

Stitteneder, Tanja

**Article**

## Wirkung wissenschaftlich analysieren: Methoden zur Evaluation sozialer Innovationen und gemeinwohlorientierter Organisationen

ifo Schnelldienst

**Provided in Cooperation with:**

Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

*Suggested Citation:* Stitteneder, Tanja (2025) : Wirkung wissenschaftlich analysieren: Methoden zur Evaluation sozialer Innovationen und gemeinwohlorientierter Organisationen, ifo Schnelldienst, ISSN 0018-974X, ifo Institut - Leibniz-Institut für Wirtschaftsforschung an der Universität München, München, Vol. 78, Iss. 02, pp. 39-44

This Version is available at:

<https://hdl.handle.net/10419/314785>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

Tanja Stitteneder

# Wirkung wissenschaftlich analysieren: Methoden zur Evaluation sozialer Innovationen und gemeinwohlorientierter Organisationen\*

Evaluationen spielen eine wichtige Rolle in der Politikgestaltung, denn sie können »reale Konsequenzen« haben: Sie hinterfragen alte Ideen, bieten neue Perspektiven und tragen zur Neuordnung der politischen Agenda bei (Weiss 1999)<sup>1</sup>. Obwohl neue Vorgaben und Leitlinien selten direkt auf Evaluationsergebnissen basieren, klären Evaluationen auf, vermitteln Erkenntnisse und informieren.

Politikevaluation bzw. *Policy Evaluation* ist ein strukturierter und objektiver Prozess zur Evaluation laufender oder abgeschlossener Politik- oder Reforminitiativen. Sie zielt darauf ab, die Relevanz, Zielerreichung, Effizienz, Effektivität, Auswirkungen und Nachhaltigkeit einer Initiative zu bestimmen (OECD 2020). Ähnliche Prinzipien finden auch bei der Evaluation gemeinwohlorientierter Initiativen und Organisationen Anwendung. Mit Wirkungsanalysen und Evaluationen können Veränderungen, die durch spezifische Interventionen herbeigeführt wurden, erfasst und bewertet werden (Phineo 2023a, b).

Mit der Verabschiedung der »Nationalen Strategie für Soziale Innovationen und Gemeinwohlorientierte Unternehmen« setzte die Bundesregierung (2023) ein klares Zeichen: Sie erkennt die Bedeutung dieser Akteure für den gesellschaftlichen Fortschritt an und verpflichtet sich zu deren aktiver Förderung und Weiterentwicklung. Die Strategie unterstreicht die Wichtigkeit von Wirkungsorientierung und -messung, indem sie ihr mit »Wirkungsorientierung und Wirkungsmessung als Standard etablieren« ein eigenständiges Handlungsfeld widmet (Bundesregierung 2023, S. 42). Die Strategie legitimiert somit nicht nur die Arbeit sozialer Innovator\*innen und gemeinwohlorientierter Unternehmen, sondern fördert deren zukünftige Entwicklung und Integration in das wirtschaftliche und soziale Gefüge Deutschlands. Ein Großteil der Maßnahmen sind bereits umgesetzt oder in Planung und es bleibt abzuwarten, wie die Umsetzung unter der neuen Regierung weiter vorangeht.

\* Vielen Dank an Rosieanna Sham für die gründliche Recherche und Zusammenstellung für diesen Artikel.

<sup>1</sup> Im Original lautet das Zitat: »...scholars in many countries have found that evaluation has real consequences: it challenges old ideas, provides new perspectives and helps to re-order the policy agenda.« (Weiss 1999).

## IN KÜRZE

Evaluationen untersuchen den Erfolg oder Misserfolg von Maßnahmen, Aktivitäten und Programmen. Sie bewerten die Effekte der jeweiligen Interventionen auf Zielgruppen oder die Gesellschaft allgemein und verwenden dabei methodische Ansätze wie Wirkungsmodelle, Indikatoren sowie qualitative und quantitative Analysemethoden. Dieser Artikel gibt einen Überblick über eine Auswahl verschiedener Ansätze und Analysemethoden, die in der Forschung und Politikberatung zum Einsatz kommen, und zeigt auf, wie sich einige dieser Methoden auch auf den weiteren sozialen Bereich übertragen lassen – teilweise geschieht dies bereits. Dabei ist es wichtig, die Realität gemeinwohlorientierter Initiativen und Organisationen im Blick zu behalten. Es ist klar, dass es nicht die eine perfekte Methode für alle Gegebenheiten und Organisationen gibt. Vielmehr geht es darum, den passenden Ansatz zu finden, der sowohl den Möglichkeiten der Organisation als auch der Komplexität ihrer Arbeit gerecht wird. Das Ziel ist es, Orientierung zu bieten und aufzuzeigen, wie Wirkungsanalysen und Evaluationen für soziale Innovationen und gemeinwohlorientierte Organisationen sinnvoll und mit angemessenem Aufwand gestaltet werden können.

Evaluationen können verschiedene Formen annehmen, von Ex-ante-Abschätzungen über Zwischenevaluationen bis hin zu Ex-post-Analysen. Sie können kurz-, mittel- oder langfristige Wirkungen untersuchen (Fraunhofer 2024). Für gemeinnützige Organisationen sind Evaluationen besonders wertvoll, da sie u. a.

- die Wirksamkeit der Intervention genauer bewerten,
- Lernprozesse fördern,
- zur Rechenschaftslegung beitragen,
- eine effiziente Ressourcenallokation ermöglichen,
- fundierte strategische Entscheidungen unterstützen,
- die Fördermittelakquise erleichtern.

Organisationen verfolgen unterschiedliche Ziele und Motivationen, wenn sie Wirkungsanalysen und Evaluationen durchführen und sind zudem äußerst vielfältig: Sie agieren in verschiedensten Tätigkeitsfeldern, unterscheiden sich hinsichtlich ihres Geschäftsmodells (sofern vorhanden), ihrer Größe, Organisationsstruktur und des Kontextes, den sie untersuchen. Diese Heterogenität führt zu einer breiten Palette von Herangehensweisen und Untersuchungsmethoden, die sich je nach Organisation und Kontext besser oder schlechter eignen.

**WIRKUNG STRATEGISCH DENKEN**

Die *Theory of Change* (ToC 2023) dient als Planungsinstrument, um den Weg von Aktivitäten zu langfristigen Zielen systematisch und nachvollziehbar zu gestalten. Der Prozess wird dabei rückwärts durchdacht und wurde von Organisationen wie Phineo (2023a, b) und der Bertelsmann Stiftung (2010) in ihrer Anwendung weiterentwickelt (vgl. Abb. 1). Zunächst werden die langfristigen gesellschaftlichen Ziele (*Impact*) definiert, um auf die kurz- und mittelfristigen Effekte auf die Zielgruppen (*Outcomes*) sowie die direkten Ergebnisse (*Outputs*) zu schließen. Abschließend werden die notwendigen Voraussetzungen und Aktivitäten (*Inputs*) identifiziert, die zur Erreichung dieser Ziele erforderlich sind:

- *Impact*: Langfristige, gesellschaftliche Wirkungen der Intervention. Beispiel: Beitrag zur Reduzierung des Fachkräftemangels durch erfolgreiche Integration von Geflüchteten in den Arbeitsmarkt.
- *Outcome*: Mittelfristige Effekte, wie Wissenssteigerungen, Verhaltensänderungen oder Kompetenzsteigerungen bei der Zielgruppe. Beispiel: Geflüchtete sind in der Lage, sich erfolgreich für Arbeit in Deutschland zu bewerben, und werden angestellt.
- *Output*: Unmittelbare, quantifizierbare Ergebnisse der Aktivitäten. Beispiel: Anzahl erreichter Geflüchteter, die erfolgreich an Schulungen zu Deutschland- und branchenspezifischem Wissen teilgenommen haben.

- *Input*: Eingesetzte Ressourcen (personell, finanziell/materiell, zeitlich) für die Umsetzung einer Maßnahme. Beispiel: Personelle, finanzielle und zeitliche Ressourcen für die Vorbereitung und Durchführung von Schulungen zur Arbeitsmarktintegration von Geflüchteten.

**EVALUATIONSMETHODEN**

Für die Auswahl der passenden Analysemethoden müssen die angestrebten Ziele und Aktivitäten (und welche Kriterien oder Indikatoren sie erfassen) gut definiert sein, denn sie bilden das Fundament für die folgende Untersuchung. Ein solides Verständnis dieser Grundlagen hilft, um beurteilen zu können, welcher Ansatz sich für den jeweiligen Kontext eignet. Grob lassen sich die Methoden in quantitative, qualitative und eine Mischung beider Ansätze unterteilen.<sup>2</sup>

**Quantitative Evaluationsmethoden**

Quantitative Evaluationsmethoden sind besonders geeignet, um den langfristigen und möglichst kausalen<sup>3</sup> Impact von Interventionen zu messen und zu analysieren. Diese Methoden erfordern in der Regel viel Zeit, Fachkenntnisse und oft auch ein entsprechendes Budget. Sie eignen sich vor allem für Organisationen, die ihre Wirkung auf einer breiteren und langfristigeren Ebene nachweisen und evidenzbasierte Entscheidungen treffen möchten.

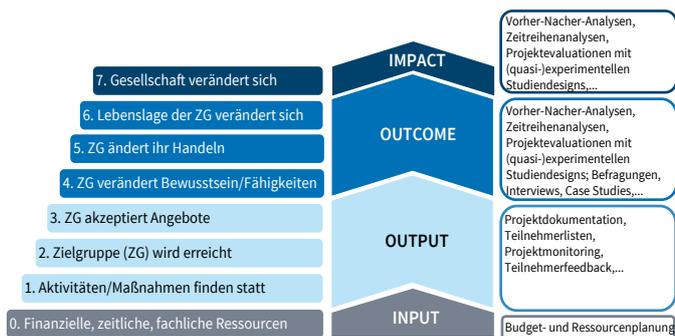
**Experimentelle Methoden**

An der Spitze der quantitativen Methoden stehen randomisierte kontrollierte Studien (*Randomized Controlled Trials* – RCTs). Sie gelten oft als »Goldstandard« unter den Evaluationsmethoden. Teilnehmende werden ohne ihr Wissen zufällig in eine Interventionsgruppe und eine Kontrollgruppe eingeteilt (sog. Randomisierung). Die Interventionsgruppe erhält die Maßnahme (d. h. die Intervention/das Treatment), während die Kontrollgruppe diese nicht erhält. Nach einem festgelegten Zeitraum werden die Ergebnisse beider Gruppen verglichen, um die Wirkung zu ermitteln. Mit dieser Methode sind hohe Objektivität und interne Validität zu erwarten. Die Studie von Resnjanskij et al. (2024) untersuchte mittels eines RCTs die Auswirkungen eines Mentoring-Programms auf benachteiligte Jugendliche in Deutschland. Die Forschenden analysierten, ob Mentoring familiäre Nachteile beim Übergang von der Schule in den Beruf ausgleichen kann. Die Ergebnisse zeigen, dass das Programm posi-

<sup>2</sup> Die folgenden Ausführungen beziehen sich neben den angegebenen Quellen im Text auf Gertler et al. (2016), Loi und Caetano Rodrigues Jorge Rodrigues Ferro (2012) und Revillard (2023).

<sup>3</sup> Ein kausaler Zusammenhang beschreibt eine eindeutige Ursache-Wirkungs-Beziehung zwischen zwei Variablen. Dies bedeutet, dass eine bestimmte Handlung oder ein Ereignis (die Ursache) direkt zu einem spezifischen Ergebnis oder einer Veränderung (die Wirkung) führt.

Abb. 1  
Stufen der Wirkung nach Input, Output, Outcome und Impact\*



\* Nach der Wirkungstreppe von Phineo und dem IOOI-Modell. Quelle: Phineo (2023a, b); Bertelsmann Stiftung (2010); Darstellung des ifo Instituts.

tive Effekte auf verschiedene Aspekte des Lebens der Jugendlichen hatte, einschließlich ihrer schulischen Leistungen, sozialen Fähigkeiten und Zukunftsperspektiven. Ebenfalls beobachtet wird eine Verbesserung der Beziehungen zu den Eltern, was möglicherweise auf eine Milderung familiärer Nachteile hindeutet.

RCTs sind besonders geeignet, wenn es darum geht, die direkte Wirkung einer Intervention nachzuweisen. Sie erfordern jedoch eine sorgfältige Planung, ethische Überlegungen und oft erhebliche Ressourcen. Darüber hinaus ist die präzise Identifikation der zugrundeliegenden Ursache-Wirkungs-Mechanismen von entscheidender Bedeutung. Werden diese nicht korrekt erfasst, besteht die Gefahr, dass die Messung am eigentlichen Ziel vorbeigeht und irrelevante oder irreführende Ergebnisse liefert. Ein wegweisendes Beispiel ist die Arbeit von Banerjee, Duflo und Kremer, die dafür mit dem Nobelpreis ausgezeichnet wurden (The Committee for the Prize in Economic Sciences in Memory of Alfred Nobel 2019). Durch den Einsatz von RCTs gelang es ihnen, den Fokus auf praxistaugliche Maßnahmen zur Armutsbekämpfung zu setzen. Ihre Forschung belegt unter anderem, dass Vertragslehrkräfte kosteneffizient die Lernergebnisse verbessern, Subventionen den Zugang zur Gesundheitsvorsorge fördern und Mikrokredite geringere Entwicklungseffekte haben als oft angenommen.

### Quasi-experimentelle Methoden

Wenn RCTs aus ethischen oder praktischen Gründen nicht durchführbar sind, kann auf quasi-experimentelle oder nicht-experimentelle Methoden der kontrafaktischen Wirkungsanalyse zurückgegriffen werden. Dabei werden die Ergebnisse der Intervention mit den Ergebnissen verglichen, die erzielt worden wären, wenn die Intervention nicht durchgeführt worden wäre. Dies soll idealerweise ermöglichen, den Anteil der beobachteten Veränderungen zu identifizieren, der auf die Wirkung der Intervention zurückzuführen ist. Dabei wird berücksichtigt, dass beobachtete Ergebnisse nicht nur durch die Intervention, sondern auch durch externe Faktoren (z. B. eine globale Pandemie) beeinflusst werden können. Ein »kontrafaktisches Szenario« ist also ein hypothetisches »Was-wäre-wenn«-Szenario: Was wäre passiert, wenn die Intervention nicht stattgefunden hätte? Diese Betrachtung hilft dabei, die spezifische Wirkung der Intervention isoliert von anderen Faktoren zu verstehen. Wenn man beispielweise die Auswirkungen eines Bildungsprogramms untersuchen möchte, könnte man kontrafaktisch untersuchen, wie sich die Teilnehmenden entwickelt hätten, wenn sie nicht am Programm teilgenommen hätten.

Eine häufig verwendete Methode ist die *Difference-in-Differences-Analyse* (DiD), die eine Interventions- und eine Kontrollgruppe über die Zeit vergleicht (Fougère und Jacquemet 2023). DiD interpretiert den

Unterschied in der Veränderung zwischen den Gruppen als Wirkung der Maßnahme. Allerdings setzt diese Methode voraus, dass zwischen den Gruppen keine versteckten Unterschiede existieren.

Andere quasi-experimentelle Ansätze wie das *Regression-Discontinuity-Design* (RDD), das *Matching-Verfahren* und der *Instrumentvariablen-Ansatz* (IV) sind ebenso effektive Methoden zur Analyse kausaler Effekte, ohne dass eine *randomisierte* Zuweisung zu einer Interventions- oder Kontrollgruppe erforderlich ist. Das RDD nutzt eine festgelegte Schwelle, um Teilnehmende von Nicht-Teilnehmenden zu unterscheiden, und vergleicht deren Ergebnisse, um den Einfluss eines Programms, Projekts o.Ä. zu bewerten.

Beim Matching wird eine Vergleichsgruppe gebildet, die den Teilnehmenden in relevanten statistischen Merkmalen möglichst ähnlich ist. Ein Beispiel hierfür ist das Programm *Aid to Families with Dependent Children* (AFDC)<sup>4</sup> (Friedlander und Robins 1995). In dieser Studie basiert eine der vier Vergleichsgruppen auf statistischem Matching, wobei die beobachteten Hintergrundmerkmale jedes Mitglieds der Programmgruppe mit denen aller potenziellen Kandidat\*innen verglichen werden. Das Programmgruppenmitglied, das den Vergleichsgruppenmitgliedern am ähnlichsten war, wird ausgewählt. Dieses Verfahren ermöglicht die Bildung einer Kontrollgruppe, die der Interventionsgruppe in entscheidenden Merkmalen ähnlich ist, ohne dass eine zufällige Zuweisung erfolgt. Dadurch können potenzielle Selektionsverzerrungen minimiert und die Programmeffekte präziser geschätzt werden.

Die Instrumentvariablen-Methode wird verwendet, um kausale Zusammenhänge zu identifizieren, indem sie eine exogene Variable (Instrument) nutzt, die mit der erklärenden Variable korreliert, aber nicht direkt die abhängige Variable beeinflusst, um Verzerrungen durch unbeobachtete Störfaktoren zu umgehen.

### Weitere quantitative Analysemethoden

Neben experimentellen und quasi-experimentellen Methoden können Beobachtungsstudien, die keine experimentelle Manipulation beinhalten, verwendet werden. Beobachtungsstudien können sowohl für bereits bestehende Daten (Sekundärdaten) oder speziell im Rahmen der Studie erhobene Daten (Primärdaten) angewendet werden. Aufgrund des fehlenden experimentellen Charakters sind sie jedoch nicht in der Lage, Kausalität nachzuweisen. Zwei wichtige Ansätze sind Querschnitts- und Längsschnittstudien.

Querschnittsstudien (*cross-sectional studies*) betrachten verschiedene Untersuchungsgruppen oder Stichproben zu einem spezifischen Zeitpunkt. Längsschnitt- oder Zeitreihenstudien (*longitudinal studies*) hingegen beobachten Veränderungen und Entwicklun-

<sup>4</sup> Mit dem Programm *Aid to Families with Dependent Children* (AFDC) wurden in den USA vier Programme finanziert, die für alleinerziehende Familien mit hauptsächlich weiblichen Hauptverdienerinnen durchgeführt wurden und darauf abzielten, die Erwerbstätigkeit durch Arbeits- und Trainingsangebote zu steigern.

gen von Variablen über einen längeren Zeitraum und zu mehreren Zeitpunkten. Sie lassen sich weiter unterteilen in Trendstudien (untersuchen allgemeine Veränderungen mit unterschiedlichen Stichproben) und Panelstudien (analysieren individuelle und gruppenspezifische Veränderungen einer Stichprobe zu mehreren Zeitpunkten). Trendstudien eignen sich, um Aussagen über größere Bevölkerungsgruppen zu machen, während Panelstudien individuelle Veränderungen untersuchen. Ergänzend dazu zeigen Korrelationsstudien, ob ein Zusammenhang zwischen Variablen besteht. Sie liefern also ebenso keine Informationen über Kausalität, können jedoch wertvolle Einsichten in Zusammenhänge liefern, die später durch genauere (quasi-) experimentelle Designs überprüft werden können. In der Finanzwirtschaft werden Korrelationsstudien beispielsweise häufig eingesetzt, um Beziehungen zwischen verschiedenen Marktindikatoren zu analysieren (Campbell et al. 1997). Vorher-Nachher-Befragungen können ein guter Einstieg in die quantitative Analyse mit eigener Datenerhebung sein. Dabei werden Ergebnisse einer Gruppe vor und nach der Implementierung eines Programms miteinander verglichen. Die Methode ist sinnvoll, wenn die Teilnehmenden zu unterschiedlichen Zeitpunkten befragt werden können und die Programmimplementierung klar abgegrenzt ist. Externe Einflüsse über die Zeit hinweg könnten jedoch Verzerrungen verursachen.

Bei der Erstellung von Fragebögen gilt es, einige Kernpunkte zu beachten, um aussagekräftige Ergebnisse zu erzielen (Bradburn et al. 2004): Zum einen sollten die Fragen relevant, prägnant und effizient sein, d. h., sie sind so gewählt und sprachlich formuliert, dass sie auch das erfassen, was untersucht werden soll. Um eine hohe Rücklaufquote zu gewährleisten, kann es zudem sinnvoll sein, die Anzahl der Fragen zu begrenzen. Außerdem verbessern eine logische Reihenfolge der Fragen und die Vermeidung komplexer Verzweigungen die Benutzerfreundlichkeit. Bei der Wahl des Fragetyps sollte beachtet werden, dass geschlossene Fragen – es werden Antwortmöglichkeiten zur Auswahl gestellt – die Beantwortung sowie die spätere Auswertung erleichtern, während offene Fragen – die Teilnehmenden geben Antworten in einem Freitextfeld ein – ergänzende Einsichten liefern können. Bei geschlossenen Fragen ist es zudem wichtig, alle möglichen Antwortoptionen abzubilden.

Die Wahl der Methode hängt davon ab, ob sie sich für die jeweilige Situation und Umstände eignet und ob ausreichend Ressourcen, insbesondere Fachkenntnis, vorhanden sind. Quantitative Methoden bieten Vorteile wie präzise Messungen, hohe Generalisierbarkeit und die Möglichkeit tiefgehender statistischer Analysen, die die Identifikation von Mustern und Zusammenhängen ermöglichen. Allerdings können sie den Kontext vernachlässigen (Ottmann 2021), sind oft wenig flexibel und erfordern einen hohen Ressourcen- und Zeiteinsatz. Aus diesem Grund eignen sie sich in

der Regel nur für die Analyse von Wirkungszielen auf den höheren Stufen der Wirkungstreppe.

Organisationen mit begrenzten Ressourcen, die dennoch quantitativ arbeiten möchten, können mit standardisierten Befragungen (wie Vorher-Nachher-Befragungen) der Zielgruppe und Stakeholder zu einem oder noch besser mehreren Zeitpunkten dennoch wertvolle Erkenntnisse gewinnen, auch wenn dies keine Aussagen zur Repräsentativität oder Kausalität erlaubt. Unabhängig vom gewählten Ansatz demonstrieren Organisationen mit dem Einsatz quantitativer Methoden ihr Engagement für evidenzbasierte Praxis und können ihre Wirkung auf einer soliden wissenschaftlichen Grundlage nachweisen.

### Qualitative Evaluationsmethoden

Quantitative Methoden sind zwar unverzichtbar, um das Ausmaß der Wirkung einer Maßnahme zu messen, stoßen aber oft an Grenzen, wenn es darum geht, Aufschluss über Herausforderungen, konkrete Erfolgs- oder Misserfolgskriterien oder unerwartete Effekte zu erhalten. Qualitative Methoden ergänzen quantitative Ansätze, indem sie Einblicke in komplexe Zusammenhänge, Herausforderungen und möglicherweise unerwartete Effekte und Mechanismen bieten.

### Process Tracing

*Process Tracing* (PT), ursprünglich in den Politikwissenschaften entwickelt, untersucht systematisch, wie ein bestimmtes Ergebnis entstanden ist, indem es die Schritte und Zusammenhänge innerhalb eines einzelnen Falls detailliert nachzeichnet (Collier 2011). PT eignet sich besonders für die Untersuchung komplexer Fälle, in denen mehrere Faktoren interagieren, und hilft, tiefere Einblicke in zugrundeliegende Prozesse zu gewinnen. Oxfams Arbeit in Ghana (Stedman-Bryce 2013) demonstriert den praktischen Nutzen von PT zur Untersuchung des Einflusses von Advocacy-Arbeit auf Veränderungen in der Gesundheitspolitik des Landes.<sup>5</sup> Durch die Analyse von Dokumenten, Interviews und Medienberichten konnte Oxfam zeigen, dass ihre Bemühungen maßgeblich dazu beitrugen, Regierungszusagen für eine kostenlose, allgemeine Gesundheitsversorgung zu erwirken.

### Interviews

Semi-strukturierte Interviews kombinieren strukturierte Elemente (vordefinierte Fragen) mit der Freiheit, das Gespräch spontan weiterzuentwickeln, um tiefergehende und detaillierte Informationen zu erhalten. Sie sind besonders nützlich, wenn es darum geht, komplexe Ziele, Themen oder Umsetzungsprozesse zu untersuchen sowie Zielgruppenreaktionen zu doku-

<sup>5</sup> Im Rahmen von Advocacy-Arbeit machen Personen, Institutionen und Organisationen auf die Interessen benachteiligter Gruppen aufmerksam.

mentieren (Kline 2021). Semi-strukturierte Interviews heben die Erfahrungen der Betroffenen hervor und bieten Flexibilität, zusätzliche Themen anzusprechen. Durch die Einbindung von Zitaten verleihen sie Evaluationen Authentizität und Tiefe. Interviews mit der Zielgruppe, Stakeholdern oder Wirkungsempfänger\*innen lassen sich im Vergleich relativ einfach umsetzen und eignen sich für nahezu alle Outcome- und Impactziele bzw. Stufen der Wirkungstreppe.

### Case Studies

*Case Studies* (Fallstudien) sind detaillierte, tiefgehende Untersuchungen einzelner Fälle, Ereignisse oder Phänomene, die dazu dienen, spezifische Aspekte zu analysieren und daraus allgemeine Schlussfolgerungen oder Erkenntnisse abzuleiten (Gerring 2004). Case Studies haben in der Entwicklungspolitik an Bedeutung gewonnen (Widner et al. 2020). Das Programm *Innovations for Successful Societies* (ISS) der Princeton University und der *Gender Development Index* (GDI) der Weltbank nutzen sie, um Erfolgsfaktoren in verschiedenen Bereichen – von Wasser-, Energie- und Sanitärversorgung über Gesundheitsinitiativen bis hin zu komplexen Regierungsprozessen und nationalen Entwicklungsstrategien – zu analysieren. Case Studies ermöglichen es, Kontextfaktoren zu identifizieren, Best-Practice-Lösungen zu dokumentieren und komplexe Wirkungszusammenhänge aufzudecken.

Neben den genannten Vorteilen haben qualitative Methoden auch Schwächen: Sie messen nicht das genaue Ausmaß von Wirkungen, lassen sich nur schwer wiederholen und können ressourcenintensiv sein. Zudem können Ergebnisse nur bedingt übertragen werden. Trotz dieser Herausforderungen können qualitative Methoden wertvolle Erkenntnisse liefern und quantitative Analysen sinnvoll ergänzen, insbesondere bei der Untersuchung komplexer sozialer Probleme oder in Bereichen, wo traditionelle quantitative Methoden an ihre Grenzen stoßen.

### Mixed-Methods Evaluationsansätze

Die Integration von *Mixed Methods* kann die Grenzen einzelner Forschungsansätze überwinden, indem quantitative und qualitative Methoden – entweder nacheinander oder parallel zueinander – eingesetzt werden (Pluye 2023).

Die Kombination bietet entscheidende Vorteile: Während beispielweise RCTs kausale Wirkungen quantifizieren können, decken qualitative Methoden wie Case Studies die zugrundeliegenden Prozesse und Wirkmechanismen auf. So entsteht ein umfassenderes Verständnis von Interventionen. USAID (2013) befürwortet diesen Ansatz nachdrücklich. Ein Beispiel ist die Studie von Wonodi et al. (2012) zur Impfstoffeinführung in Nigeria: Durch die Verknüpfung von Experteninterviews (qualitativ) mit einer sozialen Netzwerkanalyse (quantitativ) konnten die Forschenden

den formelle und informelle Entscheidungsprozesse detailliert rekonstruieren und die Einflussstrukturen der beteiligten Akteur\*innen aufzeigen.

### WICHTIGE ASPEKTE FÜR DIE ANALYSE UND EVALUATION

Bei der Auswahl der am besten geeigneten Methode und Herangehensweise gibt es einige wichtige Aspekte zu berücksichtigen. Viele Methoden können Kausalität nicht direkt bestimmen. Zudem arbeiten häufig verschiedene Akteure an komplexen gesellschaftlichen und ökologischen Fragestellungen, wodurch der Beitrag einer einzelnen Intervention schwer isolierbar ist. Langfristige Effekte zeigen sich zudem unter Umständen oft erst nach einer gewissen Zeit, was die Messung komplexer gestaltet und zu längeren Projektlaufzeiten führen kann. Es kann aber sinnvoll sein, sich mit anderen Akteuren, die an ähnlichen Themen arbeiten, zu vernetzen, auszutauschen und Synergieeffekte zu nutzen. Die Ansätze bzw. die Kombination mehrerer Ansätze haben unterschiedliche Stärken und liefern – sofern gut umgesetzt – wertvolle Erkenntnisse.

In jedem Fall spielen ethische Überlegungen eine zentrale Rolle in der Forschung, insbesondere bei der Datenerhebung. Die Studie von Schochet et al. (2008) zur Wirksamkeit des *Job Corps*-Programms verdeutlicht die ethischen Herausforderungen bei der Zuweisung von benachteiligten Jugendlichen zu Programm- oder Kontrollgruppen, da die Kontrollgruppe für drei Jahre keine Unterstützung erhielt. Solche Studien werfen ethische Bedenken auf, insbesondere wenn Teilnehmende aus benachteiligten Verhältnissen kommen. Eine Möglichkeit, dem zu begegnen, könnte es sein, das Programm zeitversetzt zu wiederholen oder der Kontrollgruppe alternative Unterstützungsmaßnahmen anzubieten. Allerdings ist das nicht immer möglich.

Darüber hinaus muss sichergestellt werden, dass die Studienteilnahme freiwillig ist, die Teilnehmenden über den Zweck und die Risiken der Studie informiert sind und ihre Daten vertraulich und unter Wahrung der geltenden rechtlichen Bestimmungen behandelt werden. Zudem ist zu berücksichtigen, dass Interviewende und Forschende durch ihre Interaktionen mit den Teilnehmenden das Antwortverhalten (und somit die Ergebnisse) bewusst oder unbewusst beeinflussen können. Die Wahrung der Unabhängigkeit und Objektivität ist daher essenziell, um Verzerrungen zu vermeiden und die Integrität der Forschung zu gewährleisten.

### FAZIT

Wirkungsanalysen und Evaluationen sind wertvolle Instrumente, um die Effektivität spezifischer Maßnahmen, Aktivitäten oder Programme zu beurteilen. Es stehen zahlreiche quantitative und qualitative Methoden zur Evaluation zur Verfügung. Die Wahl der geeig-

neten Methode hängt vom spezifischen Kontext und den verfügbaren Ressourcen ab. Umfassende wissenschaftliche Evaluationen erfordern in der Regel fachliche Unterstützung oder entsprechende Expertise. Dennoch können auch mit begrenzten Ressourcen und kleineren Studien wertvolle Erkenntnisse gewonnen werden, die zur Verbesserung der eigenen Arbeit beitragen und wichtige Botschaften nach außen kommunizieren. Da die Vielfalt und teilweise Komplexität einzelner Methoden zunächst Unsicherheit hervorrufen können, besteht ein pragmatischer Ansatz darin, mit einfacheren Analysemaßnahmen zu beginnen und diese schrittweise auszubauen.

Wirkungsanalysen und Evaluationen sind nicht nur für die Legitimation und Effizienz gegenüber Stakeholdern, Geldgebenden und der Öffentlichkeit von Bedeutung; sie alle erwarten nachvollziehbare Nachweise darüber, ob und wie die eingesetzten Ressourcen zu den beabsichtigten Veränderungen führen. Daneben ist es auch für die Organisation selbst wichtig zu verstehen, was wann, wie und warum wirkt – und was nicht. Lernen und kontinuierliche Weiterentwicklung stehen dabei im Mittelpunkt. Wirkungsanalysen und Evaluationen sind keine isolierten Ereignisse. Vielmehr fördern sie den kontinuierlichen Lernprozess innerhalb der Organisation. Sie sollten als integrale Bestandteile des organisatorischen Alltags verstanden werden, um eine wirkungsorientierte Steuerung der Organisation zu ermöglichen.

## REFERENZEN

- Bertelsmann Stiftung (2010), *Corporate Citizenship planen und messen mit der iooi-Methode – Ein Leitfaden für das gesellschaftliche Engagement von Unternehmen*, Gütersloh, verfügbar unter: [https://www.bertelsmann-stiftung.de/fileadmin/files/Leitfaden\\_CCMessungl.pdf](https://www.bertelsmann-stiftung.de/fileadmin/files/Leitfaden_CCMessungl.pdf).
- Bradburn, N. M., S. Sudman und B. Wansink (2004), *Asking Questions: The Definitive Guide to Questionnaire Design – For Market Research, Political Polls, and Social and Health Questionnaires, Research Methods for the Social Sciences*, John Wiley & Sons, San Francisco.
- Bundesregierung (2023), *Nationale Strategie für Soziale Innovationen und Gemeinwohlorientierte Unternehmen*, Berlin.
- Campbell, J. Y., A. W. Lo und A. C. MacKinlay (1997), *The Econometrics of Financial Markets*, Princeton University Press, Princeton, NJ.
- Collier, D. (2011), »Understanding Process Tracing«, *Political Science and Politics* 44(4), 823–830.
- Fraunhofer (2024), »Evaluation und Wirkungsforschung«, verfügbar unter: <https://www.isi.fraunhofer.de/de/competence-center/politik-gesellschaft/evaluation.html>.
- Fougère, D. und N. Jacquemet (2023), »Difference-in-Differences«, in: A. Revillard (Hrsg.), *Policy Evaluation: Methods and Approaches*, Éditions Science et bien commun, Quebec City, 43–54.
- Friedlander, D. und P. K. Robins (1995), »Evaluating Program Evaluations: New Evidence on Commonly Used Nonexperimental Methods«, *American Economic Review* 85(4), 923–937.
- Gerring, J. (2004), »What Is a Case Study and What Is It Good for?«, *American Political Science Review* 98(2), 341–354.
- Gertler, P. J., S. Martinez, P. Premand, L. B. Rowlings und C. M. J. Vermeersch (2016), *Impact Evaluation in Practice* (2nd Edition), Inter-American Development Bank and World Bank, Washington, D.C.
- Kline, T. (2021), »The Alchemy of Program Evaluation, Part 4: Semi-Structured Interviews«, Summit, verfügbar unter: <https://www.summitlc.us/blog/the-alchemy-of-program-evaluation-part-4-semi-structured-interviews>.
- Loi M. und M. M. Caetano Rodrigues Jorge Rodrigues Ferro (2012), *A Note on the Impact Evaluation of Public Policies: The Counterfactual Analysis*, EUR 25519 EN, Publications Office of the European Union, Luxembourg, verfügbar unter: <https://publications.jrc.ec.europa.eu/repository/handle/JRC74778>.
- OECD (2020), *Policy Framework on Sound Public Governance: Baseline Features of Governments that Work Well*, Paris, verfügbar unter: [https://www.oecd.org/en/publications/2020/12/policy-framework-on-sound-public-governance\\_931b05fc.html](https://www.oecd.org/en/publications/2020/12/policy-framework-on-sound-public-governance_931b05fc.html).
- Ottmann, S. (2021), »Den Kontext bei der Betrachtung von Wirkungen nicht vergessen!«, Blog Soziale Wirkung, verfügbar unter: <https://blog.soziale-wirkung.de/2021/08/12/kontext-wirkung/#more-218>.
- Phineo (2023a), »Unterscheidung: Wirkungsanalyse, Monitoring oder Evaluation«, Skala Campus, verfügbar unter: <https://www.skala-campus.org/artikel/unterschied-monitoring-evaluation-wirkungsanalyse/>
- Phineo (2023b), »Wirkung und gesellschaftlicher Impact lassen sich messen«, verfügbar unter: <https://www.phineo.org/magazin/was-ist-soziale-wirkung>.
- Pin, C. (2023), »Semi-structured Interview«, in: A. Revillard (Hrsg.), *Policy Evaluation: Methods and Approaches*, Éditions Science et bien commun, Quebec City, 131–140.
- Pluye, P. (2023) »Mixed Methods«, in: A. Revillard (Hrsg.), *Policy Evaluation: Methods and Approaches*, Éditions Science et bien commun, Quebec City, 195–204.
- Resnjanskij, S., J. Ruhose, S. Wiederhold, L. Woessmann und K. Wedel (2024), »Can Mentoring Alleviate Family Disadvantage in Adolescence? A Field Experiment to Improve Labor-Market Prospects«, *Journal of Political Economy* 132(3), 1013–1062.
- Revillard, A. (Hrsg.) (2023), *Policy Evaluation: Methods and Approaches*, Éditions science et bien commun, Quebec City.
- Schochet, P. Z., J. Burghardt und S. McConnell (2008), »Does Job Corps Work? Impact Findings from the National Job Corps Study«, *American Economic Review* 98(5), 1864–1886.
- Stedman-Bryce, G. (2013), *Effectiveness Review: Health for All: Towards Free Universal Health Care in Ghana, Evaluation Report*, Oxfam GB, verfügbar unter: <https://policy-practice.oxfam.org/resources/effectiveness-review-health-for-all-towards-free-universal-health-care-in-ghana-306376/>.
- The Committee for the Prize in Economic Sciences in Memory of Alfred Nobel (2019), »Understanding Development and Poverty Alleviation«, Scientific Background on the Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2019, The Royal Swedish Academy of Science, Stockholm.
- Theory of Change - TOC (2023), »What is Theory of Change?«, verfügbar unter: <https://www.theoryofchange.org/what-is-theory-of-change/>.
- USAID (2013), »Technical Note: Conducting Mixed-Method Evaluations (Version I)«, verfügbar unter: [https://2017-2020.usaid.gov/sites/default/files/documents/1870/Mixed\\_Methods\\_Evaluations\\_Technical\\_Note.pdf](https://2017-2020.usaid.gov/sites/default/files/documents/1870/Mixed_Methods_Evaluations_Technical_Note.pdf).
- Weiss (1999), »The Interface between Evaluation and Public Policy«, *Evaluation* 5(4), 468–486.
- Widner, J., Woolcock, M. und D. Ortega Nieto (Hrsg.) (2020), *The Case for Case Studies: Methods and Applications in International Development*, World Bank Group, Washington, D.C.
- Wonodi C.B., L. Privor-Dumm, M. Aina, A. M. Pate, R. Reis, P. Gadhoke und O. S. Levine (2012), »Using Social Network Analysis to Examine the Decision-making Process on New Vaccine Introduction in Nigeria«, *Health Policy Plan* 2, ii27–38.