





# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
1.1	Kleine Nichtregierungsorganisationen . . . . .	3
1.2	Ingenieure ohne Grenzen . . . . .	5
1.3	Fragestellung und Relevanz der Arbeit . . . . .	6
<b>2</b>	<b>Evaluation</b>	<b>7</b>
2.1	Dimensionen von Evaluation . . . . .	8
2.2	Systematisierung durch Evaluationsmodelle . . . . .	14
2.3	Charakterisierung von Evaluationsansätzen . . . . .	17
2.4	Kritik an Evaluation . . . . .	19
<b>3</b>	<b>Exemplarische Evaluationsansätze</b>	<b>20</b>
3.1	Methodenorientierte Evaluation nach Rossi . . . . .	21
3.2	Bewertungsorientierte Evaluation nach Scriven . . . . .	24
3.3	Nutzenorientierte Evaluation nach Fetterman . . . . .	28
<b>4</b>	<b>Fallstudie Ingenieure ohne Grenzen</b>	<b>31</b>
4.1	Einführung des Fallstudiengegenstands . . . . .	31
4.2	Aufbau, Methodik und Ziele der Fallstudie . . . . .	32
4.3	Konzeption und Durchführung der Onlineumfrage . . . . .	35
4.4	Konzeption und Durchführung der Interviews . . . . .	38
4.5	Auswertung der Fallstudie . . . . .	40
4.5.1	Quantitative Auswertung der Onlineumfrage . . . . .	40
4.5.2	Qualitative Auswertung der Interviews . . . . .	47
4.6	Ergebnisse der Fallstudie . . . . .	52
4.7	Reflexion der Fallstudie . . . . .	54
<b>5</b>	<b>Synthese und Analyse</b>	<b>55</b>
5.1	Einführung der DeGEval-Standards für Evaluation . . . . .	55
5.2	Synthese eines bedarfsgerechten Evaluationsansatzes . . . . .	56
5.3	Analyse des beschriebenen Evaluationsansatzes . . . . .	62

<b>6</b>	<b>Fazit und Ausblick</b>	<b>65</b>
	<b>Literaturverzeichnis</b>	<b>68</b>
<b>A</b>	<b>Anhänge</b>	<b>74</b>
A.1	Metaskript . . . . .	75
A.2	DeGEval Standards . . . . .	77
A.3	Rohdaten Onlineumfrage . . . . .	79

# 1 Einleitung

In Deutschland waren im Jahr 2015 rund 2.000 Nichtregierungsorganisationen (NROs) beim *Dachverband Entwicklungspolitik und Humanitäre Hilfe deutscher Nichtregierungsorganisationen e.V.* (VENRO) registriert. Viele dieser Organisationen sind kleine NROs mit geringen Ressourcen [6]. Als Ressourcen werden für die vorliegende Arbeit finanzielle Mittel, Zeit, Wissen und Personal angenommen. Diese kleinen NROs haben trotz begrenzter Ressourcen einen Projektmittelumsatz, im Fall von Ingenieure ohne Grenzen e.V. (IOG) zum Beispiel waren es im Jahr 2018 rund €1,2 Mio. [18]. Professionelle Evaluationen werden jedoch nur selten in diesem Kontext durchgeführt. Daraus lässt sich schließen, dass die NROs Projektmittel primär in der Projektarbeit umsetzen, die Ergebnisse dieser Arbeit (positiver und negativer Art) jedoch selten auf einer wissenschaftlichen Basis nachgewiesen, hinterfragt oder zum Lernen genutzt werden. Dies ist nicht zwangsweise auf mangelndes Interesse oder Unwillen zurückzuführen, sondern kann eine Vielzahl von anderen Gründen haben: Evaluation hat sich erst zu Beginn des 21. Jahrhunderts in Deutschland etabliert, im Zusammenhang mit dem Aufkommen des *New Public Managements*. In Gesetzen oder wichtigen Programmen wird daher erst wenig darauf hingewiesen oder gar gefordert. Zudem ist Evaluation in Deutschland kein geschützter Begriff, das heißt die Qualität eines Evaluators, und damit das ihm oder ihr entgegengebrachte Vertrauen, ist für einen Laien nur schwer einzuschätzen [39, S.41]. Ein weiterer Grund für die geringe Anzahl von Evaluationen bei kleinen NROs im Speziellen kann die schwierige Finanzierungssituation sein [35]: Kleine NROs haben in der Regel keine Abteilung für Qualitätsmanagement, sondern sind auf die Förderung entsprechender Aktivitäten angewiesen. Insbesondere Evaluationen lassen sich aber nur in wenigen Fällen durch Projektmittelgeber finanzieren, wodurch entsprechende Vorhaben von vielen NROs auf einen späteren Zeitpunkt verschoben oder ganz unterlassen werden. Themen wie Evaluation oder auch Monitoring rücken somit oft in den Hintergrund, weil andere Hindernisse, die zum Teil die Existenz der NRO oder ihrer Projektarbeit bedrohen, primär adressiert werden müssen. Ab einer bestimmten Größe einer NRO lässt sich ein Zuwachs von Evaluationsvorhaben feststellen. Dies legt nahe, dass der Bedarf an Evaluation erfüllt wird sobald die notwendigen Rahmenbedingungen es zulassen [26]. Aktuell gibt es einige Anbieter auf dem Evaluationsmarkt, die Evaluationslösungen zur „Durchführung

in Eigenregie“ anbieten. Diese werben mit günstigen Konditionen und belastbaren Ergebnissen, trotz etwaig geringem Vorwissen der Durchführenden. Es ist jedoch unklar, ob jedes Produkt für jede NRO geeignet ist um in diesen Rahmenbedingungen auch die gewünschten Ergebnisse verlässlich zu produzieren.

Ziel der Arbeit ist es einen Evaluationsansatz für entwicklungspolitische Projekte zu entwickeln, der insbesondere für kleine NROs mit begrenzten Ressourcen nutzbar ist. Dies wird beispielhaft an der Organisation IOG entwickelt. Mittels einer Fallstudie werden Bedarfe aufgenommen und deren Übereinstimmung mit bestehenden Evaluationslösungen überprüft.

**Struktur der Arbeit.** Die vorliegende Arbeit ist in sechs Kapitel strukturiert. Das Kapitel 1 bildet die Einleitung, in der kleine Nichtregierungsorganisationen beschrieben werden, ebenso wie der Verein Ingenieure ohne Grenzen e.V. Des Weiteren wird die Fragestellung und Relevanz der Arbeit vorgestellt. Das Kapitel 2 gibt eine theoretische Einführung und stellt dabei die Dimensionen von Evaluation vor. Verschiedene Ansätze werden systematisiert und die Charakteristika von Evaluationsansätzen vorgestellt. Schließlich werden übliche Kritikpunkte an Evaluation beschrieben. Das Ziel dieses Kapitels ist es eine theoretische Grundlage zu schaffen, auf der das folgende Kapitel aufbauen kann. In Kapitel 3 werden entsprechend der vorgestellten Systematisierung drei Evaluationsansätze diskutiert, in der Systematisierung verortet und hinsichtlich ihres Ressourcenaufwands bewertet. Anhand einer aktuellen Anwendung des jeweiligen Ansatzes wird aufgezeigt in welcher Art und Form dieser im Rahmen von Entwicklungszusammenarbeit relevant ist. Das Zusammenspiel aus Theorie, Methodik und Anwendung soll hierbei nachvollziehbar und transparent werden, um im Folgenden Aussagen zur Anwendbarkeit einer Evaluationslösung bei IOG treffen zu können. Das Kapitel 4 beschreibt die Durchführung einer Fallstudie innerhalb des Vereins Ingenieure ohne Grenzen e.V. Diese hat zum Ziel die Bedarfe von IOG an Evaluation und ihre Ressourcen diese durchzuführen festzustellen. In Kapitel 5 werden Evaluationsstandards eingeführt und ein Evaluationsansatz beschrieben, der die Ergebnisse aus Kapitel 4 mit den bestehenden Evaluationsansätzen aus Kapitel 3 zusammenführt. Das Ergebnis ist ein Evaluationsansatz, der konkret auf die Situation, den Bedarf und die Ressourcen von IOG angepasst ist. Dieser wird zum Schluss in Kapitel 5 kritisch in Bezug auf seine Aussagekraft reflektiert. In Kapitel 6 wird ein abschließendes

Fazit der Arbeit gezogen und weiterführende Fragestellungen diskutiert.

Die vorliegende Arbeit soll dazu beitragen Evaluationsforschung auf die Gruppe kleiner NROs aufmerksam zu machen. Es werden Lösungsansätze entwickelt, indem vorhandene Ansätze exemplarisch auf den Bedarf einer kleinen NRO spezialisiert werden.

## **1.1 Kleine Nichtregierungsorganisationen**

In der Literatur ist keine trennscharfe Definition für *kleine NROs* gegeben, die eine allgemeingültige Akzeptanz erreicht hat. Dornberger beschrieb im Jahr 2000 kleine NROs als solche, die einen Jahresumsatz von weniger als einer Million Euro haben [10, S. 57], dies taucht jedoch nicht in anderen Quellen auf. Auch VENRO nutzt diesen Begriff in aktuellen Stellungnahmen, ohne eine allgemeingültige Definition bereitzustellen [44].

*Kleine NROs* schließt im Rahmen dieser Arbeit solche Organisationen ein, die eine geringe Anzahl von hauptamtlichen Mitarbeitenden beschäftigen und vornehmlich in kleinen Teams an Projekten arbeiten. Die kleinen NROs streben lokal begrenzte Veränderungen an und fokussieren diese auf eine eng begrenzte Gruppe von Teilnehmenden oder Profitierenden. Die Finanzierung solcher NROs ist in der Regel nicht konstant gesichert, sondern muss regelmäßig erneuert werden, zum Beispiel durch Projektanträge bei Förderorganisationen oder durch das Aufbringen eines Spendenvolumens [26]. Typische Probleme kleiner NROs umfassen daher eine fortlaufende Akquise von Geldmitteln um die Organisation handlungsfähig zu halten. Weiterhin ist eine fortlaufende Suche und Einarbeitung von ehrenamtlich Mitarbeitenden, Qualifikation von langfristig Mitarbeitenden und die grundsätzliche Organisationsentwicklung in vielen Fällen herausfordernd. Insbesondere Organisationsentwicklung ist hierbei sehr spezifisch für verschiedene Entwicklungsstadien einer NRO, da sie sich mit der wachsenden Anzahl von Projektmitteln und Mitarbeitenden verändert. Übliche Themen dieser Organisationsentwicklung können Wissensmanagement, Qualitätsmanagement und Nachhaltigkeit sein. Es handelt sich nicht um eine trennscharfe Definition, sondern vielmehr um ein Kontinuum, dem eine Organisation in einem Stadium der Entwicklung vollständig entspricht und zu einem anderen Zeitpunkt nur noch in Teilen oder gar nicht mehr entspricht.

**Ressourcenknappheit.** Ressourcenknappheit ist ein Merkmal kleiner NROs. Sie lässt sich in verschiedene Mangelsituationen differenzieren, die finanzierungsbedingt zeitlich parallel oder auch unabhängig voneinander auftreten können. Es handelt sich in den meisten Fällen um einen Mangel an Personal, Wissen, Zeit oder Geld. In der Regel sparen kleine NROs durch den Einsatz von ehrenamtlich Mitarbeitenden Personalkosten. Dieses Ersparnis geht jedoch mit der Beobachtung einher, dass es zu einer erhöhten Fluktuation von Mitarbeitenden kommen kann. Erfahrene Mitarbeitende verlassen die Organisation und Unerfahrene rücken nach. Der entstehende Nachteil für die kleine NRO liegt darin, dass ein konstanter Abfluss von Wissen resultiert. Dieses Wissen umfasst neben der Erfahrung in der Entwicklungszusammenarbeit auch spezifisches innerhalb der NRO generiertes Wissen, das nur durch die erneute Erarbeitung wiederhergestellt werden kann. Die Aufgabe der Qualifizierung neuer Mitarbeitenden wird in der Folge notwendig und weitere Ressourcen müssen investiert werden.

Die Ressource Zeit lässt sich in der Regel nur bedingt durch eine Organisation steuern oder sparen, da Fristen und Projektlaufzeiten von außen einwirken, zum Beispiel durch Geldgeber.

Die Ressource Geld ist ebenfalls abhängig von externen Faktoren, wie zum Beispiel der Spendenbereitschaft von Mitgliedern und Mäzenen, ebenso wie von der erfolgreichen Antragsstellung auf Fördergelder und Projektmittelbewilligungen. Letzteres wird mit wachsender Erfahrung zwar planbarer, aber in keinem Fall verlässlich. Kleine NROs sind hierbei weniger resilient gegenüber Schwankungen im Mittelzufluss als größere NROs, da in vielen Fällen keine finanziellen Reserven aufgebaut werden, respektive können, und nur sehr wenige Einsparpotentiale in den Arbeitsprozessen vorhanden sind.

Bei Mängeln von einer oder mehrerer Ressourcen muss eine Organisation reagieren. Kürzungen können bei organisatorischen Abläufen gemacht werden oder müssen mit Hilfe des Auslagerns von Aufgaben an ehrenamtlich Mitarbeitende kompensiert werden. Insbesondere bei einem Mangel der Ressource Geld können diese Entscheidungen notwendig werden. An der Projektumsetzung werden im Allgemeinen keine oder nur in Ausnahmefällen Abstriche gemacht. Evaluationen sind durch ihre zum Teil ressourcenintensive Durchführung ein Bereich der sich anbieten gestrichen zu werden, weil ihr Fehlen keinen direkten Einfluss auf die Projektarbeit nimmt und viele Ressourcen freigeben kann. Zudem sind diese bei Projektmittelquellen kleiner NROs, wie zum Beispiel Spendenmitteln,

nicht explizit gefordert, sodass Evaluationen nicht umsetzungsrelevant sind.

## 1.2 Ingenieure ohne Grenzen

Ein Beispiel für eine solche NRO ist *Ingenieure ohne Grenzen e.V. (IOG)*, die sich 2003 gegründet hat. Die gemeinnützige, private Hilfsorganisation realisiert entwicklungspolitische Projekte unabhängig von religiösen, politischen oder ethischen Gesichtspunkten und war seit Gründung in rund 30 Ländern aktiv [18, S.24]. In 30 Regionalgruppen organisierten sich 2018 rund 1000 Ehrenamtliche und arbeiteten in der Projektumsetzung, im Fundraising oder in der Öffentlichkeitsarbeit [18, S.4]. Der Verein wird von einem ehrenamtlichen Vorstand sowie einer hauptamtlichen Geschäftsstelle geleitet, die rund 12 Mitarbeitende in Voll- und Teilzeit umfasst. Die mithilfe von lokalen Partnern partizipativ umgesetzten Projekte verfolgen das Ziel einer Kapazitätsstärkung in den jeweiligen Projektregionen. Eine nachhaltige Verstetigung und Instandhaltung der Projektergebnisse soll durch fortwährenden Einbezug der profitierenden Teilnehmenden und den Aufbau von Ownership<sup>1</sup> während der Vorbereitung, Umsetzung und Nachbereitung gewährleistet werden [18, S.4]. Die Finanzierung von IOG setzt sich aus Spendenmitteln von Förderern und Unternehmen, ebenso wie aus Zuwendungen von Stiftungen zusammen. Eine Finanzierung aus Projektmittelanträgen wie sie von vergleichbaren anderen Organisationen genutzt wird, wird nicht in Anspruch genommen. Dies ist ein Beispiel, in dem sich IOG von der oben beschriebenen allgemeinen NRO unterscheidet. Es bietet IOG einige Möglichkeiten, da das oben beschriebene Problem der Unförderbarkeit von Evaluation nicht relevant ist. Gleichzeitig wird für einen Anteil der Spendengelder eine Zweckbindung zugesagt. Eine solche Zweckbindung ist ein Versprechen an alle Spender, dass die finanzielle Unterstützung ohne Abzüge in die Projektarbeit vor Ort fließt. Das schließt zum Beispiel administrative Ausgaben oder Werbungskosten des Vereins aus. Im Jahr 2018 wiesen von rund 700.000 Euro Spenden rund 400.000 Euro eine solche Zweckbindung auf. Etwa 300.000€ waren zweckungebunden und konnten daher potentiell für alle vom Verein als sinnvoll erachteten Ausgaben genutzt werden, zum Beispiel auch für

---

<sup>1</sup>Die deutsche Übersetzung des Begriffes Ownership ist Eigentümerschaft. Im Bereich der Entwicklungszusammenarbeit hat er sich darüber hinausgehend in der Diskussion um die Identifikation der Menschen mit einem sie betreffenden Vorhaben etabliert. Diese gilt als Vorbedingung für die Nachhaltigkeit und den Erfolg einer Maßnahme.

Evaluation [18, S.30]. Von 2016 bis 2018 lässt sich ein Bestreben und Erfolg der Organisation feststellen den Anteil zweckungebundener Spenden zu erhöhen, sodass flexibler auf Finanzierungsbedarfe von Projekten eingegangen werden kann. In erster Linie werden dadurch Spendenmittel flexibler in der Projektarbeit einsetzbar, in zweiter Linie können aber auch unterstützende Arbeiten wie Evaluationen davon profitieren.

### **1.3 Fragestellung und Relevanz der Arbeit**

Evaluation hat einen etablierten Status in der Entwicklungszusammenarbeit. Das BMZ empfiehlt zum Beispiel Zuwendungsempfängern explizit die Beauftragung externer Evaluatoren unter dem Hinweis auf die hohen fachlichen und methodischen Anforderungen, die mit der Durchführung einer Evaluation einhergehen. Gleichzeitig wird auf die Unabhängigkeit eines externen Evaluators hingewiesen, die die Objektivität der Ergebnisse unterstützen kann [2, S. 129]. Weiterhin stimmen VENRO Mitgliedsorganisationen mit Beitritt einer Selbstverpflichtung zu wirkungsorientiert zu handeln. Insbesondere Evaluation und darauf anzuwendende Kriterien werden definiert, um erzielte Wirkungen dokumentieren zu können. Hierbei werden außerdem externe Evaluatoren empfohlen um die Objektivität zu erhöhen [43]. Zusätzlich dazu beschreibt der Verein IOG sich selbst in seinem Onlineauftritt als eine NRO, die ein hohes Maß an Qualität in ihren Projekten umsetzt und die Ergebnisse von Erfolgsüberprüfungen auch in die aktuellen Tätigkeiten einfließen lassen möchte [20]. Eine solche externe Evaluation zu beauftragen geht jedoch mit einem finanziellen Aufwand einher, der viele NROs und insbesondere kleine NROs davon abhält diesen Empfehlungen zu entsprechen [25]. Spezifisch empfiehlt Phineo 3-10 Prozent des Projektbudgets einzuplanen, um eine aussagekräftige Evaluation in Auftrag geben zu können [25]. Diese Maßgabe übersteigt die finanziellen Mittel kleiner NGOs in vielen Fällen. Kleine NROs befinden sich daher in einem Spannungsfeld zwischen der Anforderung Evaluationen in Auftrag zu geben und der begrenzten Ressource Geld.

Die vorliegende Arbeit befasst sich mit ebendieser Situation, indem anhand einer spezifischen kleinen NRO überprüft wird inwiefern eine Finanzierung von externer Evaluation möglich ist und welche Teile einer Evaluation gegebenenfalls auch intern erfolgen können, ohne dabei den Qualitätsanspruch an die Ergebnisse einschränken zu müssen.

## 2 Evaluation

Evaluation ist Teil der angewandten Sozialforschung. In vielen Kontexten wird sie in einem Atem mit Monitoring eingeführt, weshalb Evaluation nachfolgend in Abgrenzung zu Monitoring beschrieben wird. Monitoring ist ein Instrument mit dessen Hilfe nachprüfbar ist, ob durchgeführte Maßnahmen entsprechend der vorliegenden Planung richtig durchgeführt wurden. Evaluation beschäftigt sich dem entgegen mit der Frage, ob die richtigen Maßnahmen durchgeführt wurden. Diese Unterscheidung erscheint zuerst marginal, in Bezug auf entwicklungspolitische Projekte und Entwicklungszusammenarbeit im Allgemeinen können die Unterschiede zwischen Evaluation und Monitoring jedoch groß sein. Beispielhaft möchte ich ein fiktives Projekt beschreiben: Ein Brunnenbau in Kenia. Nach viel Recherche wird ein idealer Standort für einen neuen Brunnen festgestellt. Ein deutsches Team reist aus und baut diesen Brunnen, mit einer sehr tiefen Reichweite und hoher Effizienz. Laut Monitoring wurden alle Ziele entsprechend der vorliegenden Planung erreicht. In der Folge der Baumaßnahme sinkt jedoch der Grundwasserspiegel, wodurch weniger tief reichende Brunnen in der Umgebung weniger oder kein Wasser mehr fördern können. Betroffene müssen nun weite Wege auf sich nehmen um vom neuen Brunnen Wasser zu holen. Eine Evaluation ist methodisch befähigt einen nicht-intendierten Effekt wie diesen systematisch zu erfassen, wodurch sie insbesondere in der Entwicklungszusammenarbeit als wichtiges Werkzeug von Planern und Durchführenden gleichermaßen geschätzt wird.

Anwendungsgebiete von Evaluation sind über Entwicklungszusammenarbeit hinausgehend sehr divers und umfassen zum Beispiel wissenschaftlichen Nutzen in der Soziologie und in den Bildungswissenschaften, ebenso wie praktischen Nutzen im Qualitätsmanagement von Wirtschaftsunternehmen. Die allgemeinste Grundlage aller Arten von Evaluation besteht aus drei Schritten: Sammeln von Informationen, Bewerten von Informationen und letztlich das Treffen einer Entscheidung [40].

Verschiedene Arten der Evaluation finden vornehmlich Anwendung in den Bereichen Bildung, Umwelt, Qualitätsmanagement und internationale Zusammenarbeit. Die vorliegende Arbeit fokussiert sich hauptsächlich auf den zuletzt genannten Bereich.

## 2.1 Dimensionen von Evaluation

Die Dimensionen von Evaluation lassen sich in drei verschiedene unterteilen, abhängig von der Art, den Zielen und des Nutzens von Evaluation. Zur Erstellung von neuen und zur Einordnung von bereits erfolgten Evaluationen ist es hilfreich diese anhand der drei Dimensionen zu analysieren und pro Dimension die jeweiligen Schwerpunkte und Spezifikationen zu betrachten.

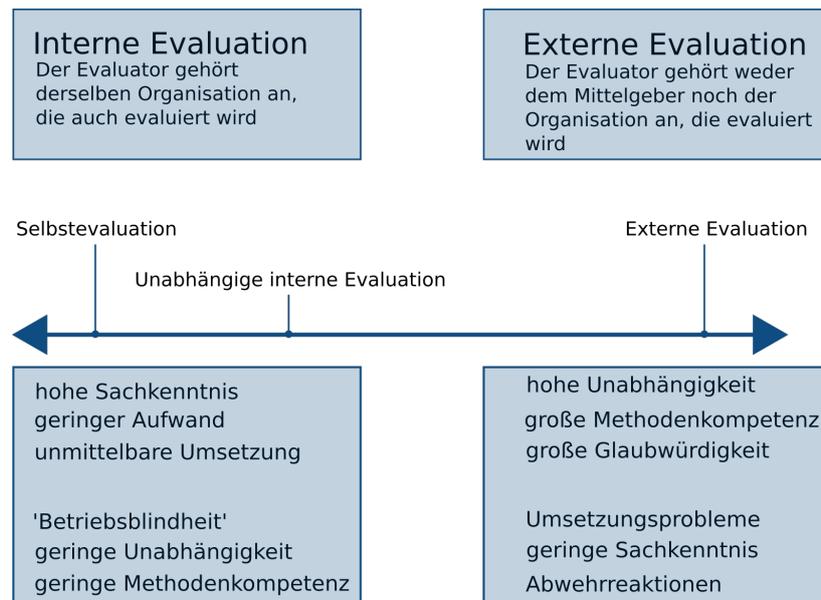
Arten 	Ziele 	Nutzen 
Alltagsevaluationen Wissenschaftliche E. Summative Evaluationen Formative Evaluationen Distanzverhältnis - Interne Evaluationen - Selbstevaluationen - Externe Evaluationen Zeitverhältnis - Ex-ante - Ongoing - Ex-post	Erkenntnis Kontrolle Legitimation Entwicklung	Programmdurchführung Gesamtbilanz Hinterfragen der Maßnahme Kausalitätsproblem

**Abbildung 1:** Die Dimensionen von Evaluation lassen sich in drei verschiedene unterteilen: Arten, Ziele und Nutzen. Jede dieser Dimensionen umfasst weitere Spezifikationen anhand derer eine Evaluation ausgerichtet werden kann.

**Arten von Evaluation.** Die diversen Einsatzgebiete von Evaluation gehen darauf zurück, dass das Einsatzfeld, nämlich die Bewertung eines Sachverhalts, sowohl im Alltag als auch in beruflichen Feldern existiert. Die erstgenannte Form von Evaluation lässt sich als Alltagsevaluation beschreiben. Die Bewertung muss hierbei nicht auf definierte objektive Kriterien zurückgehen, sondern kann auch rein intuitiv erfolgen. Bei Evaluationen mit Anwendung in Wissenschaft oder Wirtschaft handelt es sich in der Regel um wissenschaftliche Evaluationen.

Es gibt verschiedene Faktoren, anhand derer sich Evaluationen weiter voneinander abgrenzen lassen. Untenstehend werden die drei Faktoren summativ vs. formativ, Distanzverhältnis und Zeitraum beschrieben. Die wissenschaftlich-professionelle Evaluation bemüht sich um eine objektive, sachbezogene Bewertung, indem sowohl *summativ* beobachtete Veränderungen, als auch *formativ* generierte Daten genutzt werden können um eine Bewertung herbeizuführen [39]. Diese erste Unterteilung enthält unter dem for-

mativen Aspekt eine diagnostische Beurteilung zu einem bestimmten Zeitpunkt, während der summative Aspekt eine Art Bilanzierung abgeschlossener Vorgänge darstellt. Insbesondere im Bereich der Erziehungswissenschaften wird diese Unterteilung genutzt: Kursbegleitende Lernfortschrittskontrollen sind formative Beurteilungen, während es sich bei Zeugnissen um summative Beurteilungen handelt [22].



**Abbildung 2:** Das Spektrum verschiedener Distanzverhältnisse zwischen Evaluator und Evaluationsgegenstand reicht von der internen bis zur externen Evaluation, obenstehend sind Vor und Nachteile der beiden Pole beschrieben.

Das Distanzverhältnis zwischen Evaluator und Evaluationsgegenstand beschreibt die zweite Unterteilung. Zur Einordnung von Evaluationen ist es daher notwendig festzustellen wer diese durchführt. Hierbei kann es sich um einen unabhängigen Außenstehenden handeln oder um eine Person, die Teil des zu evaluierenden Systems ist. Vorteile einer *externen Evaluation* sind Objektivität, die durch den Abstand zum Evaluationsobjekt entsteht, und eine starke Professionalisierung und Methodenkompetenz, die man von einem beauftragten Dienstleister erwarten kann. Diese Kombination kann zu einem hohen Maß an Glaubwürdigkeit führen, welche notwendig ist um zum Beispiel die Umsetzung der Empfehlungen einer Evaluation anzustoßen. Nachteile einer externen Evaluation können beispielsweise mangelndes Vertrauen der Stakeholder<sup>2</sup> in den Evaluationsprozess und

<sup>2</sup>Der Begriff Stakeholder umfasst in diesem Zusammenhang sowohl IOG als Adressat einer Evaluation, das heißt als Organisation, die die Evaluationsergebnisse nutzen möchte, ebenso wie die Betroffenen, die in

dessen Ergebnis sein. Zusätzlich kann die geringe Sachkenntnis des zu evaluierenden Systems zu einer längeren Bearbeitungszeit der Evaluation führen. Als weiterer Nachteil externer Evaluationen werden häufig die höheren Kosten genannt. Insbesondere für NROs, die mit geringen finanziellen Ressourcen arbeiten, kann dieser Faktor bereits entscheidend sein, um von einer externen Evaluation abzusehen. Hierbei ist es wichtig zu betrachten, dass eine interne Evaluation oder Selbstevaluation ebenfalls Kosten verursacht, indem zum Beispiel hauptamtliches Personal mit der Betreuung oder teilweisen Betreuung der Evaluation beauftragt werden muss. Diese Kosten sind anderer Natur, sollten jedoch im Vergleich zu den finanziellen Kosten einer externen Evaluation nicht vernachlässigt werden. Eine *interne Evaluation* bietet im Gegensatz dazu die Vorteile, dass sich ein Evaluator von intern mit hoher Wahrscheinlichkeit gut im System auskennt und daher einschätzen kann, welche Daten im Kontext relevant sind und daher erhoben werden müssen. Die Datenerhebung einer Evaluation kann unter diesen Umständen zügig und mit geringem Aufwand durchgeführt werden. Im Allgemeinen lässt sich beobachten, dass Evaluatoren von intern ein hohes Vertrauen entgegengebracht wird, obwohl zum Teil eine mangelnde evaluationsbezogene Methodenkompetenz vorliegt. Dies kann zu fehlerhaft durchgeführten Evaluationen und nicht belastbaren Aussagen führen. In der Regel haben Evaluatoren von intern wenig Distanz zum Evaluierungsgegenstand, wodurch es zu subjektiven Einschätzungen kommen kann [38, S.61]. Eine bloße Einteilung in intern oder extern wird jedoch der Vielfalt der durchgeführten Evaluationen nicht gerecht. Nach Stockmann/Meyer [39, S.82] sind Evaluationen nicht bipolar einzuordnen, da der Bereich zwischen intern und extern als Spektrum begriffen werden soll. Eine Einordnung von Evaluationen kann daher auch eher intern oder eher extern sein. Diese Perspektive ermöglicht die Differenzierung von internen Evaluationen, Selbstevaluationen, unabhängigen internen Evaluationen und externen Evaluationen, die andernfalls nicht darstellbar wäre.

Die dritte Unterteilung von Evaluationen bezieht sich auf den Zeitraum, in dem diese durchgeführt wird. Evaluationen die im Vorfeld eines Projekts durchgeführt werden nennt man *ex-ante Evaluationen*. Abhängig von der Art des Evaluationsgegenstands kann Evaluation zum Beispiel Teil der Projektplanung sein und bereits im Planungsstadium genutzt werden. Ergebnisse einer solchen *ex-ante* Evaluation können daher die Planung aktueller den jeweiligen Projektkontexten zu verorten sind.

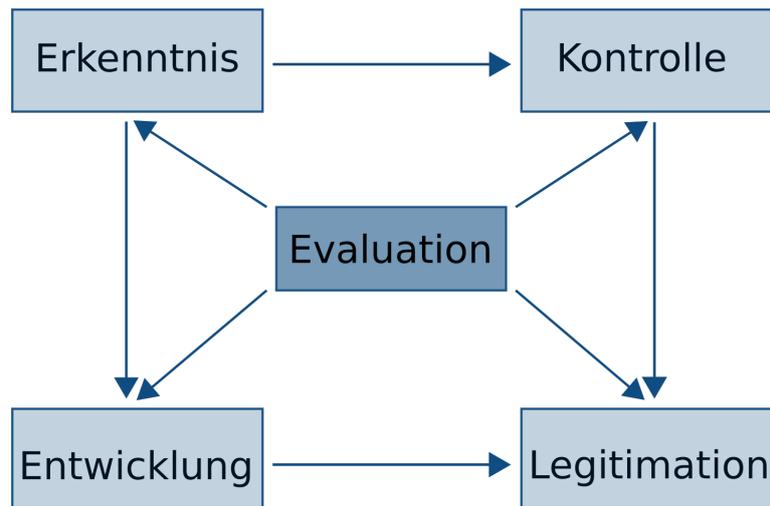
Projekte maßgeblich beeinflussen. Wird die Evaluation eines Projekts parallel zu dessen Umsetzung durchgeführt heißt sie *ongoing Evaluation*. Solche Evaluationen haben in der Regel das Ziel eine Aussage über den Status der Umsetzung treffen zu können. Der Vorteil von Evaluationen zu diesem Zeitpunkt liegt darin, dass ggf. in der Projektsteuerung Veränderungen vorgenommen werden können. Solche Evaluationen werden zum Beispiel als Teil des Monitorings durchgeführt. Ein großer Anteil von Evaluationen wird erst nach Abschluss der Intervention durchgeführt. Diese *ex-post Evaluationen* sind insbesondere in der Entwicklungszusammenarbeit zu finden, da sie in besonderem Maße Aussagen über die Wirkung von Interventionen treffen können.

**Ziele von Evaluation.** Es lassen sich nach Stockmann vier Evaluationsziele unterscheiden [38, S. 37], die sich trotz ihrer inhaltlichen Verwobenheit analytisch voneinander trennen lassen, vergleiche Abbildung 3. Diese Ziele von Evaluation sind

1. die Gewinnung von Erkenntnissen,
2. die Ausübung von Kontrolle,
3. die Legitimation der durchgeführten Intervention,
4. das Anstoßen von Entwicklungsprozessen.

Die Entscheidung welches dieser vier Ziele primär von einer Evaluation verfolgt werden soll ist maßgeblich für die Ausgestaltung der Evaluation. Der Grad von Partizipation, Repräsentativität und die Verwertung der Ergebnisse lassen sich in Abhängigkeit von der Zielsetzung bestimmen, ebenso wie die Auswahl der einzusetzenden Methoden [2, S.130].

Die *Gewinnung von Erkenntnissen* bezieht sich hierbei vor allem darauf wissenschaftlich fundierte Ergebnisse zu erzielen, die in relevante Forschung einfließen können. Diese Ergebnisse können sich zum Beispiel auf den Programmablauf, auf die Bedarfe der Zielgruppe und deren Erfüllung, oder auch die Rahmenbedingungen der Intervention beziehen. Die Informationen zu diesen Bezugsthemen werden im Rahmen einer Evaluation gesammelt und anhand von definierten Bewertungskriterien analysiert. Um die Ergebnisse dessen wissenschaftlich nutzbar zu machen muss auch die Evaluation, die diese produ-



**Abbildung 3:** Darstellung der vier Evaluationsziele, frei nach Stockmann [39, S. 73].

ziert, wissenschaftlichen Kriterien entsprechen, sowohl in der Erstellung der Evaluation, als auch in der Umsetzung spezifischer Methoden [38, S.37].

Die *Ausübung von Kontrolle* basiert ebenfalls auf der Sammlung und Bewertung von Daten zu Bezugsthemen. Es wird eine rationale Grundlage geschaffen, auf der zum Beispiel ein Management Entscheidungen treffen kann. Anders als bei dem Ziel Erkenntnis geht es vielmehr darum ob die Ziele, die in der Planung einer Intervention definiert werden, in der Umsetzung auch erreicht werden konnten. Vielfach werden zur Bewertung der Erreichung dieser Ziele Erfolgskriterien wie Effektivität, Effizienz, Akzeptanz und Nachhaltigkeit verwendet, die auf die DAC-Kriterien zurückführen.<sup>3</sup>

Die *Legitimation* durchgeführter Programme oder Projekte kann ebenfalls das Ziel von Evaluation sein. Hierbei wird die im Rahmen einer Evaluation erhobene Datenbasis genutzt, um zu belegen mit welchem Input welcher Output und welche Wirkung generiert wurde. Auf diesem Weg lassen sich als ausführende Organisation zum Beispiel die Mittelverwendung gegenüber der Mittelgeber nachweisen und Kennzahlen zu Effizienz und Effektivität ermitteln. Evaluationen die deutlich nach dem Projektabschluss stattfinden können zudem Aussagen über Nachhaltigkeit treffen, die in Bezug auf die Wirkung von Projekten und Programmen häufig als Ziel formuliert wird. Insbesondere Mittelge-

<sup>3</sup>Die Evaluierungskriterien nach DAC (Development Assistance Committee) sind vollständig: Relevanz, Effektivität, Effizienz, Entwicklungspolitische Wirkungen, Nachhaltigkeit. Sie wurden in einem Prozess der OECD erarbeitet und sind international als Kriterien zur Bewertung von Projekten der Entwicklungszusammenarbeit anerkannt [2, S.141].

ber die einen Wettbewerb um ihre Mittel verzeichnen sind in der Position ebensolche Nachweise über Output und Wirkung zu verlangen und anhand dessen eine Selektion vorzunehmen [39, S. 74f]. Große NROs haben in der Folge einen Wettbewerbsvorteil, da kleine NROs sich bei Auftreten eines Ressourcenmangels oft gegen das Durchführen einer Evaluation entscheiden und ebendiese Nachweise nicht erbringen können.

Das Evaluationsziel *Entwicklung* umfasst insbesondere Lernprozesse einer Organisation aus der durchgeführten Evaluation. Der Dialog zwischen verschiedenen Stakeholdern ist hierbei ein zentrales Instrument, um transparent zu bilanzieren mit welchen Erfolgen Projekte oder Programme umgesetzt wurden. Stakeholder können hierbei ausführende Organisationen, lokale Partner und von der Maßnahme Betroffene sein, ebenso wie Mittelgeber [39, S.75]. Auch wenn mit jeder Evaluation direkte oder indirekte Kontrolle verbunden ist, steht im Zentrum dieses Evaluationsziels das Lernen, das heißt die Ergebnisse sollen für die (Weiter-)Entwicklung von Projekten und Programmen genutzt werden.

**Nutzen von Evaluation.** Der Nutzen, den eine professionell durchgeführte Evaluation geben kann, zeigt sich in verschiedenen Bereichen und beschreibt allgemein die vorgeesehenen oder tatsächlich eingetretenen Folgen einer Evaluation. Dieser kann durch ein Umsetzen von Evaluationsempfehlungen oder ein Lernen aus gewonnenen Ergebnissen erzeugt werden und ist daher direkt abhängig davon in welcher Art eine Evaluation durchgeführt wurde. Weiterhin kann auch der Evaluationsprozess selbst positive Auswirkungen auf eine Organisation oder deren Prozesse haben [1].

Der Nutzen einer ex-ante Evaluation ist beispielsweise die Bewertung der Rahmenbedingungen einer Programmdurchführung. Das Hauptaugenmerk liegt in dieser Phase auf der Identifikation von Problemen, die sich in der Umsetzung ergeben können und um die Prüfung ob Zeitpläne realistisch und erfüllbar aufgestellt wurden. Mögliche Indikatoren für diese Fragestellungen sind zum Beispiel die Akzeptanz der Intervention bei verschiedenen Stakeholdern, die Qualifizierung des mit der Durchführung beauftragten Personals, ebenso wie die technische und finanzielle Ausstattung.

Ein weiterer Nutzen von Evaluation liegt darin eine Gesamtbilanz der erzielten Wirkungen aufstellen zu können. Mit Hilfe eines Soll-Ist-Vergleichs können geplante und erreichte Wirkungen miteinander verglichen werden, wodurch die Zielerreichung gemessen wird. Evaluationen gehen in ihrer Betrachtung noch darüber hinaus und erfassen viele,

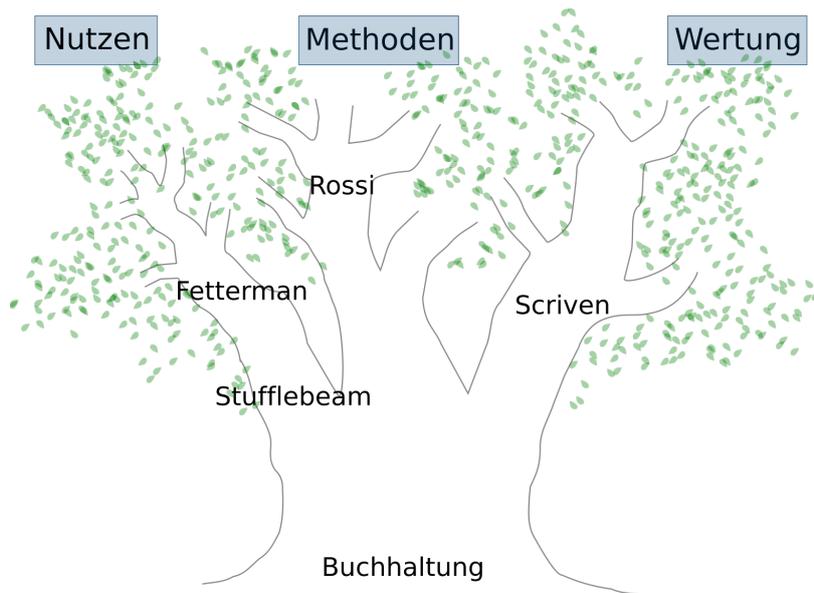
bzw. möglichst alle Wirkungen, die durch eine Programmintervention ausgelöst wurden. Diese emergenten Wirkungen können über die geplanten Wirkungen hinausgehen oder völlig anderer Natur sein. Man spricht auch von intendierten und nicht-intendierten Wirkungen. Zur Bewertung einer Intervention muss bewertet werden, ob in der Gesamtheit positive oder negative Effekte überwiegen.

Nicht zuletzt sollen Evaluationen die Kausalität von Intervention und Wirkung nachweisen, was in vielen Fällen zu einem Kausalitätsproblem führt [4]. Die Realität, in der die Intervention stattfindet, wird durch viele Faktoren wie Regierungsentscheidungen oder Naturkatastrophen beeinflusst. Die Zuordnung welches Ereignis zu einer (positiven oder negativen) Veränderung dieser Realität geführt hat und welchen Einfluss die Intervention darauf hatte lässt sich nur schwer zweifelsfrei nachweisen. Evaluationen sind ein Versuch diese Zuordnungen durch datenbasierte Schlussfolgerungen herzustellen.

## **2.2 Systematisierung durch Evaluationsmodelle**

Die historischen Wurzeln der Evaluationsforschung reichen zurück in die USA des 19. Jahrhunderts. Die Regierung beauftragte externe Inspektoren damit die Umsetzung öffentlicher Regierungsprojekte im Bereich der Bildung und des Gesundheitswesens zu bewerten [39, S.24]. Wichtige Impulse zur Ausbildung einer Evaluationskultur gab zum Beispiel die Einrichtung eines Planungs-, Programmgestaltungs- und Haushaltsplanungssystems im Verteidigungsministerium der USA im Jahr 1965. Dieses wurde schrittweise sowohl auf alle Ministerien als auch den gesamten Regierungsapparat ausgeweitet, mit dem Ziel die Systemeffektivität, Systemeffizienz und Mittelallokation zu optimieren. Trotz des beginnenden Rückbaus dieses Systems im Jahr 1971 war ein Evaluationsmarkt mit Auftraggebern im öffentlichen und privaten Sektor entstanden, in dessen Folge auch die Evaluationsforschung stark vorangetrieben wurde [39, S.27]. Etwa zehn Jahre später als in den USA begann eine vergleichbare Modernisierungsdebatte in Europa. Unter dem Titel des *New Public Managements* wurde eine Erhöhung der administrativen Effizienz verfolgt und zur Erreichung dieser Effizienz wurden binnenstrukturelle Managementreformen eingesetzt [39, S.28]. Die darauf folgende Entwicklung der Evaluation und Evaluationsforschung brachte eine Vielzahl methodischer Ansätze und Verfahren hervor.

Alkin und Christie schlugen 2004 mit ihrem Baummodell eine Kategorisierung ver-



**Abbildung 4:** Das Baummodell nach Alkin und Christie zeigt die Systematisierung von Evaluationsansätzen entsprechend der Nutzen-, Werte- oder Methodenorientierung, frei nach [39, S.125].

schiedener Evaluationsansätze vor [3]. Sie legen kein zeitliches Ordnungsprinzip zugrunde, sondern verweisen auf inhaltliche Parallelen einzelner Ansätze, die sie dementsprechend in Haupt- oder Nebenästen eines gemeinsamen Baumes anordnen. Die Wurzel des entstehenden Evaluationsbaumes sehen sie in der Rechnungslegung und der empirischen Sozialforschung, während sich weiter oben drei Hauptäste entwickeln, die als *Methodenast*, *Nutzungsast* und *Wertungsast* bezeichnet werden. Diesen werden Evaluationsansätze zugeordnet, die methoden-, nutzungs- oder wertungsorientiert sind. Von diesen Hauptästen ausgehend werden weitere thematische Verzweigungen von kleineren Ästen repräsentiert. Es ist kritisch anzumerken, dass die Unterscheidung zwischen Evaluationstheoretikern und anderen Forschern, insbesondere im Vergleich zwischen den drei Hauptästen, inkonsequent ist. Diese Schwäche ist besonders auffällig bei dem Versuch dieses amerikanische Modell auf die europäische Evaluationslandschaft zu übertragen, da keine gemeinsame Entwicklung vorliegt, sondern eine vielschichtige, die durch sektorale und nationale Unterschiede, sowie die Diversität der beteiligten Disziplinen begründet ist [39, S.125].

Mit Hilfe einer neu formulierten Ausgangsfrage entwickelten Stockmann und Meyer das *Funktionsmodell* [39, S.144]. Sie legen die vier Evaluationsziele zugrunde: Erkenntnis, Kontrolle, Legitimation und Entwicklung. Diesen werden Evaluationskonzepte und -ansätze zugeordnet, wobei die Evaluationsziele zueinander keinen Anspruch auf Trenn-

schärfe erheben, sondern vielmehr als einander überlappende Dimensionen betrachtet werden. Entsprechend der oben vorgestellten Kategorien lassen sich bedeutende Akteure der Evaluationsforschung, Evaluationstheoretiker und andere diesen vier Polen und den Grenzgebieten zueinander zuordnen. Der Mehrwert des Funktionsmodells liegt jedoch in seiner vielseitigen Anwendbarkeit. Neben der personenbezogenen Systematisierung ist zum Beispiel auch eine politikfeldbezogene Systematisierung möglich, die einen Vergleich historischer Entwicklungen der Evaluationspraktiken, ebenso wie einen Vergleich der Evaluationspraktiken unterschiedlicher Politikfelder möglich macht. Diese Vergleichbarkeit ermöglicht, anders als andere Modelle, die Darstellung von Weiterentwicklungen einzelner Konzepte durch die unterschiedlichen Autoren selbst und stellt in der Folge inhaltliche Veränderungen und Schwerpunkte in den Vordergrund [39, S.152].

Die *fachgebietsbezogene Systematisierung* kann ebenfalls durch das Funktionsmodell dargestellt werden und ist somit ein weiterer Nachweis für dessen Flexibilität. Diese ist im Rahmen der vorliegenden Arbeit am Beispiel des Fachgebiets Ökonomie von besonderem Interesse, da die Wurzeln bedeutender Methoden von Evaluationen im Bereich der Entwicklungszusammenarbeit, wie zum Beispiel logische Modelle, aufgezeigt werden. Die Systematisierung bezieht sich implizit auf das oben beschriebene Baummodell nach Alkin und Christie [3], indem sie die Einflüsse des Rechnungswesens auf Evaluation aufzeigt und weitere ökonomisch geprägte Fachbereiche einordnet, wie zum Beispiel Wirtschaftsberichterstattung, Projekt- und Programmplanungswerkzeuge, ebenso wie Qualitätsmanagementinstrumente [39, S.153]. Bei einer Betrachtung der vier Evaluationsziele können im nächsten Schritt elementare Instrumente zugeordnet werden, wieder am Beispiel der Ökonomie: Die Erkenntnis über Effizienz und Effektivität kann durch Berichtssysteme gewonnen werden. Diese nutzen Indizes, Vergleichszahlen und Zeitreihen und gleichen damit mehr einem Monitoringsystem, als einer Evaluation. Die Kontrolle wird durch Bilanzierung erreicht, zum Beispiel mit Hilfe von Soll-Ist-Vergleichen, die eine Aussage darüber treffen können ob ein vordefiniertes Ziel erreicht wurde. Die Legitimation kann durch Audit- und Zertifizierungsverfahren erreicht werden. Diese bauen im Gegensatz zur oben genannten Bilanzierung in den meisten Fällen auf einer freiwilligen Teilnahme an Prüfverfahren auf und werden in der Regel von extern durchgeführt. Das Lernen und die Entwicklung im Allgemeinen haben einen besonderen Stellenwert in diesem Fachgebiet, da eine kontinuierliche Qualitätssteigerung erstrebt wird. Auch aus diesem Grund wur-

de die Entwicklung moderner Planverfahren vermehrt gefördert und nahm besonderen Einfluss auf das Evaluationsziel Lernen [39, S. 155].

Monitoring- und Evaluationssysteme sollen integrativ zur Überwachung des Projektfortschritts genutzt werden und zu Modellen des kontinuierlichen Lernens und Weiterentwickelns werden, insbesondere in der Entwicklungszusammenarbeit. Der Ursprung dieser logischen Modelle lässt sich nicht eindeutig feststellen, Einflüsse sind aber zum Beispiel aus der betrieblichen Ablauforganisation oder der Informatik festzustellen. In der Regel geben diese Auskunft über den aktuellen Ressourceneinsatz für ein Projekt und den Status der Zielerreichung, bzw. einer Teilzielerreichung. Bereits seit 1997 nutzt die Weltbank das *Log-Frame Modell*, welches ein logisches Modell ist das 1969 für die US Agency for International Development erstellt wurde [42], vergleiche Kapitel 3.1. Zum Teil bedingt durch diese Verwurzelung in der Entwicklungszusammenarbeit konnte es sich gegen andere logische Modelle durchsetzen und wird in abgewandelter Form auch von der Gesellschaft für internationale Zusammenarbeit (GIZ), damals Gesellschaft für technische Zusammenarbeit (GTZ), genutzt. Die Projektentwicklung wird mit Hilfe von vordefinierten Indikatoren beobachtet, die in ihrer Gesamtheit eine Aussage über die Zielerreichung zum jeweiligen Zeitpunkt geben. Darüber hinaus soll über diese klassische Frage des Monitorings aber auch eine klassische Frage der Evaluation beantwortet werden: Ist das Ziel auf das hingearbeitet wird auch erreichbar? Mithilfe dieser Sinnfrage kann sowohl während des Projektverlaufs eine Veränderung von Planungsannahmen erfolgen, als auch die Zielformulierung selbst aktualisiert werden. Diese zu jedem Zeitpunkt des Projekts erhebbaren Daten ermöglichen es schwellenlos aus der Projektpraxis zu lernen und Prozesse, ebenso wie Organisationen weiterzuentwickeln.

### **2.3 Charakterisierung von Evaluationsansätzen**

Entsprechend des Funktionsmodells werden in der vorliegenden Arbeit Evaluationsansätze nach den oben beschriebenen vier Evaluationszielen unterschieden. Nach dieser systematischen Verortung müssen im zweiten Schritt grundsätzliche Charakteristika bestimmt werden, um ein klares Bild der benötigten Evaluation zu erzeugen [39, S.67]. Diese Charakteristika werden in Kapitel 4.3 zur Grundlage der Befragung von Mitgliedern des Vereins IOG, um darauf aufbauend Aussagen darüber treffen zu können wie eine konkrete

Evaluation für sie aussehen könnte. Diese Charakteristika sind:

1. Evaluationsgegenstand,
2. Ziele und Aufgaben,
3. Kriterien der Bewertung,
4. Evaluator,
5. Forschungsparadigma.

Im Folgenden werden diese näher beschrieben.

Erstens muss der Evaluationsgegenstand klar beschrieben sein, um den Umfang des Vorhabens abschätzen zu können. Dieser kann im Rahmen der Entwicklungszusammenarbeit zum Beispiel ein Projekt oder auch ein ganzes Programm sein.

Zweitens müssen Ziele und Aufgaben einer Evaluation herausgearbeitet werden. Dazu kann zum Beispiel ein Evaluationsziel aus Abbildung 3 gewählt werden. Diese Entscheidung bestimmt maßgeblich die Gestaltung und Methodik der durchzuführenden Evaluation, da sie zum Beispiel auch die Adressaten der Ergebnisse definiert. Eine Evaluation mit dem Ziel der Kontrolle und dem Adressat Geldgeber muss anders aufgebaut und durchgeführt werden, als zum Beispiel mit dem Ziel der Erkenntnis und dem Adressat Forschungsgemeinschaft.

Drittens muss festgelegt werden anhand welcher Kriterien diese Evaluation eine Bewertung vornehmen soll. Im Bereich der Entwicklungszusammenarbeit haben sich bereits Evaluationskriterien etabliert, auch wenn diese nicht als verbindlich angesehen werden. Es handelt sich um Kriterien, die vom Development Assistance Committee (DAC) der Organization for Economic Cooperation and Development (OECD) erarbeitet wurden. Diese umfassen Relevanz, Effektivität, Effizienz, Wirkung und Nachhaltigkeit. Durch den herausgestellten Charakter dieser Kriterien im Umfeld der Entwicklungszusammenarbeit werden sie in der in Kapitel 4.3 beschriebenen Onlineumfrage als Antwortmöglichkeiten bei Frage Nr. 8 genutzt.

Viertens muss festgelegt werden welche Person oder Personengruppe die Evaluation durchführt, vergleiche Kapitel 2.1. Dies bestimmt den Einfluss der verschiedenen Perspektiven durch verschiedene Evaluatoren, ebenso wie den Ressourcenaufwand der zur

Durchführung aufgebracht werden muss. Eine Entscheidung für ein Distanzverhältnis ist daher eine Schlüsselentscheidung für kleine NROs im Bezug auf Evaluation.

Fünftens stellt sich die Frage mit welchem Forschungsparadigma evaluiert werden soll. Auf der einen Seite lässt sich Evaluation als empirisch-wissenschaftliches Verfahren betrachten, das einer Forschungslogik folgt die einen kritisch-rationalen Charakter hat und alle empirischen Forschungsmethoden wie Umfragen, Experteninterviews oder Beobachtungen für einsetzbar hält. Eine solche Evaluation ist klar der angewandten Sozialforschung zuzuordnen. Auf der anderen Seite lässt sich Evaluation als sozial-konstruktivistisches Verfahren betrachten, das einer Forschungslogik folgt die die Realität als sozial-konstruiert beschreibt. In der Folge sind parallel existierende, zueinander in Konflikt stehende Perspektiven möglich. Evaluation muss in diesem Kontext einen qualitativen Ansatz verfolgen, um möglichst viele Sichtweisen der Realität einzubeziehen. Wichtig ist hierbei auch zu beachten, dass einzelne Stakeholdergruppen verschiedene Machtstellungen haben können, sodass eine Überbetonung einzelner Wahrnehmungen droht. Evaluation hat in diesem Sinne den Anspruch Unterschiede herauszuarbeiten und zu einem transformatorischen Prozess zugunsten benachteiligter Gruppen beizutragen [37, S. 11].

Diese fünf Charakteristika bieten ein Hilfsmittel um auf einer abstrakten Ebene verschiedene Evaluationsansätze zu beschreiben und miteinander zu vergleichen. Sie sind insbesondere hilfreich für Organisationen die eine Evaluation planen, da spezifische Anforderungen kategorisiert und der Entscheidungsprozess für einen Evaluationsansatz unterstützt wird.

## **2.4 Kritik an Evaluation**

Neben den obenstehend beschriebenen Zielen und Nutzen die Evaluation erreichen und bringen kann gibt es auch Kritik. Insbesondere in Deutschland zeigt sich, dass durch die mangelnde Institutionalisierung keine unabhängige Qualitätskontrolle von Evaluationen und Evaluierenden stattfindet. Die Bezeichnung *Evaluation* ist zum Beispiel kein geschützter Begriff und auch die Verpflichtung zum Einsatz bestehender Standards oder anderer Qualitätsmerkmale ist aktuell nicht geplant. Die harmonisierte Definition von Mindeststandards wäre an dieser Stelle insbesondere wichtig, da diese mit der Definition

und Einigung auf internationale, verbindliche Standards für Evaluation einhergeht. Die Deutsche Gesellschaft für Evaluation (DeGEval) wurde 1997 gegründet und setzt sich als Fachverband, ebenso wie andere in Europa, dafür ein verbindliche Qualitätsstandards und feste Prinzipien der Vorgehensweise bei Evaluationen zu etablieren. Ein Ergebnis des Verbands sind die Standards für Evaluation, die 2004 verabschiedet wurden (vgl. Kapitel 5.1), jedoch noch keine große Verbreitung erfahren haben [39, S.40]. Bemühungen wie diese führten in den vergangenen Jahren zu einem Fortschritt der Professionalisierung, sind jedoch noch nicht europaweit harmonisiert. Ein Risiko, das mit der oben genannten Kritik einhergeht, ist zum Beispiel die Gefahr von unprofessionellen Evaluatoren, die mit schlechten Methoden unwissenschaftliche Ergebnisse produzieren und den Ruf von Evaluation allgemein beschädigen. Ein weiteres Risiko ist die Erstellung mangelhafter Evaluationen von intern. Interne Evaluationen können sehr professionell gemacht werden, jedoch kann es aufgrund von Interessenskonflikten oder Ressourcenknappheit zu Verkürzungen oder oberflächlichen Betrachtungen von Sachverhalten kommen. Die Aussagen einer so vorgenommenen Evaluation sind nicht belastbar und können ebenfalls zu einer negativen Wahrnehmung von Evaluation beitragen.

### **3 Exemplarische Evaluationsansätze**

Im Rückgriff auf die oben beschriebene Systematisierung durch Evaluationsmodelle nach Alkin und Christie lassen sich drei Kategorien festlegen: Die methodenorientierten, nutzenorientierten und bewertungsorientierten Evaluationsansätze. Im Folgenden werden Evaluationsansätze aus dem jeweiligen Ast vorgestellt, die von repräsentativen Vertretern der jeweiligen Kategorie verfasst wurden. Der jeweilige Ansatz wird weiterhin im Funktionsmodell nach Stockmann/Meyer verortet, auf seinen Ressourceneinsatz überprüft und die Durchführung mit verschiedenen Distanzverhältnissen wird bewertet. Schließlich wird der jeweilige Ansatz mit einer aktuell genutzten Evaluationspraxis verglichen. Diese folgend vorgestellten Evaluationsansätze bilden die Grundlage der Analyse in Kapitel 5, in dem deren Anwendbarkeit auf den ermittelten Bedarf von IOG überprüft wird.

### 3.1 Methodenorientierte Evaluation nach Rossi

*Methodenorientierte Evaluationen* nutzen ein methodisches Vorgehen um die Wirkung einer Intervention nachzuweisen oder zu widerlegen. Abhängig von der erreichten Wirkung einer Intervention lassen sich des Weiteren Aussagen über die Güte der Intervention selbst schließen. Im Allgemeinen nutzen Evaluatoren dieser Strömung wissenschaftliche Methoden und/oder Expertenwissen, um eine *Programmtheorie* zu formulieren: Programmtheorien sind die dem Programm zugrunde liegenden Annahmen darüber, wie das Programm Veränderungen oder Stabilisierungen innerhalb des lokalen Kontexts erreichen soll [46]. Sie basieren auf Mechanismen mit nachgewiesener Wirksamkeit, zum Beispiel aus vergleichbaren, vorausgegangenen Interventionen. Anhand dieser werden in der Folge Kriterien und Messoperationen der Evaluation ausgerichtet, um im Ergebnis belegen zu können welche Intervention in welcher Intensität gewirkt hat [29, S.30].

In den siebziger Jahren trug Peter Rossi mit seiner Arbeit zur Erstellung einer gemeinsamen, methodisch orientierten Evaluationstheoretischen Tradition bei, die bis dahin nicht in dieser Form existierte [39, S.115] [28]. Die in Zusammenarbeit mit Freeman erarbeitete Theorie ist bis heute aktuell und wird auch als Basis heutiger Evaluationen genutzt [7] [21] [30]. Entsprechend des oben beschriebenen Vorgehens basiert auch die von ihm beschriebene *Theory driven evaluation* darauf eine Programmtheorie zu entwickeln. Diese muss nicht zwingend den wissenschaftlichen Maßstäben an eine Theorie genügen, sondern vielmehr die durch Akteure einer Intervention geplante Wirkungskette darstellen.<sup>4</sup> Im Kern handelt es sich daher um ein logisches Modell, vgl. Kapitel 2.2, das die Intervention und die darauf aufbauenden Wirkungen beschreibt, ohne die Notwendigkeit einer wissenschaftlichen Basis dieser Annahmen [39, S.116]. Ein übliches Beispiel in Projekten der Entwicklungszusammenarbeit ist ein Log-Frame<sup>5</sup>, der zum Beispiel im Rah-

---

<sup>4</sup>Mit seiner Einstellung bezieht sich Rossi insbesondere auf Entscheidungsträger als Adressaten von Evaluation, aber nicht nur diese. Im Vergleich zu anderen Vertretern, wie zum Beispiel Stufflebeam, etabliert Rossi eine moderate Haltung gegenüber allen Nutzergruppen. Stufflebeam fokussiert seine Theorien der Evaluation im Gegensatz zu Rossi allein auf Entscheidungsträger, da der Hauptzweck von Evaluation seiner Meinung nach der Informationsgewinn für diese ist. Nach Stufflebeam gehen Evaluationen über das Messen des Erfolgs einer Intervention hinaus und bewerten die Nützlichkeit eines Vorgehens. Ein Beispiel für Evaluation nach Stufflebeam ist das CIPP-Modell, das er 1983 veröffentlichte [41].

<sup>5</sup>Log-Frame bezeichnet hier ein Planungsdokument, das seine theoretische Grundlage im Log-Frame-Modell hat. Dieses beschreibt systematisch den geplanten Input, Output, Outcome und Impact eines Pro-

men eines Planungsworkshops zu Beginn des Projekts erstellt wird. Im ersten Schritt einer Evaluation nach Rossi wird dieser Log-Frame im wissenschaftlichen Kontext verortet und bewertet oder gegebenenfalls entsprechend wissenschaftlicher Standards aufbereitet. Auf diesem Weg wird eine Evaluierbarkeit der Projekte und insbesondere der dahinterliegenden Programmtheorien überprüft und sichergestellt. Darauf folgend werden die durch die Programmtheorie spezifizierten Ziele und Wirkungen mit den tatsächlichen Bedürfnissen, die durch das übergeordnete Programm adressiert werden sollen, verglichen [28]. Mögliche Informationsquellen für diesen Vergleich sind zum Beispiel Stakeholder der Intervention, die die Plausibilität und Machbarkeit der Programmtheorie bewerten können. Die Ergebnisse können zur Verbesserung der Konzeption der Intervention genutzt werden.

Eine Einbindung methodischer Verfahren in das Grundkonzept der Evaluation als direkte Folge der Evaluationstheorie bildet einen Meilenstein der Evaluationsforschung. Die strikte Bindung von Evaluationsforschung an ein bestimmtes Fachgebiet wurde überwunden und eine allgemeingültige Evaluationsforschung ist möglich. Die interdisziplinäre Anwendbarkeit der methodenorientierten Evaluationen trägt in großem Maße dazu bei, dass Evaluationsforschung als eigenständiges Fachgebiet verstanden werden kann [39, S.117].

**Verortung im Funktionsmodell Stockmann/Meyer.** Im Funktionsmodell nach Stockmann/Meyer ist Rossi nahe der Erkenntnis verortet. Diese Ausrichtung ergibt sich aus der starken Ausrichtung hin zur Grundlagenforschung, die ein Merkmal Rossis ist. In diesem Sinne sind Rossis Arbeiten weniger aus Interesse an Evaluationsgegenständen entstanden, sondern sind vielmehr die Dokumentation einer anwendungsbezogenen Feldforschung. Im Gegensatz zu anderen Kollegen, die im Funktionsmodell ebenfalls der Erkenntnis zugeordnet werden, geht Rossi über die Erkenntnistheorien hinaus und baut eine an Programmtheorien orientierte Evaluationsforschung auf, die oben beschrieben wurde. Ein wichtiges Merkmal dieser ist, dass vermehrt neben dem Erkenntnisgewinn auch partizipatorische und legitimierende Elemente thematisiert und in Konzeptionen integriert werden. Die grundlegende Philosophie der Rossi mit seinen Forschungen folgt ist hierbei von einem positivistischen Ideal gelenkt, sodass in der Folge die genutzten Methoden in

---

jekts [27].

der Mehrzahl quantitativer Natur sind.

**Distanzverhältnis des Evaluators.** Eine Evaluation nach Rossi ist eher extern angesiedelt, in der Folge kommt es zu einer großen Distanz zwischen Evaluator und Evaluationsgegenstand, bzw. Organisation. Durch das Ziel des Erkenntnisgewinns muss die Evaluation zwingend auf einer wissenschaftlichen Basis aufgebaut sein und mittels wissenschaftlicher Methoden und Schlussfolgerungen Ergebnisse gewinnen. Eine interne Evaluation oder Selbstevaluation mit diesem Anspruch ist möglich, jedoch muss ein hohes Maß an Qualifikation und Erfahrung der Mitarbeitenden erreicht werden, um wissenschaftlich gesicherte Erkenntnisse zu produzieren.

**Ressourcenaufwand.** Der Ressourcenaufwand der externen Evaluation nach Rossi ist direkt vergleichbar mit dem einer Evaluation nach Scriven, vgl. Kapitel 3.2, da es sich bei beiden Ansätzen um Evaluationen handelt, die idealerweise von einem externen Evaluator durchgeführt werden. Diese grundsätzliche Entscheidung führt zu einem erhöhten Bedarf an finanziellen Mitteln, dabei aber zu einem reduzierten Bedarf an Wissen über Evaluation innerhalb der Organisation. Die Auslastung des Personals ist bei der Umsetzung einer Evaluation nach Rossi gering aber vorhanden, da auch eine externe Evaluation zu einem Mindestmaß von Mitarbeitenden der Organisation begleitet werden sollte. Die Ressource Zeit wird bei einer Evaluation nach Rossi geschont, da die Durchführung durch den externen Evaluator in der Regel deutlich zügiger abgeschlossen ist als eine interne oder Selbstevaluation, die neben dem Alltagsgeschäft von internem Personal durchgeführt werden muss. Gleichzeitig ist es jedoch wichtig zu beachten, dass die Ausschreibung eines Evaluationsauftrages durch die Organisation durchaus Vorarbeit und Fachwissen benötigt, das sich ggf. im Selbststudium angeeignet werden muss.

**Aktuelle Anwendung.** In der aktuellen Evaluationslandschaft sind am ehesten die Evaluationen der Gesellschaft für internationale Zusammenarbeit (GIZ) mit der Evaluation nach Rossi vergleichbar. Der starke Bezug zum Auftraggeber als Planer, Begleiter und Ergebnisempfänger einer Evaluation trifft sowohl für Rossi als auch für die GIZ zu. In der Arbeit mit lokalen Partnern wurde in der Vergangenheit keine formale Förderung und Kapazitätsstärkung in Bezug auf Evaluation durchgeführt, sodass eine Beteiligung

der Partner nur punktuell, zum Beispiel als Informationsquelle, stattfindet. Diese Rolle entspricht der oben beschriebenen Evaluation nach Rossi. Weiterhin liegt nach Rossi die Rolle des Adressaten der Evaluation beim Auftraggebenden. Im Beispiel werden beide Rollen von der GIZ besetzt, was Rossis Ansatz entspricht [8]. Eine Ursache für diese Rollenverteilung ist das Verhältnis zwischen lokalen Partnern und GIZ. Während Partnerorganisationen von kirchlichen Trägern zum Teil bereits seit Jahrzehnten mit diesen zusammenarbeiten, gibt es bei der GIZ eine deutlich stärkere Fluktuation von Partnern, die zu großen Teilen politisch bedingt ist. Als Regierungsorganisation (RO) ist die GIZ ein ausführendes Instrument der deutschen Außenpolitik, sodass sich verändernde politische Situationen sich mittelbar oder unmittelbar auch auf Projekte und Partner auswirken. Das Verhältnis zwischen lokalen Partnern und GIZ ist durch die zum Teil erst kurzfristige gemeinsame Arbeit weniger von Vertrauen geprägt, was einen starken Bezug der Partner zu einer Evaluation erschwert. In der Umsetzung von Evaluationen setzt die GIZ daher aktuell nach Selbsteinschätzung auf einen schwachen Partnerbezug und die Durchführung von Evaluationen durch externe Evaluatoren [8] [2, S.120].

### **3.2 Bewertungsorientierte Evaluation nach Scriven**

*Bewertungsorientierte Evaluationen* haben das Ziel, Meinungen und Erwartungen aller Beteiligten und Stakeholder einer Intervention abzubilden. Diese moderne Auffassung entspricht auch der bereits 1991 formulierten Definition von Michael Scriven [33]:

*„Evaluation ist der Prozess der Bestimmung der Leistung, des Wertes und der Güte von Dingen, und Evaluierungen sind die Produkte dieses Prozesses“<sup>6</sup>*

Die Hauptaufgabe von Evaluation ist nach Scriven die Beurteilung des Evaluationsgegenstands und seiner Bestandteile. Dies ist ein signifikanter Unterschied zur oben beschriebenen methodenorientierten Evaluation, die vielmehr den Erkenntnisgewinn in den Vordergrund stellt. Scriven sieht Evaluatoren in der Pflicht Erkenntnisse über den Evaluationsgegenstand nicht nur zu gewinnen, sondern diese zu interpretieren. Die entsprechende Deutungshoheit über die Ergebnisse sieht er beim Evaluator.

Weiterhin lässt sich dieser Ansatz in Bezug auf die Methodik abgrenzen: Evaluation

---

<sup>6</sup>zu deutsch: „Evaluierung ist der Prozess der Bestimmung der Leistung, des Wertes und der Güte von Dingen, und Evaluierungen sind die Produkte dieses Prozesses“ (übersetzt durch die Autorin).

soll nach Scriven nicht bei der Zielsetzung eines Programms ansetzen, sondern explizit nach allen entstandenen oder entstehenden Wirkungen suchen und diese bewerten. Dies umfasst insbesondere auch die nicht-intendierten Wirkungen. In der Begrenzung der Evaluation auf die Ziele eines Programms besteht nach Scriven die Gefahr Nebeneffekte von Interventionen zu übersehen oder unterschätzen [39, S.122]. Der Evaluationsansatz nach Scriven wird daher auch zielfrei (englisch: goal-free) genannt und hat sechs Kennzeichen [32]:

1. Die Ziele einer Intervention sind irrelevant für die Evaluation.
2. Evaluation hat die Aufgabe sowohl intendierte, als auch nicht-intendierte Effekte einer Intervention auszumachen und zu untersuchen.
3. Gefundene Effekte werden mit den festgestellten Bedürfnissen der Zielgruppe einer Intervention verglichen und auf Übereinstimmung untersucht.
4. Die untersuchte Intervention und deren Wirkung werden ausschließlich auf der Grundlage der erhobenen Daten bewertet.
5. Während der Entwicklung von Evaluationsfragen soll kein oder nur eingeschränkter Kontakt mit den Verantwortlichen der zu untersuchenden Intervention erfolgen. Eine Beeinflussung durch deren Sichtweisen soll auf diese Weise verhindert werden.
6. Die Bewertung der gesammelten Daten erfolgt immer im Hinblick auf die tatsächlichen Folgen und Wirkungen einer Intervention, nicht im Hinblick auf die geplanten Ziele der Intervention. Diese können durch Nebenwirkungen und nicht-intendierte Folgen von der tatsächlichen Wirkung abweichen.

Eine Evaluation, die diese Kennzeichen aufweist, hat klare Vorteile: Die Bestimmung von Programmzielen steht in der Regel am Beginn eines Programms und in der Folge auch am Beginn der Evaluation. Dies ist im Allgemeinen aufwändig, weil für eine abschließende Bestimmung von Zielen verschiedene Perspektiven erfasst und in begründeter Weise zueinander gewichtet werden müssen. Dies ist in der Evaluation nach Scriven nicht notwendig, weil der Ansatz zielfrei ist. Ein weiterer Vorteil ist, dass Evaluatoren nur schwer

zu beeinflussen sind, da der Kontakt zu den Durchführenden der Intervention deutlich geringer ausfällt als bei anderen Evaluationsansätzen [39, S.123].

In der Beschreibung Scrivens fokussiert sich Evaluation auf die Analyse der Bedürfnisse der Zielgruppe, die neben den intendierten und nicht-intendierten Effekten und Wirkungen der Intervention eine zentrale Rolle in den oben beschriebenen Kennzeichen einnimmt. Zur Datenerhebung und Bewertung werden vier Methoden eingeführt [34]: das *Scoring*, das *Ranking*, das *Grading* und das *Apportioning*. Während jede dieser Methoden das Ziel der vergleichenden Bewertung hat, unterscheiden sich die Verfahren maßgeblich. Während im *Scoring* dem Evaluationsgegenstand Zahlen zugewiesen werden, die entsprechend einer Skala Aussagen über Qualität und Effekt der Intervention widerspiegeln, konzentriert sich das *Ranking*-Verfahren vielmehr auf einen Vergleich unterschiedlicher Aspekte des Evaluationsgegenstands untereinander. Mittels *Grading* lassen sich Klassen aus Informationsmerkmalen feststellen, sodass Gemeinsamkeiten oder Unterschiede aufgezeigt werden können. Das *Apportioning* teilt Ressourcen oder Effekte einer Intervention systematisch auf die Teilaspekte der Intervention oder der Handlungsalternativen dieser auf.

In der Praxis kann jedoch festgestellt werden, dass der oben skizzierte Ablauf eine Idealvorstellung ist und ein Evaluator zwangsläufig mit der Zielvorstellung eines Programms oder einer Intervention konfrontiert sein wird. Dies wurde von späteren Evaluationstheoretikern wie Brandtstädter [5] und Weiss [45] angemerkt.

**Verortung im Funktionsmodell Stockmann/Meyer.** Im Funktionsmodell nach Stockmann/Meyer ist Scriven zwischen Legitimation und Entwicklung verortet. Scriven versteht sich in der Hauptsache als Auftragsforscher, sodass der Akteur des Auftraggebers eine zentrale Rolle einnimmt. Dieser wird oft als Geldgeber beschrieben, zum Beispiel ein staatlicher, der eine Legitimation der für eine Intervention getätigten Ausgaben erwartet. Dies wird in der Regel durch einen Nachweis der Wirkungen der Intervention erbracht, bedingt durch die Konstellation durch einen unabhängigen Evaluator. Das Lernen der Betroffenen spielt hierbei zwar eine untergeordnete Rolle, ist aber dennoch ein Ergebnis von Evaluationen die entsprechend der Vorstellung Scrivens durchgeführt werden.

**Distanzverhältnis des Evaluators.** In der Folge der oben beschriebenen Haltung ist eine Evaluation nach Scriven eindeutig durch einen externen Evaluator durchzuführen. Eine interne Evaluation kann dem Anspruch der Legitimation nur schwer genügen, sowohl im Hinblick auf Objektivität, als auch im Hinblick auf die Qualifikation und Erfahrung der Durchführenden.

**Ressourcenaufwand.** Der Ressourcenaufwand einer solchen externen Evaluation ist in erster Linie finanzieller Natur. Dennoch muss auch eine externe Evaluation durch Mitarbeitende der Organisation begleitet werden um Rückfragen zu beantworten und Datensätze zugänglich zu machen. In Bezug auf die Ressourcen Zeit und Wissen von Mitarbeitenden ist diese Art der Evaluation sehr günstig. Die weiteren Hinweise aus Kapitel 3.1 im Bezug auf Ressourcenaufwand gelten aufgrund der Ähnlichkeit auch hier.

**Aktuelle Anwendung.** In der aktuellen Evaluationslandschaft sind am ehesten die Evaluationen kirchlicher NROs wie zum Beispiel Caritas oder Brot für die Welt mit der bewertungsorientierten Evaluation nach Scriven vergleichbar. Als kirchennahe NROs arbeiten beide mit kirchlichen Trägern als Partner vor Ort. Der besondere Zugang, der durch diese Verwandtschaft zustande kommt, führt in vielen Fällen zu einer langjährigen Zusammenarbeit und dementsprechend zu einem großen Maß an Vertrauen in die lokalen Partner und deren Strukturen. Diese Grundlage ermöglicht eine Evaluation wie sie oben beschrieben wurde, weil ein dezentraler Charakter, das heißt eine lokale Durchführung der Evaluation mit einem Fokus auf die Bedürfnisse der Zielgruppe, besondere Erfordernisse hat. Lokale Partner die langjährig mit einer NRO arbeiten nehmen in der Regel eine deutlich selbstbewusstere Haltung ein, was eine Analyse der Bedürfnisse unabhängig von den Zielen der durchgeführten Maßnahme einfacher gestaltet. Lokale Partner, die erst seit kurzem mit einer NRO arbeiten können dazu neigen Aussagen im Sinne der NRO zu machen, auch wenn diese Aussagen ihren eigentlichen Bedürfnissen entgegenstehen oder nicht vollständig entsprechen. Die Suche nach nicht-intendierten Effekten kann in einem solchen Umfeld deutlich erschwert werden. Diese Umstände sind insbesondere bei der vorgestellten Evaluation nach Scriven wichtig, da die Bedürfnisse der Zielgruppe eine zentrale Rolle einnehmen, während die Zieldefinition der durchführenden NRO keine oder nur eine untergeordnete Rolle einnehmen soll.

Evaluationen bei Brot für die Welt verfolgen in erster Linie die Ziele Lernen und Legitimieren. Die Verantwortung für eine Planung, Beauftragung und Durchführung liegt direkt bei den lokalen Partnern, die bei Bedarf von Mitarbeitenden von Brot für die Welt beraten werden [11].

### 3.3 Nutzenorientierte Evaluation nach Fetterman

*Nutzenorientierte Evaluationen* bewerten die Nützlichkeit eines Vorgehens [39, S.146]. Sie bilden einen Gegenpol zur in Kapitel 3.1 beschriebenen Evaluationsforschung, die hauptsächlich an Programmtheorien und Erkenntnissen arbeitet. Ein bekannter Vertreter der nutzenorientierten Evaluation ist Fetterman, der 1993 seine Theorie der *Empowerment Evaluation* publizierte [12]. Fettermans Evaluationsphilosophie entlehnt sich der modernen Bewegungen des Empowerment<sup>7</sup>, das heißt in seinem Ansatz liegt der Fokus auf dem Lernen, der Entwicklung der von Maßnahmen Betroffenen und in der Konsequenz der Stärkung ebendieser [39, S.146]. Prinzipien wie zum Beispiel die Verbesserung von Programmumsetzungen, der Aufbau von Ownership der durch Maßnahmen betroffenen Menschen und eine Inklusion ebendieser Menschen in Planung, Umsetzung und Verstärkung der Projekte spiegeln dies wieder [13]. In seiner Grundform funktioniert der Evaluationsansatz in drei Schritten: Aktuelle Aufgaben definieren; Ihren gegenwärtigen Stand überprüfen; Für die Zukunft planen. Mit Hilfe einer methodischen Vielfalt, die qualitative und quantitative Methoden einsetzt, ist der kollaborative Ansatz fachgebietsübergreifend einsetzbar. Die zentrale Rolle des Evaluators wird durch einen *Critical Friend* [12] besetzt. Dieser sollte ein methodisches Wissen im Fachgebiet der Evaluation besitzen, die Aufgabe ist jedoch nicht die aktive Umsetzung, sondern vielmehr eine Begleitung der Organisation, die die oben genannten Schritte ausführt. Der Critical Friend soll in diesem Prozess konstruktives Feedback geben und sicherstellen, dass die Evaluation organisiert, gründlich und ehrlich durchgeführt wird. Um die Haltung, die hinter diesem Evaluationsansatz steht, noch stärker hervorzuheben verweist Fetterman selbst darauf, dass Evaluation keine Dienstleistung für die Auftraggebenden sein soll, sondern vielmehr in Anwaltschaft der Betroffenen agieren muss. Dieser Unterschied hat das Potential in

---

<sup>7</sup>Empowerment, oder Selbstbefähigung auf deutsch, beschreibt die Stärkung von Autonomie und Eigenmacht und wird im entwicklungspolitischen Kontext oft genutzt um Hilfestellungen zu beschreiben, die die Aneignung von Selbstbestimmung zum Ziel haben.

der realen Anwendung zu einem Interessenskonflikt zu führen, da Evaluatoren in der Regel sämtliche Kommunikation zu Auftrag, Auftragsrahmen und nicht zuletzt auch das Honorar mit dem Auftraggeber aushandeln. Der Ansatz Fettermans umgeht diesen potentiellen Interessenskonflikt jedoch, indem er Selbstevaluationen und interne Evaluationen nicht nur thematisiert, sondern empfiehlt [12]. Als Folge dieser Empfehlung wird jedoch das Selbstverständnis und die grundsätzliche Daseinsberechtigung eines Evaluators durch den vorliegenden Evaluationsansatz in Frage gestellt. Die Abgrenzung von bewertender Evaluation zum diagnostischen Charakter eines Assessments wird unklarer, weil der Evaluator seine Rolle als letztendlich bewertende Instanz verliert. Dasselbe Phänomen tritt ebenso in Bezug auf die Abgrenzung zu Organisationsentwicklung und allgemein dem Lernen in Gruppen auf [12].

**Verortung im Funktionsmodell Stockmann/Meyer.** Im Funktionsmodell nach Stockmann/Meyer ist Fetterman sehr nahe an der Entwicklung verortet. Diese Einordnung erscheint plausibel, aufgrund der Zentrierung von Lernen und Entwicklung der durch Maßnahmen Betroffene. Andere Ziele werden dagegen klar zurückgestellt, insbesondere Legitimation und Kontrolle, die durch Evaluation gewonnen oder ausgeübt werden können. Ein Erkenntnisgewinn ist durch die selbstzentrische Herangehensweise von Fettermans Ansatz nicht ausgeschlossen, durch die Abstriche bei der Professionalisierung ist jedoch zu erwarten, dass weniger wissenschaftlich basierte Erkenntnisse und dafür vermehrt anekdotisches Wissen erzielt wird.

**Distanzverhältnis des Evaluators.** Der vorliegende Ansatz nach Fetterman bildet die theoretische Grundlage der meisten heute genutzten internen oder Selbstevaluation. Der Grund dafür liegt darin, dass Fetterman in der grundsätzlichen Anlage seiner Theorie und des vorliegenden Ansatzes interne und Selbstevaluationen explizit mitgedacht und ermutigt hat. Andere Evaluationsansätze kennen diese nur als Sonderfälle, oft im Zusammenhang mit einem Mangel an finanziellen Ressourcen. Externe Evaluationen lassen sich unter dieser Theorie kaum oder nicht verorten.

**Ressourcenaufwand.** Eine Evaluation nach Fetterman ist, im Vergleich zu den oben genannten Beispielen, finanziell deutlich ressourcenschonender. Ein großer Faktor hierfür

ist der bewusste Verzicht auf einen externen Berater und damit einhergehende Honorarkosten. Die Kosten der Beauftragung von Mitarbeitenden kann zwar ebenfalls finanzielle Kosten im Sinne von Lohnkosten bedeuten, sobald die Arbeit von ehrenamtlich Mitarbeitenden übernommen wird fallen aber auch diese weg. Etwaige Reisekosten zum Besuch der aktuellen oder durchgeführten Maßnahme können anfallen, denkt man den Ansatz von Fetterman jedoch konsequent weiter ist eine Durchführung der Evaluation durch Partnerorganisationen oder die von einer Maßnahme Betroffenen im Sinne des Empowerments ebenfalls denkbar. Die Ressource Wissen muss hierbei für jeden Durchführenden ausgebaut werden, da es unwahrscheinlich ist, dass das Expertenwissen zu Evaluation, insbesondere in Bezug auf den Ansatz Empowerment Evaluation in ausreichender Menge vorhanden ist. Der Aufbau dieses Wissens kann zum Beispiel durch Selbststudium oder externe Fortbildungen erreicht werden, muss jedoch in den zeitlichen Ablauf der Evaluation eingeplant werden. Der Zeitaufwand der Evaluation selbst hängt zudem direkt von dem zuvor genannten Faktor des Personals ab, das heißt je mehr Aufgaben die Mitarbeitenden parallel intern erledigen müssen, desto mehr Zeit wird benötigt, um eine Evaluation abzuschließen.

**Aktuelle Anwendung.** TolaData ist eine Monitoring- und Evaluationssoftware, die speziell für die Arbeit von NROs entwickelt wurde. Mit dem Slogan „Simplify the way you monitor and evaluate your projects“ wird für eine einfache Art der Evaluation und ein kontinuierliches Monitoring geworben. Beide Bereiche werden nicht weiter voneinander getrennt, sondern zusammengehörig und ineinandergreifend behandelt. Die Plattform bietet die grundsätzlichen Funktionen Informationen zu sammeln, diese mithilfe eines Log-Frames zu analysieren und im letzten Schritt Ergebnisse grafisch zusammenzufassen. Das Ziel der Entwickler war es hierbei Organisationen zu unterstützen professioneller, effizienter und transparent zu sein.

Der Preis dieser online basierten Software liegt für NROs zwischen €49 und €469 monatlich, abhängig von der Anzahl der Nutzerlizenzen.

Methodisch beschränkt sich TolaData auf die Nutzung des Log-Frames, der den Weg von der Zielformulierung bis zum Output beschreibt. Die Indikatoren, die zur Messung von Veränderungen genutzt werden sollen, können als Teil des Log-Frames definiert werden. Die Auswahl der Indikatoren und die Bestimmung ihrer Aussagekraft wird hierbei

allein dem Nutzer überlassen. Gemessene Ergebnisse können in TolaData eingetragen und verwaltet werden. Aussagen wie

*„This [defining indicators] also helps to verify for example, if the output really brought about the changes in the next outcome level.“<sup>8</sup>*

zeigen eine starke Vereinfachung der realen Problemstellungen in der Evaluation. Betrachtungen von Kausalitätsproblemen werden durch diese Vereinfachung obsolet, was wiederum das Entstehen von inhaltlich schwach begründeten Wirkungsketten zur Folge haben kann und die Aussagekraft der Ergebnisse, die mit TolaData produziert werden, in Frage stellt. Des Weiteren wird der Nutzer bei wichtigen Entscheidungen wie der Auswahl der Messmethoden nicht an eine Methodik herangeführt, es gibt keine vorgeschlagenen Werkzeuge, um Daten zu erheben.

## **4 Fallstudie Ingenieure ohne Grenzen**

In der folgenden Fallstudie werden sowohl die Erwartungen von Mitarbeitenden einer kleinen NRO an Evaluation, als auch deren für Evaluation verfügbaren Ressourcen ermittelt. Der Verein IOG dient hierbei als Beispiel für kleine NROs, da er wie in Kapitel 1.2 beschrieben, typische Merkmale kleiner NROs aufweist.

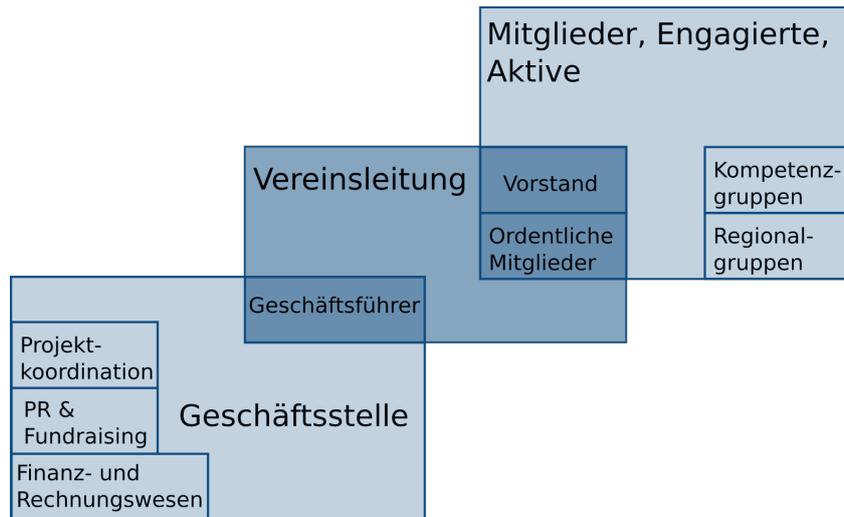
### **4.1 Einführung des Fallstudiengegenstands**

Der Verein IOG wurde bereits in Kapitel 1.2 vorgestellt. In Abbildung 5 ist ein Organigramm des Vereins dargestellt, das einzelne Gruppen des Vereins und ihre Schnittpunkte miteinander zeigt. Für die vorliegende Arbeit wurden darauf aufbauend in Zusammenarbeit mit einzelnen, nicht repräsentativen Mitarbeitenden von IOG die folgenden Gruppen definiert.

1. Projektbearbeitende mit Bezug zu Projektarbeit.
2. Projektbearbeitende mit Bezug zu Unterstützungsaufgaben (z.B. PR oder Fundraising).

---

<sup>8</sup>„Dies [die Definition von Indikatoren] hilft auch dabei, zum Beispiel zu überprüfen, ob der Output wirklich die Veränderungen auf der nächsten Ergebnisebene bewirkt hat“, übersetzt durch die Autorin



**Abbildung 5:** Organigramm der kleinen NRO Ingenieure ohne Grenzen e.V.

3. Mitarbeitende in Kompetenzgruppen.
4. Vorstand und ordentliche Mitglieder (OM), und
5. Mitarbeitende der Geschäftsstelle.

Der Verein IOG wurde für die vorliegende Arbeit ausgewählt, da er einen oder mehrere Ressourcenmängel aufweist, zum Beispiel in Bezug auf die Finanzierung. Dies geht entsprechend der Beschreibung kleiner NROs in Kapitel 1.1 auf ein nicht stetiges Einkommen zurück. Dieser Ressourcenmangel ist typisch für kleine NROs, weshalb IOG ein passendes Objekt für eine Fallstudie in der vorliegenden Arbeit darstellt. Weitere Ressourcenmängel wie zum Beispiel Personal oder Wissen können vorliegen, es muss jedoch im Rahmen der Fallstudie überprüft werden, inwiefern dies bei IOG der Fall ist. Weiterhin besteht die Vermutung, dass die Ergebnisse der folgenden Fallstudie durch die Wahl einer typischen kleinen NRO auch auf andere NROs übertragbar sein können. Diese Vermutung wird in Kapitel 6 näher diskutiert.

## 4.2 Aufbau, Methodik und Ziele der Fallstudie

Das Ziel der Fallstudie ist es verlässliche Daten darüber zu erheben welche Vorstellungen und Erwartungen Mitarbeitende von IOG an Evaluation haben und welche Ressourcen dem Verein zur Durchführung einer Evaluation zur Verfügung stehen. Diese Daten werden mittels eines zweistufigen Prozesses erhoben, der quantitative und qualitative Er-

hebungsmethoden nutzt, um eine möglichst aussagekräftige Datenbasis zu schaffen. Zur theoretischen Orientierung und Einordnung der Antworten dienen hierbei maßgeblich die in Kapitel 2.1 vorgestellten Ziele von Evaluation. Mithilfe der erhobenen Daten soll eine grundsätzliche Einordnung des Ansatzes erfolgen, entsprechend Kapitel 3. Mittels der zugrundeliegenden Theorie kann dann ein Evaluationsansatz und eine Methodik gewählt und entwickelt werden, die auf die Bedürfnisse und Erwartungen von IOG zugeschnitten ist.

**Fragestellung.** Die Erwartungen von Mitarbeitenden können verschieden sein, abhängig zum Beispiel von der Perspektive, die Mitarbeitende im Verein einnehmen, der Erfahrung die bereits in Entwicklungsprojekten gesammelt wurde oder auch der konkreten Arbeit, die von Befragten im Verein ausgeübt wird. Diverse Perspektiven sollen in der vorliegenden Fallstudie erfasst werden, um sie im Folgenden gegeneinander abzuwägen und konkurrierende Positionen, ebenso wie Gemeinsamkeiten, festzustellen.

Neben den Erwartungen und Wünschen, die Mitarbeitende an Evaluation haben können sollen auch Daten über die Bereitschaft des Einsatzes von Ressourcen gesammelt werden. Auf diesem Weg soll eine Vereinbarkeit von Erwartungen und realistischem Aufwand beibehalten werden, um den Kompromiss zwischen idealen Vorstellungen und real existierenden Hindernissen feststellen zu können. Ziel ist hierbei nicht die Festlegung eines Kompromisses, sondern das Aufzeigen von realistischen Erwartungen innerhalb der Grenzen, in denen IOG arbeitet.

**Methoden der Datenerhebung.** Die zur Beantwortung der oben beschriebenen Fragestellung notwendigen Daten bilden die unverzichtbare Basis, auf deren Grundlage Empfehlungen zur Evaluationsdurchführung gegeben werden können. Aus diesem Grund ist es von besonderer Bedeutung, dass die Erhebung dieser Daten wissenschaftlichen Standards entspricht um in der Folge so belastbar wie möglich zu sein. Die zur Erhebung der Daten genutzte Methodik muss mit Bedacht gewählt werden, um eine Erhebungsverzerrung möglichst gering zu halten. Gefahren einer Erhebungsverzerrung sind bei quantitative Methoden wie standardisierten Fragebögen zum Beispiel die Informationsselektion durch den Forscher und die künstliche Situation in der die Daten erhoben werden. Im Vergleich bieten qualitative Messverfahren, wie zum Beispiel Interviews, die Möglich-

keit, dass Teilnehmende Antworten und Informationen geben, die in der Konzeption der Datenerhebung nicht erwartet aber dennoch hilfreich und wichtig sind. Gleichzeitig kann auch hier eine Erhebungsverzerrung auftreten, zum Beispiel durch eine schlechte Performance des Interviewers und seiner Fragetechnik. Diese kann bspw. durch Suggestivfragen vorgefertigte Meinungen und Erwartungen in das Gespräch einbringen und zu Verzerrungen führen. Der Interpretationsspielraum bei der Auswertung ist zudem deutlich größer als bei qualitativen Datenerhebungen und -auswertungen. Die Messgenauigkeit in der Datenerhebung ist jedoch nicht an die Entscheidung zwischen qualitativen oder quantitativen Methoden gebunden, sondern vielmehr abhängig von der Vermeidung der oben beschriebenen Störfaktoren, die zu Erhebungsverzerrungen führen. Es gibt jedoch kein fehlerfreies Messinstrument in der realen Anwendung, das heißt jedes Verfahren ist mit spezifischen Fehlern versehen [39, S.210]. Aus diesem Grund werden zur Beantwortung der oben beschriebenen Fragestellung notwendigen Daten in einem zweistufigen Verfahren erhoben, durch eine Kombination und Integration von quantitativer und qualitativer Datenerhebung. Auf diesem Weg können die Vorteile der qualitativen Onlineumfrage, ebenso wie die Vorteile von Interviews genutzt werden. Zeitgleich können aber auch die Nachteile der jeweiligen Methoden ausgeglichen werden und etwaige Erhebungsverzerrungen können durch die verbleibende Methode aufgedeckt und entsprechende Daten geringer gewichtet werden. Diese Kombination und Integration von qualitativen und quantitativen Methoden geht zurück auf Schulz (1977) [31] und wird weiter entwickelt im Vertiefungsmodell nach Moschner und Anschütz [24], dessen konkrete Anwendung folgend beschrieben wird:

Im ersten Schritt werden aus der Vereins- und Arbeitsstruktur von Ingenieure ohne Grenzen e.V. Gruppen abgeleitet. Mittels einer quantitativen Umfrage werden Daten innerhalb einer großen Stichprobe gesammelt, wobei Teilnehmende aus allen Gruppen enthalten sein sollen. Das erfasste Datenmaterial wird quantitativ ausgewertet. In einem zweiten Schritt erfolgt eine qualitative Studie, mit deren Daten das vorliegende Material stichhaltig überprüft und interpretiert werden kann. Auf diesem Weg lassen sich Unterschiede in den erfassten Daten erkennen und Hinweise auf Erhebungsverzerrungen können aktiv adressiert werden. Die so erreichte Vertiefung der Ergebnisse bietet zudem eine große Wahrscheinlichkeit neben erwarteten auch unerwartete Effekte und Aussagen abzubilden oder zum Beispiel quantitativ erhaltene Ergebnisse mit Fallbeispielen zu untermauern [24].

### 4.3 Konzeption und Durchführung der Onlineumfrage

Der quantitative Anteil der Fallstudie wird mit Hilfe eines onlinebasierten Fragebogens durchgeführt. Diese Befragungsmethode erlaubt ein hohes Maß an Beteiligung, da die Einschätzungen und Perspektiven von nahezu beliebig vielen Teilnehmenden verarbeitbar sind. Gleichzeitig erlaubt die Methode jedoch durch ihren nicht-reaktiven Charakter nur ein Mindestmaß an Interaktion. Der daraus entstehende Vorteil ist eine reduzierte Gefahr von Kommunikationsstörungen und fehlerhafter Informationstransfers [39, S.207]. Gleichzeitig besteht der Nachteil, dass die vorformulierten Fragen Antworttendenzen provozieren und in der Folge zu einem verzerrten Datenbild führen können. Dieses Risiko wurde im Rahmen der Fallstudie durch Pre-Tests<sup>9</sup> reduziert, wird jedoch in Kapitel 4.7 diskutiert.

**Erstellung des Fragebogens und Pre-Test.** Entsprechend der Publikationen von Bortz und Döring wurden in der vorliegenden Fallstudie Fragen mit Antwortvorgaben offenen Fragen vorgezogen. Zur Formulierung der Fragen wird sich insbesondere auf die acht Anforderungen bezogen, die F. J. Fowler Jr. 2001 publizierte [14]. Demnach müssen Fragen so formuliert werden, dass sie valide sind, eine kognitive Grundlage bieten, die soziale Situation beachten, die Datenqualität sichern, in einer Umfrage handhabbar, variabel im Modus und in der Sprache und schließlich auch effizient sind. Der Fragebogen zur vorliegenden Fallstudie nutzt weiterhin hauptsächlich geschlossene Fragen. Dies bietet den Vorteil, dass den Antworten auf eine geschlossene Frage größere Zuverlässigkeit, also Reliabilität, zugeschrieben werden kann [9] [23, S.74]. Zusätzlich wird durch den einheitlichen Bezugsrahmen die Validität erhöht und das Antwortverhalten kann im Sinne der Forschungsfrage begrenzt werden. Auch auf formaler Ebene lassen sich geschlossene Fragen im Gegensatz zu offenen Fragen eindeutiger vergleichen und eine statistische Bearbeitung ist durch quantifizierbare Antwortmöglichkeiten möglich. Bei einer entsprechenden Stichprobenauswahl und Stichprobengröße können die Ergebnisse in der Folge als repräsentativ gelten. Ein Nachteil der geschlossenen Fragen ist die mögliche Verzerrung der Antworten. Dies kann zum Beispiel durch unpräzise Formulierungen in den Fra-

---

<sup>9</sup>Pre-Tests entstammen der empirischen Sozialforschung und werden genutzt um Erhebungsinstrumente zu testen. Das Ziel dieser Vortests ist die Qualität der Fragen und in der Folge die Qualität der erhobenen Daten zu verbessern.

gen, unbewusste Beeinflussung durch die Reihenfolge oder eine unpassende Wahl der Skalen passieren. Nicht vermeidbare Nachteile dieser Umfrageform sind die geminderte Spontaneität und der Verlust der eigenen Ausdrucksweise der Teilnehmenden, da Befragte keine Möglichkeit erhalten die Antworten in ihren eigenen Worten zu geben [23, S.75]. Im Hinblick auf die vorliegende Fallstudie wird ein besonderes Augenmerk auf die Formulierung der Antwortmöglichkeiten gelegt, um vermeidbare Nachteile auszuschließen. Die Ausgestaltung des Fragebogens umfasst die Formulierung der Fragen. Diese werden abhängig von den Daten die mit ihnen erzeugt werden unterteilt. Man spricht in diesem Zusammenhang auch von ihrem Skalenniveau.

Eine Frage mit der nur die Verschiedenheit von Objekten abgebildet wird bewegt sich zum Beispiel auf Nominalniveau, es handelt sich um Auswahlfragen. Werden neben der Verschiedenheit auch Hierarchien zwischen den Objekten abgefragt handelt es sich um eine Rangordnungsfrage auf Ordinalniveau. Werden mit der Beantwortung der Frage Distanzen zwischen den Objekten abgebildet handelt es sich um Intensitätsfragen auf Intervallniveau, bzw. metrischem Niveau [23, S.79f].

Der erstellte Fragebogen wird informell mit drei Test-Teilnehmenden besprochen, um das Risiko einer fehlschlagenden Umfrage zu minimieren und die Qualität der Fragen zu verbessern. Ein Fokus dieser Prüfung liegt dabei auf der Verständlichkeit, potentiellen Missverständnissen und der Suggestion von Antworten. Nach dem Pre-Test wurden minimale Veränderungen in der Formulierung vorgenommen.

**Die Fragen.** Im ersten Teil des Fragebogens werden Erfahrungen mit Evaluation und Entwicklungszusammenarbeit abgefragt. Durch die sehr diverse Mitgliederschaft des Vereins IOG liegt die Vermutung nahe, dass die Erfahrungen und das Vorwissen der Mitglieder ebenfalls sehr divers sind. Fragen zu Erfahrung, Interesse und eine Einschätzung der Wichtigkeit von Evaluation sollen daher in der Auswertung eine Einordnung der weiteren Antworten möglich machen.

Im zweiten Teil des Fragebogens werden Daten und Aussagen über die Grundlagen von Evaluationsansätzen gesammelt. Wie in Kapitel 2.3 beschrieben, lassen sich diese Ansätze mithilfe von fünf Entscheidungen gut voneinander abgrenzen. Diese Abgrenzung soll in der vorliegenden Arbeit eine Basis darstellen, auf der eine spezifische Empfehlung für IOG entwickelt wird. Es ist hierbei offen, ob ein bereits bestehender und verfügba-

rer Ansatz genutzt wird, oder eine Kombination von verschiedenen Ansätzen entwickelt werden muss. Die Fragen zur Eingrenzung des Ansatzes beziehen sich auf den Evaluationsgegenstand, die Ziele und Aufgaben von Evaluation, die Kriterien anhand derer eine Bewertung erfolgen soll und die Frage wer evaluieren soll.

Im dritten Teil des Fragebogens erfolgt eine allgemeine Stichprobenbeschreibung. Diese fragt Informationen wie das Geschlecht und die Dauer der Aktivität im Verein ab und dient der Einordnung der gegebenen Antworten. Für die vorliegende Arbeit sind diese Daten von eher untergeordnetem Interesse, dennoch wird in der Auswertung überprüft, ob sich aufgeschlüsselt nach diesen Kenndaten Besonderheiten abzeichnen. Vor der Umfrage wurden mit der Beratung von Teilnehmenden der Kompetenzgruppe Qualitätsmanagement Gruppen definiert, in denen sich die Teilnehmenden selbst verorten sollten, um die Ergebnisse der Onlinebefragung mit den Ergebnissen der Interviews vergleichen zu können. Die Teilnehmenden ordnen sich genau einer Gruppe zu.

Untenstehend ist eine Übersicht über alle Fragen der Onlineumfrage angefügt, für eine Übersicht aller Fragen und Antwortmöglichkeiten vgl. A.3.

1. Wie viel Erfahrung hast Du bereits in der Entwicklungszusammenarbeit gesammelt?
2. Wie viel Erfahrung hast Du mit Evaluationen gesammelt?
3. Wie viel Interesse hast Du persönlich an Evaluation?
4. Welche Wichtigkeit hat Evaluation Deiner Meinung nach für Entwicklungszusammenarbeit?
5. Welche Wichtigkeit soll Evaluation Deiner Meinung nach für die Arbeit von IOG haben?.
6. Was genau soll Deiner Meinung nach bei IOG evaluiert werden?
7. Mit welchem Ziel sollen Projekte bei IOG evaluiert werden?
8. Welche Schwerpunkt in der Evaluation eines Projektes sind Deiner Meinung nach die Wichtigsten?
9. Evaluationen bei IOG sollen. . .

10. Wie viel Budget darf Deiner Meinung nach für Evaluation verwendet werden?
11. Welches Geschlecht hast Du?
12. In welcher Gruppe engagierst Du dich hauptsächlich bei IOG?
13. Wie lange bist Du bereits aktiv bei IOG?

**Durchführung.** Für die beschriebene Onlineumfrage wird der Fragebogen mithilfe der kostenpflichtigen, onlinebasierten Weblösung survio.com über einen Emailverteiler an Mitglieder von IOG verschickt. Survio wurde ausgewählt, da es ein paralleles Ausfüllen von mehreren Teilnehmenden ermöglicht und Rohdaten zur Auswertung zugänglich gemacht werden. Insbesondere der zweite Punkt ist ein Unterschied zu anderen verbreiteten Weblösungen, da diese zum Teil nur ausgewertete Ergebnisse präsentieren. Es wird erhofft eine ausreichend hohe Teilnehmeranzahl für die vorliegende Umfrage zu erhalten und somit die Aussagen und Perspektiven möglichst vieler Mitarbeitender von IOG aufnehmen zu können. Nach zehn Tagen wird mittels einer Erinnerungs-Email erneut auf den Fragebogen aufmerksam gemacht. Die Auswahl der Stichprobe ist daher nicht-probabilistisch, genauer: die Auswahl hat Anteile einer theoretischen Stichprobe, weil die Gruppe gezielt ausgewählt wurde nach dem Merkmal des Engagements bei IOG. Gleichzeitig hat die Auswahl Anteile einer ad-hoc Stichprobe, da die Umfrageteilnehmenden die sind, die gerade zur Verfügung standen und nicht weiter gefiltert werden [47].

#### **4.4 Konzeption und Durchführung der Interviews**

Nach der Durchführung einer Onlineumfrage wie der oben beschriebenen ist nach dem Vertiefungsmodell nach Moschner und Anschutz [24] ein qualitativer Schritt hilfreich, um die erhobenen Daten zu überprüfen und eine Intensivierung der festgestellten Ergebnisse zu erreichen. Dieser qualitative Schritt wird in der vorliegenden Arbeit durch Interviews von Einzelpersonen vorgenommen. Zur Vergleichbarkeit der Interviews untereinander wird ein Leitfaden erstellt, der die Gespräche semi-strukturiert.

**Erstellung des Leitfadens und Pre-test.** Der Leitfaden selbst umfasst dabei alle Themenbereiche des Onlinefragebogens, die darin bewusst geschlossen formulierten Fragen

werden offener formuliert, um den Gesprächsfluss möglichst natürlich zu erhalten. Dieser ist im Anhang einsehbar A.1.

Die Logik der aufeinanderfolgenden Fragekategorien wird beibehalten, es sind Fragen enthalten die insbesondere die persönlichen Erfahrungen mit Entwicklungszusammenarbeit und Evaluation erfragen. Des Weiteren ist ein fachlicher Teil enthalten. In diesem werden erneut die fünf Charakteristika von Evaluationsansätzen abgefragt, die in Kapitel 2.3 eingeführt wurden. An dieser Stelle wird die tatsächliche Vertiefung der in der Onlineumfrage festgestellten Tendenzen und Ergebnisse erwartet, weil für Teilnehmende die Möglichkeit besteht Überlegungen auszudrücken die sich hinter den Antworten verbergen, die im Onlinefragebogen gegeben wurden. Dieser vertiefende Teil nimmt den größten Raum im Leitfaden ein, im Interview wird die Erzählung der Interviewten zudem mit Vertiefungsfragen und Aufrechterhaltungsfragen ermuntert. Zuletzt wird die Stichprobe mit Hilfe von zwei geschlossenen Fragen in Bezug zu IOG eingeordnet. Die Pre-Tests wurden analog zum Pre-Test der Onlineumfrage durchgeführt, vgl. 4.3.

**Durchführung.** Die Mitarbeitenden von IOG werden per Email gebeten sich für ein Interview bereitzuerklären. Die persönlichen Ansprachen, Emailverteiler und Listen sind hierbei dieselben wie für die Onlineumfrage. Rückmeldungen der einzelnen Interessenten kommen per Email zurück und Termine werden nach angegebenen Präferenzen der Interessenten vergeben. In Antwort auf ihre jeweilige Bereitschaftserklärung werden Termine bestätigt und Einwahldaten in ein Onlinemeeting geteilt. Sechs der zehn Interviewten haben sich dafür entschieden das Onlinemeeting zu nutzen, vier baten darum das Interview per Telefon durchführen zu können. Diese unvorhergesehene Besonderheit ist problematisch, da ein Interview am Telefon nicht mit einem Interview in einem Onlinemeeting, insbesondere einem mit Bildübertragung, gleichzusetzen ist. Ein Interview mit Bildübertragung ist deutlich reaktiver, das heißt Interviewer und Befragte kommunizieren neben der Tonspur auch nicht-verbal über Gesten und Mimik. Interviews mit und ohne Bildübertragung sind daher nur eingeschränkt miteinander vergleichbar. Um diese Vergleichbarkeit wieder herzustellen wurden die Onlinemeetings ohne Bildübertragung durchgeführt, alle Interviews wurden als mp3-Dateien aufgezeichnet.

Bei der Fragestellung wird darauf geachtet die im Leitfaden formulierten offenen Fragen möglichst natürlich in den Gesprächsverlauf einzubringen. Zudem wird auf Wertun-

gen verzichtet und Suggestionen von möglichen Antworten werden vermieden. Der oben beschriebene Leitfaden wird in allen Interviews angewandt und nur dann verändert, falls ein organischer Gesprächsfluss durch ein Beharren auf der Struktur unterbrochen würde. Angestrebt wird daher eine möglichst fließende Abfolge der Fragethemen in einem natürlichen Gespräch, auch wenn ein semi-strukturiertes Interview als reaktives Verfahren das Risiko von Kommunikationsfehlern birgt. Diese können Missverständnisse zwischen Interviewendem und Interviewtem oder Fehlinterpretationen des Interviewenden umfassen, die in der Folge die Ergebnisse verfälschen können [39, S. 210].

Zudem sollen die Gesprächspartner zu Erzählungen ermuntert werden, die über die bloße Fragestellung hinausgehen. Im Sinne des Überprüfens von gefundenen Ergebnissen und Vertiefens von unklaren Aussagen werden die Fragen des Leitfadens mit Vertiefungsfragen nach Helfferich ergänzt. Weiterhin werden Aufrechterhaltungsfragen genutzt um Impulse für assoziative Gedanken zu geben, ohne ein neues Thema vorzugeben [17, S. 104].

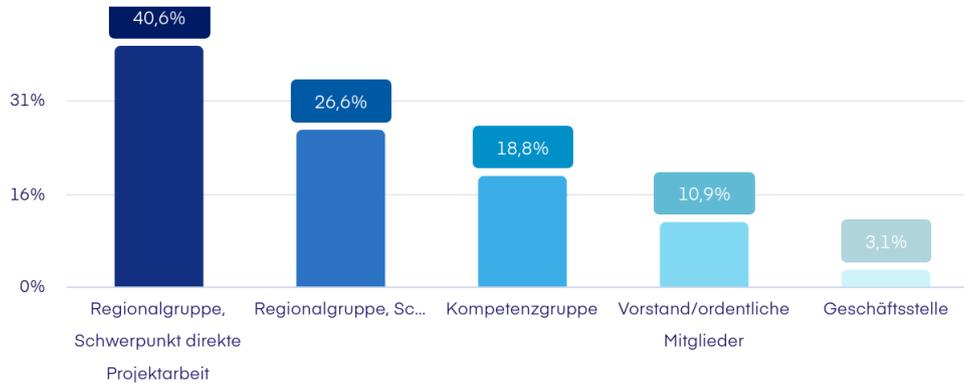
## **4.5 Auswertung der Fallstudie**

### **4.5.1 Quantitative Auswertung der Onlineumfrage**

Der Onlinefragebogen wurde von 65 Menschen vollständig ausgefüllt, kein Teilnehmender hat den Fragebogen unvollständig abgebrochen. Das Ausfüllen des Fragebogens hat für die meisten Teilnehmenden zwischen 5 und 33 Minuten gedauert. Die Antworten der Befragten werden im Folgenden nicht losgelöst voneinander ausgewertet, sondern im Kontext des Sinnzusammenhangs diskutiert. Zur Auswertung von Hierarchiefragen werden Mittelwert, Median und Standardabweichung der Antworten berechnet.

**Einordnung der Stichprobe.** An der Umfrage haben 65 Menschen teilgenommen, davon waren 57% männlich, 38% weiblich und 5% haben Anderes oder Keine Angabe gewählt. Eine Mehrheit von 59% der Befragten war bereits drei Jahre oder länger bei IOG aktiv, während eine Minderheit von 5% weniger als 6 Monate aktiv involviert war, vgl. 6. Die Zeiträume wurden in den Antwortmöglichkeiten dieser Frage vergleichsweise kurz gewählt, weil bereits bekannt war, dass IOG eine Fluktuation von ehrenamtlich Mitarbeitenden hat. Die vertretene Stichprobe zeigt diese Besonderheit des Vereins nicht auf, was

darauf hinweisen kann, dass die Aufforderung zur Teilnahme am Fragebogen nicht alle Gruppen ausreichend eingeschlossen hat.

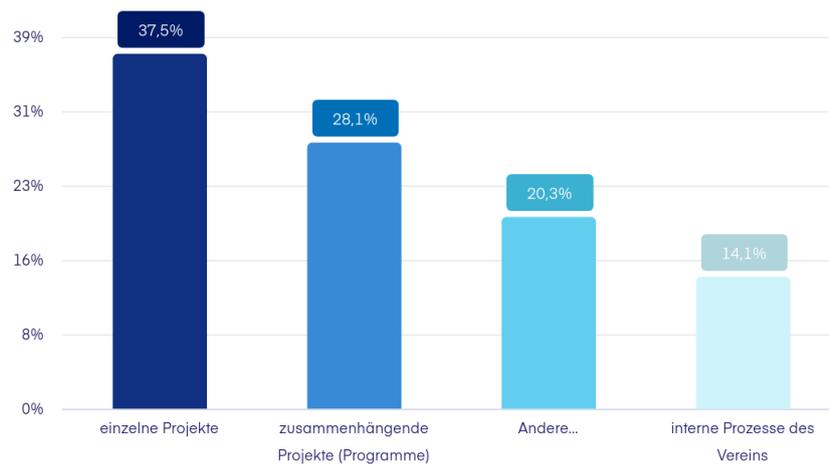


**Abbildung 6:** Die Verteilung der Antworten bei Frage 12 entspricht der üblichen Verteilung bei IOG. Die Umfrage bildet daher die Perspektive jeder Gruppe ab, entsprechend ihrer relativen Größe im Verein.

Im Ergebnis ist festzustellen, dass die Gruppen relativ zu ihrer Größe repräsentativ in der Umfrage vertreten sind. Die Mitarbeitenden der Regionalgruppe mit Schwerpunkt auf direkter Projektarbeit bilden mit 43% die größte Gruppe, gefolgt von Mitarbeitenden der Regionalgruppe mit Schwerpunkt auf unterstützenden Arbeiten. An dritter Stelle folgen die Mitarbeitenden der Kompetenzgruppen mit 19%, gefolgt von Mitarbeitenden des Vorstandes, bzw. ordentlichen Mitgliedern des Vereins mit 10%. Die kleinste Anzahl von Teilnehmenden im Fragebogen sind mit 3% Mitarbeitende der Geschäftsstelle, diese bilden auch im Verein allgemein die zahlenmäßig kleinste Gruppe.

**Evaluationsgegenstand.** Mit einer Mehrheit von 38% sprechen sich Mitarbeitende von IOG dafür aus, dass einzelne Programme evaluiert werden sollten. An zweiter Stelle wurden mit 27% Programme gewählt, während bereits an dritter Stelle mit 21% Anderes gewählt wurde, vgl. 7.

Andere ist hierbei im Fragebogen als Freitextfeld eingefügt, sodass Teilnehmende ihre Erwartungen in eigenen Worten wiedergeben können. Diese Möglichkeit wurde von 13 Menschen wahrgenommen. Von diesen Menschen sprechen sich neun für eine Evaluation aller vorgeschlagenen Bereiche aus. Diese Antwortmöglichkeit wurde in der Erarbeitung des Fragebogens bewusst nicht vorgeschlagen, um eine überhöhte Erwartung an Evaluati-



**Abbildung 7:** Die Verteilung der Antworten bei Frage 6 „Was genau soll deiner Meinung nach bei IOG evaluiert werden?“ zeigt, dass eine Mehrheit der Befragten einzelne Projekte evaluieren würde.

on nicht zu unterstützen und Teilnehmende zu einer Reflexion über Prioritäten anzuregen. Zwei Teilnehmende sprechen sich in eigenen Worten für eine Evaluation interner Strukturen aus und spezifizieren diese weiter. Zwei weitere Antworten sprechen sich sowohl für die Evaluation von Projekten als auch für die Evaluation von Programmen aus.

Zusammenfassend ist festzustellen, dass sich eine Mehrheit dafür ausspricht einzelne Projekte zu evaluieren.

**Ziele einer Evaluation.** Die Ziele und Aufgaben einer Evaluation wurden in Frage 7 mittels einer Hierarchiefrage abgefragt. Das heißt Teilnehmende sind aufgefordert die Antwortmöglichkeiten in eine Reihenfolge zu bringen, von wichtigstem Ziel zu unwichtigstem. Die Ursache dieser Entscheidung liegt darin, dass die Ziele zueinander nicht eindeutig abzugrenzen sind und eine Evaluation direkt oder indirekt in den meisten Fällen mehr als ein Ziel erreichen wird. Dennoch ist es unabdingbar eine primäre Richtung zu definieren, da eine Evaluation sämtlicher Bestandteile und Schwerpunkte der Projekte im Kontext kleiner ressourcenarmer NROs nicht realistisch umzusetzen ist. Die vorgegebenen Antwortmöglichkeiten orientieren sich an den Evaluationszielen, die in Kapitel 2.1 beschrieben wurden, das heißt sie umfassen Legitimation, Entwicklung, Erkenntnis und Kontrolle.

Zur Quantifizierung der Antworten einer Hierarchiefragen wird der Rang der Antwort-

ten berechnet. Der Rang  $rg_A^{x_i}$  einer Antwortmöglichkeit  $A$  eines Teilnehmenden  $x_i$  ist ein ganzzahliger Wert, der sich aus der Reihenfolge (Anordnung) der gegebenen Antwortmöglichkeiten ergibt. Gibt es  $n$  verschiedene Antwortmöglichkeiten zu einer Hierarchiefrage, so erhält die zuerstgenannte (die wichtigste) Antwort  $A_1$  den Rang  $n$ , d.h.,  $rg_{A_1}^{x_i} = n$ . Der nächsten Antwort wird Rang  $n - 1$  zugewiesen, usw. Die zuletztgenannte Antwort erhält Rang 1. Bei Frage 7 gibt es zum Beispiel vier verschiedene Antwortmöglichkeiten, sodass der zuerstgenannten Antwort der Rang 4 zugewiesen wird. Intuitiv repräsentiert der Rang einer Antwort ihre Wichtigkeit für den Teilnehmenden (je höher desto wichtiger).

Der mittlere Rang  $\bar{r}g_A$  einer Antwortmöglichkeit ergibt sich aus dem arithmetischen Mittel der jeweiligen Ränge aus den Umfrageantworten  $x_1, \dots, x_n$  der verschiedenen Teilnehmenden, siehe Formel (1). Analog ist der mediane Rang  $\tilde{r}g_A$  als der Median der jeweiligen Ränge definiert. Ist der Bezug zur Antwort aus dem Kontext ersichtlich wird auch  $\bar{r}g$  bzw.  $\tilde{r}g$  statt  $\bar{r}g_A$  bzw.  $\tilde{r}g_A$  verwendet. Die Standardabweichung der Ränge  $\sigma$  ist wie üblich definiert.

$$\bar{r}g_A = \frac{1}{n} \sum_{i=1}^n rg_A^{x_i} \quad (1)$$

Der mittlere Rang, der mediane Rang und die Standardabweichung der Antwortmöglichkeiten wurden für Frage 7 berechnet, vgl. Tabelle 1. Auf dieser Basis sind die Antworten der Teilnehmenden miteinander vergleichbar. Im Ergebnis dieser Berechnung ist abzulesen, dass die Ziele Kontrolle und Legitimation vergleichbar oft als wichtige Ziele eingeordnet wurden, die mittleren Ränge liegen bei 2,9 (Legitimation) und 2,8 (Kontrolle). Die medianen Ränge liegen ähnlich nah beieinander, sodass diese zwei Ziele als gleichwertig für die Teilnehmenden angesehen werden.

Das Ziel Entwicklung folgt an dritter Stelle mit einem mittleren Rang von 2,5. Das Ziel Erkenntnis folgt mit deutlichem Abstand auf Platz vier, der mittlere Rang beträgt 1,8. Hier tritt die stärkste Abweichung zwischen medianem und mittlerem Rang auf. Mit Hilfe der Standardabweichung lässt sich daher schließen, dass es wenige aber starke Ausreißer in den Antworten gab. Zusammenfassend ist festzustellen, dass sich die Mehrheit der Teilnehmenden dafür ausspricht, dass eine Evaluation im gleichen Maße das Ziel der Legitimation und das Ziel der Kontrolle verfolgen soll. Im Bezug auf das Ziel Erkenntnis ist eine Polarisierung der Antworten erkennbar.

Antwort	Mittl. Rang $\bar{r}_g$	Med. Rang $\tilde{r}_g$	Standardabw. $\sigma$
Erkenntnis	2,49	3	1,06
Kontrolle	2,83	3	1,04
Legitimation	2,92	3	1,05
Entwicklung	1,76	1	0,96

**Tabelle 1:** Quantitative Auswertung der Hierarchiefrage 7: „Mit welchem Ziel sollen Projekte bei IOG evaluiert werden?“.

**Kriterien der Bewertung.** Ein wichtiger Faktor zur Konkretisierung eines Evaluationsvorhabens ist die Definition von Kriterien der Bewertung. Mittels einer weiteren Hierarchiefrage, Frage 8, haben Teilnehmende aus in der Entwicklungszusammenarbeit üblichen Bewertungskriterien ausgewählt. Die Grundlage der Antwortmöglichkeiten waren hierbei die DAC-Kriterien, ergänzt um Kriterien wie Partizipation und Gender- und Diversitysensibilität. Es handelt sich hierbei um übliche Bewertungskriterien in der Entwicklungszusammenarbeit, die zum Beispiel auch von den in Kapitel 2.1 vorgestellten Evaluationsansätzen genutzt werden. Die Berechnung erfolgt analog zu Frage 7. Das Kriterium mit dem höchsten mittleren Rang ist ein reibungsloser Projektablauf mit 6,9, dicht gefolgt von einem guten Kosten-Nutzen-Verhältnis mit 6,7. An dritter Stelle folgt sehr nah das Kriterium der gender- und diversitätsensible Planung mit 6,6. Alle weiteren Antwortmöglichkeiten folgen mit einem auffälligen Abstand von 0,8 und die Standardabweichung der drei oben genannten Kriterien ist vergleichsweise klein, vgl. Tabelle 2. Ein Beispiel aus dem unteren Bereich ist die Antwortmöglichkeit Nachhaltigkeit, die eine vergleichsweise kleine Standardabweichung aufzeigt. Gleichzeitig sind sowohl der mediane als auch der mittlere Rang gering, worin sich zeigt, dass die Teilnehmenden im Konsens dafür gestimmt haben, dass Nachhaltigkeit kein Bewertungskriterium für Projekte von IOG sein soll.

Im oberen Bereich zeigt der reibungslose Projektablauf einen Median von 8, dieser zeigt dass es Ausreißer gab. Da die Standardabweichung weiterhin gering ist lässt sich schließen, dass es sich um wenige Ausreißer handelt.

Zusammenfassend ist festzustellen, dass sich eine Mehrheit dafür ausspricht, dass insbesondere die Kriterien des reibungslosen Projektablaufs, des Kosten-Nutzen-

Antwort	Mittl. Rang $\bar{r}_g$	Med. Rang $\tilde{r}_g$	Standardabw. $\sigma$
Gender	6,56	7	2,38
Partizipation	4,43	4	2,38
Nachhaltig	2,76	2	1,97
Ziele	3,56	4	2,19
Ressourcen	5,14	5	1,75
Ziele 2	3,29	3	2,16
Ziele 3	5,68	6	2,21
Prozesse	6,89	8	1,98
Kosten/Nutzen	6,70	7	2,25

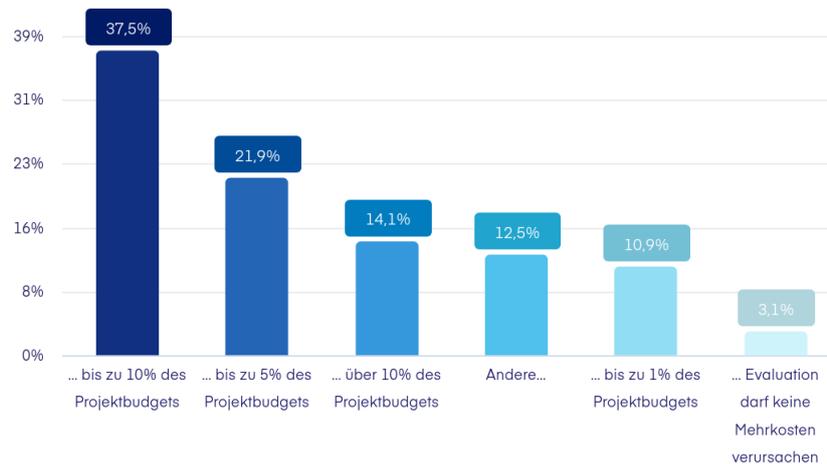
**Tabelle 2:** Quantitative Auswertung der Hierarchiefrage 8: „Welche Schwerpunkt in der Evaluation eines Projektes sind Deiner Meinung nach die Wichtigsten?“.

Verhältnisses und der gender- und diversitätsensiblen Planung zur Bewertung von Projekten genutzt werden.

**Evaluator.** Zur Erstellung einer Evaluation ist es notwendig zu wissen wer evaluieren soll. Hierbei kann es sich um einen externen Evaluator, einen internen oder eine Selbstevaluation handeln (vgl. Kapitel 2.1). Diese Frage ist nicht konkret in den Fragebogen eingegangen. Der Grund für diese Entscheidung liegt darin, dass eine Antwort der Teilnehmenden allein nicht aussagekräftig ist. Insbesondere die Überlegungen, die zu der Entscheidung geführt haben, müssen hierbei betrachtet werden um sicherzustellen, dass die Antwort aus einer Abwägung der Vor- und Nachteile der verschiedenen Distanzverhältnisse resultiert. Diese Abwägungen können jedoch nicht pauschal erwartet werden, da die Stichprobe der Umfrage ihre Erfahrung mit Evaluation zu 55% als keine oder wenig beschreibt, vgl. Frage 2. Um diese Schwierigkeit in der Fragestellung zu umgehen wurden stattdessen mehrere limitierende Ressourcen, die zur Entscheidungsfindung beitragen, abgefragt. In Bezug auf die Ressource Wissen gaben 55% der Teilnehmenden an keine oder wenig Erfahrung mit Evaluation zu haben, 38% gaben eine mittlere Erfahrung mit Evaluation an. Dieses vorhandene Wissen im Verein kann genutzt werden, um intern oder selbst zu evaluieren, zeigt jedoch auch Risiken auf.

Aufgrund methodischer oder sonstiger Fehler in der Evaluation ist es möglich, dass Daten erhoben werden die nicht belastbar sind, oder Ergebnisse formuliert werden die

sich nicht ausreichend auf erhobene Daten stützen. Die Akzeptanz einer intern oder selbst durchgeführten Evaluation kann darunter leiden, sollten die damit Beauftragten zu wenig methodisches oder allgemeines Wissen über Evaluation haben. Eine Umsetzung der Empfehlungen aus der Evaluation kann dadurch erschwert oder verhindert werden.



**Abbildung 8:** Die Verteilung der Antworten bei Frage 10 suggeriert, dass eine andere Skala eine andere Verteilung nach sich gezogen hätte. Zur tieferen Klärung der inhaltlichen Frage wird in den Interviews ein Fokus darauf gelegt.

In Bezug auf die Ressource Geld wurde erfragt wieviel Budget, bezogen auf das Budget eines spezifischen Projektes, für Evaluation aufgebracht werden sollte. Mit einer Mehrheit von 38% gaben Teilnehmende an, dass bis zu 10% des Projektbudgets für eine Evaluation verwendet werden dürften. Es ist auffällig, dass mit 15% ein signifikanter Anteil der Teilnehmenden auch mehr als 10% des Projektbudgets aufwenden würden. Der Schluss liegt nahe, dass die Skala der Antwortmöglichkeiten unpassend gewählt wurde, weil die Verteilung der Antworten über den Antwortmöglichkeiten nicht Normalverteilt ausfällt, sondern sich 52% der Antworten im Bereich der höheren Budgets akkumulieren. Die Aussagekraft dieser Antworten muss daher mehr als Tendenz gewertet werden und weniger als eindeutige Antwort im Sinne der Antwortmöglichkeit „bis zu 10% des Projektbudgets“. Aus diesem Grund ist es insbesondere wichtig diesen Punkt innerhalb der Interviews tiefergehend zu erfragen.

Zusammenfassend ist festzustellen, dass eher wenig Wissen zur selbstständigen Durchführung einer Evaluation vorhanden ist. Gleichzeitig besteht tendentiell eine Bereitschaft 10% des Projektbudgets oder mehr für die Evaluation desselben aufzubringen,

was eine externe oder teil-externe Evaluation ermöglichen würde.

**Ergebnis.** Das Ergebnis der Onlineumfrage bei IOG zeigt, dass die Mehrheit der Mitarbeitenden von IOG die Evaluationsziele Legitimation und Kontrolle verfolgen möchte. Die Kriterien der Bewertung sollen insbesondere der reibungslose Projektablauf, das Kosten-Nutzen-Verhältnis und eine gender- und diversitätsensible Planung sein. Die Rolle des Evaluators konnte innerhalb der Onlineumfrage nicht abschließend geklärt werden. Die Ziele Legitimation und Kontrolle legen eine externe oder teil-externe Evaluation nahe, da diese unabhängiger Ergebnisse produzieren kann als eine interne oder Selbstevaluation. Durch die Bereitschaft 10% des Projektbudgets oder mehr finanzielle Ressourcen für Evaluation zu mobilisieren erscheint dies auch möglich, es überrascht jedoch diese Position von einer ressourcenarmen NRO wie IOG festzustellen. Es wird daher eine Vertiefung innerhalb der Interviews angestrebt um zu überprüfen, inwieweit die Bereitschaft finanzielle Ressourcen zu investieren tatsächlich gegeben ist und in welchem Maße.

#### **4.5.2 Qualitative Auswertung der Interviews**

Um belastbare Aussagen erzeugen zu können müssen qualitative Inhaltsanalysen den drei Gütekriterien der qualitativen Forschung nach Steinke [36] entsprechen: Transparenz, Reichweite, Intersubjektivität. Die Interviews der vorliegenden Fallstudie werden untenstehend qualitativ ausgewertet. Die Fallstudie entspricht dabei dem Kriterium der Transparenz, da eindeutig nachzuvollziehen ist wie viele Menschen aus welchen Gruppen des Vereins IOG befragt wurden. Es ist ebenfalls eindeutig nachzuvollziehen welche Fragen gestellt wurden und wie diese festgelegt wurden, vgl. A.1. Sie entspricht dem Kriterium der Reichweite, da für alle Interviews derselbe Leitfaden genutzt wurde und die Gespräche damit untereinander vergleichbar sind. Bei einer erneuten Durchführung ist daher davon auszugehen, dass die Ergebnisse reproduzierbar sind. Sie entspricht dem Kriterium der Intersubjektivität zum Teil, da die Durchführende alle Gespräche als Außenstehende mit den Teilnehmenden durchführt und somit keine gefestigte Meinung zu Fragen bezüglich des Vereins IOG besitzt. Im Rahmen der vorliegenden Masterarbeit ist jedoch keine zweite Person hinzugezogen worden um dieselben Inhaltsanalysen mit gleichem, bzw. ähnlichem Ergebnis durchzuführen. Eine solche Prüfung der gesammelten Ergebnisse kann bei weiteren Arbeiten zu diesem Thema das Kriterium der Intersubjektivität

vität deutlich stärken und wird in Kapitel 6 weiterführend diskutiert.

Das im Anhang A.1 gezeigte Metaskript gibt eine Übersicht über alle Interviews und deren grobe Inhalte. Das Metaskript enthält Notizen, Zitate und Hinweise zu den Interviews, diese wurden jedoch nicht systematisch transkribiert und werden in der Folge auch nicht zu Auswertungszwecken herangezogen. Zitate aus den Interviews werden zum Teil in der Auswertung genannt und als solche gekennzeichnet. Angegebene Zeitmarken ermöglichen ein leichtes nachvollziehen der Interviews, die dieser Arbeit als mp3-Dateien beiliegen.

**Einordnung der Stichprobe.** Es wurden zehn Interviews durchgeführt, deren Dauer zwischen 33 und 53 Minuten lag. Drei Teilnehmende sind unter drei Jahren bei IOG aktiv, zwei Teilnehmende sind 4 Jahre bei IOG aktiv und fünf Teilnehmende sind über 7 Jahre bei IOG aktiv. Diese Verteilung ist nicht repräsentativ für den Verein, der zu großen Teilen aus Studierenden besteht und von hoher Fluktuation geprägt ist. Drei der zehn Teilnehmenden haben Erfahrung in Organisationen der Entwicklungszusammenarbeit außerhalb von IOG. Diese Erfahrungen wurden in der Regel bei größeren Organisationen gemacht, zum Beispiel der GIZ oder Brot für die Welt.

Teilnehmende wurden, ebenso wie in der Onlineumfrage, aufgefordert sich in Gruppen einzuordnen, vgl. 4.3. Im Gegensatz zur Onlineumfrage gab es im Interview die Möglichkeit einen Werdegang zu beschreiben um ein tieferes Verständnis für die Perspektiven der Teilnehmenden zu erlangen. Fünf Teilnehmende beschrieben daraufhin ihre Zugehörigkeit zu mehreren Gruppen. Auch wenn es möglich ist sich zu jedem Zeitpunkt eindeutig zuzuordnen beziehen sich viele Erfahrungen auf Rollen die man früher im Verein übernommen hat. Eine Negierung dieser Entwicklungen im Sinne der Vereinfachung verfälscht Schlüsse, die sich aus den Aussagen ziehen lassen, weshalb die Abfrage im Interview offener gestaltet wurde.

Zwei Teilnehmende arbeiten in der Geschäftsstelle in Berlin und ein Teilnehmender ist ordentliches Mitglied im Verein. Des Weiteren wurden folgende Rollen angegeben: Vier Teilnehmende sind oder waren mit Bezug zu Unterstützungsaufgaben betraut, drei Teilnehmende sind oder waren aktiv in die Projektumsetzung involviert und sechs der Teilnehmenden sind aktiv in Kompetenzgruppen. Die Vielfalt der Rollen der einzelnen Teilnehmenden resultiert aus der oben beschriebenen langen Aktivität im Verein. Alle

sechs Teilnehmenden die heute in Kompetenzgruppen aktiv sind, haben vorhergehend andere Rollen ausgefüllt und darin die Erfahrung gesammelt, die notwendig ist um sich aktiv in die Prozesse einer Kompetenzgruppe einbringen zu können. Es wäre wünschenswert bei einer erneuten Durchführung von Interviews ein breiteres Spektrum von kurzfristig involvierten Teilnehmenden zu befragen um Vergleiche der inhaltlichen Standpunkte ziehen zu können. Die Onlineumfrage zeigt hier ein deutlich repräsentativeres Bild.

**Ressourcenmangel.** Während sich die Onlineumfrage nur in einer Frage und nur in Bezug auf Geld mit dem Mangel an Ressourcen beschäftigt, ist dies einer der Schwerpunkte in den Interviews. Alle Teilnehmenden bestätigen, dass sie bereits Erfahrungen mit Ressourcenmangel innerhalb des Vereins IOG gemacht haben. In der ersten Frage zum Thema, vgl. Metaskript A.1 Frage 1, sind keine Antwortmöglichkeiten vorgeschlagen. Die Teilnehmenden antworten dennoch sehr einheitlich, dass ein Mangel an finanziellen Ressourcen, ebenso wie ein Mangel an Wissen ihre Arbeit beeinflusst. Ein bestehender Mangel an Geld ist bei einer kleinen NRO wie IOG erwartbar, nur ein Teilnehmender erklärte, dass Geld zwar mangelte, aber anders als Wissen nicht die entscheidende Ressource sei, vgl. Metaskript A.1 Interview 7. Der Mangel an Wissen wurde von verschiedenen Teilnehmenden verschieden beschrieben. Zum Einen handelt es sich um mangelndes Wissen im Sinne von Erfahrung im Bereich der Entwicklungszusammenarbeit oder im Bereich der technischen Umsetzung von Projekten. Zum Anderen handelt es sich um mangelndes Wissen über lokale Gegebenheiten im Partnerland, Informationen über die Lage vor Ort oder auch um das Querschnittthema Wissensmanagement, also ein zugänglich-machen von Gelerntem und Best-Practices. Insbesondere das Thema des Wissensmanagements wurde in vielen Interviews angesprochen, sowohl als Antwort auf die Frage nach Ressourcenmangel, als auch assoziativ bei der Beantwortung anders gelagerter Fragen. Es ist wahrscheinlich, dass das Thema den Interviewten aufgrund aktueller Bemühungen der Kompetenzgruppe Qualitätsmanagement sehr präsent ist. Dennoch lässt sich aus der Intensität und Häufigkeit mit der es angemerkt und eingefordert wurde schließen, dass es ein wichtiges Thema für die Teilnehmenden ist und der Mangel in der Vergangenheit zu Frustration führte. Weiterhin wurde ein Personalmangel angesprochen, insbesondere in der Phase nach der Projektimplementierung. Hierbei handelt es sich nach Aussagen von zwei Teilnehmenden um einen Zeitpunkt zu dem üblicherweise unerfahrene Mitarbeiten-

de bei IOG ausscheiden, vgl. Metaskript A.1 Interview 7. Die Projektabwicklung und nachhaltige Betreuung der Projekte ist in der Folge erheblich erschwert.

**Ziel einer Evaluation.** Das Ziel oder die Ziele einer Evaluation wurden in der Onlineumfrage entsprechend der theoretischen Herleitung zur Auswahl gestellt. In den Interviews war es daher wichtig die Frage nach den Zielen oder Schwerpunkten frei zu formulieren, um die festgestellten Ergebnisse tatsächlich zu Prüfen. Die Teilnehmenden haben mit großer Mehrheit das Ziel Lernen genannt, in den meisten Fällen war es ein Lernen in Bezug auf IOG selbst, die Prozesse des Vereins und dessen Kapazitätsstärkung durch Mitarbeitendenqualifizierung. Vier Teilnehmende haben darüber hinaus ein gemeinschaftliches Lernen mit den Partnerorganisationen und eine lokale Kapazitätsstärkung als Ziel genannt und ein Teilnehmender hat die Kontrolle von Partnerorganisationen als Ziel benannt, vgl. Metaskript A.1 Interview 8. Diese Verteilung ist deutlich anders als die Ergebnisse der Onlineumfrage ausgefallen, dennoch lässt sich feststellen, dass in den Umschreibungen der Ziele und Schwerpunkte einer Evaluation der Wunsch formuliert wurde die Planung einer Intervention mit der Umsetzung zu vergleichen. Zur Bewertung wurden Erfolgskriterien wie Effizienz, Akzeptanz oder technische Funktionalität genannt, ebendiese Kriterien sind nach Kapitel 2.1 typische Kriterien des Evaluationsziels Kontrolle.

Zusammenfassend lässt sich feststellen, dass die Mehrheit der Interviewten eine Evaluation mit dem primären Ziel des Lernens und dem sekundären Ziel der Kontrolle wünscht.

**Kriterien der Bewertung.** Ein sehr prominentes Kriterium der Bewertung ist die technische Ausführung der Maßnahmen. Da der Verein IOG einen besonderen Fokus auf bauliche und technische Maßnahmen hat ergibt es Sinn, dass diese auf ihre Umsetzung, Funktionalität und Instandhaltung geprüft und bewertet werden sollen. Die Rangfolge dieses Kriteriums in Bezug zu anderen wie Nachhaltigkeit, Ownership und Wirkung ist jedoch umstritten. Es gibt eine polarisierte Verteilung, die auch in zwei Interviews als vereinsweites Phänomen beschrieben wurde, vgl. Metaskript A.1 Interview 3 und 10: Während ein Teil der Mitarbeitenden die technische Maßnahme als Mittelpunkt des Projekts wahrnimmt, geht ein anderer Teil davon aus, dass technische Maßnahmen nur ein Instrument

sind um Wirkungen und entwicklungspolitische Ziele zu erreichen. Im erstgenannten Fall ist es eine logische Konsequenz, dass die technische Maßnahme auch im Mittelpunkt einer Evaluation stehen sollte, als Kriterien der Bewertung wurden für diesen Fall Qualität der Bauausführung, Instandhaltung und Funktionalität der Anlage vorgeschlagen.

Im Fall der Evaluation der Wirkung einer Maßnahme wurde insbesondere die Feststellung und Bewertung von intendierten und auch nicht-intendierten Wirkungen gewünscht. Als Kriterien, die hierbei zur Messung herangezogen werden sollten, wurden Nachhaltigkeit und Ownership vorgeschlagen. Die deutliche Teilung der Mitarbeitenden von IOG ist an dieser Stelle nachweisbar und wurde auch in drei der Interviews benannt und reflektiert, eine Vereinigung der Positionen ist an dieser Stelle nicht möglich.

**Evaluator.** Die Frage wer eine Evaluation durchführen soll und in der Folge ob es sich um eine interne, externe oder Selbstevaluation handeln wird, wurde nicht explizit im Interview gefragt. Es wurden stattdessen, vergleichbar mit der Onlineumfrage, die limitierenden Ressourcen abgefragt, woraus Gespräche über Evaluation resultierten, die in drei Fällen auch explizit die verschiedenen Arten von Evaluatoren benannten, vgl. Metaskript A.1 Interview 6, 4 und 10. Insgesamt wurde die Vorstellung eine Evaluation von externen Evaluatoren durchführen zu lassen als unwahrscheinliche oder unpassende Option beschrieben. Die Mehrzahl der Teilnehmenden beschrieb Evaluation als einen Prozess, der durch Mitarbeitende, im besten Fall im Rahmen einer Selbstevaluation durchgeführt werden würde. Auf diesem Weg erhoffen sich die Teilnehmenden das Vertrauen des lokalen Partners in den Prozess zu erreichen, welches sich in einer offenen Beantwortung der Fragen widerspiegeln würde. Dennoch gab es, wie auch bereits in der Onlineumfrage, eine breite Bereitschaft Geld für Evaluation auszugeben. Da die Skala der Antwortmöglichkeiten bei der entsprechenden Frage im Onlinefragebogen unpassend gewählt wurde, wurde diese explizit in die Interviews mit aufgenommen.

Im Ergebnis konnten sich Teilnehmende finanzielle Investitionen zwischen 5 und 30% des Projektbudgets vorstellen, viele verwiesen aber auch darauf, dass die Evaluation einzelner Projekte durch die zum Teil geringen Projektbudgets nicht lohnenswert wäre. Sechs Teilnehmende schlugen die Evaluation von Projektbündeln vor, im besten Fall von solchen, die thematisch und/oder örtlich nah beieinander liegen.

**Ergebnis.** Das Ergebnis der Interviews bei IOG ergibt, dass eine Minderheit der an den Interviews Teilnehmenden eine Kontrolle im Sinne der Überprüfung der Umsetzungsleistung der NRO IOG vornehmen möchte. Eine Mehrheit der Teilnehmenden fordert Lernen als grundsätzliches Ziel der Evaluation, zum Teil in Bezug auf IOG, zu einem kleineren Teil in Bezug auf die Partnerorganisationen. Kriterien der Bewertung sollen hierbei auf der technischen Seite die Funktionalität und Instandhaltung sein, auf der wirkungsorientierten Seite soll es einen Fokus auf Nachhaltigkeit und Ownership geben. Es bietet sich an die Evaluationsvorhaben entsprechend der kommunizierten Bedarfe voneinander zu trennen, sodass auf der einen Seite eine Prüfung der technischen Maßnahmen stattfindet und auf der anderen Seite eine Aufnahme von intendierten und nicht-intendierten Wirkungen möglich ist. Auf diesem Wege ließe sich eine klare Trennung von interner oder Selbstevaluation im technischen Bereich und externer Evaluation im Wirkungsbereich realisieren, von der sowohl die Qualität profitieren könnte, die aber gleichermaßen die finanziellen Ressourcen des Vereins schont.

#### **4.6 Ergebnisse der Fallstudie**

Der Bedarf von IOG umfasst primär eine Evaluation mit dem Ziel des Lernens, der Unterschied zwischen den Ergebnissen der Onlineumfrage zum Thema Ziele und Schwerpunkte und dem hier beschriebenen konnte durch die Erklärungen, die in den Interviews abgegeben wurden eingeordnet und harmonisiert werden. Als sekundäres Ziel wurde in beiden Erhebungen Kontrolle genannt. Beim Evaluationsgegenstand soll es sich laut Onlineumfrage um einzelne Projekte handeln, in den Interviews wurde vermehrt von Projektbündeln gesprochen. Letzteres erscheint vor dem Hintergrund begrenzter Ressourcen sinnvoll, da eine Bündelung nach logistischen und/oder inhaltlichen Gesichtspunkten zu einer Effizienzsteigerung in der Auseinandersetzung mit diesen führen kann.

Die Umsetzung einer solchen Evaluation sollte hierbei schonend mit der Ressource Geld umgehen, das heißt sie soll regelmäßig stattfinden und dabei langfristig für IOG finanzierbar sein. Weiterhin wurde sowohl in der Onlineumfrage als auch in den Interviews ein Mangel an Wissen in Bezug auf Entwicklungszusammenarbeit beschrieben, sodass diese Kapazitäten vor der Durchführung einer internen oder Selbstevaluation aufgebaut werden müssen.

Die Kriterien zur Bewertung einer Evaluation wurden in den Interviews in zwei Bereiche unterteilt: Zum Einen in die Evaluation technischer Maßnahmen, zum Anderen in die Evaluation der Wirkung von Maßnahmen. Die Evaluation technischer Maßnahmen hat eine technische Bewertung der implementierten Anlage zum Ziel, das heißt die Kriterien der Bewertung wären, wie in den Interviews beschrieben, die Qualität der Bauausführung, Instandhaltung und Funktionalität der Anlage. Die Evaluation der Wirkung einer Maßnahme hat zum Ziel alle intendierten und nicht-intendierten Wirkungen zu erfassen und im lokalen Kontext zu bewerten. Entsprechende Kriterien wurden ebenfalls in den Interviews genannt, insbesondere Nachhaltigkeit und Ownership. Diese bestehende Teilung ist für eine NRO wie IOG sinnvoll, da die Mehrheit der Mitarbeitenden einen technischen beruflichen Hintergrund hat. Das vorhandene Wissen um technische Umsetzungen und Maßnahmen ist in der Folge deutlich größer als in vergleichbaren kleinen NROs. Die Fallstudie zeigt insbesondere ein Interesse daran die technischen Maßnahmen zu prüfen und deren Umsetzung zu verbessern, was sich durch den oben beschriebenen fachlichen Fokus erklären lässt. Die Wirkungsfrage ist dagegen in ihrem Ursprung soziologischer Natur und daher für Menschen mit einem technischen beruflichen Hintergrund eher fachfremd. Die Trennung dieser beiden Bereiche in zwei Evaluationsvorhaben erscheint vor diesem Hintergrund sinnvoll. Die Ressource Wissen ist im Bereich der technischen Umsetzung enorm groß und kann daher auch im Rahmen einer Evaluation genutzt werden, um andere Ressourcen zu entlasten. Die Externalisierung der Evaluation von technischen Maßnahmen kostet Geld, während der Mehrwert im Vergleich zur Evaluation von Wirkungen klein ist. Die Trennung der beiden Vorhaben ermöglicht aber sowohl die Anwendung verschiedener Kriterien der Bewertung, ebenso wie die Auswahl verschiedener Evaluatoren um ressourcenschonend vorzugehen. Der Evaluator kann im Falle der Evaluation technischer Maßnahmen von intern sein oder eine Selbstevaluation durchführen. Im Falle der Evaluation von Wirkungen ist ein externer Evaluator durch den Mangel an methodischem Wissen und Erfahrung mit Entwicklungszusammenarbeit im Verein anzuraten.

Bei beiden Evaluationen ist es möglich und empfehlenswert interne Prozesse zu betrachten, die sowohl die technische Qualität der Maßnahme, als auch deren Wirkung beeinflussen.

## 4.7 Reflexion der Fallstudie

Das Ziel der Fallstudie ist es verlässliche Daten darüber zu erheben welche Vorstellungen und Erwartungen Mitarbeitende von IOG an Evaluation haben. Dieses Ziel wurde mit Hilfe von wissenschaftlichen Methoden und Auswertungsverfahren erreicht. In der quantitativen Auswertung der Onlineumfrage ist aufgefallen, dass es bei Frage Nr. 10 zu einer unerwarteten Verteilung der Antworten kam. Hierbei wirkt sich ein methodischer Nachteil der Onlineumfrage aus, die vorformulierten Antwortmöglichkeiten beeinflussen das Ergebnis. Im vorliegenden Fall ist es die Begrenzung der Antwortmöglichkeiten auf „bis zu 10% des Projektbudgets“, die zu einer Antwortverteilung führt aus der nicht ablesbar ist wieviel Prozent die Teilnehmenden tatsächlich investieren würden. Diese lässt darauf schließen, dass die Skala der Antwortmöglichkeiten unpassend gewählt wurde und die Aussage der Antworten bei einer anderen Skala gegebenenfalls anders ausgefallen wäre. Zur Prüfung dieser These wurde die Frage erneut in den Interviews gestellt. Es wurde festgestellt, dass die Bereitschaft Geld für Evaluationen auszugeben in der Tat in einem höheren Bereich lag, als es implizit in den Antwortmöglichkeiten der Onlineumfrage angenommen wurde (bis zu 30% des Projektbudgets).

Es ist wahrscheinlich, dass die Verzerrung der Ergebnisse der Onlineumfrage durch die erneute Erhebung mittels Interview behoben werden konnte.

Neben diesem Beispiel kam es auch im Hinblick auf das Ziel von Evaluation durch die Nutzung von zwei verschiedenen Verfahren zu einem tieferen Verständnis für die Erwartungen der Mitarbeitenden von IOG, da der Wunsch nach Kontrolle eingeordnet und weiterhin der Wunsch nach Entwicklung und Lernen festgestellt werden konnte. Es ist dennoch möglich, dass trotz dieses zweistufigen Verfahrens der Datenerhebung nicht alle Perspektiven der Mitarbeitenden abgebildet werden konnten. Insbesondere die Teilnehmenden an den Interviews weisen in Bezug auf die Dauer ihres Engagements eine für IOG untypische Verteilung auf. Es ist daher zu vermuten, dass die Perspektive der Mehrheit der Mitarbeitenden (die nach einer Schätzung eines Mitarbeitenden der Geschäftsstelle durchschnittlich ein Engagement von zwei Jahren aufweisen) unterrepräsentiert wurde. Die qualitative Auswertung der Interviews ließe sich methodisch im Hinblick auf das Gütekriterium der Intersubjektivität der qualitativen Inhaltsanalyse verbessern. In der vorgestellten Durchführung ist keine zweite Person in die Durchführung oder Auswertung

der gesammelten Daten einbezogen worden, was zu einer Erhöhung der Objektivität der Ergebnisse führen würde.

## 5 Synthese und Analyse

Im folgenden Kapitel wird der in Kapitel 4.6 vorgestellte Bedarf von IOG an Evaluation mit den in Kapitel 3 dargestellten Evaluationsansätzen zusammengebracht. Es entsteht ein auf den Bedarf des Vereins angepasster Evaluationsansatz. Zur Analyse dieses Ansatzes werden im Folgenden zuerst die Evaluationsstandards der Gesellschaft für Evaluation eingeführt, anhand derer Folgen von ressourcenbedingten Verkürzungen und Risiken in der Durchführung aufgezeigt werden.

### 5.1 Einführung der DeGEval-Standards für Evaluation

Die DeGEval (Gesellschaft für Evaluation e.V.) wurde 1997 gegründet. Als Zusammenschluss von Institutionen und anderen Akteuren aus dem Bereich der Evaluation verfolgt sie das Ziel der Professionalisierung von Evaluation, die Verknüpfung unterschiedlicher Perspektiven miteinander und bietet eine Plattform für Informationen und Austausch über Evaluation [16]. Die Standards für Evaluation wurden 2002 von der DeGEval veröffentlicht, 2012 wurde ein erster Revisionsprozess angestoßen, der unter Einbezug der Mitglieder durchgeführt wurde. Diese zweite Fassung der Standards wurde 2016 verabschiedet und enthält neben diesen auch ein Glossar der Fachbegriffe, die im Rahmen von Evaluation und Evaluationsforschung genutzt werden. Die Fachsprache der vorliegenden Arbeit bezieht sich in hohem Maße auf dieses Glossar.<sup>10</sup>

Die Veröffentlichung von Standards durch die DeGEval ist ein Werkzeug, um das Ziel der Professionalisierung erreichen zu können. Durch Standardisierung sollen Aussagen über die Qualität von professionellen Evaluationen möglich gemacht und somit auch zur Transparenz über Evaluation beitragen. Aus der Perspektive der Evaluierenden kann eine Standardisierung zur Orientierung, Planung und Durchführung von Evaluationen genutzt werden, ebenso wie zur Aus- und Weiterbildung jüngerer KollegInnen [15, S. 5].

---

<sup>10</sup>Dieses Glossar ist online abrufbar unter <https://www.degeval.org/glossar-der-standards-fuer-evaluation/>.

**Standards für Evaluation nach DeGEval.** Die gesamten Standards sind im Anhang A.2 einsehbar.

Den Kern der Standards bilden vier Gruppen, denen insgesamt 25 Einzelstandards zugeordnet sind. Die Standardgruppen lauten Nützlichkeit, Durchführbarkeit, Fairness und Genauigkeit. Das Ziel der Nützlichkeitsstandards ist sicherzustellen, dass die Evaluation sich an klaren Evaluationszwecken und dem Informationsbedarf der Nutzenden orientiert [15]. Dieser Standardgruppe sind acht Standards untergeordnet, zum Beispiel die Identifizierung der Beteiligten und Betroffenen (N1) und die Klärung der Evaluationszwecke (N2).

Die Standardgruppe der Durchführbarkeit soll sicherstellen, dass die Planung und Ausführung der Evaluation realistisch, durchdacht und kosteneffizient ist. Es sind drei Standards zugeordnet, darunter Angemessene Verfahren (D1) und Effizienz von Evaluation (D3).

Das Ziel der Standardgruppe Fairness ist sicherzustellen, dass respektvoll und fair gegenüber allen Beteiligten und betroffenen Gruppen evaluiert wird. Es sind fünf Standards zugeordnet, darunter die Offenlegung von Berichten und Ergebnissen (F5) und die Umfassende und faire Prüfung (F3).

Die Standardgruppe der Genauigkeit soll sicherstellen, dass die Ergebnisse einer Evaluation gültig und nachvollziehbar sind, dies soll durch wissenschaftliche Methoden und wissenschaftliches Arbeiten sichergestellt werden. Dieses Ziel hat daher einen besonders engen Bezug zur Evaluationsforschung. Es sind neun Standards zugeordnet, darunter die Kontextanalyse (G2) und Angabe von Informationsquellen (G4).

## **5.2 Synthese eines bedarfsgerechten Evaluationsansatzes**

Ausgehend von dem in Kapitel 4 beschriebenen Bedarf von IOG an Evaluation bieten sich verschiedene Evaluationsansätze aus Kapitel 3 an. Es empfiehlt sich die Evaluation technischer Maßnahmen von der Wirkungsevaluation zu trennen. Evaluationen technischer Maßnahmen lassen sich in Zusammenarbeit mit lokalen Partnerorganisationen oder kurzen Ausreisen auch von unerfahrenen Mitarbeitenden realisieren. Es bestehen bereits ein erhebliches Wissen und entsprechende Strukturen darüber nach welchen Kriterien und mit Hilfe welcher Indikatoren die technische Nutzbarkeit der Maßnahmen ermittelt wer-

den können. Der Zugang zu diesem Wissen kann in kurzen Workshops und standardisierten Formularen gegeben werden, sodass die Ressource Wissen innerhalb des Vereins und innerhalb der Partnerorganisationen vergrößert wird. Die Ressource Geld wird geschont, da diese Vorgehensweise im Vergleich zu anderen Ansätzen sehr günstig ist.

Die Evaluation von Ownership, Nachhaltigkeit und Wirkung ist dagegen mit einem großen methodischen Wissen und viel Erfahrung in der Entwicklungszusammenarbeit und in der Umsetzung von Evaluationen verbunden, die nur schwer im Selbststudium zu erlangen sind. Es ist unwahrscheinlich, dass der Verein IOG innerhalb der nächsten fünf Jahre eine entsprechende Kapazität aufbauen kann. Um eine methodische und verlässliche Aussage über Wirkungen, intendierter und nicht-intendierter Natur, zu erhalten ist es empfehlenswert, dass diese Art der Evaluation externalisiert wird. Diese Entscheidung umfasst einen erheblichen Ressourcenaufwand, insbesondere im Hinblick auf die Ressource Geld. Andere Ressourcen, wie zum Beispiel Personal und Zeit werden mit dieser Herangehensweise geschont, da ein vergleichsweise geringer Personalaufwand notwendig ist um den Evaluierenden mit bestehenden Daten und Informationen zu versorgen.

**Evaluation von Technik.** Wie in Kapitel 4.6 beschrieben, kann die Evaluation technischer Komponenten in den Projekten von IOG durch Mitarbeitende von IOG intern oder im Rahmen einer Selbstevaluation bewertet werden. Im Sinne des Evaluationsansatzes nach *Fetterman* kann das in der Organisation in hohem Maße vorhandene Wissen in die Bewertung der Maßnahme einfließen, in Zusammenarbeit mit und in Anwaltschaft für die von der Maßnahme Betroffenen. Mit Wissen ist in diesem Fall insbesondere das technische Verständnis und die Erfahrung in der Instandhaltung technischer Anlagen gemeint, das die Mitarbeitenden von IOG in vielen Fällen entweder aus dem praktischen Berufsalltag oder theoretischer Art aus dem Studium besitzen. Diese Ressource sollte entsprechend der selbst formulierten Mission von IOG angewendet werden:

*Unsere Mission ist es, mit technischem Wissen die Lebensbedingungen der Menschen zu verbessern, um das Zusammenwachsen der Welt zu fördern [19].*

Zur Anwendung einer solchen Evaluation ist eine interne Kapazitätenstärkung notwendig, sodass die Rolle des Critical Friend durch einen Mitarbeitenden besetzt werden kann. Voraussetzungen für die Übernahme einer solchen Rolle im Evaluationsprozess sind Erfahrung in der Entwicklungszusammenarbeit und ein grundsätzliches metho-

disches Wissen in Bezug auf Evaluation. An erster Stelle ist dieses Wissen notwendig um die vorhandenen Daten zu bewerten und daraus Hinweise zu entwickeln. Im zweiten Schritt ist auch die Vermittlung dieser konstruktiven Rückmeldungen an Durchführende der Maßnahme, sowohl innerhalb von IOG als auch innerhalb der Partnerorganisation, eine anspruchsvolle Aufgabe. Diese erfordert ein hohes Maß an interkultureller Kommunikation und methodischer Kompetenz im Hinblick auf Wissensvermittlung.

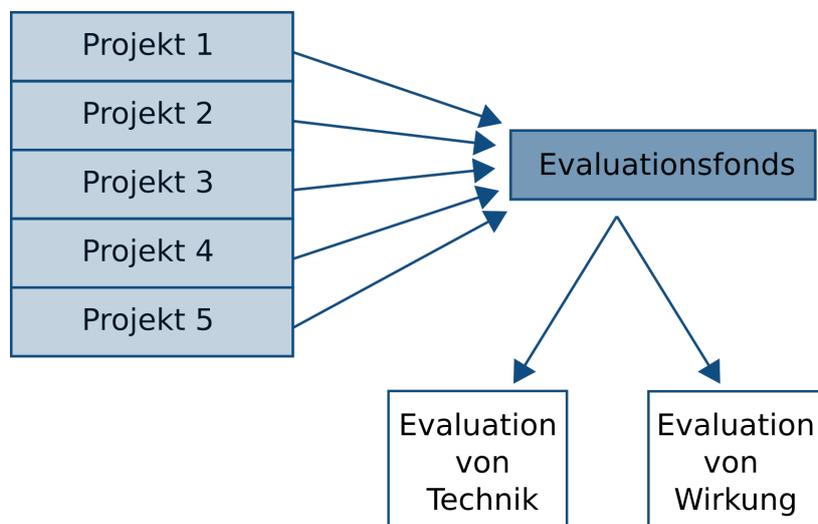
**Evaluation von Wirkung.** Die Evaluation von Wirkung erfordert ein höheres Maß an Vorwissen bezüglich Evaluation und Entwicklungszusammenarbeit, sodass eine Evaluation nach Fetterman in Bezug auf IOG nur in begrenztem Maße zu belastbaren Daten und Aussagen führen könnte. Passender erscheint die zielfreie Evaluation nach *Scriven*, mit einem starken Bezug zu lokalen Kapazitäten. Wie bereits in Kapitel 3.2 erläutert legt diese einen besonderen Fokus darauf sowohl intendierte als auch nicht-intendierte Wirkungen festzustellen. Insbesondere in den Interviews wurde dies als Erwartung an Evaluation mehrfach formuliert. In der Ausführung empfiehlt sich eine Durchführung, wie sie zum Beispiel von Brot für die Welt aktuell vorgenommen wird: Lokale Evaluatoren werden beauftragt vor Ort Untersuchungen vorzunehmen. Das heißt konkret in Bezug auf IOG, dass es sich um eine Evaluation von extern handelt.

Um den Nebeneffekt der Kapazitätenstärkung innerhalb von IOG zu erreichen und das Vertrauen der lokalen Partner in die lokalen Evaluatoren zu stärken empfiehlt es sich zusätzlich unerfahrene Mitarbeitende von IOG zur Evaluation zu entsenden. Diese können die Besuche der Partnerorganisation begleiten und die Datenerhebung des Evaluators unterstützen, die methodische Leitung obliegt hierbei zu jeder Zeit dem lokalen Experten.

Während Brot für die Welt diesen Ansatz insbesondere aufgrund von langjährigen Partnerschaften mit den meisten ihrer Partnerorganisationen nutzen kann, bietet sich bei IOG ein gemischtes Feld von Partnerorganisationen. Solche die bereits längerfristig mit IOG zusammengearbeitet haben und in der Folge auch lokal eine herausgehobene Stellung haben werden als Ankerpartner bezeichnet. Zu Beginn empfiehlt es sich die Wirkung von Projekten dieser Ankerpartner zu evaluieren, zum Einen um ähnlich wie Brot für die Welt vom bestehenden Vertrauensverhältnis profitieren zu können und zum Anderen um die neuen Prozesse der Evaluationsbeauftragung und -begleitung einzuüben.

**Evaluationsfonds** Da bereits ein Mangel finanzieller Ressourcen für den Verein IOG festgestellt wurde, ist es neben der festgestellten Bereitschaft Geld für Evaluation auszugeben auch notwendig zu erfragen woher dieses Geld kommen kann.

Ein Evaluationsfonds ist eine Vermögensreserve, die gemeinsam von allen Regionalgruppen und Projekten aufgebaut wird und zweckgebunden für Evaluation eingesetzt wird. Die Verwaltung dieses Fonds sollte zentral gesteuert sein, zum Beispiel im Hauptamt in der Geschäftsstelle oder im Ehrenamt in der Kompetenzgruppe Qualitätsmanagement. Die finanzielle Grundlage wird durch einzelne Projekte gelegt, diese zahlen einen



**Abbildung 9:** Ein vereinsübergreifend finanzierter Evaluationsfonds kann zur Mittelquelle für alternierende Evaluationen von Technik und Wirkung werden.

kleinen Prozentsatz ihres Budgets pro Jahr ein. Sie ermöglichen dem Verein gemeinsam die Möglichkeit, dass zentral gesteuert die Evaluation einzelner Maßnahmen oder Projektbündel stattfinden kann, werden dadurch aber nicht zwangsläufig selbst einer Evaluation unterzogen. Aktive Mitarbeitende und insbesondere Projektdurchführende sind eingeladen ihre Projekte für eine Evaluation vorzuschlagen, Gründe können hierbei sowohl positive als auch negative Eindrücke aus den Bereichen Technik, Wirkung, Partnerorganisation und interne Prozesse sein. Ausgerichtet an strategischen Zielen des Vereins entscheidet die Geschäftsstelle und/oder die Kompetenzgruppe Qualitätsmanagement, inspiriert durch die Vorschläge aber nicht darauf beschränkt, welche Projekte zu einem Bündel zusammengefasst und evaluiert werden können.

Wird ein Evaluationsfonds auf der Basis eines Solidaritätsprinzips aufgebaut, das

heißt durch Beiträge aller Projekte finanziert, wird es notwendig ein angemessenes Abgabesystem zu etablieren. Als kurze Machbarkeitsstudie dieses Vorschlags wird das Abgabesystem folgend skizziert und am Beispiel des Jahres 2018 berechnet.

Die erste Annahme ist, dass eine prozentuale Abgabe zu einer kalkulierbaren und vertretbaren finanziellen Last der einzelnen Regionalgruppen führt. Vorgeschlagen wird eine pauschale Abgabe von 3% auf alle Ausgaben.

Zieht man das Jahr 2018 als Beispiel heran wurden in diesem Jahr €422.330 im Rahmen von Projekten ausgegeben[18]. Berechnet man 3% dieser Ausgaben ergibt das Einnahmen für den Evaluationsfonds von rund €12.670. Eine solche finanzielle Ressource kann Evaluation in mehrfacher Hinsicht ermöglichen: Einerseits können Ausreisen für interne technische Evaluationen finanziert werden, andererseits aber auch externe Wirkungsevaluationen. Die Entscheidung dafür welche Art der Evaluation durch den Fonds unterstützt wird muss hierbei jedes Jahr unter strategischen und kontextuellen Gesichtspunkten neu verhandelt werden.

Entscheidende Kriterien sind hierbei die Folgenden: Technische Evaluationen können von intern erfolgen und gegebenenfalls auch mit Erkundungsreisen für neue Projekte zusammengefasst werden. Ist dies möglich und sinnvoll, so kann mehr Geld aus dem Evaluationsfonds für eine Wirkungsevaluation von extern investiert werden. Bei logistisch ungünstiger Projektlage kann es sinnvoll sein den Evaluationsfonds komplett oder anteilig für technische Evaluationen von intern zu investieren.

Für eine technische Evaluation von intern müssten Reise- und Aufenthaltskosten finanziert werden. Auf der Basis informeller Gespräche mit Mitarbeitenden von IOG werden folgende Annahmen für Kosten genutzt: Hin- und Rückflug €2.000, Hotelkosten €300, Verpflegungspauschale €240 und lokale Transportkosten €150 für einen Aufenthalt von 10 Tagen. In der Summe erhält man €2.690. Inklusiv einer Rücklage für unvorhergesehene Kosten lässt sich ein finanzieller Aufwand von €3.000 pro evaluierendem Mitarbeitenden von IOG vor Ort feststellen.

Entsprechend des Beispieljahres 2018 hätten mittels der nur in diesem Jahr eingezahlten Evaluationsbeiträge der einzelnen Projekte 2 Projektbündel von jeweils zwei Mitarbeitenden für 10 Tage im Rahmen einer internen technischen Evaluation vor Ort besucht werden können.

Um eine sichere Reise zu ermöglichen, Unterstützung bei der Evaluation zu erhalten

und zeitgleich eine Qualifikation von Personal zu erreichen ist es empfehlenswert einen erfahrenen Mitarbeitenden von IOG zusammen mit einem unerfahrenen Mitarbeitenden zur Evaluation zu entsenden. Unerfahrene Mitglieder von IOG sind in der Regel Studierende, die sich in mehrerer Hinsicht dafür eignen zu einer Evaluation entsendet zu werden: Es handelt sich um eine Personengruppe die mehr Zeit in den Verein investieren kann als andere Personengruppen, zum Beispiel durch freiere Zeiteinteilung im Studium und die Möglichkeit eine längere Zeit am Stück, zum Beispiel in den Semesterferien, für IOG zu arbeiten. Gleichzeitig handelt es sich bei dieser Personengruppe, wie bereits beschrieben, in der Regel auch um unerfahrene Mitarbeitenden. Insofern ist der Mehrwert der Teilnahme an einer Evaluation umso größer, weil ein erster Kontakt mit Entwicklungszusammenarbeit erfolgen kann ohne die Verantwortung für eine Implementierung zu tragen. Das Lernen aus den Fehlern Anderer trägt zudem aktiv zur Verbreitung von Wissen innerhalb des Vereins bei, wenn beispielsweise ein Mitarbeitender aus der Regionalgruppe Aachen ein Projekt der Regionalgruppe Berlin evaluiert.

Die zweite Annahme ist, dass ein durchschnittliches Projekt bei IOG €20.000 kostet. Entsprechend einer Schätzung von *Phineo*, einer Beratungsorganisation für NROs im entwicklungspolitischen Kontext, werden 3-10% des Projektbudgets für eine externe Evaluation veranschlagt [25]. Im Fall von IOG empfiehlt es sich auch hier mit Projektbündeln zu arbeiten. Bei einem hohen Satz von 10% wäre entsprechend der obigen Berechnung 2018 eine externe Evaluation von sechs durchschnittlich großen Projekten möglich gewesen. Die zusätzliche Ausreise von Mitarbeitenden von IOG zur Kapazitätsstärkung ist hierbei nicht eingerechnet.

Strebt der Verein IOG an sowohl den Fokus Technik, als auch den Fokus Wirkung zu evaluieren muss entsprechend der vorgestellten Finanzierung alternierend vorgegangen werden. In einem Jahr kann intern in Bezug auf Technik evaluiert werden, im nächsten Jahr kann extern in Bezug auf Wirkung evaluiert werden. Auf diese Art und Weise kann eine Evaluation von Technik ebenso stattfinden wie eine Evaluation von Wirkung, da ein personalintensiver Ansatz mit einem finanzintensiven Ansatz kombiniert wird. Unabhängig von der Finanzierung erscheint dies in Bezug auf die Größe des Vereins IOG sinnvoll, da jede Art der Evaluation mit einem erheblichen Maß an Vorbereitungs- und Begleitungsarbeit einhergeht, das neben der bisher anfallenden Arbeiten notwendig wird. In jedem Fall ermöglicht ein solidarisch befüllter Evaluationsfonds den Aufbau einer Evalua-

tionskultur, die sich Dank der Bemühungen der Regionalgruppen auf eine vorhersehbare Finanzierung stützt.

### **5.3 Analyse des beschriebenen Evaluationsansatzes**

Die oben beschriebenen Standards wurden als Maximalstandards beschrieben, in dem Bewusstsein, dass sie nicht in jeder Evaluation in vollem Umfang umsetzbar sind. Dies zeigt sich in der natürlichen Gegensätzlichkeit einzelner Standards, zum Beispiel Durchführbarkeit im Gegensatz zu Genauigkeit. Eine teilweise oder vollständige Nichterfüllung einzelner Standards sollte daher dokumentiert und begründet werden. Eine qualitativ hochwertige Evaluation zeichnet sich dadurch aus, dass nachvollziehbar ist, dass und inwiefern der Versuch gemacht wurde die Standards unter den konkreten Bedingungen möglichst weitgehend umzusetzen. In der vorliegenden Arbeit wird keine konkrete Evaluation überprüft, sondern die generelle Anwendbarkeit des beschriebenen Evaluationsansatzes unter Einbezug der zur Verfügung stehenden Ressourcen. Aus diesem Grund wird im Folgenden metaevaluiert inwiefern die Ansätze auch in der für IOG angepassten und zum Teil reduzierten Form anwendbar sind. Anwendbar heißt hierbei, dass mit großer Wahrscheinlichkeit Ergebnisse erzielt werden, die den Standards für Evaluation entsprechen.

**Evaluation von Technik.** Für die Evaluation technischer Maßnahmen, die durch IOG im Rahmen von entwicklungspolitischen Projekten durchgeführt werden, wird obenstehend eine Evaluation nach Fetterman empfohlen. Die notwendige Kapazitätenstärkung, die bereits in Kapitel 5.2 angesprochen wurde erscheint im Hinblick auf die DeGEval Standards umso wichtiger, um eine Evaluation mit hoher Aussagekraft herstellen zu können.

Mit einem Fokus auf die Standardgruppe der Nützlichkeit sind die Anforderungen der Einzelstandards bereits von vielen Mitarbeitenden von IOG gut erfüllbar. Der Einzelstandard N3, der sich mit der Kompetenz und Glaubwürdigkeit des Evaluators befasst muss hierbei jedoch kritisch gesehen werden. Eine Evaluation von intern oder als Selbstevaluation durchzuführen birgt in jedem Fall das Risiko, dass Glaubwürdigkeit und Akzeptanz durch fehlende Qualifikationen sinken. Diesem Einzelstandard soll daher mit einer vorhergehenden Qualifikation der mit der Durchführung Beauftragten entsprochen werden.

Mit einem Fokus auf die Standardgruppe der Durchführbarkeit sind insbesondere die Einzelstandards D2 und D3 kritisch zu sehen. Das diplomatische Vorgehen, das im Einzelstandard D2 gefordert wird, erfordert ein hohes Maß an interkultureller Kommunikation und besetzt zudem eine Schlüsselfunktion. Denn nur mit dem entsprechenden Maß an diplomatischem Vorgehen kann bei IOG und bei Partnerorganisationen der Grundstein für die Implementierung der Ergebnisse einer Evaluation gelegt werden. Eine Evaluation, deren Ergebnisse und Empfehlungen nicht beachtet und umgesetzt werden hat für das Evaluationsziel Lernen zum Beispiel keinen Wert. Der Einzelstandard D3 fordert eine Effizienz von Evaluationen. Dies kommt dem Verein IOG entgegen, weil impliziert wird, dass auch der Einsatz von Ressourcen nur insoweit erfolgen soll, wie der Nutzen der Evaluation dies rechtfertigt. Obwohl diese Forderung einen ressourcenschonenden Umgang beinhaltet ist es wahrscheinlich, dass sie erst umgesetzt werden kann sobald die Prozesse der Evaluation routiniert erfolgen.

Im Hinblick auf die Standardgruppe Fairness muss insbesondere die Erreichung des Einzelstandards F4 geprüft werden. Sofern die Evaluation durch Mitarbeitende von IOG durchgeführt wird, muss sichergestellt werden, dass auch alle anderen Sichtweisen von Beteiligten und Betroffenen beachtet werden. Dieses Risiko der einseitigen Betrachtung von Daten und Kontext besteht zwar in jeder Evaluation, die von intern geleitet wird oder im Rahmen einer Selbstevaluation erfolgt, es ist jedoch umso größer, je unerfahrener ein Evaluator ist. Die größten Risiken für die Evaluation von Technik bei IOG, wie sie oben beschrieben wurde, bestehen in der Einhaltung der Standardgruppe der Genauigkeit. Insbesondere die Einzelstandards G2, G5, G6 und G7 haben die Gefahr nicht ausreichend beachtet zu werden. Der Einzelstandard G2 fordert eine detaillierte Analyse des Kontextes, innerhalb dessen eine Maßnahme implementiert wurde. Eine solche Bewertung einzelner Gegebenheiten erfordert neben der Einordnung der festgestellten Faktoren auch eine Wertung und das Erstellen von Relationen der Faktoren zueinander. Dazu sind neben einem hohen Maß an entwicklungspolitischem Wissen auch Kenntnisse von wissenschaftlichen Methoden des Schlussfolgerns, auch im Hinblick auf Kausalitätsprobleme, notwendig. Die Einzelstandards G5 und G7 beziehen sich beide auf Informationen, die in eine Evaluation eingehen und in dieser analysiert werden. Sowohl Erhebungsverfahren, als auch Datenquellen sollen durch den Evaluator auf ihre Reliabilität geprüft werden, entsprechend der Gütekriterien empirischer Forschung. Diese muss sich ein Mitarbeitender von

IOG voraussichtlich im Rahmen seiner Qualifizierung für die Übernahme von Evaluations-tätigkeiten erarbeiten und deren routinierte Umsetzung erlernen. Die Analyse solcher reliabler Daten soll nach Einzelstandard G7 angemessen und systematisch erfolgen, eine konkrete wissenschaftliche Theorie wird dem nicht zugeordnet, wodurch die Hürde für eine Qualifizierung in diesem Fall weniger hoch erscheint, als es noch in Bezug auf G5 war.

Der Einzelstandard G5 fordert in Bezug auf gesammelte Informationen eine systematische Prüfung auf mögliche Fehler. Auch hier ist keine konkrete wissenschaftliche Basis oder Methode genannt, dennoch ist der Anspruch an einen ungeübten Evaluator hoch, um einzuschätzen was ein Fehler in den Daten ist, wie man diesen in den Rohdaten erkennen und im besten Fall aus diesen entfernen kann. Weiterhin fordert der Einzelstandard G6 eine nahezu wissenschaftliche Aufbereitung und Prüfung der gesammelten Informationen. Dies erfordert neben dem notwendigen methodischen Wissen zur Aufbereitung auch ein Mindestmaß an Erfahrung im Umgang mit Daten um fehlerhafte Informationen zu erkennen.

Zusammenfassend lässt sich sagen, dass die Anforderungen an einen Evaluator entsprechend der DeGEval Standards hoch sind. Auch wenn diese, wie eingangs beschrieben, die Maximalstandards für Evaluation darstellen erscheint es sinnvoll diese als Ziel anzustreben. Ein besonderes Augenmerk sollte in der Qualifikation der mit Evaluations-tätigkeiten beauftragten Mitarbeitenden von IOG auf die oben näher beschriebenen Einzelstandards gelegt werden, da das Risiko besteht diese nicht erfüllen zu können. Sollte es dazu kommen, dass Evaluationen erstellt werden ohne die oben beschriebene vorherige Qualifikation und Kapazitätenstärkung, so sind die Ergebnisse dieser nicht reliabel und sollten in der Folge nicht zur Basis von Entscheidungen genutzt oder als Lernhilfe innerhalb des Vereins oder einer Partnerorganisation genutzt werden. Ist eine entsprechende Qualifikation hingegen erfolgt, stellt der vorgestellte Ansatz eine Möglichkeit dar Evaluationen durchzuführen deren Ergebnisse einem hohen Qualitätsanspruch genügen, die Ressource Geld schonen und zum Lernen der Organisation beitragen.

**Evaluation von Wirkung.** Die Evaluation von Wirkung, die entsprechend der Ressourcen und des Bedarfs von IOG externalisiert werden soll, wird im Folgenden nur zum Teil mit den Evaluationsstandards der DeGEval verglichen. Da die Einhaltung dieser Stan-

dards zum größten Teil in der Verantwortung der lokal engagierten Evaluatoren liegt muss angenommen werden, dass diese mit Hilfe ihres methodischen Wissens eine qualitative Evaluation liefern. Bei der Auswahl von Evaluatoren sollte in jedem Fall auf die Einhaltung der Standards hingewiesen werden und bereits in der Ausarbeitung der Terms of Reference <sup>11</sup> als Vertragsinhalt können diese Standards als Grundlage dienen. Ein gemeinsames Verständnis des Evaluationsauftrags trägt in jedem Fall dazu bei eine Evaluation zu erhalten, die dem Verständnis und den Erwartungen von IOG entspricht. Unabhängig von der Wahl des Evaluators muss jedoch entsprechend der Standardgruppe Nützlichkeits der Einzelstandard N2 erfüllt werden, der eine Klärung der Evaluationszwecke fordert. Während der Zweck bei der Evaluation von Technik sehr deutlich abzugrenzen ist, gibt es bei der Evaluation von Wirkung ein breites Feld von möglichen Schwerpunkten, die gewählt werden können. Entsprechend der Fallstudie können die Ziele Lernen und Kontrolle zur groben Orientierung dienen, die insbesondere in der Onlineumfrage eindeutig gewählten Kriterien des reibungslosen Projektablaufs, des Kosten-Nutzen-Verhältnisses und der gender- und diversitysensiblen Planung könnten zum Beispiel spezifisch in die Terms of Reference aufgenommen werden. Bei einer Evaluation durch einen externen Evaluators muss in der Vorarbeit der Beauftragung eine Klärung dieser Zwecke durch IOG möglichst konkret erfolgen, um ein verwertbares Ergebnis zu erzeugen.

## **6 Fazit und Ausblick**

Das in Kapitel 1.3 beschriebene Ziel der vorliegenden Arbeit war es einen Evaluationsansatz für entwicklungspolitische Projekte zu entwickeln, der insbesondere für kleine NROs mit begrenzten Ressourcen nutzbar ist. Dies wurde beispielhaft an der Organisation IOG entwickelt, mittels einer Fallstudie wurden Bedarfe aufgenommen und deren Übereinstimmung mit Evaluationsansätzen geprüft. Weiterhin wurden Risiken der auf den Bedarf und die vorhandenen Ressourcen von IOG angepassten Evaluationslösungen aufgezeigt und anhand der Standards für Evaluation nach DeGEval geprüft. Abschließend lässt sich feststellen, dass eine externe Evaluation von Wirkungen für den Verein IOG finanzier-

---

<sup>11</sup>Terms of Reference oder TOR sind ein Schlüsseldokument in der Beauftragung externer Dienstleister. Hierin sind alle Aspekte der Durchführung, Ziele der Evaluation, Rolle und Verantwortlichkeiten des Evaluators und für die Evaluation verfügbare Ressourcen definiert.

bar ist, jedoch nicht in jedem Jahr für jedes Projekt, sondern nur für ein ausgewähltes Projektbündel alle zwei Jahre. Eine interne oder Selbstevaluation von Technik erfolgt innerhalb des in Kapitel 4.6 vorgestellten Rahmens ebenfalls alle zwei Jahre. Bei einer Implementierung von 17 Projekten in 2018 und der in Kapitel 4.6 hergeleiteten Anzahl von durchschnittlich sechs Projekten in einem Projektbündel wird alle zwei Jahre ein Drittel der implementierten Projekte in Bezug auf ihre Wirkung evaluiert. Dieses Verhältnis erscheint in Bezug auf die Größe des Vereins, die finanziellen Ressourcen und personellen Kapazitäten der Begleitung in Bezug auf beide Schwerpunkte realistisch und aussagekräftig. Das Ziel der Arbeit wurde damit in vollem Umfang erreicht.

**Wissensmanagement.** Der Umgang mit erarbeiteten Informationen und den Ergebnissen von Evaluationen ist von hervorgehobener Bedeutung, wenn es darum geht aus Evaluationen zu lernen. Für die vorliegende Arbeit war es von untergeordneter Bedeutung, da die Betrachtung über den Rahmen der Arbeit hinausgeht. Es wurden bewusst keine Fragen dazu in Onlineumfrage und Interviews gestellt. Dennoch wurde das Thema in beiden Erhebungen durch Teilnehmende aufgebracht und eingefordert. Das lässt sich zum Einen auf aktuelle Aktivitäten der Kompetenzgruppe Qualitätsmanagement zurückführen, die das Thema Wissensmanagement bearbeitet. Zum Anderen beziehen sich viele Aussagen auf Erfahrungswerte, auf Situationen in denen mangelndes Wissensmanagement zu Mehrarbeit und Frustration geführt hat. Im Sinne der Nutzbarkeit von Evaluationsergebnissen bildet die Verfügbarkeit von Wissen einen Grundpfeiler, da andernfalls Evaluationsergebnisse der Mehrheit vorenthalten bleiben und benannte Erwartungen wie Lerneffekte und Qualitätssteigerung nicht eintreten werden. Eine Betrachtung des aktuellen Wissensmanagements geht über den Rahmen der vorliegenden Arbeit hinaus, sollte aber in zukünftigen Betrachtungen thematisiert werden. Vorstellbar für eine solche Arbeit ist eine Aufnahme der vorhandenen Strukturen, eine Analyse unter Einbezug der aktuellen Bestrebungen und schließlich eine Bewertung dieser.

**Partner- und Wirkungsverständnis.** In den Ergebnissen der Onlineumfrage und insbesondere in den Interviews mit Mitarbeitenden von IOG ist deutlich geworden, dass das Partner- und Wirkungsverständnis innerhalb des Vereins und unter den Mitarbeitenden nicht harmonisch verankert ist, sondern verschiedene Mitarbeitende deutlich unter-

schiedliche Sichtweisen auf ihre Arbeit mit den Partnerorganisationen haben. Auch die Wahrnehmung der eigenen Wirkung und Intervention vor Ort divergiert deutlich, was zu großen Teilen auf das verschieden lang- oder kurzfristige Engagement bei IOG zurückzuführen ist, ebenso wie auf die sehr verschieden großen Erfahrungen im Bereich der Entwicklungszusammenarbeit allgemein. Es ist in jedem Fall empfehlenswert einen internen Dialog darüber aufzunehmen, um eine Harmonisierung des Verständnisses zu erreichen und dieses auch mit offiziellen Zielen und Statements des Vereins in Einklang zu bringen.

**Analyse nach DeGEval Standards.** Die Analyse der Evaluationsansätze nach DeGEval Standards kann den Umfang einer weiteren Masterarbeit haben, insbesondere wenn man diese anhand erster durchgeführter Beispiele bewertet. Eine wichtige weiterführende Fragestellung ist hierbei, ob es sich im Falle der Evaluation von Technik um ein Self-Assessment, vgl. Kapitel 3.1, oder um eine interne oder Selbstevaluation handelt. Dies ist zu großen Teilen abhängig von der Qualifizierung, welche die die Evaluation begleitenden Mitarbeitenden erhalten und wie diese später in der Rolle des Critical Friend agieren. Zur Prüfung können beispielsweise die in Kapitel 5.1 vorgestellten DeGEval Standards genutzt werden. Insbesondere die als kritisch näher diskutierten Einzelstandards sollten hierbei auf ihre Einhaltung mit dem verfügbaren Wissen geprüft werden.

**Verallgemeinbarkeit.** Die vorliegenden Ergebnisse sind eingeschränkt auch auf andere kleine NROs übertragbar, da die festgestellten Ressourcenmängel bei IOG (Geld und Wissen) beispielhaft für kleine NROs sind. Das Alleinstellungsmerkmal des Vereins, das insbesondere die Möglichkeit der internen oder Selbstevaluation von technischen Maßnahmen eröffnet, ist das im Verein sehr verbreitete Fachwissen um technische Maßnahmen. Es ist sehr ungewöhnlich für einen Verein in der Entwicklungszusammenarbeit diese Kompetenzen kostenlos durch ehrenamtliche Mitarbeitende zur Verfügung gestellt zu bekommen und in der Folge selbstständig Infrastrukturprojekte durchführen zu können. Die Möglichkeit vergleichbare Projekte ebenfalls von intern evaluieren zu können ist daher nicht verallgemeinbar, sondern vielmehr eine Option für IOG, die aus der speziellen Ausrichtung des Vereins und der Ressourcen seiner ehrenamtlich Mitarbeitenden resultiert. Die Finanzierung externer Evaluationen mit Hilfe von selbstverwalteten Evaluationsfonds ist dagegen ein Werkzeug, das auch andere kleine NROs anwenden können.

## Literatur

- [1] M. C. Alkin und S. Taut. “Unbundling evaluation use”. In: *Studies in Educational Evaluation* 29.1 (2002), S. 1–12. DOI: 10.1016/S0191-491X(03)90001-0.
- [2] A. Borrmann und R. Stockmann. *Evaluation in der deutschen Entwicklungszusammenarbeit*. Hrsg. von R. Stockmann. Bd. 8. Sozialwissenschaftliche Evaluationsforschung. Münster: Waxmann, 2009.
- [3] R. Boruch und H. Chen. *Evaluation Roots: Tracing Theorists’ Views and Influences*. Hrsg. von M. C. Alkin. SAGE Publications, 2004.
- [4] P. Bouman. “Kausalität und Funktionalzusammenhang in der Soziologie Max Webers”. In: *Zeitschrift für die gesamte Staatswissenschaft / Journal of Institutional and Theoretical Economics* 105.3 (1948), S. 463–475.
- [5] J. Brandtstädter. “Evaluationsforschung: Probleme der wissenschaftlichen Bewertung von Interventions- und Reformprojekten”. In: *Zeitschrift für Pädagogische Psychologie* 4.4 (1990), S. 215–227.
- [6] Bundesministerium für wirtschaftliche Zusammenarbeit. *Akteure der bilateralen Zusammenarbeit, Nichtregierungsorganisationen*. Online verfügbar unter [https://www.bmz.de/de/ministerium/wege/bilaterale\\_ez/akteure\\_ez/nros/index.html](https://www.bmz.de/de/ministerium/wege/bilaterale_ez/akteure_ez/nros/index.html), Stand 22.09.2020. 2020.
- [7] A. Danish, F. Champagne und R. Blais. “Theoretical analysis of policies to improve the recruitment and retention of rural physicians”. In: *Australian Journal of Rural Health* 28.5 (2020), S. 427–433. DOI: 10.1111/ajr.12666.
- [8] Deutsche Gesellschaft für internationale Zusammenarbeit. *Das Evaluierungssystem der GIZ, Zentrale Projektevaluierung im BMZ-Geschäft*. Online verfügbar unter [https://www.giz.de/de/downloads/04\\_GIZ\\_EVAL\\_ZPE\\_BMZ-Geschäft.pdf](https://www.giz.de/de/downloads/04_GIZ_EVAL_ZPE_BMZ-Geschäft.pdf), Stand 12.12.2020. 2018.
- [9] N. Döring und J. Bortz. *Forschungsmethoden und Evaluation*. 5. Aufl. Springer-Lehrbuch. Springer, 2016. DOI: 10.1007/978-3-642-41089-5.

- [10] U. Dornberger. *Auf eigenen Beinen stehen – Finanzierungsmöglichkeiten Entwicklungspolitischer Nichtregierungsorganisationen*. Hrsg. von Stiftung Nord-Süd-Brücken und Eine-Welt-Haus Jena. Online verfügbar unter [https://www.sept.uni-leipzig.de/fileadmin/user\\_upload/Publications/3\\_Dornberger\\_2000\\_\\_b\\_.pdf](https://www.sept.uni-leipzig.de/fileadmin/user_upload/Publications/3_Dornberger_2000__b_.pdf), Stand 15.11.2020. Stiftung Nord-Süd-Brücken, 2000.
- [11] Evangelisches Werk für Diakonie und Entwicklung e.V., Brot für die Welt. *Evaluationskonzept*. Online verfügbar unter [https://www.brot-fuer-die-welt.de/fileadmin/mediapool/2\\_Downloads/Sonstiges/Evaluationskonzept\\_BfdW\\_DKH\\_2017.pdf](https://www.brot-fuer-die-welt.de/fileadmin/mediapool/2_Downloads/Sonstiges/Evaluationskonzept_BfdW_DKH_2017.pdf), Stand 20.11.2020. 2020.
- [12] D. Fetterman, S. Kaftarian und A. Wandersman. *Empowerment Evaluation: Knowledge and Tools for Self-Assessment and Accountability*. SAGE Publications, 2015. DOI: 10.4135/9781483387079.
- [13] D. Fetterman und A. Wandersman, Hrsg. *Empowerment Evaluation Principles in Practice*. Guilford Press, 2005.
- [14] F. J. Jr. Fowler. “Why it is easy to write bad questions”. In: *ZUMA Nachrichten* 25.48 (2001), S. 49–66.
- [15] Gesellschaft für Evaluation e.V. *Standards für Evaluation (Langfassung)*. Online verfügbar unter [https://www.degeval.org/fileadmin/Publikationen/DeGEval\\_Standards\\_fuer\\_Evaluation\\_-\\_Erste\\_Revision\\_\\_2016\\_.pdf](https://www.degeval.org/fileadmin/Publikationen/DeGEval_Standards_fuer_Evaluation_-_Erste_Revision__2016_.pdf), Stand 18.09.2020. 2016.
- [16] Gesellschaft für Evaluation e.V. *Wir über uns*. Online verfügbar unter <https://www.degeval.org/ueber-uns/>, Stand 25.07.2020. 2020.
- [17] C. Helfferich. *Die Qualität qualitativer Daten – Manual für die durchführung qualitativer Interviews*. 4. Aufl. VS Verlag für Sozialwissenschaften, 2011.
- [18] Ingenieure ohne Grenzen e.V. *Jahresbericht 2018*. Online verfügbar unter <https://ingenieure-ohne-grenzen.org/sites/default/files/files/downloads-thumbnail/jahresbericht2018.pdf>. Berlin, 2018.
- [19] Ingenieure ohne Grenzen e.V. *Vision, Mission, Werte und Leitbild*. Online verfügbar unter <https://ingenieure-ohne-grenzen.org/de/ueber-iog/vision-mission-leitbild>, Stand 09.09.2020. 2020.

- [20] Ingenieure ohne Grenzen e.V. *Wie wir arbeiten*. Online verfügbar unter <https://ingenieure-ohne-grenzen.org/de/projekte/projektablauf>, Stand 09.09.2020. 2020.
- [21] K. Krieger, D. Gibbs und S. Cutbush. “Evaluating Human Trafficking Service Programs: What Can Be Learned from Domestic Violence Service Program Evaluations”. In: *Journal of Human Trafficking* (2020). DOI: 10.1080/23322705.2020.1784686.
- [22] U. Maier. “Formative Assessments – Ein erfolgversprechendes Konzept zur Reform von Unterricht und Leistungsmessung?” In: *Zeitschrift für Erziehungswissenschaft* 13.2 (2010), S. 293–308.
- [23] W. Möhring und D. Schlütz. *Die Befragung in der Medien- und Kommunikationswissenschaft: Eine praxisorientierte Einführung*. Studienbücher zur Kommunikations- und Medienwissenschaft. VS Verlag für Sozialwissenschaften, 2010.
- [24] B. Moschner und A. Anschütz. “Kombination und Integration von qualitativen und quantitativen Forschungsmethoden in einem interdisziplinären Forschungsprojekt”. In: *Didaktik der Informatik – Möglichkeiten empirischer Forschungsmethoden und Perspektiven der Fachdidaktik*. Hrsg. von I. Diethelm u. a. Gesellschaft für Informatik e.V., 2010.
- [25] PHINEO gemeinnützige AG. *Was kostet eine Evaluation?* Online verfügbar unter <https://www.wirkung-lernen.de/wirkungsanalyse/vorbereiten/kosten-evaluation/>, Stand 09.09.2020. 2020.
- [26] D. Rodrigues. *Small Organisations: A Challenge for Monitoring and Evaluation*. Online verfügbar unter <http://www.developmentresearch.eu/?p=478>, Stand 28.06.2020. 2019.
- [27] L. Rosenberg, L. Posner und E. Hanley. *Project Evaluation and the Project Appraisal Reporting System*. Abschlussbericht, vorgelegt bei US Agency for International Development (USA AID), contract no. csd-2510. 1970.
- [28] P. H. Rossi, H. E. Freeman und S. Rosenbaum. *Evaluation: A Systematic Approach*. SAGE Publications, 1982.

- [29] W. Schlicht. *Gesundheit systematisch fördern: Von der Absicht zur Realisierung. essentials*. Springer, 2018. DOI: 10.1007/978-3-658-20961-2.
- [30] J. Schmitt u. a. “Evaluation einer medienpädagogischen Maßnahme im Schulunterricht zur Prävention der Wirkung extremistischer Online-Propaganda: Ein Überblick”. In: *Propaganda und Prävention*. Springer, 2020, S. 61–79. DOI: 10.1007/978-3-658-28538-8\_4.
- [31] W. Schulz. “Zum Stellenwert qualitativer Untersuchungsmethoden in der empirischen Forschung”. In: *Österreichische Zeitschrift für Soziologie* 2.5 (1977), S. 63–68.
- [32] M. Scriven. “Die Methodologie der Evaluation”. In: *Evaluation. Beschreibung und Bewertung von Unterricht, Curricula und Schulversuchen*. Hrsg. von C. Wulf. Bd. 8. Erziehung in Wissenschaft und Praxis. R. Piper & Co. Verlag, 1972, S. 60–91.
- [33] M. Scriven. *Evaluation Thesaurus*. SAGE Publications, 1991.
- [34] M. Scriven. “The Logic of Evaluation”. In: *Proceedings of the Ontario Society for the Study of Argumentation Conference*. University of Windsor, 2007, S. 1–16.
- [35] D. Sriskandarajah. *Five reasons donors give for not funding local NGOs directly*. Online verfügbar unter <https://www.theguardian.com/global-development-professionals-network/2015/nov/09/five-reasons-donors-give-for-not-funding-local-ngos-directly>, Stand 09.11.2020. 2020.
- [36] I. Steinke. “Gütekriterien qualitativer Forschung”. In: *Qualitative Forschung. Ein Handbuch, 9. Auflage*. Hrsg. von U. Flick, E. von Karsdorff und I. Steinke. Rowohlt, 2010, S. 319–331.
- [37] R. Stockmann. *Was ist eine gute Evaluation? Einführung zu Funktionen und Methoden von Evaluationsverfahren*. Bd. 9. Arbeitspapiere. Centrum für Evaluation, Universität des Saarlandes, 2004.
- [38] R. Stockmann. *Handbuch zur Evaluation*. 6. Aufl. Münster: Waxmann, 2007.
- [39] R. Stockmann. *Evaluation. Eine Einführung*. Leverkusen: Verlag Barbara Budrich, 2010.

- [40] R. Stockmann und W. Meyer. *Evaluation von Nachhaltigkeit*. 3. Auflage. Kaiserslautern: Studienbrief Nr. 1020 des Fernstudiengangs Nachhaltige Entwicklungszusammenarbeit der TU Kaiserslautern., 2017.
- [41] D. Stufflebeam. “The CIPP model for program evaluation”. In: *Evaluation models*. Bd. 6. Evaluation in Education and Human Services. Springer, 1983, S. 117–141. DOI: 10.1007/978-94-009-6669-7\_7.
- [42] Team Technologies. *The World Bank Logframe Handbook, A Logical Framework Approach to Project Cycle Management*. The World Bank, 2008.
- [43] Verband Entwicklungspolitik und Humanitäre Hilfe deutscher Nichtregierungsorganisationen (VENRO). *Verhaltenskodex Transparenz Organisationsführung und Kontrolle*. 2018.
- [44] Verband Entwicklungspolitik und Humanitäre Hilfe deutscher Nichtregierungsorganisationen (VENRO). *Stellungnahme Kofinanzierung*. Online verfügbar unter <https://venro.org/publikationen/detail/stellungnahme-kofinanzierung>, Stand 09.09.2020. 2020.
- [45] C. H. Weiss. *Evaluierungsforschung: Methoden zur Einschätzung von sozialen Reformprogrammen*. Hrsg. von M. Küchler. Bd. 10. Studienbücher zur Sozialwissenschaft. VS Verlag für Sozialwissenschaften, 1974. DOI: 10.1007/978-3-322-85590-9.
- [46] A. von Werthern. “Zusammenfassendes zum Verständnis von Programmtheorie”. In: *Theoriebasierte Evaluation: Entwicklung und Anwendung eines Verfahrensmodells zur Programmtheoriekonstruktion*. Wiesbaden: Springer Fachmedien Wiesbaden, 2020, S. 261–265. DOI: 10.1007/978-3-658-27579-2\_20.
- [47] M. Wirtz, Hrsg. *Dorsch - Lexikon der Psychologie*. 19. Aufl. Hogrefe, 2020.



## A Anhänge



Frage 8: Welche Ressource für Evaluation einsetzen?	IOG intern, Wissensmanagement, Umfragedata, Mitarbeiter, Subevaluati...	Indirekt Zielgruppen von Nachfolgeprojekten, direkt IOG intern durch Lernen	Projektbesichtigte, Mitarbeitende von lokalen POs durch Kapazitätsklärung, Mitarbeiter von IOG	IOG intern, indirekt durch Lerneffekte von IOG auch die Begünstigten	IOG intern durch lernen, indirekt neue Projektpartner.	IOG intern, Lernende sind u.U. schon weg, wenn eine Evaluation stattfindet. Indirekt Zielgruppe des nächsten Projekts, Zielgruppe des evaluierten Projekts, wenn nachgeprüft werden kann	Stakeholder, weil sie die Ergebnisse der Evaluation in der Praxis umsetzen und daraus lernen und dann Dinge besser machen können	Partner, IOG intern, über übermäßige Kontrolle sein, sondern eine Möglichkeit	Bisher dachte ich IOG, um zu lernen, (...) Sicher profitieren aber auch die Partnerorganisationen, die IOG unterstützen. Sie haben dasselbe Potential von einer Evaluation zu profitieren wie auch IOG.
Frage 9: Wieviel Geld für Evaluation investieren?	29.95 - 29.95 Es gibt viel Personal, wenig Wissen und wenig Geld. Evaluationsysteme sollten Wissen im Anzeile muss werden darüber innerhalb IOG aufgebaut werden	18.07 - 21.05 Indirekt Zielgruppen von Nachfolgeprojekten, direkt IOG intern durch Lernen	22.43 - 27.44 Geld ist wichtig aber nicht erscheidend. Mitarbeiter, die arbeiten mit erfahrenen Studierenden vorhanden. Personal bezahlt, knapp Personal gibt es viel. Wissen kann knapp sein, aber kann auch vorhanden sein	26.33 - 29.40 IOG intern, indirekt durch Lerneffekte von IOG auch die Begünstigten	22.01 - 24.06 IOG intern durch lernen, indirekt neue Projektpartner.	23.06 - 26.13 IOG intern, Lernende sind u.U. schon weg, wenn eine Evaluation stattfindet. Indirekt Zielgruppe des nächsten Projekts, Zielgruppe des evaluierten Projekts, wenn nachgeprüft werden kann	32.03 - 35.12 Stakeholder, weil sie die Ergebnisse der Evaluation in der Praxis umsetzen und daraus lernen und dann Dinge besser machen können	36.54 - 41.48 Partner, IOG intern, über übermäßige Kontrolle sein, sondern eine Möglichkeit	24.54 - 27.23 Bisher dachte ich IOG, um zu lernen, (...) Sicher profitieren aber auch die Partnerorganisationen, die IOG unterstützen. Sie haben dasselbe Potential von einer Evaluation zu profitieren wie auch IOG.
Frage 10: Wieviel Geld für Evaluation investieren?	31.50 - 35.44 Einzelentscheidungen, zwischen 10 und 30 Prozent.	24.09 - 27.31 20 Prozent, investiert in Personal, Reisekosten, ggf. lokale Evaluatoren könnte uns zu 10% bringen.	27.44 - 28.48 E. lohnt in diesem Fall nicht. Projektbündel lohnt sich aber. 10% von 50.000€ wäre denkbar.	32.17 - 32.30 Pauschale Antwort nicht möglich und nicht nötig. Evaluation soll gut gemacht werden, egal was es kostet (0% bei digital und Ehrenamt, 95% bei extern).	26.12 - 28.31 Keine Einzelevaluationen, besser Projektbündel	26.13 - 29.24 Pauschal lässt sich das nicht sagen, im Mittel schätze ich 10-25%.	31.05 - 32.03 Geld ist kritisch	41.48 - 44.24 Studierende einsetzen, die haben Zeit, Wissen muss generiert werden	27.23 - 29.49 Das kommt auf den Nutzen an. Wenn 10 andere Projekte investiert werden, dann ist der Prozentsatz investiert von (kurze Unterbrechung wg. Verbindungsschwierigkeiten) Realistisch wird es eher auf 5-10% hinauslaufen.
Frage 11: Warum ist die Investitionslage momentan so?	35.44 - 43.43 0%, aber gelegentlich wird per Ausreizen auch ein alles investiert. Es gibt ein Problem, dass es Kosten von etwa 10% des Projektbudgets. Wenn wir eine (...) Zentrale bauen ist das natürlich gut, aber es ist ein Risiko. Ein Verständnis für regelmäßige Evaluationen muss so aufbereitet sein, dass jeder die Risiken selbst durchdringen kann.	27.31 - 32.23 Kosten reduzieren, wenig langfristige Partnerschaften, daher wenig zusammenhängende Projekte, die man langfristig evaluieren kann. Wissensmanagement muss verbessert werden. Projektkosten pro einzelnes Projekt sind nicht gering. Sponsoring ist schwer, die Projektkosten sind nicht gering. Projektkosten sind nicht gering. Projektkosten sind nicht gering. Projektkosten sind nicht gering.	29.13 - 31.48 Aktuell gibt es keine Evaluationskosten	32.30 - 36.20 Wir müssen deutlich mehr evaluieren.	32.03 - 39.29 Evaluation macht nicht soviel Spaß. Projektkosten sind nicht mehr aktiv im Ver...	46.32 - 50.50 Spendern lässt sich nicht erklären, dass 30% Evaluation bezahlt werden muss.	44.24 - 46.32 20-30% ?	44.24 - 46.32 20-30% ?	29.49 - 31.13 Keine Idee, wurde eher an andere Projektleitende angeknüpft, um kleine zusätzlichen Kosten zu betreiben
Frage 12: Wie lange aktiv?	43.43 - 43.51 anderthalb Jahre	34.44 - 34.57 Geschäftsstelle Berlin	32.44 - 32.59 zweiteinhalb Jahre	37.56 - 36.53 8 Jahre	31.43 - 33.02 10 Jahre	38.20 - 38.36 4 Jahre	52.30 - 52.56 7 Jahre	35.50 - 36.11 6 Jahre	33.57 - 35.50 Kompetenzgruppe, Projektbearbeitender mit Bezug zu Projektumsetzung

## A.2 DeGEval Standards

	<b>Nützlichkeit</b>	<b>Durchführbarkeit</b>	<b>Fairness</b>	<b>Genauigkeit</b>
<b>N 1 Identifizierung der Beteiligten und Betroffenen</b>	Die am Evaluationsgegenstand oder an der Evaluation Beteiligten sowie die von Evaluationsgegenstand oder Evaluation Betroffenen sollen vorab identifiziert werden, damit deren Interessen und Informationsbedürfnisse geklärt und so weit wie möglich bei der Anlage der Evaluation berücksichtigt werden können. Es soll deutlich bestimmt sein, welche Zwecke mit der Evaluation verfolgt werden, so dass die Beteiligten und Betroffenen Position dazu beziehen und die Evaluierenden einen klaren Arbeitsauftrag verfolgen können.	<b>D 1 Angemessene Verfahren</b> Evaluationsverfahren, einschließlich der Verfahren zur Beschaffung notwendiger Informationen, sollen so gewählt werden, dass einerseits die Evaluation professionell und den Erfordernissen entsprechend umgesetzt wird und andererseits der Aufwand für die Beteiligten und Betroffenen in einem adäquaten Verhältnis zum intendierten Nutzen der Evaluation gehalten wird.	<b>F 1 Formale Vereinbarungen</b> Die Rechte und Pflichten der an einer Evaluation beteiligten Parteien (was, wie, von wem, wann getan werden soll und darf) sollen schriftlich festgehalten werden.	<b>G 1 Beschreibung des Evaluationsgegenstandes</b> Sowohl das Konzept des Evaluationsgegenstands als auch seine Umsetzung sollen genau und umfassend beschrieben und dokumentiert werden.
<b>N 2 Klärung der Evaluationszwecke</b>	Wer Evaluationszwecke mit der Evaluation verfolgt werden, so dass die Beteiligten und Betroffenen Position dazu beziehen und die Evaluierenden einen klaren Arbeitsauftrag verfolgen können.	<b>D 2 Diplomatisches Vorgehen</b> Evaluierungen sollen so geplant und durchgeführt werden, dass eine möglichst hohe Akzeptanz der verschiedenen Beteiligten und Betroffenen in Bezug auf Vorgehen und Ergebnisse der Evaluation erreicht werden kann.	<b>F 2 Schutz individueller Rechte</b> Evaluierungen sollen so geplant und durchgeführt werden, dass Rechte, Sicherheit und Würde der in eine Evaluation einbezogenen Personen geschützt sind.	<b>G 2 Kontextanalyse</b> Der Kontext des Evaluationsgegenstandes soll ausreichend umfassend und detailliert analysiert sowie bei der Interpretation von Ergebnissen berücksichtigt werden.
<b>N 3 Kompetenz und Glaubwürdigkeit des Evaluators/der Evaluatorin</b>	Wer Evaluierungen durchführt, soll fachlich und methodisch kompetent sein, damit für die Evaluation und ihre Ergebnisse ein Höchstmaß an Glaubwürdigkeit und Akzeptanz erreicht wird.	<b>D 3 Effizienz von Evaluation</b> Der Aufwand für Evaluation soll in einem angemessenen Verhältnis zum Nutzen der Evaluation stehen.	<b>F 3 Umfassende und faire Prüfung</b> Evaluierungen sollen die Stärken und die Schwächen des Evaluationsgegenstandes möglichst fair und umfassend prüfen und darstellen.	<b>G 3 Beschreibung von Zwecken und Vorgehen</b> Zwecke, Fragestellungen und Vorgehen der Evaluation, einschließlich der angewandten Methoden, sollen so genau dokumentiert und beschrieben werden, dass sie nachvollzogen und beurteilt werden können.
<b>N 4 Auswahl und Umfang der Informationen</b>	Auswahl und Umfang der erfassten Informationen sollen die adäquate Beantwortung der zu untersuchenden Fragestellungen zum Evaluationsgegenstand ermöglichen und die Informationsbedürfnisse der Auftraggebenden und weiterer Beteiligter und Betroffener berücksichtigen.	<b>F 4 Unparteiische Durchführung und Berichterstattung</b> Die Evaluation soll unterschiedliche Sichtweisen von Beteiligten und Betroffenen auf Gegenstand und Ergebnisse der Evaluation beachten. Der gesamte Evaluationsprozess sowie die Evaluationsberichte sollen die unparteiische Position der Evaluierenden erkennen lassen.	<b>G 4 Angabe von Informationen</b> Die im Rahmen einer Evaluation genutzten Informationsquellen sollen hinreichend genau dokumentiert werden, damit die Verlässlichkeit und Angemessenheit der Informationen eingeschätzt werden können.	

<b>N 5 Transparenz von Werthaltungen</b>	Werthaltungen der Beteiligten und Betroffenen, die sich in deren Perspektiven und Annahmen manifestieren und einen Einfluss haben auf die Evaluation und Interpretation ihrer Ergebnisse, sollten transparent dokumentiert werden, um Evaluationsergebnisse besser einordnen zu können.
<b>N 6 Vollständige Information und Klarheit der Berichterstattung</b>	Evaluationsberichte sollen alle wesentlichen Informationen zur Verfügung stellen und für ihre Adressatinnen und Adressaten verständlich und nachvollziehbar sein.
<b>N 7 Rechtzeitigkeit der Evaluation</b>	Evaluationsvorhaben sollen so rechtzeitig begonnen und abgeschlossen werden, dass ihre Ergebnisse in anstehende Entscheidungs-, Verbesserungs- oder sonstige Nutzungsprozesse einfließen können.
<b>N 8 Nutzung und Nutzen der Evaluation</b>	Planung, Durchführung und Berichterstattung einer Evaluation sollen die Beteiligten und Betroffenen dazu ermuntern, die Evaluation mitzutragen und ihre Ergebnisse zu nutzen.

<b>F 5 Offenlegung von Ergebnissen und Berichten</b>	Evaluationsergebnisse und -berichte sollen allen Beteiligten und Betroffenen soweit wie möglich zugänglich gemacht werden.	Erhebungsverfahren und Datenquellen sollen so gewählt werden, dass die Zuverlässigkeit der gewonnenen Daten und ihre Gültigkeit bezogen auf die Beantwortung der Evaluationsfragestellungen nach fachlichen Maßstäben sichergestellt sind. Die fachlichen Maßstäbe sollen sich an den Gütekriterien der empirischen Forschung orientieren.
<b>G 5 Valide und reliable Informationen</b>		Die in einer Evaluation gesammelten, aufbereiteten, analysierten und präsentierten Informationen sollen systematisch auf Fehler geprüft werden.
<b>G 6 Systematische Fehlerprüfung</b>		Qualitative und quantitative Informationen einer Evaluation sollen nach fachlichen Maßstäben angemessen und systematisch analysiert werden, damit die Fragestellungen der Evaluation beantwortet werden können.
<b>G 7 Angemessene Analyse qualitativer und quantitativer Informationen</b>		Die in einer Evaluation getroffenen wertenden Aussagen sollen auf expliziten Kriterien und Zielwerten basieren. Schlussfolgerungen sollen ausdrücklich und auf Grundlage der erhobenen und analysierten Daten begründet werden, damit sie nachvollzogen und beurteilt werden können.
<b>G 8 Begründete Bewertungen und Schlussfolgerungen</b>		Meta-Evaluationen evaluieren Evaluationen. Um dies zu ermöglichen, sollen Evaluationen in geeigneter Form dokumentiert, archiviert und soweit wie möglich zugänglich gemacht werden.
<b>G 9 Meta- Evaluation</b>		







## **Eigenständigkeitserklärung**

Ich versichere, dass ich diese Masterarbeit selbstständig und nur unter Verwendung der angegebenen Quellen und Hilfsmittel angefertigt und die den benutzten Quellen wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Berlin, \_\_\_\_\_

\_\_\_\_\_

Giulia Steen