse	Identifikation/Bestätigung kausaler Prozesse oder Ketten Unterstützende Faktoren und Mechanismen, die je nach Kontext relevant sind
Fallbasiert	<i>Interpretativ</i> : Naturalistisch, theoriebasiert, ethnographisch <i>Strukturiert</i> : Konfiguration, Prozessverfolgung, Kongruenzanalyse, qualitativ vergleichende Analyse, Analyse innerhalb von Fällen, Simulation und Netzwerkanalyse	Vergleich innerhalb und zwischen Fällen unterschiedlicher Kombinationen von Kausalfaktoren Analytische Verallgemeinerung auf theoretischer Grundlage
Partizipativ	<i>Normative Designs</i> : Partizipative oder demokratische Evaluierung, <i>Empowerment</i> -Evaluierung <i>Agency Design</i> : Praxislernen, Politikdialog, kollaborative Handlungsforschung	Validierung durch die Teilnehmenden, durch ihre Handlungen und Erfahrungseffekte, die von Programmen erzielt wurden Adoptionsrate, Personalisierung und Zielverpflichtung
Bewertung und Synthese	Metaanalyse, narrative Synthese, realistische Synthese	Sammlung und Aggregation innerhalb einer Anzahl an Perspektiven (statistisch, theoriebasiert, ethnographisch, etc.)

Quelle: Adaptiert und übersetzt von Stern et al. (2012)

Evaluierungen in der öffentlichen EZ in Deutschland liefern im Wesentlichen das BMZ (Referat für Evaluierung und Forschung), DEval, die Stabstelle Evaluierung der GIZ sowie die Evaluierungsabteilung der KfW Entwicklungsbank. Bereits in den frühen 1970er Jahren begann das BMZ damit, ein anspruchsvolles System aufzubauen, um Entwicklungspolitik zu evaluieren. Aus diesem Grund wurden 1971 ein Inspektionsreferat und später ein Evaluierungsreferat (*Zentrale Erfolgskontrolle*) eingerichtet (Wollmann, 2017). Mittlerweile agiert das Evaluierungsreferat des BMZ weitgehend als Lotse für das gesamte Evaluierungssystem und beauftragt die Implementierung von Evaluierungen. Das DEval leistet unabhängige und strategisch relevante Evaluierungen für alle deutschen Entwicklungsaktivitäten. Die GIZ führt strategische und Unternehmensbewertungen durch. Jährlich nimmt sie etwa 100 Projektevaluierungen vor, erstellt aber auch Studien zu spezifischen Themen. Sie dienen dazu, aus den Erfahrungen über die Breite des

2 Vgl. <http://www.3ieimpact.org>

Programmspektrums zu lernen, und zwar gestützt auf die Ergebnisse zahlreicher Evaluierungen, Überprüfungen und Fortschrittsberichte. Die GIZ führte auch ein Management-Response-System ein. Es soll den Nachfolgeprozess von Empfehlungen stärken, die auf Evaluierungen beruhen. Die KfW konzentriert sich stärker auf projektspezifische Ex-post- und Wirkungsevaluierungen. Sie führt auch Analysen zu spezifischen Themen durch, die sich auf verfügbare Ex-post-Evaluierungsberichte stützen. Sowohl die GIZ als auch die KfW haben verschiedene Mechanismen eingeführt, um evaluierungsbasiertes Wissen stärker für künftige Planungen und Entscheidungsfindungen zu nutzen. Beide Durchführungsorganisationen haben in ihren jeweiligen Organisationen zudem Anreize geschaffen, um die Evaluierungskultur zu festigen (OECD, 2015). Im Gegensatz zu den USA, wo Michael Kremer und Esther Duflo viele RCTs in Kooperation mit NROen implementierten (vgl. etwa Banerjee & Duflo, 2011), ist die Zusammenarbeit mit NROen in Deutschland eher begrenzt.

In aller Kürze: Evaluierungen sind eines der am meisten anerkannten Instrumente, um über Erfolge – neben den Fehlschlägen – zu berichten, und um ein auf Erfahrungen gestütztes Lernen zu ermöglichen.

3.6 Systematische Reviews und Synthesen

Einzelne Studien sind häufig nur von begrenztem Interesse für politische Entscheidungsträger*innen. Dies gilt vor allem bei geringen Stichprobengrößen. Systematische Reviews bündeln die Ergebnisse zahlreicher Studien und können auf verschiedenen Studientypen beruhen, auch auf Evaluierungen (Waddington et al., 2012). Da sie Fragen dazu ansprechen, ob und wie Interventionen über ein konkretes Umfeld hinaus funktionieren, liefern sie die stärkste Basis zur Politikgestaltung. Sie umfassen auch Methoden wie narrative Reviews, Metaethnographie, realistische Synthese und systematische Überblicksdarstellungen. Die gestiegene Anzahl von Reviews – vor allem von systematischen Reviews – und Synthesestudien, hat einen neuen Impuls geliefert, um sogenannte Repositorien des Wissens zu schaffen, evidenzproduzierende Organisationen zweiter Ordnung. Beispiele für derartige Einrichtungen sind etwa die *Cochrane Collaboration*, die *Campbell Collaboration*, das *Evidence for Policy and Practice Information and Coordinating Centre* und 3ie. Letztere ist besonders für den Bereich Entwicklung relevant (White, 2014).

EBM wird häufig als Referenzfall für die Nützlichkeit von Evidenz bei Entscheidungsfindungen betrachtet. Methoden, die in der Humanmedizin und in den Naturwissenschaften genutzt werden, wie etwa RCTs, werden als Blaupause für die Bereitstellung von Evidenz in anderen Disziplinen angesehen. In hierarchischen Systemen, die die Evidenz klassifizieren, die von EBM verwendet wird, gilt die Schaffung von systematischen Reviews mit hoher Qualität als führende Methode. Allerdings bestehen im EZ-Bereich viele Probleme mit der Evidenzsynthese. Das liegt vor allem daran, dass sich vorhandene EZ-Studien in ihrem Fokus und Ansatz, bei den verwendeten Methoden, den genutzten Variablen sowie in der Qualität der Daten und Ergebnisse unterscheiden. Da systematische Reviews in der EZ normalerweise lediglich auf einer geringen Anzahl von eingeschlossenen Studien beruhen, verfügen sie über eine wesentlich geringere Erklärungskraft (Mallett, Hagen-Zanker, Slater, & Duvendack, 2012).

3.7 Zwischenfazit

Daten sind *die* wesentliche Ressource, um Evidenz zu generieren. Diesbezüglich gibt es hinsichtlich der Abdeckung der SDGs immer noch gewaltige Lücken, besonders mit Blick auf Länder mit niedrigem Einkommen (LICs). Die Datenverfügbarkeit unterscheidet sich auch bezüglich der Politikfelder.

Bis vor Kurzem war in der EZ überzeugende Evidenz nur bis zu einem gewissen Maß für bestimmte Kontexte verfügbar. Und zwar vor allem auf der Projektebene bzw. in einigen Sektoren wie Gesundheit und Bildung. Monitoringdaten und Informationen wurden gesammelt, um die Ex-post-Bewertung nachverfolgen zu können. Umfangreichere ergebnisorientierte Managementsysteme verfügten häufig nur über ein geringes Leistungsvermögen. Seit dem Jahrtausendwechsel hat sich die Verfügbarkeit von Daten und Informationen stark verbessert. Monitoringdaten liefern umfangreiche Leistungsdaten. Allerdings sind Monitoringsysteme besonders in Partnerländern noch weit davon entfernt, perfekt zu sein. EOM und ergebnisorientierte Ansätze sind heute eine solide Basis für operative Evidenz. Verschiedene Evaluierungstypen sind hilfreich, um unterschiedliche Kategorien analytischer und operativer Evidenz verfügbar zu machen.

4 Die Nutzung von Evidenz in der Entwicklungszusammenarbeit

4.1 Wahrnehmung

Nach der Jahrtausendwende herrschten noch eine starke Ablehnung und Skepsis gegenüber der Nutzung von Wirkungsevaluierungen in der EZ (Faust, 2017). Die tatsächliche Nachfrage vonseiten der politischen Entscheidungsträger*innen nach mehr und besserer Evidenz, um ihre Entscheidungen zu erleichtern, war damals gering. Das lag daran, dass Evidenz für sie nur begrenzt relevant war. Seit damals hat sich die Wahrnehmung von EZ-Akteur*innen gewandelt und der Wert von Evidenz ist für sie gestiegen.

Mehrere Faktoren sind dafür verantwortlich, dass Evidenz auf den Agenden der OECD/DAC-Mitgliedsstaaten nach oben gerückt ist. Erstens wurde der Wert von Evidenz in einigen Mitgliedsstaaten bereits höher gewichtet. Im Vereinigten Königreich hatte sich schon zuvor die politische Meinung dazu geändert, wie bedeutend Ergebnismessungen und -management sind. Zum Ende der 1990er Jahre hatte die Regierung von Tony Blair den Ausdruck ‚evidenzbasierte Politikgestaltung‘ populär gemacht (Banks, 2009). Seine Regierung definierte ganz klar die erstrebenswerten Ziele der EZ (*Department for International Development*, 2011). Auch von der Nachfrageseite her (politische Entscheidungsträger*innen) gab es einen Schub hin zur stärkeren Nutzung von Evidenz, um Ergebnisse darzustellen. Gleichzeitig wurde Evaluierungen mehr Aufmerksamkeit geschenkt. Selbst wenn dies durch die politischen Entscheidungsträger*innen insbesondere aus strategischen Gründen erfolgte, um Evaluierungen als Instrumente zur retrospektiven Rechtfertigung von bereits getroffenen Entscheidungen zu nutzen (Stewart & Smith, 2015), beförderte dieses veränderte Verständnis doch einen kulturellen Wandel: bei Treffen zwischen Partner*innen verschob sich der Schwerpunkt weg vom Austausch über ausgegebenes Geld hin zu Ergebnissen und Wirkungen. Das Vereinigte Königreich richtete das ICAI ein, um an das Parlament zu berichten und die geleistete ODA zu überprüfen (OECD, 2014b). Seit einigen Jahren sind Ergebnisse und Rechenschaft auch ein wichtiger Schwerpunkt in der US-amerikanischen Entwicklungszusammenarbeit. Der *Performance and Results Modernization Act* der Regierung von US-Präsident Barack Obama legte den Grundstein für ihre stärkere Ausrichtung auf Ergebnisse (OECD, 2016b).

Delegierte und Vertreter*innen dieser Pionierstaaten waren es, die das Thema auf die Agenda der internationalen Plattformen setzten. Sie wurden dabei durch die Aktivitäten ihrer Durchführungsorganisationen und -institute unterstützt. Das Bekanntwerden ernster ‚Schlupflöcher‘ in der EZ war der Anstoß, um 3ie zu gründen. Es folgten bilaterale Geber wie DFID und USAID, die in ihrer Politik der Wirkungsorientierung zunehmende Aufmerksamkeit schenkten (Levine & Savedoff, 2016). Insbesondere das DFID (Vereinigtes Königreich) finanzierte viele RCTs und hat die internationalen Debatten stark beeinflusst. Allerdings ist es unklar, ob die korrespondie-

renden Ergebnisse in der Forschungsdomäne verbleiben, oder ob sie auch dazu genutzt werden, um die eigenen DFID-Programme zu verbessern. Im Gegensatz dazu machten die Niederlande Ausschreibungen, bevor sie ihre eigenen Projekte implementierten. Deshalb sind ihre Evaluierungsergebnisse für die Umsetzung von Nutzen. Die *Development Research Group* evaluiert gestützt auf ihren eigenen Fonds für strategische Wirkungsevaluierungen die Wirkungen ausgewählter Projekte der Weltbank und anderer multilateraler Entwicklungsbanken.

Internationale Organisationen und Plattformen, einschließlich der OECD, fungieren als wichtige Plattformen für Diskussionen über die ‚Ergebnis-Agenda‘ und ihre Elemente. Produkte der OECD wie die *DAC Peer Review Reports* spiegeln das Bedürfnis wider nach einer Stärkung der entsprechenden Systeme der Geberländer, ihrer Prozesse und Aktivitäten.

Eine stärkere Wahrnehmung von Evidenz ist auch in Partnerländern ein wichtiges Thema. Deshalb haben einige EZ-Geber*innen, Organisationen und Netzwerke Programme gestartet, die der Bewusstseinsbildung und dem Kapazitätsaufbau bei Entscheidungsträger*innen dienen (Stewart, 2015).

Sie arbeitet eng mit politischen Entscheidungsträger*innen in Entwicklungsländern zusammen. Um eine ‚Kultur evidenzbasierter Politikgestaltung‘ zu festigen, organisieren sie Workshops, um die Nachfrage zu beleben. In diesen Trainingsprogrammen zur Kapazitätsbildung erklären Referent*innen, welche Art von Evidenz durch Wirkungsevaluierung verfügbar ist, wie sie bei der Politik- und Programmausgestaltung informativ unterstützen kann und warum eben diese Form von Evidenz wichtig ist. Diese Veranstaltungen helfen dabei, dass die Wahrnehmung von und die Kenntnisse über den Wert von Evidenz in der Politikgestaltung steigen. Zusätzlich werden ‚Praxen der Wirkungsevaluierung‘ um ‚Politikfelder‘ herum organisiert, um Organisationen in Entwicklungsländern darin zu unterstützen, Evaluierungen für konkrete Programme zu implementieren (Monaghan, 2013).

Auf ähnliche Weise hat DFID 2013 das Programm ‚*Building Capacity to Use Research Evidence*‘ (BCURE)³ aufgesetzt, um mit diesem Thema über zwölf Länder zu erreichen. Im Laufe der letzten vier Jahre hat BCURE die Nutzung von Evidenz durch Entscheidungsträger*innen gefördert. Dies war ein wichtiger Beitrag, um Entwicklungswirkungen zu verbessern. Tabelle 4 zeigt einen Rahmen mit drei Kategorien zu den Evidenztypen, die von den entsprechenden politischen Entscheidungsträger*innen an drei Standorten der Studie genutzt wurden (Rinnert & Brower, 2017).

Das *Africa Evidence Network*⁴ besteht aus Forscher*innen, Praktiker*innen und politischen Entscheidungsträger*innen aus Universitäten, Regierungseinrichtungen und NROen, die in Afrika arbeiten, aber auch daran interessiert sind, Evidenz für Entscheidungen bereitzustellen.

Auch das BMZ erkennt die Bedeutung an, die darin liegt, Evidenz in Gesundheits- und Sozialthemen zu berücksichtigen. Seit 2004 wird die ‚*Health Practice Collection*‘ veröffentlicht, die eine Gemeinschaftsinitiative von GIZ und KfW darstellt. Das sind von Expert*innen geschriebene Fallstudien und Kurzdarstellungen zum Thema Evidenz, die Ergebnisse der von Deutschland unterstützten Programme aufführen und die während ihrer Implementierung generiert wurden.⁵

3 Vgl. <https://bcureglobal.wordpress.com/>

4 Vgl. <http://www.africaevidencenetwork.org>

5 Vgl. <http://health.bmz.de/ghpc/index.html>

Tabelle 4: BCURE-Rahmen zum Wert genutzter Evidenz

	Transparente Nutzung	Eingebettete Nutzung	Instrumentelle Nutzung
Beschreibung	Gesteigertes Verständnis und transparente Nutzung der (Beweislage der) Evidenz durch die politischen Entscheidungsträger*innen	Infolge der Evidenz wird keine direkte Handlung ausgelöst. Allerdings wird die Evidenznutzung eingebettet in Prozesse, Systeme und Arbeitskultur	Wissen über robuste Evidenz wird genutzt, um die Politik oder ein Programm darüber zu informieren
Beispiele	BCURE Vaka Yiko: Mehrere Gesprächsrunden wurden abgehalten, um die Lücke zwischen Forschung und Politikgestaltung zum Klimawandel in Kenia zu schließen und Entscheidungsträger*innen darin zu unterstützen die volle Beweislage an Evidenz zum Klimawandel im Land zu würdigen	BCURE Harvard: Die Forscher*innen arbeiteten direkt mit Fachleuten der Regierung zur Schaffung eines Dashboards, das ausgestaltet wurde, um über 50 Indikatoren, die als wesentlich für die Evaluierung des Mahatma Gandhi <i>National Rural Employment Act</i> aufgefasst wurden, an einer Stelle zu bündeln	BCURE University of Johannesburg: In Südafrika wurde die Evidenzkarte, die vom Ministerium für Planung, Monitoring und Evaluierung veröffentlicht wurde, direkt in die Entscheidungsfindung für die Informationsschrift zu menschlichen Siedlungen eingepflegt
Umfang: Das Feld der politischen Entscheidungsträger*innen, die von der Reform betroffen wurden – ist deren Wirkung über Akteur*innen hinaus weitreichend?	+++ zwischenstaatlich, Policy Teams und Länderbüros	+ Ein Ministerium der Kommunalverwaltung	+++ Politik auf nationaler Ebene
Tiefe: Wirkung des Wandels, wie umfangreich ist die Reform? Gibt es substanzielle Veränderungen gegenüber der früheren Praxis?	+ Keine tiefgründigen Veränderungen in der Praxis, die direkt BCURE zugeordnet werden könnten, aber ein Beitrag zur Initiierung nachfolgender Handlungen	++ Schaffung eines Evidenz-Tools mit unmittelbarer Nutzung; 150.000 Klicks im ersten Jahr	++ Die Siedlungspolitik erreicht möglicherweise einen großen Anteil der Bevölkerung. Allerdings wurde die Gesamtwirkung auf Basis der Ergebnisse von Monitoring und Evaluierung bislang noch nicht ermittelt
Nachhaltigkeit: Wie nachhaltig ist die Veränderung bezüglich der Nutzung von Evidenz?	+ Einmalige Besprechungen, aber mit dem Potenzial, weitere Veränderungen zur Nutzung von Evidenz anzustoßen	++ Evidenz legt nahe, dass dies zu anhaltenden Veränderungen führt	++ Evidenz wird für mehrere politische Entscheidungen genutzt, um mit dem Potenzial weitere politische Entscheidungen zu beeinflussen

Quelle: Adaptiert und übersetzt von Rinnert und Brower (2017)

4.2 Einbeziehung in Entscheidungsprozesse

Evidenz in Entscheidungsprozessen zu berücksichtigen, kann einen Mehrwert gegenüber meinungsbasierten Entscheidungen haben, wenn zwei kritische Annahmen erfüllt sind: Erstens müssen die Qualität und Stärke der verfügbaren Evidenz solide sein. Zweitens muss eine umfangreiche Menge an Evidenz zu einem Problem oder einem Umfeld vorliegen. Wenn es nur für bestimmte Kontexte Evidenz gibt, wird sie wahrscheinlich nur von geringerer Bedeutung sein. Dies war bis vor Kurzem meistens der Fall. Beispielsweise wies 2005 die OECD/DAC-Begutachtung der individuellen Anstrengungen Deutschlands im Bereich der Entwicklungszusammenarbeit darauf hin, dass Deutschlands Fähigkeiten nur schwach ausgeprägt seien, aussagekräftige Informationen nachzuverfolgen und darüber zu berichten (OECD, 2006). Im selben Jahr schlussfolgerten die Autor*innen des Berichts *“When Will We Ever Learn?”*, der von der *Evaluation Gap Working Group* veröffentlicht wurde, dass es nur wenig bis keine aussagekräftige verfügbare Evidenz gibt, aus der hervorgeht, dass Entwicklungsprogramme tatsächlich funktionieren. Veröffentlichte Evaluierungsberichte zeigten deutlich, dass Geld ausgegeben wurde und welche direkten Aktivitäten oder Dienste erbracht wurden. Sie zeigten aber nicht, ob diese geleisteten Dienste für die Zielgruppen tatsächlich nützlich waren (Gaarder, 2014). Es gab zudem auch zu wenig qualitative Studien (Savedoff et al., 2006). Es war das wesentliche Ergebnis eines Workshops über die rigorose Wirkungsanalyse in der EZ, der 2008 vom Deutschen Institut für Entwicklungspolitik (DIE) veranstaltet wurde, dass das verfügbare methodologische Wissen begrenzt und die damit zusammenhängenden Einsichten ziemlich knapp waren. Dennoch haben sich die Dinge verändert und weiterentwickelt.

Seit dem Workshop 2008 hat sich die Lage deutlich verbessert. Die Anzahl der Wirkungsanalysen hat signifikant zugenommen, die deutsche Forscher*innen und operative Abteilungen – einschließlich die der internationalen Zusammenarbeit – durchgeführt haben. Die gesamte Evaluierungslücke schließt sich, auch wenn sie bislang noch nicht vollständig geschlossen ist. Inzwischen existiert eine Vielzahl qualitativ hochwertiger und für die Politik relevanter Forschung. Gleiches gilt für Evidenz über verschiedene Aspekte der EZ. Das liegt an einer gestiegenen Nachfrage und an der Finanzierung derartiger Produkte.

Evidenz spielt eine Rolle! Es gibt nun zahlreiche Beispiele für die Nutzung von Evidenz in der EZ (Carden, 2009; Newman et al., 2014; Young, 2006). Die konkrete Antwort auf die Frage, wie relevant Evidenz ist, hängt allerdings davon ab, wie Evidenz definiert wird. In welcher Weise – und in welchem Ausmaß – die bereitgestellte Evidenz zu Veränderungen in Politik oder Praxis beigetragen hat, ist auch in Zahlen schwer auszudrücken. Evidenz wird von Entscheidungsträger*innen unterschiedlich genutzt, etwa für analytische und/oder für operative Zwecke.

Ergebnisinformationen werden zum Zweck der Rechenschaft oder als Management-Instrument auf verschiedenen Ebenen genutzt: auf der Ebene der Maßnahme, des Landes oder des Unternehmens. Gegenwärtig richten viele DAC-Mitglieder einen stärkeren Fokus auf den ersten Aspekt, also den der Maßnahme. So soll dargestellt werden, dass Entwicklungszusammenarbeit in den Projekten, Aktivitäten und Programmen tatsächlich funktioniert. Das System der Vereinten Nationen und die meisten Agenturen der internationalen Zusammenarbeit nutzen derzeit einige Varianten von EOM-Systemen und -Rahmen, meistens eher auf der Organisations- oder programmweiten Ebene. Stark nachgefragt sind Ergebnisinformationen zu Rechenschaftszwecken und für die Kommunikation.

Eine neuere Befragung der OECD galt der Messung und dem Management von Ergebnissen und richtete sich an Ministerien und Entwicklungsagenturen. In den Antworten wurde an erster Stelle als Zweck der Evidenznutzung am häufigsten die ‚Nachverfolgung von Fortschritten‘ genannt (Tabelle 5). Informationen über das Ergebnis werden allerdings auch für Entscheidungsprozesse meistens auf der Programm- oder auf der organisatorischen Ebene verwendet. Informationen über Outputs, *Outcomes* oder Wirkungen von EZ-Interventionen werden eher weniger dazu eingesetzt, Analysen oder Erfolgserklärungen durchzuführen (OECD, 2014a).

Tabelle 5: Wie Ergebnisinformationen genutzt werden

Niveau	Meistgenannte erste Priorität	Meistgenannte zweite Priorität
Agentur/ Unternehmen	Für Rechenschaft gegenüber dem Parlament/der Gesetzgebung (11)	In externer Kommunikation (11)
Land	Nachverfolgung von Fortschritten (9)	Für Entscheidungsfindungen (7)
Sektor/ Programm	Nachverfolgung von Fortschritten (13)	Für Entscheidungsfindungen (8)
Projekt	Nachverfolgung von Fortschritten (15)	Für Entscheidungsfindungen (11)

Quelle: Adaptiert und übersetzt von OECD (2014a)

Gemäß einer unlängst durchgeführten DFID-Befragung über die Nutzung von Evidenz im Ministerium, stimmten 60% der Mitarbeitenden darin überein, dass sie in den letzten drei Jahren zugenommen habe (United Kingdom, 2014).

In den USA werden alle Ergebnisse der länderspezifischen Entwicklungsstrategien in Wirkungsrahmen abgebildet. Sie sind in Entwicklungsziele, intermediäre und sub-intermediäre Ergebnisse sowie nach Erfolgsindikatoren angeordnet. Diese Ergebnisse dienen als Grundlage dafür, Projekte und Evaluierungen auszugestalten.

In Deutschland nutzen GIZ und KfW routinemäßig Ergebnisinformationen, um das Programm-Management zu verbessern (Deutsche Gesellschaft für internationale Zusammenarbeit, 2013; KfW Development Bank, 2017).

Wie gegenwärtig verfügbare Daten und Informationen verwendet werden, die durch das Ergebnismanagement generiert werden, hängt von der intendierten Nutzung und den Erfolgsindikatoren ab. In theoretischer Hinsicht ist EOM ein mächtiges Instrument. Allerdings gibt es in der Praxis zahlreiche Schwierigkeiten, die den Informationswert solcher Systeme begrenzen. Beispielweise zeigt der OECD *Managing and Results Survey* aus dem Jahr 2013, dass viele Ministerien und Entwicklungsagenturen bei der Verwendung von Ergebnissystemen auf Herausforderungen stießen. Neben anderen Ergebnissen gaben 5 (13) von 28 Antwortenden die Antwort 100% (häufig) bei der Aussage „unvollständige Wirkungsrahmen (Fehlen von Ausgangswerten oder Unterzielen, unklare Ergebnisse)“; 3 (15) von 28 Antwortenden wählten zu 100% (häufig) die Aussage: „Schwierigkeiten bei der Verbindung zwischen Budgets und Ergebnisinformationen“; und 3 (15) von 28 Antwortenden bestätigten zu 100% (oft) die Aussage: „Schwierigkeiten bei der Auswahl angemessener Indikatoren, die Ergebnisse auf dem korrekten Niveau messen“. Derartige Schwierigkeiten entstehen vielfach, weil die Komplexität im Umgang Probleme bereitet.

Bestehende Richtlinien, institutionelle Dokumente und Anleitungen liefern umfangreiche Informationen, um Evidenz bei der Politikgestaltung zu nutzen. Das bestehende deutsche EOM hat einige Schwächen. Während Länderstrategien die Ziele der gesamten deutschen EZ und spezifische Programme für jedes Partnerland festsetzen, arbeitet das BMZ gegenwärtig an einer systematischeren Messung und Berichterstattung der Ergebnisse.

Nach Angaben des jüngsten OECD/DAC *Peer Review Report* war das BMZ nicht in der Lage, Ergebnisinformationen und Evidenz aus anderen Quellen vollständig zur strategischen Planung und Kommunikation einzusetzen. Da übergreifende Indikatoren und Unterziele auf der organisatorischen Ebene fehlen, ist es unklar, wie das BMZ die Ergebnisinformationen aus der Programmimplementierung und aus anderen Evidenzquellen in die strategische Planung und öffentliche Kommunikation einspeist. Im Unterschied dazu nutzen GIZ und KfW jeweils routinemäßig die Ergebnisinformationen auf den Projekt- und Programmebenen, um das übergreifende korrespondierende Management zu verbessern. Das BMZ schreibt der GIZ und der KfW vor,

eine Ergebnismatrix anzuwenden und Monitoring zu implementieren. In beiden Organisationen sind gute Monitoringsysteme vorhanden. Sie beruhen primär auf Indikatoren und Daten der eigenen Systeme der Partnerländer, was den Bedarf an zusätzlichem Monitoring und Berichterstattungen verringert. Zusätzlich dazu beinhalten die meisten Programme der technischen Kooperationen eine spezifische Komponente, um die Monitoring- und Evaluierungssysteme der Partnerländer zu stärken (OECD, 2015).

Im Kontext der Projektzyklen besteht etwas Freiraum, um Evidenz klarer zu berücksichtigen. So ergänzte beispielsweise USAID eine ausdrückliche Erklärung in der operativen Politik ihres Programmzyklus dazu, wo und wie die eingesetzten finanziellen Mittel für Entwicklungszusammenarbeit auf Analysen und Folgerungen beruhen sollten, die von Evidenz unterstützt werden (United States Agency for International Development, 2017). Die 'Leitlinien für die bilaterale Finanzielle und Technische Zusammenarbeit mit Kooperationspartnern der deutschen Entwicklungszusammenarbeit' des BMZ (Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung, 2008) setzen fest, dass wirtschaftliche, sozio-ökonomische, geschlechtsspezifische und ökologische Wirkungen bei der Auswahl von Projekten berücksichtigt werden sollten. Unlängst wurde eine Verfahrensreform auf BMZ-Ebene implementiert. Nachweise bestehender Evidenz bei Projektvorschlägen und die Unterbreitung von Projektalternativen werden seitdem stärker betont. Allerdings gibt es immer noch Raum für Verbesserungen.

Evidenz, die aus Evaluierungen hervorgeht, hat nun einen hohen Stellenwert, wenn sie den Nachweis liefert, dass eine laufende Intervention erfolgreich ist. Beispielsweise wurde im Fall eines sozialen Geldtransferprogramms (*social cash transfer programme*) von DFID der Nachweis erbracht, dass diese Politik einen positiven Effekt auf das Leben armer Menschen hat. Infolgedessen wurde das Programm aufgestockt (Barr et al., 2016). Als Erfolg gewertet wurde das gemeinsame ergebnisorientierte Finanzierungsprogramm EOF4MNH für die Unterstützung der Gesundheit von Müttern und Neugeborenen des Gesundheitsministeriums von Malawi und der KfW (White-Kaba, 2017). Die Regierung von Malawi erklärte, dass sie willens sei, das Programm nach dem Auslaufen der deutschen Finanzierung fortzuführen. Deshalb werde das Programm in die nationale Strategie des Gesundheitssektors mit aufgenommen. Allerdings gibt es in der Literatur unzureichende Informationen darüber, welche Folgen es für Projekte oder Programme hat, wenn positive Beweise oder negative Implikationen fehlen. Wahrscheinlich ist die Literatur diesbezüglich verzerrt.

Partnerländer nutzen Evidenz bei der Prioritätensetzung für Gesundheitsprogramme in Ländern mit niedrigem Einkommen (LIC). Um die Verteilung finanzieller Mittel zu steuern, wird Evidenz zur Krankheitslast genutzt und zur Kostenwirksamkeit von ausgewählten Maßnahmen; beispielsweise die globale Krankheitslast und die KWA. Beim Austausch über Evidenz erwähnen politische Entscheidungsträger*innen selten, dass sie besondere Literaturübersichten verwenden. Hinweise zur Evidenz fallen eher als rhetorisches Mittel in der Argumentation darüber, was priorisiert werden sollte. Vielen Einrichtungen der Vereinten Nationen und bilateralen Geberländern hilft quantitative Evidenz dabei, (instrumentelle) Entscheidungen über Prioritäten zu treffen. Im Unterschied dazu schätzen NROen auch den Wert von qualitativen Informationen. Ansätze mit gemischten Methoden (*mixed methods*) werden heutzutage häufig als Mittel verstanden, angemessene Hintergrundinformationen einzubinden. Da allerdings bilaterale Geberorganisationen und nationale Organisationen dazu verpflichtet sind, die Regierungsprioritäten zu implementieren, werden Prioritäten politisch gesetzt und Evidenz insbesondere symbolisch verwendet, um bereits von Politiker*innen getroffene Entscheidungen zu untermauern (Kapiriri, Sinding, & Arnold, 2017). Wie etwa im Fall von Simbabwe, erkennen Partnerländer bisweilen auch die Bedeutung und den zusätzlichen Nutzen von Evidenz. So bat die Regierung von Simbabwe um eine Ex-post-Evaluierung der Wirkungen eines Programms, das AIDS-Waisen unterstützt (KfW Development Bank, 2016).

4.3 Barrieren und unterstützende Faktoren

Auf konzeptioneller und theoretischer Ebene stimmen viele der beteiligten Parteien darin überein, dass es notwendig ist, verfügbare Evidenz einschließlich Informationen, Daten, Statistiken, Messgrößen und Messungen bei der Politikgestaltung zu nutzen. Dennoch begrenzen einige wesentliche Barrieren die praktische Nutzung von Evidenz. Bei diesen Themen handelt es sich nicht um spezifische Aspekte der EZ (vgl. z.B. Oliver, Innvær, Lorenc, Woodman, & Thomas, 2014).

1. *Politische Ökonomie*

Herkömmlicherweise wurde die Verbindung zwischen Forschung und Politik als ein linearer oder rationaler Prozess verstanden, in dessen Verlauf viele Forschungsbefunde von der ‚Forschungssphäre‘ hin zur ‚Politikssphäre‘ verlagert werden, die sich dann auf die Entscheidungen der politisch Verantwortlichen auswirken (Young, 2006). Neu gewonnene Evidenz wird aber wahrscheinlich keine direkten Folgen für viele Entscheidungen haben. Vernunftgründe ignorieren die Unterschiede zwischen Rhetorik und praktischer Wirklichkeit sowie zwischen Eigeninteressen und der Wissensbeschaffung im EZ-Sektor (McNulty, 2012). Politische Entscheidungsträger*innen sind oftmals politische Akteur*innen. Deshalb sind Werte, politische Überzeugungen und Interaktionen mit dem politischen System relevant. In der Praxis ist Evidenz nur einer von zahlreichen anderen Faktoren, die auf politische Prozesse einwirken (Parkhurst, 2016).

2. *Erfahrungen versus Evidenz*

Einwände dürften auch vonseiten der Entscheidungsträger*innen und Verwaltungsmitarbeitenden kommen –, weil sie Innovationen allgemein ablehnend gegenüberstehen oder weil sie finden, dass die gewonnene Evidenz nicht mit ihren eigenen Erfahrungen korrespondiert.

3. *Timing und günstige Gelegenheiten*

Manchmal wird verfügbare Evidenz bei der Entscheidungsfindung nicht beachtet, weil sie nicht dann vorliegt, wenn sie benötigt wird. Bamberger (2015) zeigt auf, dass Ergebnisse manchmal präsentiert werden, nachdem die entsprechenden Entscheidungen bereits getroffen wurden. Oder Berichte werden zu früh veröffentlicht, bevor die Öffentlichkeit begonnen hat, sich mit dem Thema näher zu befassen. Ähnliches verdeutlichen die Ergebnisse einer Befragung vom *Accountability Office* der US-Regierung zum organisationalen Erfolg und zu Aspekten des Managements. Nur etwa 15% der USAID Manager*innen bestätigten demnach ausdrücklich, dass sie Zugang zu Erfolgswissen hatten, als sie sie benötigten, um ihre Abläufe oder die technische Arbeit auszuführen (United States Government Accountability Office, 2017).

4. *Offenheit gegenüber Evidenz*

Politische Entscheidungsträger*innen, die sich offen dafür zeigen, Evidenz zu nutzen, schränken ihren Entscheidungsraum ein. In solchen Fällen bedarf es besonderer Rechtfertigungen, wenn Evidenz bei Entscheidungen nicht berücksichtigt wird. So etwa, wenn verfügbare Evidenz im Konflikt zu den eigenen ideologischen Vorstellungen steht.

5. *Widersprüchliche und inkonsistente Evidenz*

Politische Entscheidungen zu treffen, stellt in vielerlei Hinsicht eine Herausforderung dar: aufgrund von Unsicherheiten, einem komplexen Umfeld und wegen unterschiedlicher Weltanschauungen. Diese Aspekte und die Beschaffenheit der Realität bestimmen das politische Universum und folglich die Nachfrage nach Evidenz. Häufig gibt es keine einfachen Lösungen, da sich akademische Ergebnisse unterscheiden. Forscher*innen können verschiedenen ‚Schulen‘

angehören und unterschiedliche Forschungsmethoden einsetzen. Forscher*innen haben auch einen unterschiedlichen disziplinären Hintergrund und teilweise sich stetig verändernde Blickwinkel. Es gibt verschiedene Arten von Evidenz, die oftmals miteinander konkurrieren. Widersprüchliche und manchmal inkonsistente Evidenz kann ihre Akzeptanz erschweren, um als Ressource beim Treffen von Entscheidungen zu dienen. Außerdem kann sich gewonnene praktisch-technische Evidenz auch entsprechend der Kapazitäten von Verwaltungen unterscheiden. Ergänzend dazu kann soziale Evidenz aufgrund bestimmter sozialer Bewertungen und Interessen unterschiedliche Formen annehmen.

6. *Richtige und falsche Fragen*

Es gibt Anhaltspunkte dafür, dass die Akzeptanz der Forschung oder Evidenz vor allem davon abhängt, ob sie als Lösung für ein bestehendes Problem wahrgenommen wird oder nicht (Young, 2006). Einige politische Entscheidungsträger*innen sind nicht der Meinung, dass Forschungsfragen für sie relevant sind (Fourie, 2017). Bei komplexen Evaluierungen ist es auch denkbar, dass falsche Fragen und irrelevante Ergebnisse präsentiert werden, die für politische Entscheidungsträger*innen uninteressant sind. Manchmal konzentrieren sich Evaluierungen auch zu eng nur auf die Wirkungen von Programmen oder es bestehen Schwierigkeiten, bei komplexen Programmen Ursache und Wirkung zuzuordnen.

7. *Ergebnisdarstellung*

Selbst bei Wirkungsevaluierungen, die aus Gründen der Rechenschaft erfolgen oder als Instrument für effektivere Entwicklungsziele dienen, sind die Art und Weise relevant, wie Ergebnisse präsentiert werden. Evaluierungsergebnisse müssen so dargestellt werden, dass politische Entscheidungsträger*innen sie als bedeutsam einstufen. So weist White (2014) darauf hin, dass Politik sowohl vom Produkt als auch vom Prozess beeinflusst wird. Das Produkt, zum Beispiel eine Studie, sollte sämtliche Evaluierungsfragen behandeln, die für politische Entscheidungsträger*innen von Interesse sind. Nach Stern et al. (2012) sind dies folgende Fragen:

- Inwieweit kann der Maßnahme eine spezifische (Netto-)Wirkung zugeordnet werden?
- Hat die Maßnahme eine Veränderung bewirkt?
- Auf welche Weise hat die Maßnahme etwas bewegt?
- Wird die Maßnahme auch in anderen Zusammenhängen funktionieren?

Politische Entscheidungsträger*innen lesen keine akademischen Zeitschriften. Sie sind nicht an t-Statistiken, Wahrscheinlichkeitswerten oder an anderen Formen interessiert, statistische Signifikanz zu präsentieren. Sie wollen über die Bedeutung der Ergebnisse und die damit zusammenhängenden Kosten informiert werden. Darüber hinaus wollen Forscher*innen dazu beitragen, dass politische Entscheidungsträger*innen ihre Ergebnisse kennen, wenn sie dazu angehen sind, Politik zu beeinflussen. Die politische Wirkung wird größer ausfallen, wenn sich politische Entscheidungsträger*innen und Programm-Manager*innen von Anfang an engagiert haben, und zwar beginnend mit der Formulierung der Evaluierungsfragen. Allerdings muss die Einschätzung der politischen Entscheidungsträger*innen, warum sie Evidenz nutzen oder nicht, nicht notwendigerweise mit der objektiven Realität übereinstimmen.

8. *Wert der Evidenz im Zusammenhang mit bestehenden politischen Rahmenwerken*

Falls die Empfehlungen – die auf einer Studie wie etwa einer Evaluierung beruhen – nicht in den politischen Rahmen der Vergabebehörde passen, wird die Evidenz wahrscheinlich nicht berücksichtigt werden. Dies war der Fall bei einer DEval-Studie über allgemeine Budgethilfe für

Subsahara-Afrika (Krisch, Schmitt, & Dörr, 2015). Das BMZ zog den Schluss, dass der bestehende politische Rahmen der Regierung den Nutzen einiger Ergebnisse einschränken würde.

9. *Bestehende Evidenzkultur*

Auf nationaler Ebene ist die ‚Evidenzkultur‘ und das Interesse der politischen Entscheidungsträger*innen, Evidenz zu nutzen, unterschiedlich ausgeprägt. Es gibt Hinweise darauf, dass in Ländern wie dem Vereinigten Königreich eine ‚Evidenzkultur‘ entwickelt werden konnte – eine Kultur, die offener für die Wahrnehmung und Nutzung von Evidenz ist. Damit ist gemeint, die Verwendung von Evidenz in bestehende organisatorische Strukturen und Prozesse einzubinden (Baker & Salib, 2017). Dafür werden ein politischer Schub oder Vermittler*innen benötigt. So scheint es beispielsweise in Lateinamerika eine ausgeprägtere Nutzungskultur von Evidenz zu geben, die sich auf Wirkungsevaluierungen stützt, und zwar als Folge des Programms PROGRESA in Mexiko. Aber auch in Afrika nimmt das Interesse an evidenzbasierter Politik zu, so etwa in Nigeria (Uneke et al., 2011).

Mit dem Ziel, die Einbindung von Evidenz bei Entscheidungsfindungen zu verbessern, erhielt der Ausdruck ‚Wissensübersetzung von der Forschung in die Praxis‘ mehr Aufmerksamkeit. Es wurden viele Modelle und Ansätze dafür entwickelt, wie politische Entscheider*innen motiviert werden können, verfügbare Evidenz in die Entscheidungsfindung einzubeziehen (Howlett & Craft, 2013; Young, Ashby, Boaz, & Grayson, 2002). Tatsächlich hängt die Nützlichkeit der vorgeschlagenen Ansätze entscheidend vom Umfeld ab. Das Konzept, das der nachfolgenden Fallstudie zugrunde liegt, ist nur ein (erfolgversprechender?) Ansatz.

4.4 **Zwischenergebnis**

In jüngster Zeit war eine veränderte Einstellung von politischen Entscheidungsträger*innen zu beobachten. Einige Regierungen und internationale Einrichtungen haben damit begonnen, EBMP auf die Agenda zu setzen. Dies hat sich wiederum auf die Generierung und das Angebot an Evidenz ausgewirkt.

Es gibt Anzeichen dafür, dass die Bedeutung von Evidenz in der EZ zugenommen hat. Ihre Wahrnehmung auf der Ebene der politischen Entscheidungsträger*innen in Geberländern und – zu einem geringeren Ausmaß – in Partnerländern hat sich deutlich verbessert. Praktisch-technische Evidenz wird zu operativen Zwecken in Projekten und akademische Evidenz zu analytischen Zwecken genutzt, um die Politikgestaltung und die Auswahl von Einsatzstrategien zu steuern. Im Gegensatz dazu wird soziale Evidenz bei praktischen Entscheidungen auf der Ebene von Verwaltungsbehörden und Regierungsorganisationen oft nicht direkt berücksichtigt. Allerdings wirkt sich soziale Evidenz manchmal direkt auf politische Entscheidungsprozesse aus, und ihr Beitrag kann Entscheidungen beeinflussen.

Politische Entscheider*innen sind oftmals politische Akteur*innen. Deshalb wird verfügbare Evidenz häufig nicht für die Umsetzung wohlfahrtssteigernder Politik verwendet, sondern eher als Instrument zur Rechtfertigung politischer Aussagen, Absichten und zur Umsetzung bereits beschlossener politischer Maßnahmen.

Ob Evidenz berücksichtigt wird oder nicht, hängt von den Rahmenbedingungen des Entscheidungsproblems und dessen Umfeld ab. Um die Chancen zu verbessern, dass Evidenz bei der politischen Entscheidungsfindung berücksichtigt wird, müssen die Hindernisse abgebaut werden. Die Einführung einer evidenzbasierten Kultur und angemessener Konzepte für die Übertragung von Evidenz, scheinen sehr wichtig zu sein.

5 Fallstudie: Der Kopenhagen-Konsens

5.1 Hintergrund und Ansatz

Das *Copenhagen Consensus Centre* ist ein internationaler Thinktank, der 2002 gegründet wurde.⁶ Seit 2004 wurden mehrere ‚Projekte‘ durchgeführt, die sich meist auf Entwicklungsprobleme auf regionaler, nationaler und globaler Ebene konzentrieren, wie etwa in Andhra Pradesh, Bangladesch, Haiti oder auf Ausgaben der Entwicklungszusammenarbeit in Übersee. Einige Projekte bezogen sich direkt auf die größten Entwicklungsprobleme der Welt (Lomborg, 2007b) oder die Post-2015-Entwicklungsagenda (Lomborg, 2015) und andere legten besonderes Augenmerk auf thematische Fragen wie beispielsweise HIV (Lomborg, 2012) oder den Klimawandel (Lomborg, 2007a). Alle nutzen akademische Evidenz zumindest, um Prioritäten zu setzen. Aus diesem Grund und wegen einiger besonderer Merkmale der Art und Weise (siehe unten), wie die beteiligten Personen diskutieren und kommunizieren, wurde der Kopenhagen-Konsens (KK) als Fallstudie ausgewählt.

Auf Grundlage der Finanzierung, die von Regierungen und Stiftungen geleistet wurde, werden optimale Lösungen für die untersuchten Probleme unter Nutzung des ökonomischen Konzepts der KNA ermittelt, einer routinemäßig eingesetzten Bewertungstechnik, die auch für Ex-post-Evaluierungen verwendet wird. Bei einer KNA wird der Nettonutzen einer Investition berechnet, indem die erwarteten Erträge alternativer politischer Maßnahmen (= Preis-Leistungs-Verhältnis, *value for money*) verglichen werden. Schlussendlich beruht die zugrunde liegende Logik darauf, die effizienteste Alternative auszuwählen. Aus einer rationalen Perspektive kann der KK-Ansatz daher als optimaler Ansatz für die Bereitstellung von Evidenz für politische Entscheidungen angesehen werden. Die meisten Projekte wurden auf ähnliche Weise unter Nutzung der KK-Methodik durchgeführt.

5.2 Bangladesch-Prioritäten

Beispielsweise war es angesichts begrenzter Ressourcen und Zeit das Ziel der ‚Bangladesch-Prioritäten‘, intelligentere Lösungen für die bestehenden Herausforderungen in diesem Land zu finden. Bei dem gesamten Projekt handelte es sich um ein Planspiel. Die angenommene jährliche Haushaltslinie für die Regierung belief sich auf 30 Milliarden US\$, wobei drei Milliarden US\$ als Entwicklungshilfe von externen Organisationen beigetragen wurden. Politische Entscheidungsträger*innen, internationale Geber*innen, NROen und Unternehmen wurden als Hauptakteur*innen angesprochen. Das Projekt dauerte tatsächlich mehr als ein Jahrzehnt (Lomborg, 2017).

Das Experiment in Bangladesch begann mit einer Analyse des siebten Fünfjahresplans des Landes als Ausgangspunkt für die Diskussion über die nationale Entwicklung. Alle der 20 einbezogenen Themenbereiche wurden in die Analyse aufgenommen. In Zusammenarbeit mit dem *Bangladesh Rural Advancement Committee* – eine der weltweit größten entwicklungsorientierten NROen –, der Schwedischen Internationalen Agentur für Entwicklungszusammenarbeit und der dänischen Botschaft in Dhaka, wurden mehr als 800 Mitarbeitende aus der Regierung von Bangladesch, Akademiker*innen, aus Thinktanks, NROen, dem Privatsektor und aus Entwicklungsorganisationen eingeladen, ihre Empfehlungen zu formulieren. Insgesamt wurden mehr als 1.000 Ideen eingereicht, darunter viele Themen, die auch im offiziellen Plan erwähnt wurden. Dazu zählten Infrastruktur, Bildung und öffentliche Gesundheit.

Die laufenden Untersuchungen wurden in den größten Tageszeitungen von Bangladesch präsentiert, und die Vorschläge wurden in Foren für ländliche Dörfer und in Jugendforen bewertet.

6 Vgl. <http://www.copenhagenconsensus.com/our-story>

Gestützt auf Hintergrundstudien wurden 20 Gesprächsrunden mit Expert*innen aus Bangladesch abgehalten, um die Vorschläge zu kommentieren und sie anhand von Kriterien wie Potenzial, politische Unterstützung und verfügbare Daten zu bewerten. Auf Grundlage dieser Vorgehensweise wurden 76 erfolgversprechende Vorschläge identifiziert. Um sie alle zu bewerten, wurden KNA verwendet. Im Fall der ‚Erhaltung von Feuchtgebieten in den Sundarbans‘ wurde der erwartete Nutzen mit nahezu fünf Milliarden US\$ bei Kosten von 1,4 Milliarden US\$ berechnet. Dies bedeutet einen Nutzen von beinahe drei US\$ für jeden einzelnen ausgegebenen US\$. Die meisten der Maßnahmen wurden zwischen 1 und 29 bewertet. Beispielsweise wurde ein Programm zur frühkindlichen Bildung mit 18 bewertet, was bedeutet, dass jeder investierte Dollar 18US\$ einbringen würde. Allerdings gab es gewisse Interventionen – darunter die bedingungslose Vergabe von Bargeld an arme Bangladeschi und die Impfungen gegen Gebärmutterhalskrebs – die niedrigere Bewertungen bis hinunter zu 1 erhielten. Die beste Lösung, ein digitales Vergabeverfahren, brachte einen Nutzen von 663 US\$ für jeden ausgegebenen Dollar (The Economist, 2016). Ausgangspunkt für diese empfohlene Maßnahme waren die enormen Ineffizienzen im öffentlichen Beschaffungswesen, Korruption eingeschlossen.

5.3 Würdigung

Frühere und laufende KK-Projekte haben große Aufmerksamkeit erzielt. Das Konzept übt eine starke Anziehungskraft aus. Aufgrund des Produktdesigns und der entsprechenden Öffentlichkeitskampagnen – einschließlich Büchern in zugänglichen Sprachen, wissenschaftlichen Aufsätzen in renommierten Zeitschriften (*Nature*, *The Lancet* etc.), Artikeln in Magazinen (*The Economist* etc.) sowie Beiträgen in wichtigen Zeitungen – konnten viele Akademiker*innen und politische Interessensvertreter*innen die Ergebnisse des KK verfolgen. Auch für viele Menschen, die über keine Fachexpertise verfügen, ist die angewandte Methode leicht zu verstehen, weil die verfügbaren Optionen als Reihenfolge aufgelistet werden. KK-Projekte arbeiten zudem mit einem hohen Grad an Transparenz. Viele Hintergrundinformationen sind auf der Internetseite des *Copenhagen Consensus Centre* zu finden. Expert*innen für die jeweiligen Gebiete präsentieren die Ergebnisse. Darunter sind manchmal auch Nobelpreisträger*innen. Für viele Menschen ist diese Evidenz plausibel. Allgemein betrachtet scheint der KK-Ansatz ein erfolgreiches Modell zu sein, um bestehende Evidenz in die Politikberatung zu übersetzen, und zwar als Resource für die Entscheidungsfindung.

Selbst im Fall des KK ist es nicht möglich, den Erfolg dieses Ansatzes vollständig zu bewerten. Nichtsdestotrotz weist das KK-Modell einige zentrale Eigenschaften auf, die mit Blick auf die Wissensübertragung interessant sind:

- a) Die Einbindung einiger hochrangiger Expert*innen in thematische Fragen. Einige von ihnen sind Nobelpreisträger*innen und somit Träger*innen eines Preises mit exzellentem Ansehen;
- b) die Verwendung eines methodischen Ansatzes, der in den Wirtschaftswissenschaften anerkannt ist, routinemäßig im öffentlichen Sektor verwendet wird und der auch für Nicht-Ökonom*innen leicht verständlich ist;
- c) die Neubewertung der relevanten Evidenz und die entsprechende transparente Dokumentation der Bewertung;
- d) die Einbeziehung alternativer Auffassungen;
- e) die hierarchische Einordnung der Optionen nach ihrem Preis-Leistungs-Verhältnis;
- f) eine exzellente Marketing-Kampagne, in der die Empfehlungen von Expert*innen abgesichert werden.

Dennoch wird das KK-Konzept sehr kontrovers eingeschätzt.

Erstens ist es ein Projekt, das von Ökonom*innen dominiert wird. Bei den eingeladenen Wissenschaftler*innen handelt es sich meist um Forscher*innen, die auf dem Gebiet arbeiten. Nach Bjorn Lomborg, dem Leiter des *Copenhagen Consensus Centre*, besteht ein „Bedürfnis danach, dass Ökonom*innen globale Prioritäten setzen“ (Lomborg, 2004a). Eine solche Aussage wird für viele Wissenschaftler*innen aus anderen akademischen Disziplinen nicht akzeptabel sein. In einer Welt des Pluralismus sollten die Prioritäten der Welt nicht durch ‚ökonomischen Imperialismus‘ definiert werden. Um mit den Worten von Jeffrey Sachs zu sprechen: „Nur Ökonomen allein sind nicht genug“ (Sachs, 2004).

Zweitens erkennen viele Wissenschaftler*innen den akademischen Wert der Hintergrundberichte an, die als wichtiger Input für eine KNA dienen, aber sie stimmen mit einigen der in der KNA verwendeten Modelle und Annahmen nicht überein. So lehnten beispielsweise einige Akademiker*innen, die im Bereich der Umweltwissenschaften arbeiten, die KK-Ergebnisse bei der Bewertung des Klimawandels ab. Das Projekt machte Schlagzeilen, als es einen politischen Rahmen zur Kontrolle des Klimawandels im Kontext globaler Prioritäten ablehnte (Lomborg, 2004b). Auf der Grundlage einer neueren KK-Analyse über sinnvolle Lösungen für den Klimawandel folgerte Lomborg, dass es nicht notwendig sei, die CO₂-Emissionen in naher Zukunft nennenswert zu reduzieren. Höhere Ausgaben für umweltbezogene Forschung und Entwicklung wären vernünftiger (Lomborg, 2010). Ein wesentlicher Punkt der entsprechenden Diskussion mit Klimawissenschaftler*innen war die Diskontierung künftiger Kosten und Nutzen (Hamide & Boland, 2006). Dabei handelt es sich um ein Werkzeug, das eingesetzt wird, um zukünftige Werte in die Gegenwart zu übertragen. Das Konzept beruht auf der Annahme, dass Menschen ungeduldig sind und die Gegenwart bevorzugen. Aufgrund der hohen Diskontierung des künftigen Nutzens von Klimaschutzmaßnahmen ist ihr Gegenwartswert sehr gering. Sie sind daher scheinbar keine gute Investition, da die Zukunft systematisch unterbewertet wird (Guo, Hepburn, Tol, & Anthoff, 2006). Allerdings argumentieren Klimawissenschaftler*innen, dass eine Verzögerung der Maßnahmen die Auswirkungen dieser Herausforderungen noch verschlimmern würde. Es wurde auch angemerkt, dass die Verwendung einer marginalen KNA nicht angemessen ist, um das Problem zu lösen, und dass das genutzte Klimamodell nicht geeignet war (Zenghelis, 2010). Ein anderes Argument war, dass eine KNA nicht in der Lage sei, die wesentlichsten Vorteile des Klimaschutzes zu berücksichtigen und zu messen (Ackerman, 2008). Es gab auch eine Diskussion zwischen einer Gruppe von Wissenschaftler*innen, die mit dem *Club of Rome* assoziiert sind, einem interdisziplinären Thinktank, und zu denen die Autor*innen der ‚Grenzen des Wachstums‘ (Meadows, Meadows, Randers, & Behrens, 1972) zählen. Dabei wurde die Frage erörtert, wie Wachstum angemessen zu bewerten sei (Beinecke et al., 2012). Für Nicht-Akademiker*innen klingen diese Debatten seltsam. Die Kontroversen verdeutlichen aber, dass besonders bei hoher Unsicherheit mehr als nur die wissenschaftlichen Erkenntnisse einer Disziplin zu berücksichtigen sind. Es ist nicht einfach, die Auswirkungen von Maßnahmen über die Zeit hinweg zu bewerten. Der Verlauf der Zeit kann die Wirkung eines Programms verändern. Schlussfolgerungen können für die beteiligten Parteien mit unterschiedlichen Zeitpräferenzen unterschiedlich ausfallen. Für Entscheidungsträger*innen wird ein weit umfassenderes Bild benötigt, als nur die unvollständigen zusammenfassenden Messgrößen, wie sie das KK verwendet.

Drittens wurde auch kritisiert, dass das Ziel des KK darin besteht, unabhängige Finanzierungs- und Prioritätensilos abzubauen, indem gleichzeitig zwischen einer Vielzahl von Problemen und Lösungen gewählt wird. Die Verflechtung von Projekten und *Outcomes* wird überhaupt nicht berücksichtigt (Greenwood, 2010). Dieses Problem wird in der EZ seit vielen Jahren diskutiert. Dennoch setzen die Träger von EZ weiterhin auf diese Denkweise.

Viertens passen sich die KK-Rangordnungen allgemein nicht an sich bietende Gelegenheiten in Prozessen der Entscheidungsfindung an. Für gewöhnlich sind Regierungen nicht in der Lage, ihre Budgets in großem Umfang neu zu überdenken und Mittel umzuwidmen. Entscheidungen über öffentliche Ausgaben benötigen häufig mehrere Jahre. Aus diesem Grund ist die Evidenz, die das KK vermittelt, nicht immer dienlich.

6 Potenziale von Evidenz in der Entwicklungspolitik

6.1 Interpretation und Ergebnisdiskussion

Aus der Literatur lassen sich folgende Aussagen über die Bereitstellung und Nutzung von Evidenz in der EZ ableiten.

Akademiker*innen und Thinktanks debattieren seit einigen Jahren über den Wert von Evidenz und über geeignete Kanäle und Werkzeuge, wie Forschung größeren Einfluss auf die Politikgestaltung nehmen kann. Was die Politikgestaltung angeht, so ist die Evidenzkultur zwischen den einzelnen Ländern sehr unterschiedlich ausgeprägt.

In Deutschland hat die Diskussion über eine stärkere Nutzung von Evidenz in der EZ gerade erst begonnen. Im Vergleich zu anderen Ländern wie den Niederlanden oder dem Vereinigten Königreich herrscht hier zwischen den politischen Parteien ein viel breiterer Konsens, dass EZ gebraucht wird. Die Notwendigkeit, Entwicklungszusammenarbeit zu rechtfertigen, ist daher nicht so hoch (Klasen, 2017). Unabhängig davon werden nicht alle Lippenbekenntnisse tatsächlich bei der Politikgestaltung berücksichtigt.

6.1.1 Umfang der Evidenz

Die aktuelle Diskussion auf höherer Ebene über die Einbeziehung von mehr Evidenz in die EZ kreist hauptsächlich um zwei *Typen der Evidenz*: akademische und praktisch-technische Evidenz. In dieser Hinsicht stehen zwei Aspekte im Vordergrund:

1. Verbesserung der Wissensübertragung von verfügbarer Evidenz auf die Politikgestaltung und die Entscheidungsfindung (vgl. z.B. Ward, 2017). Diese Diskussion ist Teil einer breiteren Debatte in akademischen Zirkeln. Es existiert bereits viel Literatur über die Identifikation von Transferbarrieren, und zwar sowohl aufseiten der Anbieter*innen als auch aufseiten der Nutzer*innen. Es gibt viele Konzepte mit Vorschlägen, wie Verbesserungen erzielt werden können. Diese Ansätze sind wichtig, aber es ist auch eine kritische Diskussion darüber erforderlich, ob ein konkretes Konzept in einem spezifischen Kontext nützlich ist oder nicht. Darüber hinaus sind die Entwicklungszusammenhänge sehr dynamisch und können komplex sein. Aber die Aufmerksamkeit zu erhöhen und Kapazitäten aufzubauen, sind wichtige erste Schritte.
2. Verstärkte Schaffung von Evidenz, die auf Ergebnismanagement und ergebnisorientierten Ansätzen basiert. Nach den von der OECD vorgelegten Übersichten, ist die Leistung des EOM zwischen den Ländern tendenziell unterschiedlich ausgeprägt. Dies ist die Folge von Zielsystemen, die in unterschiedlichem Maße komplex sind und zu Problemen bei der Ermittlung geeigneter Indikatoren und beim Messen führen können. Es mangelt an Wissen über ergebnisorientierte Ansätze. Viele von ihnen befinden sich noch in der Pilotphase.

Es hängt vom konkreten Umfeld ab, wie soziale Evidenz und besonders die Erfahrungen der Zivilgesellschaft eingebunden werden. Es gibt beispielsweise Hinweise darauf, wonach soziale Evidenz manchmal im Gesundheitssektor genutzt wird, da Menschen als wichtige beteiligte Parteien gelten. Auf höherer Ebene wird sie berücksichtigt, indem die Zivilgesellschaft (bzw. deren Vertreter*innen) in die politikrelevante Entscheidungsfindung eingebunden wird. Die operative Ebene bietet mehr Raum, um auf soziale Evidenz zurückzugreifen.

Die Praxis der EZ zeigt, dass der ‚Ruf nach mehr akademischer Evidenz‘ oft zu eng greift. Eine starke und intensive Zusammenarbeit auf Arbeitsebene zwischen den Durchführungsorganisationen und Akademiker*innen sowie ein offener Dialog können zu einer Win-win-Situation für alle Partner*innen führen und eine valide Evidenz garantieren, die eine wichtige Quelle für das

Lernen auf operativer Ebene darstellt. In dieser Hinsicht sollte die Beziehung zwischen denjenigen, die Evidenz liefern und denjenigen, die sie nutzen, nicht als eingleisiger Prozess aufgefasst werden. Zwischen ihnen sind Feedbackschleifen sehr wichtig. Dies ist einer der wesentlichen Ansätze im Projektmanagement. Die Informationen liefern Ansatzpunkte für das Lernen und für die Anpassung von Aktivitäten. Diese Idee ist auch eine wichtige Säule von ergebnisorientierten Ansätzen.

Die laufende Debatte über Evidenz verläuft bezüglich akademischer Evidenz stark verzerrt. Für Akademiker*innen besteht ein Bedarf an mehr Forschungsgeldern, um diese Art von Evidenz zu liefern. Allerdings ist die Erbringung akademischer Evidenz möglicherweise teuer. Forschungsprojekte haben oftmals nur ein begrenztes Zeitfenster. Deshalb ist es wichtig, sie in praktische Routinen zu überführen. Allerdings können auch diese durch finanzielle Probleme eingeschränkt sein. Um die beste verfügbare Evidenz für die Entscheidungsfindung bereitzustellen, ist es wichtig, den verfügbaren Evidenzraum zu erweitern, und zwar indem der Zugang zu allen möglichen Evidenztypen, einschließlich der sozialen Evidenz, gewährleistet wird. Letztere könnten durch die Berücksichtigung sozialer Medien einbezogen werden. Dies ist auch im Interesse der politischen Entscheidungsträger*innen, die bevorzugt vielfältige Wissensquellen nutzen, wenn sie Entscheidungen treffen (Cairney & Oliver, 2017).

6.1.2 Stärke und Qualität der Evidenz

Im Laufe der Zeit hat sich die *Qualität* der verfügbaren akademischen und praktischen Evidenz auf beeindruckende Art und Weise verbessert. Inzwischen gibt es viele Evaluierungen, die auf rigorosen Methoden beruhen. Die gegenwärtige Debatte konzentriert sich auf quantitative Evidenz. In der EZ haben auch qualitative Ergebnisse einen hohen Stellenwert. Es besteht die Gefahr, dass das Konzept der Evidenzhierarchie ohne nähere Reflexion auf den Bereich der EZ übertragen wird. Aus Qualitätsgründen haben Forscher*innen ein Interesse daran, quantitative Studien mit einer hohen Zahl an Beobachtungen durchzuführen. Dennoch ist das keine Garantie dafür, dass auch kleine, benachteiligte Gruppen einbezogen werden. Diesbezüglich ist akademische Evidenz oft kaum vorhanden. Nicht zuletzt, weil es schwierig ist, benachteiligte Gruppen als solche zu identifizieren. Hier stellt sich ein großes Problem im Zusammenhang mit der Initiative, niemand zurückzulassen (*Leave No One Behind*).

Es gibt viele Unterschiede zwischen der Anwendung des EBM-Konzepts und der Nutzung von Evidenz in der EZ: Ersteres ist mit der Mikro-Ebene verknüpft, bezieht sich weitgehend nur auf eine Disziplin und Studien werden mit einer homogeneren Ausgestaltung durchgeführt. Letztere ist stärker auf der Makro-Ebene verortet und beinhaltet mehr Disziplinen. Vor allem aber sind entsprechende Studien nicht standardisiert, weil sie sich auf ein komplexeres Umfeld beziehen. Dennoch darf das absolute Qualitätsniveau der Evidenz nicht mit der von den politischen Entscheidungsträger*innen wahrgenommenen Qualitätsstärke verwechselt werden. Besonders die Bereitstellung sozialer Evidenz steht häufig in Verbindung mit Partikularinteressen. In dieser Hinsicht ist es sehr bedeutsam, auf welche Art und Weise sie den beteiligten Parteien vermittelt wird, wie deren Machtbefugnisse aussehen und wie all dies vonstattengeht.

6.1.3 Evidenzorientierung

In den letzten Jahren hat die Nutzung von Evidenz in der EZ beträchtlich zugenommen und das Nutzungsniveau ist im Vergleich zu früher substanziell angestiegen. Gegenwärtig erfolgt die Orientierung weitgehend an einem Niveau, das *evidenzunterstützt* ist. Viele Faktoren haben die Nutzung beeinflusst:

- **Das Interesse an Evidenz vonseiten politischer Entscheidungsträger*innen rührt bisweilen daher, dass sie bereits getroffene Entscheidungen rechtfertigen wollen.** Es gibt

Anzeichen dafür, wonach die Aufnahme von Evidenz in die Agenda von politischen Entscheider*innen das Ergebnis strategischen Denkens ist und nicht dem Gemeinwohl diene. Dieser Punkt wird als politikbasierte Evidenz diskutiert (Strassheim & Kettunen, 2014).

- **Evidenz wird stärker in weiter fortgeschrittenen Kontexten genutzt.** In Partnerländern wird die Nützlichkeit von Evidenz oftmals gering eingeschätzt. Häufig fehlen die erforderlichen Kapazitäten, um sie zu ermitteln, zu interpretieren und zu nutzen.
- **In manchen Bereichen ist qualitativ hochwertige Evidenz reichhaltig vorhanden, wohingegen anderswo beträchtliche Evidenzlücken existieren.** Im Verlauf der Zeit wurde in den Forschungsfeldern Gesundheit und Bildung viel Evidenz gesammelt, während andere Politikbereiche vernachlässigt wurden. Beispielsweise gibt es nur begrenzte Erkenntnisse über den Aufbau von Institutionen, was darauf zurückzuführen ist, dass RCTs in diesem Bereich nicht anwendbar sind.
- **Je näher ein Entscheidungsorgan dem politischen Raum ist, umso weniger ist es bereit, Evidenz analytisch zu berücksichtigen.** Aus Sicht des Entscheidungsorgans wird Evidenz als Werkzeug für bessere Rechenschaft und zur Rechtfertigung der eigenen Politik verstanden. Allerdings könnte diese Bereitschaft durch einen allgemeinen Anstoß erhöht werden, etwa vonseiten der OECD.
- **Die grundsätzliche Zuweisung von EZ-Geldern basiert auf politischen Prioritäten und nicht auf Kosten-Nutzen-Erwägungen,** wie es der Kopenhagen-Konsens vorschlägt. Auf den nachgelagerten Stufen wird Evidenz in einem größeren Umfang berücksichtigt. Auf der Ebene der konkreteren EZ-Maßnahmen werden die erreichten Ergebnisse berücksichtigt, etwa mit Blick auf genutzte Kanäle und implementierte Aktivitäten. Auf der Ebene der multilateralen EZ spielt das Vertrauen in die Aktivitäten der durchführenden Organe eine wesentliche Rolle. Wichtige Partner auf dieser Ebene sind etwa der Global Fund und GAVI. Sie sind Zuweisungsempfänger und nutzen Evidenz zur Rechenschaft gegenüber ihren Geber*innen.
- **Der Wettbewerb um finanzielle Mittel zwischen Durchführungsorganisationen erhöht die Evidenzorientierung.** Organisationen wie die deutsche GIZ, die im Wettbewerb mit anderen Anbietern steht, haben starke Anreize, den Erfolg ihrer Projekte durch Ergebnismangement und Wirkungsevaluierungen zu belegen. Indem sie qualitativ hochwertige Methoden verwenden, kann die Wahrnehmung der Zuverlässigkeit ihrer eigenen Aktivitäten verbessert werden. Evidenz wird zur Rechenschaftslegung und zur Unterstützung weiterer Projekte genutzt. Allerdings ist in der deutschen EZ die entsprechende Transparenz begrenzt. Leistungserbringer*innen haben keine starken Anreize, transparent zu handeln und sie sind nicht verpflichtet, alle Ergebnisse zu veröffentlichen. Gegenwärtig bestehen eher Anreize, nicht ‚zu viel‘ zu publizieren. Oft werden beispielsweise Projektvorschläge und Berichte nicht veröffentlicht. Die GIZ publiziert nur Zusammenfassungen von Evaluierungen. Personen, die an projektbezogenen Informationen interessiert sind, wissen häufig nicht, wie sie diese auf der Homepage finden oder Sie erhalten können.
- **Es besteht eine Tendenz, unerwünschte Evaluierungsergebnisse zu verbergen.** In der Forschungsdomäne ist es oftmals nicht möglich, die Ergebnisse von Studien zu veröffentlichen, die über keine signifikanten Wirkungen berichten. Gleichzeitig sorgen sich Projektleiter*innen darum, im Falle ‚schlechter Ergebnisse‘ keinen Zuschlag für künftige Projekte zu erhalten. Manchmal haben politische Entscheider*innen kein Interesse daran, kritische Studien zu veröffentlichen. Dies könnte für Forschungsinstitute zu großen Schwierigkeiten führen, die von politischen Entscheidungsträger*innen abhängen. Diese Aspekte können sich nachteilig auf das Lernen aus Erfahrungen auswirken.

- **NROen erkennen mehr und mehr, wie wichtig die Nutzung verfügbarer Evidenz ist, um ihre eigenen Forderungen zu untermauern.** Im Zeitalter des Internets und neuer Informations- und Kommunikationstechnologien, reichen moralische Appelle nicht mehr, um EZ zu begründen und zu rechtfertigen. Aufgrund ihrer großen Anzahl stehen NROen im Wettbewerb miteinander um die knapp vorhandenen Mittel. Deshalb nutzen sie für ihre Kampagnen ausgewählte Evidenz, die ihre Ansichten und ihr Handeln bestätigen soll, um so ihre Bedürfnisse abzudecken. Glaubhafte Informationen und ein guter Ruf der Akteur*innen sind dabei sehr wichtig (Green, 2017; Guijt & de Goede, 2017).

6.2 Künftiges Potenzial

Auch in der EZ wird Evidenz – ähnlich wie in den Aktivitätsbereichen Gesundheit und Bildung – zukünftig an Bedeutung zunehmen. Um die Relevanz von Evidenz für die Politikgestaltung zu verbessern, muss ihre Schaffung und Bereitstellung gestärkt werden. Art und Form der Übermittlung von Evidenz für den praktischen Gebrauch – einschließlich der Koproduktion mit politischen Entscheider*innen – müssen verbessert werden. Politische Entscheidungsträger*innen müssen ein besseres Verständnis für die Nützlichkeit von Evidenz entwickeln, zumindest aus Gründen der Rechenschaftspflicht und zur besseren Durchführung der Maßnahmen, die auf ihre eigenen politischen Ziele ausgerichtet sind.

Evidenz, die für die EZ relevant ist, beschränkt sich nicht nur auf akademische Evidenz, auch wenn Akademiker*innen in diesem Politikbereich eine zentrale Rolle spielen. Momentan gibt es für Forscher*innen keine starken Anreize, politikorientierte Beiträge zu schreiben. Das gegenwärtige System, der Rahmen und die Anreize für Akademiker*innen, das Management und die Organisation ihrer Arbeit beruhen auf Impact-Faktoren, die an das Renommee von wissenschaftlichen Zeitschriften gekoppelt sind. Als Folge der antizipierten Belohnungsstrukturen sind Forscher*innen stärker daran interessiert, theoretische Arbeiten zu publizieren, die einen begrenzten praktischen Wert haben. Auch wenn dieses System in hohem Maße von öffentlicher Regulierung losgelöst ist, können doch Belohnungen und Anreize für stärker praxisorientierte Leistungen gesetzt werden. Es ist sehr wichtig, den Aufbau von Institutionen zu stärken, damit der Nutzen von Evidenz verbessert wird (Parkhurst, 2016).

Die Tatsache, dass die politikorientierten Veröffentlichungen der Weltbank unzureichend beachtet werden, zeigt die Mängel einer Politikberatung, die auf einem linearen Modell beruht, nämlich dem zwischen Forscher*innen, die Evidenz verfügbar machen und politischen Entscheidungsträger*innen als Kunden. Um den Bekanntheitsgrad akademischer Ergebnisse zu erhöhen, müssen politikrelevante akademische Ergebnisse auf angemessene Art und Weise den politischen Entscheider*innen vermittelt werden. Politische Entscheidungsträger*innen lesen keine akademischen Arbeiten und lassen sich nicht von drei Sternen für die Darstellung statistischer Signifikanz in einer Regressionstabelle begeistern. Sie wollen über die jeweiligen Wirkungen von Maßnahmen informiert werden. Heutzutage werden Kurzdarstellungen, die auf die Politik ausgerichtet sind, oftmals von der Zielgruppe gar nicht wahrgenommen. Deshalb müssen Ergebnisse über die sozialen Medien verbreitet werden, denen eine stärkere Reichweite und entsprechende Kommunikationsstrategien zu eigen sind. Glaubwürdigkeit ist von zentraler Bedeutung, um politische Entscheider*innen dazu zu bewegen, dass sie Forschungsergebnisse anwenden (Pittore, te Lintelo, Georgalakis, & Mikindo, 2017). Da politische Entscheider*innen eine sehr heterogene Gruppe sind, muss die Evidenz stärker zugeschnitten werden. Darüber hinaus muss das Verständnis von Forschung und Politikberatung angepasst werden – und zwar hin zu einem Konzept der transformativen Forschung mit transdisziplinären Fokus –, einschließlich eines engen Austauschs mit Entscheidungsträger*innen. Soziale Akteur*innen

müssen in den Forschungsprozess durch die Mitgestaltung und die Koproduktion von Wissen eingebunden werden.⁷

Der zahlenmäßig gewaltige Anstieg von Wirkungsevaluierungen in den letzten Jahren hat nicht gereicht. Die wesentliche Herausforderung für die Wirkungsevaluierung ist es, noch mehr Studien zu erstellen. Die globale Politik sollte nicht auf einer einzigen Studie aus einem einzigen Land beruhen, sondern eher auf einer großen Anzahl von Studien, die bestätigen, ob eine Intervention funktioniert oder nicht, und wie sich die Wirkung kontextspezifisch verändert. Die jeweiligen Folgen sind normalerweise nur für spezielle Kontexte bedeutsam und können nicht verallgemeinert werden. Auch die Forderung nach mehr RCTs stößt an ihre Grenzen. Sie sind kein Allheilmittel. RCTs sind in einigen Kontexten wie etwa dem Klimawandel nicht effektiv (Leigh, 2010). Grundsätzlich sollte der von zusätzlichen Studien erwartete Wert kritisch mit den für sie anfallenden Kosten verglichen werden. Die Ökonomie der Evidenz zu berücksichtigen, ist wichtig.

Um die bestehenden Mängel der Evidenzsynthese zu überwinden, müssen die gebräuchlichen Messgrößen standardisiert werden, um so wertvolleres Wissen zu schaffen. Zusätzlich dazu sollten Standardvariablen aufgenommen werden, um den Vergleich zwischen Studienergebnissen zu verbessern. Auf dieser Basis ist es möglich, gemeinsame Evidenz zu verbreiten.

Auf Ebene der Ministerien, Verwaltungen und Durchführungsorganisationen müssen Wege gefunden werden, um den gegenwärtigen Problemen von EOM zu begegnen. Wertvoll ist der Erfahrungsaustausch auf Plattformen, der durch die OECD geleistet wird. Allerdings hat das Beispiel des Vereinigten Königreichs gezeigt, dass es hilfreich sein kann, eine Evidenzkultur zu etablieren. Dazu sind Vermittler*innen erforderlich.

Anders als Akademiker*innen vertreten politische Entscheider*innen oft unterschiedliche Auffassungen über den Wert von Evidenz. Ihr Ansatz ist häufig breiter aufgestellt, da sie verschiedene Interessen und zahlreiche Positionen ausgleichen, anstatt sich aus einer disziplinären Perspektive auf Details zu konzentrieren. Wird politischen Entscheider*innen ein umfassenderes Bild von Evidenz vermittelt, anstatt spezielle und kleinteilige Ergebnisse, die sich aus einem disziplinären Tunnelblick ergeben, ist ein erster wichtiger Schritt getan, damit der Wert von Evidenz anerkannt wird.

Um das Potenzial von Evidenz in der EZ einzuschätzen, muss die Nachhaltigkeit von Projekten und Programmen beachtet werden. Das BMZ-finanzierte Programm EOF4MNH wurde als Erfolg eingestuft (White-Kaba, 2017) und die Begleitforschung hat dies bestätigt (Brenner & De Allegri, 2016). Alle beteiligten Parteien erkannten die Bedeutung von Evidenz an und bekräftigten ihr Interesse, das Vorhaben zu unterstützen. Deshalb verlängerte das BMZ die Programmfinanzierung. Die Regierung von Malawi nahm das Vorhaben in die nationale Gesundheitsstrategie mit auf. Sie bat zudem um eine Korbfinanzierung durch Geber*innen, bis die Frage der künftigen Finanzierung geklärt sei. Malawi ist allerdings ein Land, das sehr stark von Entwicklungszusammenarbeit abhängt. Daher stellt sich die Frage, ob dort genügend finanzielle Ressourcen verfügbar sind, um das Programm künftig zu schultern. Darüber hinaus ist die vermittelte positive Evidenz abhängig von den Rahmenbedingungen und den Anreizstrukturen in der Vergangenheit. Allerdings können sich diese Konfigurationen verändern und somit auch insgesamt die vorteilhaften Ergebnisse. Es besteht zudem die Gefahr, dass EOF4MNH als Leuchtturmprojekt verstanden wird und sich die Finanzierung auf diese Initiative konzentriert. Dann stünde kein Geld für alternative Projekte zur Verfügung. Kurz gesagt: es gibt viele offene Fragen zur Nachhaltigkeit.

7 Dies ist der Fall beim German Institute of Development and Sustainability (IDOS) gGmbH und seinem Konzept für Politikberatung.

Zusammenfassend ist zu sagen, dass das Potenzial von Evidenz in der EZ stark von vielen Einflussgrößen abhängt. Um Evidenznutzung substanziell zu steigern, ist die Etablierung einer evidenzorientierten Kultur im Umfeld der Politikgestaltung eine *Conditio sine qua non*. In diesem Sinne ist es besonders hilfreich, Politikberatung auf transformative Forschung zu stützen und alle gesellschaftlichen Akteur*innen in den Prozess einzubinden, der die gemeinsame Ausgestaltung der Forschung und der Wissensproduktion betrifft.

6.3 Stärken und Schwächen

Die Diskussion über die Nutzung von Evidenz in der EZ ist relativ neu. Anders als die bisherige Literatur blickt diese Arbeit umfassender auf die Bereitstellung und den Einsatz von Evidenz auf der Grundlage eines konzeptionellen Rahmens. Dieser Veröffentlichung sind mehrere Grenzen gesetzt: Im Allgemeinen ist der Begriff 'Evidenz' – so wie er in der Literatur verwendet wird – ein eher unscharfer Ausdruck. Das Verständnis hiervon ist unterschiedlich und einige Aussagen über den Einsatz von Evidenz sind eher rhetorisch zu verstehen. Auf die Literatur gestützt war eine vollständige Bewertung im dargestellten konzeptionellen Rahmen nicht möglich. Deshalb ist eine empirische Studie erforderlich, die auf einer Befragung der beteiligten Parteien in der EZ zu ihren Wahrnehmungen und ihrer konkreten Verwendung von Evidenz beruht.

7 Folgerungen und Empfehlungen

Es war das wesentliche Ziel dieser Arbeit, Bereitstellung und Nutzung von Evidenz in der EZ zu analysieren. Ein besonderes Augenmerk lag dabei auf ihrem Potenzial in naher Zukunft – mit dem Fokus darauf, wie und in welchem Kontext Evidenz nützlich ist. Evidenz ist ein Begriff mit verschiedenen Bedeutungen und Konnotationen. Häufig wird er austauschbar mit empirischen Daten, Informationen, Statistik, Maße und Messgrößen verwendet. Oftmals aber steht Evidenz für die Fähigkeit, Folgerungen zu ziehen, die sich auf verfügbare Daten, Informationen und/oder Wissen stützen und sich auf Maßnahmen beziehen, die funktionieren. Um die Bereitstellung von Evidenz und ihre Verwendung zu veranschaulichen, wurde ein dreidimensionales Modell eingeführt. Die erste Dimension zeigt die Nutzung unterschiedlicher Typen von Evidenz; die zweite Dimension die Stärke und Qualität von Evidenz und die dritte Dimension steht für die Bedeutung von Evidenz bei der Politikgestaltung.

Vor etwa zehn Jahren war die verfügbare Evidenz darüber, was in der EZ funktioniert, von geringer Qualität. Aktivitäten und Maßnahmen in der EZ waren vorwiegend auf Prozesse und Inputs orientiert. Bestimmt wurden sie von Konzepten und Ideen dazu, wie Entwicklung funktionieren sollte. Insbesondere herrschte ein Mangel an Leistungs- und ergebnisbasierten Daten, die die Grundlagen liefern, um evidenzorientierte Informationen zu ermitteln. Evidenz war zwar verfügbar und wurde genutzt – aber vorrangig auf anekdotischem und narrativem Niveau.

Seit damals hat sich die Bereitstellung von Evidenz in der EZ verbessert. Es wurden Management-Praktiken eingeführt, die sich ausdrücklich auf Ergebnisse und ergebnisorientierte Ansätze konzentrieren. In Vorreiterländern hat sich in der EZ eine Evaluierungskultur etabliert. Anzahl und Qualität der Studien, die Evidenz erzeugen, sind gestiegen. Internationale Netzwerke und Organisationen bieten nun den Aufbau von Kapazitäten an, um mehr und bessere Evidenz liefern zu können. Es gibt Institute und Personen, die das Bewusstsein für mehr und bessere Evidenz auf Ebene der Entscheider*innen und der Verwaltung stärken und schärfen. Schulungsprogramme sind vorhanden, um Verständnis dafür zu entwickeln, wie Evidenz erzielt und nutzbar gemacht werden kann. Allerdings ist die Verwendung von Evidenz bei politischen Entscheidungen unterschiedlich ausgeprägt. Meistens spielt sie eine Rolle bei der politischen Entscheidungsfindung.

Einzelne Länder unterscheiden sich darin, wie Politik gestaltet wird. Sie unterscheiden sich in ihrer Bereitschaft zu lernen und darin, wie und in welchem Umfang Evidenz angewandt wird. Auch die politische Ökonomie darf nicht vergessen werden. Trotz internationaler Vereinbarungen über eine stärkere Wirksamkeit der Entwicklungszusammenarbeit und trotz der Lippenbekenntnisse zur ‚Bedeutung von mehr Evidenz bei der Politikgestaltung‘, sind politische Akteur*innen oftmals nicht daran interessiert, die gesamte Palette der verfügbaren Evidenz zu berücksichtigen.

Ranglisten sind willkommen, auch die Ermittlungen zum Preis-Leistungs-Verhältnis verschiedener Maßnahmen, wie sie vom *Copenhagen Consensus Centre* veranschaulicht werden. Allerdings werden solche Menüs oftmals nur für Rosinenpickerei verwendet. Bestehende Ungewissheit über die Zukunft und ernsthafte Zweifel an den zugrunde liegenden Berechnungen von Wissenschaftler*innen und Thinktanks können leicht als Entschuldigung dafür dienen, bereitgestellte Evidenz überhaupt nicht zu beachten.

Sicherlich ist mehr Evidenzorientierung besser. Doch ist es eine Illusion, sich eine Politikgestaltung vorzustellen, die vollkommen evidenzbasiert ist, und zwar einschließlich der ganzen Breite der Evidenzkomponenten und auf dem höchstmöglichen Qualitätsniveau. Der maximale Umfang, die höchstmögliche Stärke und Qualität und die entsprechende Evidenzorientierung sind aus ökonomischer Perspektive aufgrund der Opportunitätskosten gar nicht angemessen. Fundierte Evidenz zu schaffen, benötigt Zeit, aber manchmal sind schnelle Antworten im Bereich der humanitären Hilfe erforderlich. Nichtsdestotrotz wird Evidenz benötigt – selbst in diesem Fall und analog zum Gesundheitswesen – denn das Wissen, ‚was funktioniert‘, kann Leben retten (de Geoffroy, Léon, & Beuret, 2015). Die besondere Herausforderung besteht darin, die bestmögliche Evidenz auch in komplexen Notsituationen zeitnah zu liefern (Bradt, 2009).

Die Fallstudie zum Kopenhagen-Konsens illustriert auch, dass akademische Evidenz manchmal kontrovers diskutiert wird, da Akademiker*innen nicht mit einer Stimme sprechen. Vorschläge, die aus dem Blickwinkel verschiedener akademischer Disziplinen abgeleitet werden, unterscheiden sich durchaus und stehen bisweilen auch im Konflikt zueinander. Darüber hinaus haben Akademiker*innen aufgrund ihres disziplinären Backgrounds auch ein voneinander abweichendes Verständnis davon, was im konkreten Fall annehmbare Evidenz für politische Maßnahmen ist (Jasanoff, 1995). Es gibt auch Sozialwissenschaftler*innen, die Forschungsmethoden bezweifeln, die versuchen, Klarheit und Präzision zu erzielen. Radikale Sozialwissenschaftler*innen, wie etwa Law (2004), weisen darauf hin, dass Methoden nicht nur soziale Realitäten beschreiben, sondern auch helfen, sie zu schaffen. Es existiert auch bereits ein Hinterland der Realitäten: offenkundige Abwesenheit und Andersartigkeit, Echos und Muster unterschiedlicher Art und Weise, wie etwa ‚alternative Fakten‘. Sie können im Prozess der Bereitstellung von Evidenz nicht ignoriert werden.

Manchmal können akademische Ergebnisse im Widerspruch zu gewissen Werturteilen stehen. In Disziplinen erfolgen Forschungsfragen und -strategien auch nach einer gewissen Weltsicht und werden von einem bestimmten Zeitgeist beeinflusst. Es ist zudem möglich, dass Effizienz und Wirksamkeit im Konflikt mit Verteilungsfragen stehen.

Diese Analyse zeigt, dass das größtmögliche Potenzial an Evidenzorientierung ausgeschöpft werden muss. Auf der Angebotsseite sind die Strukturen zu verbessern, damit Evidenz stärker bereitgestellt werden kann. Dafür müssen hilfreiche Faktoren erkannt und anvisiert werden. Dazu gehört es, Akademiker*innen zu ermutigen, ihre Arbeit stärker auf politik- oder praxisbezogene Aspekte auszurichten und die Übertragbarkeit ihrer Ergebnisse zu verbessern. Sicherlich sollten viel mehr Evaluierungsstudien unter Verwendung internationaler Standards durchgeführt werden. Aber mehr ist nicht immer besser. Aufgrund der hohen Kosten von Evaluierungen können sie auf die Kontexte beschränkt werden, in denen es an Evidenz mangelt. Es ist auch wichtig, Ergebnisse global zu teilen und Zugangsbarrieren abzubauen. Kleinere Studien sollten vermieden werden. Die gleichen Prinzipien gelten ebenfalls für das Monitoring und für

Evidenz aus ergebnisorientierten Studien. Es besteht jedoch nicht nur Bedarf an mehr und besserer akademischer Evidenz, sondern auch an der Sammlung von Evidenz aus anderen Quellen, wie etwa aus praktischen Erfahrungen und zivilgesellschaftlichem Know-how.

Momentan wird ein großer Teil der verfügbaren Evidenz nicht für die Politikgestaltung genutzt. Hinsichtlich der potenziellen Nutzer*innen ist es wichtig, eine stärker evidenzorientierte Kultur zu etablieren. Es werden mehr Vermittler*innen benötigt, um akademische Evidenz in die Sprache der politischen Entscheidungsträger*innen und der Verwaltungen zu übersetzen. In der Vergangenheit wurden viele Modelle zur Wissensvermittlung entwickelt (Newman et al., 2014). Angesichts konkreter Rahmenbedingungen müssen geeignete Konzepte, die zum Erfolg führen, erkannt werden. Zum Beispiel die direkte Kommunikation mit allen beteiligten Parteien auf allen Projektebenen und -stufen und während aller Projektphasen, um die relevante Evidenz, die am nützlichsten ist, zu identifizieren.

Die zentrale Aussage dieser Arbeit lautet NICHT, dass jeder Forschungsbeitrag auf eine Plattform gestellt werden sollte, um dort für den nachfolgenden Einsatz in der Politikgestaltung optimiert zu werden. Forscher*innen sollten NICHT nur antizipieren, was auf den politischen Agenden steht. Sicherlich sind Themen wie die ‚Bekämpfung von Krankheiten wie HIV/AIDS, Malaria und Tuberkulose‘ von hoher Bedeutung. Doch hat dieser Fokus auf Themen mit ‚hoher Priorität‘ auch einige schwerwiegende Nebenwirkungen, nämlich dass andere Krankheiten vernachlässigt werden. Um künftige Herausforderungen, wichtige Themen und relevante Handlungsfelder zu erkennen, ist Grundlagen- und explorative Forschung wichtiger denn je. Dies gilt auch für die internationale Entwicklungszusammenarbeit. Es ist aber bedeutsam, dass Studien, die auf den politischen Gebrauch ausgerichtet sind, besser konzipiert werden; ermittelte Hindernisse auf der Angebots- und Nachfrageseite sollten gezielter angegangen werden, um die Wirkung der Forschung zu verbessern. Wir sind in der Entwicklungszusammenarbeit auf einem guten Weg. Allerdings gibt es noch viel zu tun.

Die folgenden Empfehlungen für die deutsche EZ stützen sich auf die Ergebnisse dieser hier vorliegenden Analyse:

Politikempfehlungen und gewonnene Erkenntnisse

1. **Um die Wirkung der deutschen EZ zu verbessern, ist ein starker Fokus auf Evidenz erforderlich.** Auf Grundlage der besonderen Aspekte der Problemstellung, einschließlich des Zeithorizonts, müssen die benötigten Evidenztypen ausgewählt werden. Dafür ist es hilfreich, eine starke evidenz- und ergebnisorientierte Kultur in allen Teilen des deutschen EZ-Systems zu etablieren. Ziel sollte es sein, Evidenz bei allen Entscheidungen zeitnah und effizient einzubinden.
2. **Das Bestreben, zu ermitteln, ‚was wirkt‘, ist von zentraler Bedeutung für die deutsche EZ sowie für alle, die von ihr unterstützt werden.** Deshalb sollte die Nutzung von Evidenz in allen Strategiepapieren der deutschen EZ festgeschrieben werden.
3. **Werden Entscheidungsvorschläge unterbreitet oder Entscheidungen getroffen, ist es wichtig, den genutzten Evidenztyp sowie dessen Stärke zu spezifizieren.** Die verfügbare Evidenz unterscheidet sich. Es ist wichtig, über angemessene Kriterien nachzudenken und sie zu definieren, um die Stärke der Evidenz zu bewerten; dazu zählen Validität, Zeithorizont und Zuverlässigkeit. Dies hilft dabei, herauszufinden, was die beste verfügbare Evidenz in dem jeweiligen spezifischen Kontext und zeitlichen Rahmen ist.
4. **Um evidenzorientierte Politikgestaltung zu unterstützen, können spezifische Studien hilfreich sein.** Verfügbare Evidenz muss das Rückgrat der Politikgestaltung bilden. Außer-

dem ist es hilfreich, die Akteure des deutschen EZ-Systems und ihre Organe bei der Ermittlung geeigneter Evidenzquellen zu unterstützen und Forschung in den Fällen anzustoßen, in denen Studien Evidenzlücken aufweisen.

5. **Die Stärkung des deutschen Ergebnismanagements ist wichtig für die Rechenschaftspflicht und für die Verbesserung der Entscheidungsfindung.** Ein voll funktionsfähiges Ergebnismanagement, das alle relevanten Ergebnisinformationen integrieren kann, wird für fundierte und konsistente EZ-bezogene Politik, Planung, Budgetierung und Kommunikation benötigt. Ein starkes Ergebnissystem ermöglicht ein besseres Verständnis dafür, welche Faktoren für Wirkungen im Kontext eines breiteren Entwicklungsprozesses verantwortlich sind. Deshalb könnte die Einführung einer Ergebnismatrix nach dem Vorbild anderer Geber*innen ein wichtiger Beitrag sein.
6. **Entscheidungsträger*innen in Partnerländern müssen dabei gefördert werden, Evidenz für die Politikgestaltung zu nutzen.** Informationen über die Wirkung deutscher EZ in Partnerländern sind nur unzureichend verfügbar, da Partnerländer oftmals nicht an Evaluierungen interessiert sind. Entwicklungsziele zu erreichen, ist nicht nur eine Frage des Geldes, sondern erfordert Expertise und Kapazität. Deshalb ist der Zugang zu Evidenz notwendig. Es gibt allerdings einige hohe Barrieren. Dazu zählen Schwierigkeiten im Verständnis für die Komplexität von Evidenz und mangelnde analytische Kapazitäten. Die EZ muss Kapazitätslücken auf der Programmebene adressieren. So sind etwa das Monitoring und die Evaluierung sehr komplex, wenn quantitative Daten analysiert werden müssen. Das erforderliche statistische Wissen und die benötigte Software sind oftmals nicht vorhanden.
7. **Um die Wirkung in vernachlässigten Politikbereichen zu verbessern, müssen bestehende Evidenzlücken identifiziert und adressiert werden.** Die SDGs sind stark miteinander verwoben. Um Projekte angemessen zu gestalten, bedarf es eines besseren Verständnisses der Wirkungen konkreter Maßnahmen. Beispielsweise zählt Waldschutz zu den effektivsten Ansätzen, dem Klimawandel zu begegnen. Gleichzeitig liefern landwirtschaftliche Nutzflächen und Wälder Nahrung und Brennstoffe als Existenzgrundlage für Milliarden von Menschen. Es gibt Bedenken, dass sich groß angelegte Waldschutzprogramme negativ auf die Nahrungssicherheit auswirken und auch auf andere Aspekte für Menschen in Ländern mit geringem Einkommen. Um Maßnahmen ohne nachteilige Nebenwirkungen zu ergreifen, ist eine bessere Zusammenarbeit zwischen den Durchführungsorganisationen der deutschen EZ und der Forschung erforderlich.
8. **Stärkere Nutzung von robusten Wirkungsevaluierungen.** In der deutschen EZ werden Wirkungsevaluierungen nach wie vor selten eingesetzt. Gegenwärtig ist die Zusammenarbeit zwischen Forscher*innen und den operativen Abteilungen der Durchführungsorganisationen selektiv und beruht meistens auf persönlichen Initiativen. Es gibt weiterhin Skepsis unter einigen Projektleiter*innen hinsichtlich des Werts von Wirkungsanalysen. Allerdings gibt es einen breiten Konsens in der Entwicklungspolitik, wonach die Wirkung von Interventionen bewertet werden sollte, indem rigorose Wirkungsevaluierungen verwendet werden. Deshalb ist es ein angemessenes Ziel, mehr Begleitforschung einzubeziehen. Es könnte die Motivation gestärkt werden, Wirkungsevaluierungen durchzuführen, wenn das BMZ oder DEval einen speziellen Fonds einrichten würden. Projektleiter*innen könnten sich dann um entsprechende Mittel bemühen und Anbieter*innen von Evaluierungen könnten über ein Bieterverfahren ausgewählt werden. Auf der Grundlage eigener Wirkungsanalysen sollten Durchführungsorganisationen und DEval auch stärker an systematischen Reviews arbeiten.
9. **Steigerung der Mittel für die Forschungsfinanzierung im Kontext der deutschen EZ.** Eine höhere und zielgerichtete Finanzierung von Forschung führt zu mehr verfügbarer Evidenz und zu einer größeren Wirkung der deutschen EZ.

10. **Die Durchführung von Wirkungsevaluierungen ist kein Selbstzweck.** Eine breitere Diskussion über Ergebnisse ist notwendig, um Lernen zu ermöglichen. Einrichtungen sollten öffentliche Kampagnen initiieren, die den Wert von Evidenz hervorheben. In Forschungsprogrammen sollten mehr Ressourcen und Zeit eingeplant werden, um Strategien zu verbreiten.
11. **Systematisches Lernen ist notwendig.** Erfahrungen aus Wirkungsevaluierungen werden gegenwärtig nur in ausgewählter Form bereitgestellt. Um ihre Strahlkraft zu erhöhen, müssen die Autor*innen ihre Netzwerkarbeit verbessern und die verfügbaren Ergebnisse miteinander koppeln. Diesbezüglich ist es wichtig, die institutionellen Strukturen für die Förderung systematischen Lernens zu verbessern.
12. **Evidenz als explizites Kriterium bei der Projektauswahl aufführen.** Gegenwärtig werden Projektvorschläge nach ihrer wirtschaftlichen, sozio-ökonomischen, sozio-kulturellen, geschlechtsspezifischen und ökologischen Erwünschtheit bewertet. Die unlängst implementierte Verfahrensreform hat zu einer Verbesserung in diesem Kontext beigetragen. Allerdings besteht weiterhin die Notwendigkeit, evidenzbasierte Informationen direkt zu berücksichtigen. Ergänzend dazu mangelt es weiterhin an Transparenz. Praktiken, die von anderen OECD/DAC-Mitgliedsländern angewandt wurden, können als Quelle für die Verbesserung der bestehenden Verfahren dienen.
13. **Die Nutzung von Evidenz bei der Entscheidungsfindung als institutionelle Frage berücksichtigen.** Dies könnte zur Formulierung einer internen Richtlinie für den Umgang mit den vielfältigen Rückwirkungen einer solchen Verpflichtung auf den verschiedenen Entscheidungsebenen innerhalb einer Organisation führen. Ein gut ausgestaltetes Wissensmanagement bildet eine wichtige Basis für institutionelles Lernen.

Literatur

- Achinstein, P. (2001). *The book of evidence*. Oxford: Oxford University Press.
- Ackerman, F. (2008). *Hot, it's not: Reflections on cool it, by Bjorn Lomborg*. Berlin: Springer.
- Baker, M., & Salib, M. (2017). Cracking the evidence conundrum: Four ideas to get people to use evidence. *Learning Lab*. Abgerufen von <https://usaidlearninglab.org/lab-notes/cracking-evidence-conundrum-four-ideas-get-people-use-evidence>
- Bamberger, M. (2015). Management of complexity-responsive evaluations. In M. Bamberger, J. Vaessen, & E. Raimondo (Eds.), *Dealing with complexity in development evaluation: A practical approach* (pp. 48-59). London, New Dehli, Singapore: Sage Publications.
- Bamberger, M., Vaessen, J., & Raimondo, E. (2015). *Dealing with complexity in development evaluation: A practical approach*. London, New Dehli, Singapore: Sage Publications.
- Banerjee, A. V., & Duflo, E. (2011). *Poor economics: A radical rethinking of the way to fight global poverty*. New York, NY: Public Affairs.
- Banks, G. (2009). *Evidence-based policy making: What is it? How do we get it?* Melbourne: Australian Government, Productivity Commission.
- Barr, J., Rinnert, D., Lloyd, R., Dunne, D., & Henttinen, A. (2016). *The value of evaluation: Tools for budgeting and valuing evaluations*. Abgerufen von https://assets.publishing.service.gov.uk/media/57a312d8e5274a31e0001b5a/Value_of_Evaluation_Discussion_Paper_-_Final_Version_for_Publication_03082016_clean.pdf
- Bédécarrats, F., Guérin, I., & Rouband, F. (2017). All that glitters is not gold. The political economy of randomized evaluations in development. *Development and Change*. <http://dx.doi.org/10.1111/dech.12378>
- Beinecke, F., Meadows, D., Randers, J., Harte, J., Harte, M. E., & Lomborg, B. (2012). Is growth good? Resources, development, and the future of the planet. *Foreign Affairs*, 91(5), 163-175.
- Bernal, P., Celhay, P., & Martinez, S. (2018). *Is results-based aid more effective than conventional aid? Evidence from the health sector in El Salvador*. Abgerufen von <https://publications.iadb.org/bitstream/handle/11319/8750/is-results-based-aid-more-effective-than-convention-aid.PDF?sequence=1&isAllowed=y>
- Binnendijk, A. (2000). *Results based management in the development cooperation agencies: A review of experience*. Background report, DAC Working Party on Aid Evaluation. Paris: Organisation for Economic Co-operation and Development.
- Blume, J. (2012). *Results-based approaches in German financial cooperation*. Abgerufen von <https://www.kfw-entwicklungsbank.de/Download-Center/Themenpublikationen/Results-based-approaches-in-German-Financial-Cooperation.pdf>
- Bradt, D. A. (2009). *Evidence-based decision-making in humanitarian assistance* (HPN Network Paper 67). London: Humanitarian Practice Network, Overseas Development Institute.
- Brenner, S., & De Allegri, M. (Eds.). (2016). *Final results of the EOF4MNH impact evaluation*. Prepared by the Research Team Members of the University of Heidelberg and College of Medicine. University of Heidelberg.
- Buetow, S., & Kenealy, T. (2000). Evidence-based medicine: The need for a new definition. *Journal of Evaluation in Clinical Practice*, 6(2), 85-92.
- Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung. (2008). *Leitlinien für die bilaterale Finanzielle und technische Zusammenarbeit mit Kooperationspartnern der deutschen Entwicklungszusammenarbeit* (BMZ Konzepte 165). Bonn: Autor.
- Cairney, P., & Oliver, K. (2017). Evidence-based policymaking is not like evidence-based medicine, so how far should you go to bridge the divide between evidence and policy? *Health Research Policy and Systems*, 15(1), 35.

- Cameron, D. B., Mishra, A., & Brown, A. N. (2016). The growth of impact evaluation for international development: How much have we learned? *Journal of Development Effectiveness*, 8(1), 1-21.
- Camfield, L., Duvendack, M., & Palmer-Jones, R. (2014). Things you wanted to know about bias in evaluations but never dared to think. *IDS Bulletin*, 45(6), 49-64.
- Cohen, C., & Werker, E. D. (2008). The political economy of “natural” disasters. *Journal of Conflict Resolution*, 52(6), 795-819.
- Carden, F. (2009). *Knowledge to policy: Making the most of development research*. Ottawa: International Development Research Centre.
- Clark, W. C., van Kerkhoff, L., Lebel, L., & Gallopin, G. C. (2016). Crafting usable knowledge for sustainable development. *Proceedings of the National Academy of Sciences*, 113(17), 4570-4578.
- Copenhagen Consensus Centre. (2016). *Bangladesh priorities eminent panel findings*. Abgerufen von http://www.copenhagenconsensus.com/sites/default/files/bangladesh_priorities_outcome_document_final.pdf
- Davies, H. T. O., Nutley, S. M., & Smith, P. C. (Eds.). (2000). *What works? Evidence-based policy and practice in public services*. Bristol, UK: Blackwell Publ. Ltd.
- De Geoffroy, V., Léon, V., & Beuret, A. (2015). *Evidence-based decision making for funding allocations*. Abgerufen von https://www.ghdinitiative.org/assets/files/Activities/Our%20Work/Evidence_Based_Study_Final_15_10_01.pdf
- Department for International Development. (2011). *UK aid: Changing lives, delivering results*. London: Autor.
- Deutsche Gesellschaft für internationale Zusammenarbeit. (2013). *Monitoring an evaluation unit – in focus: Knowing what works*. Eschborn: Autor.
- Dicks, L. V., Walsh, J. C., & Sutherland, W. J. (2014). Organising evidence for environmental management decisions: A “4S” hierarchy. *Trends in ecology & evolution*, 29(11), 607-613.
- Dobrow, M. J., Goel, V., & Upshur, R. (2004). Evidence-based health policy: Context and utilisation. *Social science & medicine*, 58(1), 207-217.
- Doemeland, D., & Trevino, J. (2014). *Which World Bank reports are widely read?* Abgerufen von <http://documents.worldbank.org/curated/en/387501468322733597/pdf/WPS6851.pdf>
- Donaldson, S. I., Christie, C. A., & Mark, M. M. (Eds.). (2009). *What counts as credible evidence in applied research and evaluation practice?* Thousand Oaks, CA: Sage.
- Evidence-Based Medicine Working Group. (1992). Evidence-based medicine. A new approach to teaching the practice of medicine. *Jama*, 268(17), 2420-2425.
- Eyben, R. (2015). Uncovering the politics of evidence and results. In R. Eyben, I. Guijt, C. Roche, & C. Shutt (Eds.), *The politics of evidence and results in international development: Playing the game to change the rules* (pp. 19-38). Rugby, UK: Practical Action Publishing.
- Faust, J. (2017, November 30). *(Quasi-)experimental impact evaluation in the German development cooperation: Usefulness, potential of big data and future cooperation*. DEval workshop. Bonn: German Institute for Development Evaluation.
- Fourie, W. (2017). Six barriers that make it difficult for African states to use research for policy. Abgerufen von <https://theconversation.com/six-barriers-that-make-it-difficult-for-african-states-to-use-research-for-policy-86492>
- Gaarder, M. (2014). *Beyond good intentions: Designing an evidence-based approach to aid*. Abgerufen von <http://ieg.worldbankgroup.org/blog/beyond-good-intentions-designing-evidence-based-approach-aid>
- Global Partnership for Effective Development Co-operation. (2016). *Making development co-operation more effective: 2016 progress report*. Abgerufen von [http://www.undp.org/content/dam/undp/library/development-impact/--2016%20progress%20report--Final%20\(e-book\).pdf](http://www.undp.org/content/dam/undp/library/development-impact/--2016%20progress%20report--Final%20(e-book).pdf)
- Green, D. (2017). *How can NGOs get better at using evidence to influence governments and companies?* Abgerufen von <http://oxfamblogs.org/fp2p/how-can-ngos-get-better-at-using-evidence-to-influence-governments-and-companies/>

- Greenwood, D. T. (2010). Solutions for the world's biggest problems: Costs and benefits. *Eastern Economic Journal*, 36(4), 539-541.
- Grittner, A. M. (2013). *Results-based financing: Evidence from performance-based financing in the health sector* (Discussion Paper 6/2013). Bonn: German Development Institute / Deutsches Institut für Entwicklungspolitik (DIE).
- Guijt, I., & de Goede, I. (2017). Evidence for influencing: Why we're excited about the first Oxfam Research Network conference. Abgerufen von <https://views-voices.oxfam.org.uk/methodology/research/2017/10/evidence-influencing-excited-first-oxfam-research-network-conference/>
- Guijt, I., & Roche, C. (2014). Does impact evaluation in development matter? Well, it depends what it's for! *The European Journal of Development Research*, 26(1), 46-54.
- Guo, J., Hepburn, C. J., Tol, R. S., & Anthoff, D. (2006). Discounting and the social cost of carbon: A closer look at uncertainty. *Environmental Science & Policy*, 9(3), 205-216.
- Hamaide, B., & Boland, J. J. (2006). Cost-benefit analysis is useful in evaluating climate change policies. *World Resource Review*, 18(2), 279-290.
- Hewman, K., Capillo, A., Famurewa, A., Nath, C., & Siyanbola, W. (2013). *What is the evidence on evidence-informed policy making? Lessons from the International Conference on Evidence-Informed Policy Making*. Abgerufen von http://www.inasp.info/uploads/filer_public/2013/04/22/what_is_the_evidence_on_eipm.pdf
- Hjørland, B. (2011). Evidence-based practice: An analysis based on the philosophy of science. *Journal of the Association for Information Science and Technology*, 62(7), 1301-1310.
- Holzappel, S. (2014). *The role of indicators in development cooperation: An overview study with a special focus on the use of key and standard indicators* (Studies 81). Bonn: German Development Institute / Deutsches Institut für Entwicklungspolitik (DIE).
- Holzappel, S., & Janus, H. (2015). *Improving education outcomes by linking payments to results* (Discussion Paper 2/2015). Bonn: German Development Institute / Deutsches Institut für Entwicklungspolitik (DIE).
- Howlett, M., & Craft, J. (2013). Policy advisory systems and evidence-based policy: The location and content of evidentiary policy advice. In S. P. Young (Ed.), *Evidence-based policy-making in Canada: A multidisciplinary look at how evidence and knowledge shape Canadian public policy* (pp. 27-44). Don Mills, Ontario: Oxford University Press.
- Janus, H., & Holzappel, S. (2016). *Results-based approaches in agriculture: What is the potential?* (Discussion Paper 25/2016). Bonn: German Development Institute / Deutsches Institut für Entwicklungspolitik (DIE).
- Jasanoff, S. (1987). Contested boundaries in policy-relevant science. *Social Studies of Science*, 17(2), 195-230.
- Jasanoff, S. (1995). Product, process, or programme: Three cultures and the regulation of biotechnology. In M. Bauer (Ed.), *Resistance to new technology: Nuclear power, information technology and biotechnology* (pp. 311-331). Cambridge: Cambridge University Press.
- Jasanoff, S. (2005). Restoring reason: Causal narratives and political culture. In B. Hutter & M. Power (Eds.), *Organizational encounters with risk* (pp. 209-232). Cambridge: Cambridge University Press.
- Jasanoff, S. (2006). Just evidence: The limits of science in the legal process. *The Journal of Law, Medicine & Ethics*, 34(2), 328-341.
- Kapiriri, L., Sinding, C., & Arnold, E. (2017). How do development assistance partners conceptualise and prioritise evidence in Priority Setting (PS) for health programmes relevant to low-income countries? A qualitative study. *Evidence & Policy: A Journal of Research, Debate and Practice*, 13(2), 255-273.
- Keijzer, N., & Klingebiel, S. (2017). *Realising the data revolution for sustainable development: Towards capacity development 4.0*. Abgerufen von https://www.paris21.org/sites/default/files/CapacityDevelopment4.0_FINAL_0.pdf
- Kemper, N., & Sieler, S. (2016). *Results-based approaches in international development cooperation*. Abgerufen von https://www.kfw-entwicklungsbank.de/PDF/Download-Center/PDF-Dokumente-Development-Research/2016-11-17_EK_EBA_EN.pdf

- KfW Development Bank. (2016). *Ex post evaluation Zimbabwe*. Abgerufen von https://www.kfw-entwicklungsbank.de/PDF/Evaluierung/Ergebnisse-und-Publikationen/PDF-Dokumente-R-Z_EN/Simbabwe_OVC_2015_E.pdf
- KfW Development Bank. (2017). *14th evaluation report 2015-2016*. Frankfurt am Main: Autor.
- Klasen, S. (2017, November 30). *(Quasi-)experimental impact evaluation in the German development cooperation: Usefulness, potential of big data and future cooperation*. DEval workshop. Bonn: German Institute for Development Evaluation.
- Klingebiel, S. (2012). *Results-based aid (EOZ): New aid approaches, limitations and the application to promote good governance* (Discussion Paper 14/2012). Bonn: German Development Institute / Deutsches Institut für Entwicklungspolitik (DIE).
- Klingebiel, S. (2014). *Development cooperation: Challenges of the new aid architecture*. Basingstoke: Palgrave Macmillan.
- Klingebiel, S., & Janus, H. (2014). Results-based aid: Potential and limits of an innovative modality in development cooperation. *International Development Policy*, 6(1).
- Klingebiel, S., Mahn, T., & Negre, M. (Eds.). (2016). *The fragmentation of aid: Concepts, measurements and implications for development cooperation*. Basingstoke: Palgrave Macmillan.
- Krisch, F., Schmitt, J., & Dörr, U. (2015). *Begleitende Maßnahmen der allgemeinen Budgethilfe in Subsahara-Afrika*. Abgerufen von http://www.deval.org/files/content/Dateien/Evaluierung/Berichte/DEval_Begleitende_Massnahmen_Budgethilfe_Gesamtbericht.pdf
- Lavis, J. N., Ross, S. E., & Hurley, J. E. (2002). Examining the role of health services research in public policymaking. *The Milbank Quarterly*, 80(1), 125-154.
- Law, J. (2004). *After method: Mess in social science research*. London: Routledge.
- Leigh, A. (2010). Evidence-based policy: Summon the randomistas? In Productivity Commission (Ed.), *Strengthening evidence-based policy in the Australian Federation, Volume 1: Proceedings, roundtable proceedings* (pp. 215-226). Canberra: Productivity Commission.
- Levine, R., & Savedoff, W. D. (2016). *Impact evaluation: How the wonkiest subject in the world got traction*. Abgerufen von <http://blogs.3ieimpact.org/impact-evaluation-how-the-wonkiest-subject-in-the-world-got-traction/>
- Lomborg, B. (2004a). Need for economists to set global priorities. *Nature*, 431(7004), 17.
- Lomborg, B. (Ed.). (2004b). *Global crisis, global solutions*. Copenhagen: Copenhagen Consensus Center.
- Lomborg, B. (2007a). *Cool it: The skeptical environmentalist's guide to global warming*. New York, NY: Vintage Books.
- Lomborg, B. (2007b). *Solutions for the world's biggest problems: Costs and benefits*. Cambridge: Cambridge University Press.
- Lomborg, B. (Ed.). (2010). *Smart solutions to climate change: Comparing costs and benefits*. Cambridge: Cambridge University Press.
- Lomborg, B. (2012). *Rethink HIV: Smarter ways to invest in ending HIV in sub-Saharan Africa*. Cambridge: Cambridge University Press.
- Lomborg, B. (2015). *The Nobel laureates' guide to the smartest targets for the world: 2016-2030*. Copenhagen: Copenhagen Consensus Center.
- Lomborg, B. (2017). Making government smarter: How to set national priorities. *Foreign Affairs*, 96, 90-98.
- Lundahl, M. (2013). *The political economy of disaster: Destitution, plunder and earthquake in Haiti* (Vol. 61). Oxford: Routledge.
- Mallett, R., Hagen-Zanker, J., Slater, R., & Duvendack, M. (2012). The benefits and challenges of using systematic reviews in international development research. *Journal of Development Effectiveness*, 4(3), 445-455.
- Marmot, M. G. (2004). Evidence-based policy or policy-based evidence?: Willingness to take action influences the view of the evidence – look at alcohol. *BMJ: British Medical Journal*, 328(7445), 906.

- McGillivray, M., Feeny, S., Hermes, N., & Lensink, R. (2005). *It works; it doesn't; it can, but that depends...: 50 years of controversy over the macroeconomic impact of development aid* (9291907391). Abgerufen von <https://www.wider.unu.edu/sites/default/files/rp2005-54.pdf>
- McNulty, J. (2012). Symbolic uses of evaluation in the international aid sector: Arguments for critical reflection. *Evidence & Policy: A Journal of Research, Debate and Practice*, 8(4), 495-509.
- Meadows, D. H., Meadows, D. L., Randers, J., & Behrens, W. W. (1972). *The limits to growth*. New York, NY: Universe Books.
- Miles, B., & Morse, S. (2007). The role of news media in natural disaster risk and recovery. *Ecological Economics*, 63(2), 365-373.
- Monaghan, M. (2013). Sources and resources: An interview with ... Dr Howard White by Mark Monaghan. *Evidence & Policy*, 9(1), 131-136.
- Musgrove, P. (2011). *Rewards for good performance or results: A short glossary*. Washington, DC: World Bank.
- Newman, K., Evans, W., Dercon, S., Whitty, C., Harvey, M., Vince, J., & Rimmer, J. (2014). *What is the evidence on the impact of research on international development?* London: Department for International Development.
- Nutley, S., & Webb, J. (2000). Evidence and the policy process. In H. T. O. Davies, S. M. Nutley, & P. C. Smith (Eds.), *What works? Evidence-based policy and practice in public services* (pp. 13-41). Bristol, UK: Blackwell Publ. Ltd.
- OCEBM Levels of Evidence Working Group. (2011). The Oxford 2011 levels of evidence. Abgerufen von <http://www.cebm.net/wp-content/uploads/2014/06/CEBM-Levels-of-Evidence-2.1.pdf>
- OECD (Organisation for Economic Co-operation and Development). (2006). *Peer review: Germany*. Paris: Autor.
- OECD. (2014a). *Measuring and managing results in development co-operation: A review of challenges and practices among DAC members and observers*. Paris: Autor.
- OECD. (2014b). *OECD development co-operation peer reviews: United Kingdom*. Paris: OECD Publishing.
- OECD. (2015). *OECD development co-operation peer reviews: Germany*. Paris: OECD Publishing.
- OECD. (2016a). *Evaluation systems in development co-operation: 2016 review*. Paris: OECD Publishing.
- OECD. (2016b). *OECD DAC Peer review USA*. Paris: OECD Publishing.
- OECD. (2017). *Development co-operation report 2017: Data for development*. Paris: Autor.
- OECD. (s.a.). *Results-based decision making*. Abgerufen von <http://www.oecd.org/development/results-measurement-and-management.htm>
- OECD/DAC (Development Assistance Committee). (2002). *Glossary of key terms in evaluation and results-based management*. Working Party on Aid Evaluation. Paris: Autor.
- Oliver, K., Innvær, S., Lorenc, T., Woodman, J., & Thomas, J. (2014). A systematic review of barriers to and facilitators of the use of evidence by policymakers. *BMC health services research*, 14(1), 2.
- Oliver, K., Lorenc, T., & Innvær, S. (2014). New directions in evidence-based policy research: A critical analysis of the literature. *Health Research Policy and Systems*, 12(1), 34.
- Parkhurst, J. (2016). *The politics of evidence: From evidence-based policy to the good governance of evidence*. Oxford: Routledge.
- Paul, E., Albert, L., Bisala, B. N. S., Bodson, O., Bonnet, E., Bossyns, P., . . . Eclou, D. S. (2018). Performance-based financing in low-income and middle-income countries: Isn't it time for a rethink? *BMJ Global Health*, 3(1), e000664.
- Pawson, R. (2006). *Evidence-based policy: A realist perspective*. Thousand Oaks, CA: Sage.
- Perakis, R., & Savedoff, W. (2015). *Does results-based aid change anything? Pecuniary interests, attention, accountability and discretion in four case studies* (CGD Policy Paper 52). Washington, DC: Center for Global Development.

- Pittore, K., te Lintelo, D. J., Georgalakis, J., & Mikindo, T. (2017). Choosing between research rigour or support for advocacy movements, a false dichotomy? *IDS Bulletin*, 47(6). Brighton, UK: Institute of Development Studies.
- Rinnert, D., & Brower, L. (2017). *How is evidence actually used in policy-making? A new framework from a global DFID programme*. Abgerufen von <https://blogs.oxfam.org/en/content/how-evidence-actually-used-policy-making-new-framework-global-dfid-programme>
- Roberts, D., & Khattri, N. (2012). *Designing a results framework for achieving results: A how-to guide*. Washington, DC: International Bank for Reconstruction and Development (World Bank).
- Rogers, P., Hawkins, A., McDonald, B., Macfarlan, A., & Milne, C. (2015). *Choosing appropriate designs and methods for impact evaluation*. Abgerufen von <https://industry.gov.au/Office-of-the-Chief-Economist/Publications/Documents/Impact-evaluation-report.pdf>
- Rudolph, A., von Schiller, A., & Strupat, C. (2017). *Lernen aus Evidenz – Wie Wirkungsanalysen besser genutzt werden können* (Die aktuelle Kolumne). Abgerufen von <https://www.die-gdi.de/die-aktuelle-kolumne/article/lernen-aus-evidenz-wie-wirkungsanalysen-besser-genutzt-werden-koennen/>
- Sachs, J. D. (2004). Solving global crises: Economists alone are not enough. *The Lancet*, 364(9451), 2087-2088.
- Savedoff, W. D., Levine, R., & Birdsall, N. (2006). *When will we ever learn? Improving lives through impact evaluation*. Washington, DC: Center for Global Development.
- Scott-Findlay, S., & Pollock, C. (2004). Evidence, research, knowledge: A call for conceptual clarity. *Worldviews on Evidence-Based Nursing*, 1(2), 92-97.
- Segone, M. (2007). Why evaluate? The evolution of the evaluation function. In M. Segone (Ed.), *New trends in development evaluation* (pp. 8-30). New York, NY: UNICEF Regional Office for CEE/CIS and the International Program Evaluation Network.
- Segone, M., & Ocampo, A. (Eds.). (2006). *Creating and developing evaluation organizations, lessons learned from Africa, Americas, Asia, Australasia and Europe*. Lima: desco.
- Stern, E., Stame, N., Mayne, J., Forss, K., Davies, R., & Befani, B. (2012). *Broadening the range of designs and methods for impact evaluations. Report of a study commissioned by the Department for International Development*. Abgerufen von https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/67427/design-method-impact-eval.pdf
- Stewart, R. (2015). A theory of change for capacity-building for the use of research evidence by decision makers in southern Africa. *Evidence & Policy: A Journal of Research, Debate and Practice*, 11(4), 547-557.
- Stewart, E., & Smith, K. E. (2015). "Black magic" and "gold dust": The epistemic and political uses of evidence tools in public health policy making. *Evidence & Policy: A Journal of Research, Debate and Practice*, 11(3), 415-437.
- Strassheim, H., & Kettunen, P. (2014). When does evidence-based policy turn into policy-based evidence? Configurations, contexts and mechanisms. *Evidence & Policy: A Journal of Research, Debate and Practice*, 10(2), 259-277.
- The Economist. (2016). *Developing Bangladesh: How to spend it*. Abgerufen von <https://www.economist.com/news/finance-and-economics/21698302-ambitious-attempt-work-out-best-use-scarce-resources-how-spend-it>
- Tytler, R. (2001). Dimensions of evidence, the public understanding of science and science education. *International Journal of Science Education*, 23(8), 815-832.
- Uneke, C., Ezeoha, A., Ndukwe, C., Oyibo, P., Onwe, F., Igbinedion, E., & Chukwu, P. (2011). Individual and organisational capacity for evidence use in policy making in Nigeria: An exploratory study of the perceptions of Nigeria health policy makers. *Evidence & Policy: A Journal of Research, Debate and Practice*, 7(3), 251-276.
- United Kingdom. (2014). *What works review of the use of evidence in the Department for International Development*. Evidence into Action and What Works Team. Abgerufen von https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/340971/What_Works_Review_DFIG.pdf

- United Nations Development Programme. (2007). *MDG-based national development strategies*. Poverty Group Bureau for Development Policy. New York, NY: Autor.
- United States Agency for International Development. (2017). *ADS chapter 201: Program cycle operational policy*. Abgerufen von <https://www.usaid.gov/sites/default/files/documents/1870/201.pdf>
- United States Government Accountability Office. (2017). *2017 survey of organizational performance and management issues*. Abgerufen von <https://www.gao.gov/special.pubs/gao-17-776sp/resultsall.htm>
- Vaessen, J., Raimondo, E., & Bamberger, M. (2015). Impact evaluation approaches and complexity. In M. Bamberger, J. Vaessen, & E. Raimondo (Eds.), *Dealing with complexity in development evaluation: A practical approach* (pp. 62-87). London, New Delhi, Singapore: Sage Publications.
- Vähämäki, J., Schmidt, M., & Molander, J. (2011). *Results-based management in development cooperation*. Stockholm: Riksbankens Jubileumsfond.
- Waddington, H., White, H., Snilstveit, B., Hombrados, J. G., Vojtkova, M., Davies, P., . . . Petticrew, M. (2012). How to do a good systematic review of effects in international development: A tool kit. *Journal of Development Effectiveness*, 4(3), 359-387.
- Ward, V. (2017). Why, whose, what and how? A framework for knowledge mobilisers. *Evidence & Policy: A Journal of Research, Debate and Practice*, 13(3), 477-497.
- White, H. (2014). Current challenges in impact evaluation. *The European Journal of Development Research*, 26(1), 18-30.
- White-Kaba, M. (2017). *How rewards improve health practice in Malawi: Learnings from a maternal and newborn Health Initiative*. Abgerufen von http://health.bmz.de/ghpc/case-studies/how_rewards_improve_health_practice_malawi/EOF_Malawi_6.pdf
- Witter, S., Kardan, A., Scott, M., Moore, L., & Shaxson, L. (2017). Generating demand for and use of evaluation evidence in government health ministries: Lessons from a pilot programme in Uganda and Zambia. *Health Research Policy and Systems*, 15(1), 86.
- Wollmann, H. (2017). Entwicklungslinien von Evaluation und Evaluationsforschung in Deutschland. Zwischen Vorgeschichte, Aufbruch und Konsolidierung. *Zeitschrift für Evaluation*, 16(2), 33-53.
- Young, J. (2006). Bridging research and policy: Insights from 50 case studies. *Evidence & Policy: A Journal of Research, Debate and Practice*, 2(4), 439-462.
- Young, S. P. (2013). *Evidence-based policy-making in Canada: A multidisciplinary look at how evidence and knowledge shape Canadian public policy*. Don Mills, Ontario: Oxford University Press.
- Young, K., Ashby, D., Boaz, A., & Grayson, L. (2002). Social science and the evidence-based policy movement. *Social Policy and Society*, 1(3), 215-224.
- Zenghelis, D. (2010). Book review: Smart solutions to climate change: Comparing costs and benefits, edited by Bjørn Lomborg. *Journal of Environmental Investing*, 1(2), 79-89.

Anhang 1: Nutzung von Evidenz in verschiedenen Politikfeldern – einige Beispiele

Politikfeld	Ansatz	Strang und Niveau	Erklärung
Gesundheit	Evidenzbasierte Medizin (EBM)	Strang A: Praxis Mikro-Ebene	Fokussiert sich auf die Beziehung zwischen Arzt/Ärztin und Patient*in. Die Qualität der verfügbaren Evidenz als Referenz für die Lenkung gesundheitsbezogener Maßnahmen stützt sich auf eine Hierarchie von (weitgehend quantitativen) Methoden. Randomisiert kontrollierte Studien und ihre systematischen Reviews werden als Methoden für die Bereitstellung bester Evidenz verstanden. Es besteht ein besonderer Schwerpunkt bei der Wirksamkeit. EBM zielt darauf ab, direkt auf die Praxis einzuwirken, indem praktischen Ärzt*innen empfohlen wird, schlechte Maßnahmen durch solche zu ersetzen, die die <i>Outcomes</i> verbessern. Auf Grundlage der Evidenz können Interventionsstrategien auf der Makro-Ebene identifiziert werden.
	Evidenzbasierte (öffentliche) Gesundheitspolitik	Strang B: Politikgestaltung Makro-Ebene	Nutzung der EBM, Evidenz aus dem öffentlichen Gesundheitswesen in Verbindung mit ökonomischen Studienergebnissen (falls gegeben), beispielsweise gesundheitsökonomische Evaluationen (Kosten-Wirksamkeits-Analyse; Kosten-Nutzen-Analyse). Dieser Ansatz zielt darauf ab, die Allokation finanzieller Mittel im Gesundheitswesen, einschließlich der Prioritätensetzung, zu verbessern.
Bildung	Evidenzbasierte Bildungspolitik	Strang B: Politikgestaltung Makro-Ebene	Deckt alle Aspekte der Bildung ab – von der Praxis im Klassenzimmer bis hin zur Politikgestaltung. Sie wird aus empirischer qualitativer und quantitativer Forschung gewonnen und bezeichnet die Idee, dass sich Bildungspolitik auf die beste Evidenz über die voraussichtlichen Auswirkungen stützen sollte.
	Evidenzbasierte Bildungspraxis	Strang A: Praxis Mikro-Ebene	Lehre wird ähnlich verstanden wie das Gesundheitswesen; angemessene Lehrkonzepte, die auf empirischen Einsichten basieren, werden in der Bildungspraxis implementiert.
Öffentliche Verwaltung	Ergebnisorientierte Haushaltsführung	Strang A: Praxis Mikro-/Meso-Ebene	Ergebnisorientierte Haushaltsführung verbindet die zugeteilten finanziellen Mittel mit messbaren Ergebnissen und ersetzt dabei die Ressourcenorientierung. Sie stützt sich auf verfügbare Informationen über die erwarteten Ergebnisse und zielt darauf ab, Rechenschaft, Effizienz und Transparenz zu verbessern. Diesbezüglich haben die Bereitstellung und Nutzung von Evidenz einen hohen Stellenwert im gesamten Haushalts- bzw. Programmzyklus: Programmbewertung, Budgetentwicklung, Durchführung, Überwachung, Monitoring der <i>Outcomes</i> und zielgerichtete Evaluierung.
Strafrechtspflege	Rechtsprechung	Strang A: Praxis	Evidenz ist das durch Regeln und Gesetze sanktionierte Mittel, durch das im Rahmen gerichtlicher Verfahren die Wahrheit festgestellt wird. Diesbezüglich werden Fingerabdrücke, Blut, Haare, Haut, Zeugenaussagen etc. als wichtige belastende Evidenztypen verstanden.
	Evidenzbasierte Kriminologie und Strafrechtspflege (Kriminalprävention)	Strang B: Politikgestaltung	Bei der Kriminalprävention und -kontrolle, der Behandlung von Inhaftierten und rehabilitierten Beschuldigten wird Evidenz genutzt, die sich auf quantitative und qualitative Daten und Informationen stützt.
Umwelt	Evidenzbasiertes Umweltmanagement und -politik	Strang B: Politikgestaltung	Befürwortet einen rationaleren, rigorosen und systematischen Ansatz für das Umweltmanagement. Auf diese Weise soll die Entscheidungsfindung durch die zuverlässigsten Forschungsergebnisse zu verschiedenen Umweltaspekten unterstützt werden.

Quelle: Autor.