



Wirkungsmessung für NPO

Zewo-Leitfaden für Dienstleistungen und Projekte im Inland



Schritt 3: Wirkungsmessung planen

Schon vor der Lancierung einer neuen Dienstleistung oder der Umsetzung eines Projekts muss geplant werden, wie die Wirkung später gemessen und beurteilt werden soll. Die Ergebnisse dieses Schrittes werden schriftlich festgehalten.

So wird es gemacht

Aktivitäten Es muss geklärt werden, was wem wozu beantwortet werden soll. Daraus kann abgeleitet werden, womit die Resultate der Wirkungsmessung verglichen werden und wer die Evaluation später durchführt. Die angestrebten Leistungen und Wirkungen müssen in Indikatoren überführt werden und es braucht Zielwerte für diese Messgrössen. Für jeden Indikator muss geplant werden, mit welchen Methoden die nötigen Daten erhoben werden.

Fragen Mit der Beantwortung folgender Fragen kann der dritte Schritt der Wirkungsmessung angegangen werden:

- Was wollen wir in Erfahrung bringen?
- Womit werden die Ergebnisse verglichen?
- Wer soll die Wirkungsmessung durchführen und wer ist dafür verantwortlich?
- Mit welchen Indikatoren können wir unsere Outputs (Leistungen) und Outcomes (Wirkungen) messen?
- Aus welchen Quellen können die Daten bezogen werden?
- Wie werden die Daten erhoben und wer ist dafür zuständig?
- Wie häufig werden Daten erhoben?
- Wurde bei der Formulierung des Auftrags zur Wirkungsmessung an alles gedacht?

Resultate

- Verantwortlichkeiten stehen fest.
- Indikatoren, Datenquellen, Erhebungsmethoden, Häufigkeit und Zeitpunkt der Messung sowie Bezugsgrössen für Vergleiche sind klar.
- Konzept, Plan und Auftrag zur Wirkungsmessung sind formuliert.

Beispiele

- Schuldenberatung: Ziele, Indikatoren, Soll-Werte, Messung

- Wolf: Ziele, Indikatoren, Soll-Werte, Messung

WICHTIG

Im Rahmen eines Leistungsvertrages zwischen der öffentlichen Hand und einer leistungserbringenden Non-Profit-Organisation ist es empfehlenswert, dass die Wirkungsmessung gemeinsam geplant wird:

- Indikatoren sollten so definiert werden, dass der Auftraggeber ein objektives Bild über die Wirkung einer Dienstleistung oder eines Projekts erhält.
- Bei der Definition der Indikatoren sollte umgekehrt aber auch darauf geachtet werden, dass die Erhebung der Daten eines Indikators in der Feldarbeit auf Akzeptanz stösst.
- Bei der Zusammenarbeit zwischen dem Auftraggeber und der leistungserbringenden Non-Profit-Organisation sollte ein Lernumfeld geschaffen werden, bei dem die Verbesserung der Leistung im Vordergrund steht

WICHTIG

Innerhalb dieses Schrittes und zu den vorangegangenen Schritten (Dienstleistung-/Projektziel definieren und Wirkungsmodell entwickeln) gibt es wichtige Feedbackschlaufen:

- Die Formulierung der Wirkungsziele mittels Indikatoren hilft, zu überprüfen, ob die geplanten Dienstleistungs- oder Projektziele realistisch und erreichbar sind.
- Die Planung der Datenerhebung hilft, zu überprüfen, ob die Indikatoren rechtzeitig und mit vertretbarem Aufwand gemessen werden können.
- Beides zusammen zeigt, ob die Evaluationsfrage beantwortet werden kann.



Wirkungsmessung für NPO

Zewo-Leitfaden für Dienstleistungen und Projekte im Inland



Vergleiche

Die Planung einer Messung beinhaltet immer auch die Definition des Massstabes zur Bewertung des untersuchten Gegenstandes. Die Definition des Vergleichsmassstabes bestimmt weitgehend, wie die Untersuchung angegangen werden sollte. Dabei müssen folgende Dimensionen beachtet werden:

- **Zeit**
Die Messung kann vor Aufnahme einer Tätigkeit, am Ende einer Periode oder nach Abschluss einer Tätigkeit durchgeführt werden. Soll die Entwicklung über die Zeit untersucht werden, sind zusätzliche Messungen während der Ausführung einer Tätigkeit nötig.
- **Bezugsgrösse**
Der Ist-Zustand kann mit den Zielen, der Ausgangslage oder mit einer Vergleichsgruppe verglichen werden.

Basismodelle

Mit folgenden Basismodellen können die Zielerreichung, die Veränderungen bei der Zielgruppe oder der Einfluss einer Dienstleistung oder eines Projekts gemessen werden:

<u>Soll-Ist</u>	Vor Aufnahme am Ende einer nach		
<u>Vergleich</u>	der Tätigkeit	Periode/	Abschluss
	(Ex-ante)	des Projekts	(Ex-post)
Ziele		Definition	
Zielgruppe		Messung 1	

<u>Vorher-</u>	Vor Aufnahme am Ende einer nach		
<u>Nachher</u>	der Tätigkeit	Periode/	Abschluss
<u>Vergleich</u>	(Ex-ante)	des Projekts	(Ex-post)
Zielgruppe	Messung 1	Messung 2	

<u>Vergleich mit</u>	Vor Aufnahme am Ende nach		
<u>Vergleichsgruppe</u>	der Tätigkeit	einer Periode/	Abschluss
	(Ex-ante)	des Projekts	(Ex-post)
Zielgruppe		Messung 1	
Vergleichsgruppe		Messung 2	

<u>Vorher-Nachher</u> <u>Vergleich mit</u> <u>Vergleichsgruppe</u>	Vor Aufnahme am Ende der Tätigkeit (Ex-ante)	nach einer Periode/ Abschluss des Projekts (Ex-post)
Zielgruppe	Messung 1	Messung 3
Vergleichsgruppe	Messung 2	Messung 4

WICHTIG

Eine rigorose Wirkungsmessung kann mittels einem Vorher-Nachher Vergleich mit Vergleichsgruppe oder im Rahmen einer randomisierten kontrollierten Studie¹ durchgeführt werden. Festgestellte Veränderungen können so eindeutig der Dienstleistung oder dem Projekt zugeschrieben und externe Einflüsse ausgeschlossen werden. Solche Vergleichsmodelle werden in der Praxis zunehmend von institutionellen Geldgebern verlangt. Sie sind aber sehr aufwändig: Es braucht mindestens vier Messungen. Die Auswahl der Vergleichsgruppe und das Ziehen von Stichproben sind keineswegs trivial. Insbesondere kleinere Organisationen werden kaum die Ressourcen haben, rigorose Wirkungsmessungen durchzuführen. Auch grössere Organisationen müssen sich überlegen, wann und wo es für sie Sinn macht.

An dieser Stelle wird dafür plädiert, dass auch einfachere Modelle zweckmässig sind. Sie eignen sich beispielsweise für organisationsinternes Lernen. Damit Vorher-Nachher Vergleiche möglich sind, sollte jedoch eine Messung bei der Lancierung einer neuen Dienstleistung beziehungsweise zu Beginn des Projekts (Baselinestudie) eingeplant werden. Damit Soll-Ist Vergleiche gemacht werden können, müssen klare Ziele formuliert sein.

Für die kausale Zuordnung (Attribution) der Wirkung zu einer Intervention braucht es rigorose Methoden zur Wirkungsmessung. Zur Plausibilisierung, dass eine Intervention einen Beitrag zur Erreichung übergeordneter Ziele geleistet hat (Contribution), genügen auch einfachere Ansätze. In jedem Fall muss aber klar sein, wie die Wirkungsmessung gemacht wurde und welche Schlüsse sich daraus ableiten lassen.

¹Bei einer randomisierten kontrollierten Studie (Randomized Controlled Trial) handelt es sich um eine Untersuchung, bei der man aus einer Grundgesamtheit mit identischen Bedingungen zufällig ein Sample auswählt, bei der man eine Einflussgrösse verändert. In einer kontrollierten Umgebung kann in der Folge der Effekt dieser veränderten Einflussgrösse im Vergleich zur Kontrollgruppe statistisch signifikant bestimmt werden.

ACHTUNG

In der Praxis kommt es immer wieder vor, dass lediglich der Zustand bei einer Dienstleistung am Ende einer Periode oder der Endzustand eines Projekts beschrieben wird. Für die Messung der Wirkung reicht eine Beschreibung der Zielgruppe ohne Bezug zu den Zielsetzungen, der Ausgangslage oder zu einer Vergleichsgruppe jedoch nicht aus.

Erweiterte Modelle

Komplexere Modelle mit erhöhter Aussagekraft entstehen durch die Erweiterung und die Kombination von Basismodellen.

<u>Vorher-Nachher</u> <u>und</u> <u>Soll-Ist Vergleich</u>	Vor Aufnahme der Tätigkeit (Ex-ante)	am Ende einer Periode/ des Projekts	nach Abschluss (Ex-post)
Ziele		Definition	
Zielgruppe	Messung 1	Messung 2	

<u>Ex-post</u> <u>Vergleich</u>	Vor Aufnahme der Tätigkeit (Ex-ante)	am Ende einer Periode/ des Projekts	nach Abschluss (Ex-post)
Zielgruppe	Messung 1	Messung 2	Messung 3

<u>Ex-post und</u> <u>Soll-Ist Vergleich</u>	Vor Aufnahme der Tätigkeit (Ex-ante)	am Ende einer Periode/ des Projekts	nach Abschluss (Ex-post)
Ziele		Definition	
Zielgruppe	Messung 1	Messung 2	Messung 3

<u>Ex-post Vergleich</u> <u>mit</u> <u>Vergleichsgruppe</u>	Vor Aufnahme der Tätigkeit (Ex-ante)	am Ende einer Periode/ des Projekts	nach Abschluss (Ex-post)
Zielgruppe	Messung 1	Messung 3	Messung 5
Vergleichsgruppe	Messung 2	Messung 4	Messung 6

<u>Soll-Ist Vergleich</u> <u>mit</u> <u>Vergleichsgruppe</u>	Vor Aufnahme der Tätigkeit (Ex-ante)	am Ende einer Periode/ des Projekts	nach Abschluss (Ex-post)
Ziele		Definition	
Zielgruppe		Messung 1	
Vergleichsgruppe		Messung 2	

**Vorher-Nachher
und
Soll-Ist Vergleich
mit
Vergleichsgruppe** **Vor Aufnahme am Ende nach**
der Tätigkeit einer Periode/ Abschluss
(Ex-ante) des Projekts (Ex-post)

Ziele		Definition
Zielgruppe	Messung 1	Messung 3
Vergleichsgruppe	Messung 2	Messung 4

**Ex-post- und
Soll-Ist Vergleich
mit
Vergleichsgruppe** **Vor Aufnahme am Ende nach**
der Tätigkeit einer Periode/ Abschluss
(Ex-ante) des Projekts (Ex-post)

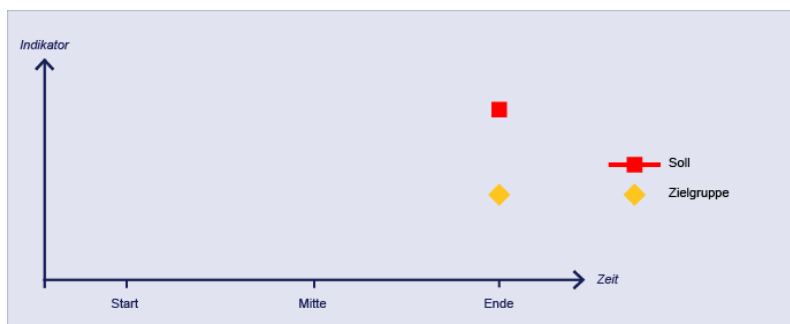
Ziele		Definition	Definition
Zielgruppe	Messung 1	Messung 3	Messung 5
Vergleichsgruppe	Messung 2	Messung 4	Messung 6

Wirkungsmessung für NPO

Zewo-Leitfaden für Dienstleistungen und Projekte im Inland



Soll-Ist Vergleich



Charakteristik

Es handelt sich um eine statische Betrachtung, bei der zu einem bestimmten Zeitpunkt die tatsächliche Situation der Zielgruppe mit der zu diesem Zeitpunkt geplanten Situation der Zielgruppe verglichen wird.

Beispiel

20% der von Verschuldung betroffenen Menschen, die den Kurs zum Thema «Schuldenbewältigung» besucht haben, konnten innerhalb von 12 Monaten ihre Verschuldung reduzieren. Das sind deutlich weniger als geplant (Soll):

Vorteile

Eine einfache Art, die Zielerreichung festzuhalten.

Der Aufwand für die Datenerhebung ist gering.

Das methodisches Know-how ist intern vorhanden oder erlernbar.

Nachteile

Es gibt keinen Bezug zur Ausgangslage: Man weiss nicht, ob und wie sich die Situation der Zielgruppe seit dem Beginn der Intervention verändert hat.

Man weiss nicht, ob und wie sich die Situation der Zielgruppe verändert hätte, ohne das Projekt.

Es ist keine Aussage zur Wirkung möglich.

Es ist keine Aussage zur Nachhaltigkeit möglich.

Eignung

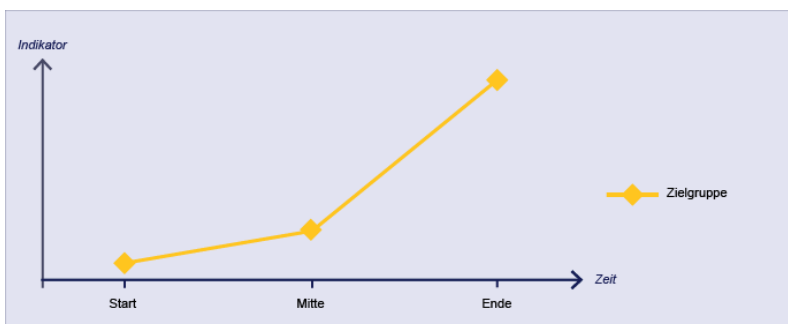
Ist bedingt geeignet für das Lenken der Organisation und das Lernen innerhalb einer Organisation. Geeignet zu Legitimationszwecken bei Zielvereinbarungen. Kommt insbesondere dann zum Einsatz, wenn Daten zur Ausgangslage und geeignete Vergleichsgruppen fehlen oder nur mit unverhältnismässig grossem Aufwand beschafft werden können. Zur Wirkungsmessung ist ein punktueller Soll-Ist Vergleich nicht geeignet.

Wirkungsmessung für NPO

Zewo-Leitfaden für Dienstleistungen und Projekte im Inland



Vorher-Nachher Vergleich



Charakteristik

Es wird die Entwicklung der Zielgruppe einer Dienstleistung oder eines Projekts beschrieben, z.B. am Anfang und am Ende einer Periode oder am Anfang und am Ende des Projekts. Zusätzliche Messungen geben Auskunft über die Entwicklung einer Tätigkeit.

Beispiel

70% der Langzeitarbeitslosen, welche Unterstützungsangebote für die Stellensuche in Anspruch genommen haben, konnten innerhalb von sechs Monaten eine neue Arbeitsstelle finden. Bei der Lancierung dieser Dienstleistung lag dieser Anteil lediglich bei 10%.

Vorteile

Es kann plausibel gemacht werden, ob eine Intervention einen Beitrag zur angestrebten Wirkung geleistet hat oder nicht.

Aufwand für die Datenerhebung ist oft vertretbar.

Methodisches Know-how ist oft intern vorhanden oder erlernbar.

Nachteile

Man weiss nicht, ob und wie sich die Situation der Zielgruppe verändert hätte ohne das Projekt. Eine eindeutige Zuordnung der Wirkung zur Intervention ist nicht möglich.

Es gibt keinen Bezug zur Zielsetzung. Man weiss nicht, ob und inwieweit die Organisation die gesetzten Ziele erreicht hat.

Aussagen zur Nachhaltigkeit des Projekts sind nicht möglich.

Eignung

Geeignet, um aus den beobachteten Veränderungen zu lernen sowie zur Legitimation, wenn eine Plausibilisierung der Wirkung genügt. Kommt dann zum Einsatz, wenn Daten zur Ausgangslage vorhanden sind oder mit vertretbarem Aufwand rekonstruiert werden können, jedoch geeignete Vergleichsgruppen fehlen oder nur mit unverhältnismässig grossem Aufwand

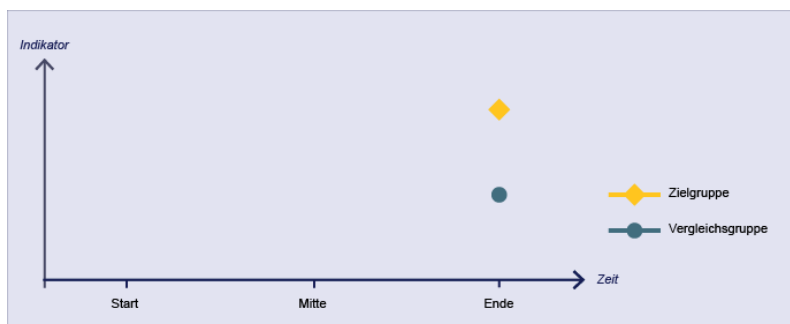
konstruiert werden können oder wenn ein Vergleich mit Vergleichsgruppen ethisch fragwürdig sind.

Wirkungsmessung für NPO

Zewo-Leitfaden für Dienstleistungen und Projekte im Inland



Vergleich mit Vergleichsgruppe



Charakteristik

Es handelt sich um eine statische Betrachtung, bei der zu einem bestimmten Zeitpunkt die Situation der Zielgruppe mit der Situation der Kontrollgruppe verglichen wird.

Beispiel

Die Privatkonkurse sind in der Region x im vergangenen Jahr um 10% zurückgegangen. In der Vergleichsregion y sind die Privatkonkurse nur um 5% zurückgegangen.

Vorteile

Es kann plausibel gemacht werden, ob eine Intervention einen Beitrag zur angestrebten Wirkung geleistet hat oder nicht.

Falls anfänglich identische Vergleichsgruppen vorhanden sind, können Aussagen zur Wirkung der Dienstleistung oder des Projekts bei der Zielgruppe gemacht werden.

Der Aufwand für die Datenerhebung (2 Messungen) ist oft noch vertretbar.

Nachteile

Es gibt keinen Bezug zur Ausgangslage: Man weiss nicht, ob und wie sich die Situation der Ziel- und der Vergleichsgruppe seit dem Beginn der Intervention verändert hat.

Identische Gruppen fehlen oft, methodisches Know-how ist erforderlich, wenn Vergleichsgruppen konstruiert werden müssen. Die Wahl der Vergleichsgruppe ist nicht trivial.

Es gibt keinen Bezug zu den gesetzten Zielen: Man weiss nicht, ob und inwieweit die Organisation ihre Ziel erreicht hat.

Es sind keine Aussagen zur Nachhaltigkeit möglich.

Eignung

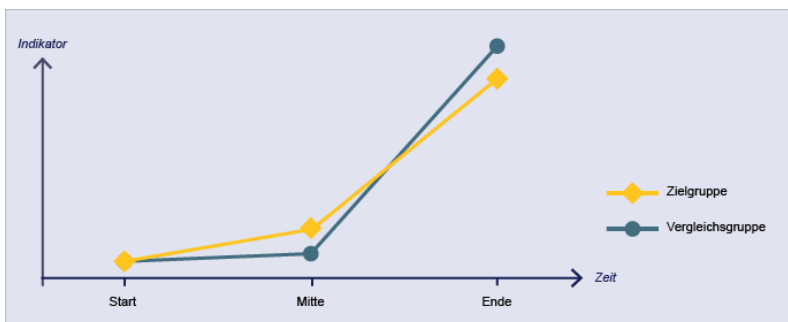
Durch seine erhöhte Aussagekraft ist ein Vergleich mit Vergleichsgruppe für Legitimationszwecke geeignet, sofern geeignete Vergleichsgruppen vorhanden sind oder konstruiert werden können und es keine ethischen Bedenken gibt. Er wird dann angewandt, wenn Daten zur Ausgangslage fehlen oder nur mit unverhältnismässig grossem Aufwand beschafft werden können.

Wirkungsmessung für NPO

Zewo-Leitfaden für Dienstleistungen und Projekte im Inland



Vorher-Nachher Vergleich mit Vergleichsgruppe



Charakteristik

Die Entwicklung der Zielgruppe wird mit jener der Vergleichsgruppe verglichen. Zusätzliche Messungen geben Auskunft über den Verlauf einer Tätigkeit.

Beispiel

In der Region x sind im vergangenen Jahr die Privatkonkurse um 10% zurückgegangen. Das ist ein Fortschritt seit Beginn der Präventionskampagne. In der Vergleichsregion y ist aber ohne eine Intervention für die gleiche Periode ein stärkerer Rückgang der Privatkonkurse festzustellen.

Vorteile

Der Intervention kann eine bestimmte Wirkung zu- oder abgesprochen werden.

Der Ansatz ist methodisch fundiert.

Nachteile

Die Datenerhebung ist aufwändig und erfordert mindestens 4 Messungen.

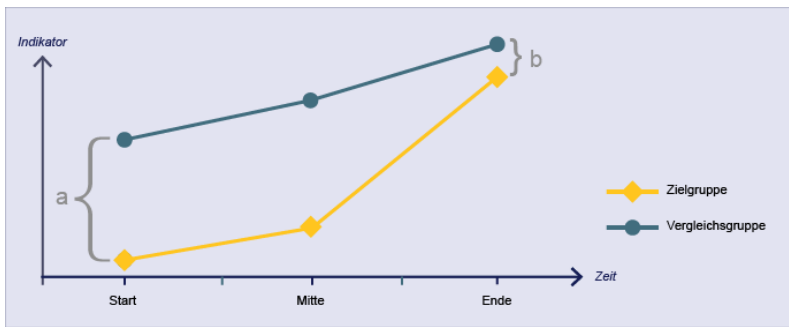
Das Vorgehen ist methodisch anspruchsvoll.

Die Zielerreichung wird nicht gemessen.

Es ist keine Aussage zur Nachhaltigkeit möglich.

Eignung

Eignet sich sehr gut zum Legitimieren, aber auch zum Lernen und Lenken der Organisation. Wird dann angewandt, wenn die Wirkung eindeutig nachgewiesen werden soll und es keine ethischen Bedenken zum Vergleich mit Vergleichsgruppen gibt. Geeignete Vergleichsgruppen müssen bestimmt oder konstruiert werden können.



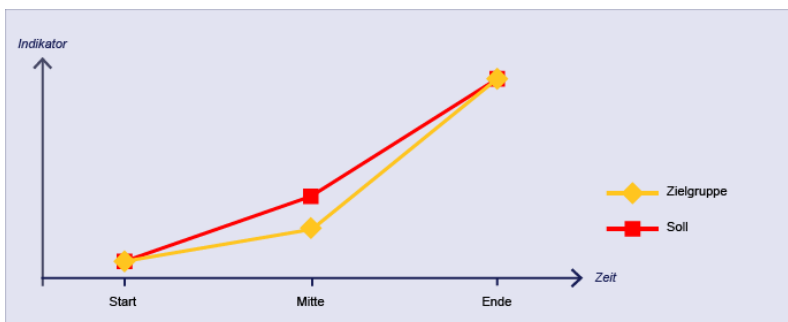
Wenn die Ausgangslage der Vergleichsgruppe nicht mit jener der Zielgruppe identisch ist, muss die Differenz zwischen der Ziel- und der Vergleichsgruppe am Anfang a) und am Ende b) der Betrachtungsperiode eines Dienstleistungsangebotes beziehungsweise des Projektes bestimmt werden. Die Aussage zur Wirkung hängt in diesem Fall davon ab, ob sich die Differenz vergrößert oder verringert hat. Dieses Verfahren wird difference-in-difference genannt.

Wirkungsmessung für NPO

Zewo-Leitfaden für Dienstleistungen und Projekte im Inland



Vorher-Nachher und Soll-Ist Vergleich



Charakteristik

Die tatsächliche Entwicklung der Zielgruppe wird mit der geplanten Entwicklung der Zielgruppe verglichen. Zusätzliche Messungen geben Auskunft über den Verlauf der betrachteten Periode eines Dienstleistungsangebots oder des Projekts.

Beispiel

80% der Schulabbrecher in der Region x finden innerhalb von 6 Monaten eine Arbeitsstelle. Das sind dreimal mehr als zu Beginn des Projekts und so viele wie angestrebt.

Vorteile

Es wird gezeigt, inwieweit die gesetzten Ziele erreicht werden konnten.

Es kann plausibel gemacht werden, ob eine Intervention einen Beitrag zur angestrebten Wirkung geleistet hat oder nicht.

Der Aufwand für die Datenerhebung (2 Messungen) ist oft vertretbar.

Das methodische Know-how ist oft vorhanden oder erlernbar.

Nachteile

Man weiss nicht, ob und wie sich die Situation der Zielgruppe verändert hätte ohne das Projekt. Eine eindeutige Zuordnung der Wirkung zur Intervention ist nicht möglich.

Es ist keine Aussage zur Nachhaltigkeit möglich.

Eignung

Ist zum Lernen und insbesondere zum Lenken der Organisation geeignet kann aber bei Zielvereinbarungen der wenn eine Plausibilisierung der Wirkung

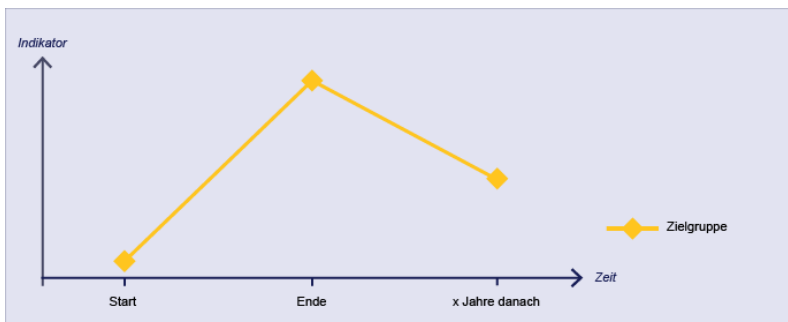
genügt, auch zu Legitimation dienen. Kommt zum Einsatz, wenn klare Ziele formuliert wurden und Daten zur Ausgangslage vorhanden sind, jedoch geeignete Kontrollgruppen fehlen oder wenn Vergleiche mit Kontrollgruppen ethisch fragwürdig sind.

Wirkungsmessung für NPO

Zewo-Leitfaden für Dienstleistungen und Projekte im Inland



Ex-post Vergleich



Charakteristik

Es wird die Entwicklung der Zielgruppe über die analysierte Periode eines Dienstleistungsangebotes oder über die Beendigung des Projekts hinaus betrachtet.

Beispiel

Zwei Jahre nach Beendigung einer Kursreihe für alleinerziehende Mütter, die in eine finanzielle Notsituationen geraten sind, geben 40% der Kursteilnehmerinnen an, dass sie ihre finanzielle Situation als stabil bezeichnen. Das sind zwar deutlich mehr als zu Kursbeginn, aber weniger als unmittelbar nach Beendigung der Kursreihe.

Vorteile

Der Aufwand für die Datenerhebung ist oft vertretbar.

Das methodische Know-how ist oft intern vorhanden oder erlernbar.

Es kann plausibel gemacht werden, ob eine Intervention einen Beitrag zur angestrebten Wirkung geleistet hat oder nicht.

Es kann gezeigt werden, ob eine Intervention nachhaltig war oder nicht.

Nachteile

Externe Einflussfaktoren können nicht ausgeschlossen werden, da der Vergleich zur Vergleichsgruppe fehlt.

Es wird nicht ersichtlich, ob die gesetzten Ziele erreicht worden sind.

Die Erkenntnisse liegen erst Jahre nach Abschluss der Intervention vor.

Eignung

Kommt dann zum Einsatz, wenn die Nachhaltigkeit einer Intervention untersucht werden soll. Durch die zeitliche Verzögerung kommt die

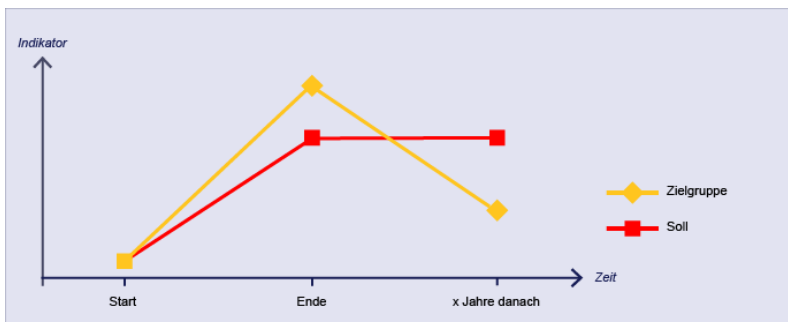
Information zur kurzfristigen Legitimation und für das unmittelbare Lenken und Lernen zu spät. Sie kann jedoch für die langfristige Entwicklung der Organisation und die strategische Ausrichtung genutzt werden.

Wirkungsmessung für NPO

Zewo-Leitfaden für Dienstleistungen und Projekte im Inland



Ex-post und Soll-Ist Vergleich



Charakteristik

Die tatsächliche Entwicklung der Zielgruppe wird mit der geplanten Entwicklung über die analysierte Periode eines Dienstleistungsangebotes beziehungsweise über das Projektende hinaus verglichen.

Beispiel

Drei Jahre nach Beendigung der Intervention finden 25% der Schulabbrecher innerhalb von 6 Monaten eine Arbeitsstelle. Das sind zwar mehr Jugendliche als vor dem Projekt, aber weniger als am Ende des Projekts und auch weniger als angestrebt.

Vorteile

Es wird gezeigt, inwieweit die gesetzten Ziele erreicht werden konnten.

Es kann plausibel gemacht werden, ob eine Intervention einen Beitrag zur angestrebten Wirkung geleistet hat oder nicht.

Es kann gezeigt werden, ob eine Intervention nachhaltig war oder nicht.

Das methodische Know-how ist vorhanden oder erlernbar.

Nachteile

Man weiss nicht, ob und wie sich die Situation der Zielgruppe verändert hätte ohne das Projekt. Eine eindeutige Zuordnung der Wirkung zur Intervention ist nicht möglich.

Eignung

Eignet sich zur Legitimation bei Zielvereinbarungen oder wenn eine Plausibilisierung der Wirkung genügt. Kommt zum Einsatz, wenn die Nachhaltigkeit einer Intervention untersucht werden soll und Vergleiche mit

Vergleichsgruppen zu aufwändig oder ethisch fragwürdig sind. Durch die zeitliche Verzögerung kommt die Information zur unmittelbaren Legitimation und für das kurzfristige Lenken und Lernen zu spät. Sie kann jedoch für die langfristige Entwicklung der Organisation und die strategische Ausrichtung genutzt werden.

Wirkungsmessung für NPO

Zewo-Leitfaden für Dienstleistungen und Projekte im Inland



Ex-post Vergleich mit Vergleichsgruppe



Charakteristik

Die Entwicklung der Zielgruppe wird mit jener der Vergleichsgruppe über die analysierte Periode eines Dienstleistungsangebotes oder über das Projektende hinaus verglichen.

Beispiel

30% der Alleinerziehenden in der Region x nutzen die Kultur-Legi, welche diverse Vergünstigungen für kulturelle Veranstaltungen bietet. Das sind mehr als vor dem Projekt aber weniger als am Ende des Projekts. Die Nutzung in der Vergleichsregion y hat sich in der gleichen Zeit von einer schwächeren Ausgangslage auf das Niveau der Zielgruppe entwickelt und verbessert sich weiter.

Vorteile

Es kann wissenschaftlich fundiert nachgewiesen werden, ob eine Intervention nachhaltig wirksam war oder nicht.

Nachteile

Die Datenerhebung ist sehr aufwändig und erfordert mindestens 6 Messungen.

Das Vorgehen ist methodisch anspruchsvoll.

Die Zielerreichung wird nicht gemessen.

Eignung

Geeignet, wenn wissenschaftlich fundiert nachgewiesen werden soll, ob mit der Intervention eine nachhaltige Wirkung erzielt worden ist. Kommt zum Einsatz, wenn es keine Bedenken gegen Vergleiche mit Vergleichsgruppen gibt und geeignete Vergleichsgruppen bestimmt oder konstruiert werden können. Die Daten zur Ausgangslage müssen allenfalls rekonstruiert werden.

Wenn die Ausgangslage der Vergleichsgruppe nicht mit jener der Zielgruppe identisch ist, muss die Differenz zwischen der Ziel- und der Vergleichsgruppe verfolgt werden. Die Aussage zur Wirkung hängt in diesem Fall davon ab, ob sich die Differenz vergrößert oder verringert. Durch die zeitliche Verzögerung kommt die Information zur kurzfristigen Legitimation und für das unmittelbare Lenken und Lernen zu spät. Sie kann jedoch für die langfristige Entwicklung der Organisation und die strategische Ausrichtung genutzt werden.

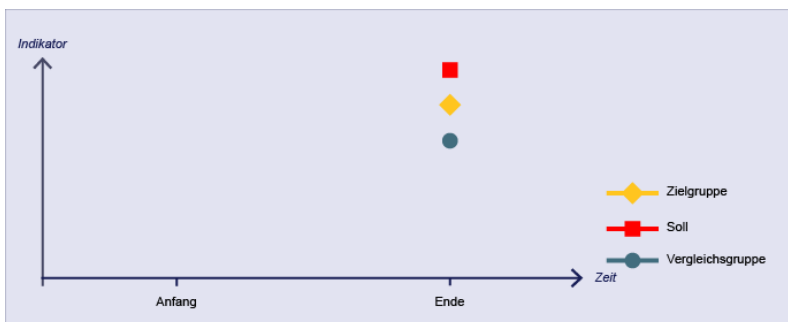


Wirkungsmessung für NPO

Zewo-Leitfaden für Dienstleistungen und Projekte im Inland



Soll-Ist Vergleich mit Vergleichsgruppe



Charakteristik

Es handelt sich um eine punktuelle Betrachtung, bei der zu einem bestimmten Zeitpunkt die Situation der Zielgruppe mit den Zielen und mit der Situation der Vergleichsgruppe verglichen wird.

Beispiel

60% der Jugendlichen ohne Lehrstelle, welche Unterstützungsangebote für die Lehrstellensuche in Anspruch genommen haben, konnten innerhalb von sechs Monaten eine Lehrstelle finden. Das sind mehr als in der Vergleichsgruppe, aber weniger als angestrebt.

Vorteile

Der Aufwand für die Datenerhebung mit 2 Messungen ist oft noch vertretbar.

Es ist eine einfache Art, die Zielerreichung festzuhalten.

Falls identische Vergleichsgruppen vorhanden sind, können Aussagen zur Wirkung der Dienstleistung oder des Projekts bei der Zielgruppe gemacht werden.

Nachteile

Es gibt keinen Bezug zur Ausgangslage. Man weiss nicht, ob und wie sich die Situation der Ziel- und der Vergleichsgruppe seit dem Beginn der Intervention verändert hat und ob sie vom selben Ausgangspunkt gestartet sind.

Identische Vergleichsgruppen fehlen oft.

Es sind keine Aussagen zur Nachhaltigkeit möglich.

Eignung

Ist nur bedingt geeignet für das Lenken und Lernen innerhalb einer

Organisation. Kann jedoch zur Legitimation bei Zielvereinbarungen geeignet sein. Wird in der Praxis dann angewandt, wenn Daten zur Ausgangslage fehlen oder nur mit unverhältnismässig grossem Aufwand beschafft werden können, jedoch geeignete Vergleichsgruppen vorhanden sind.

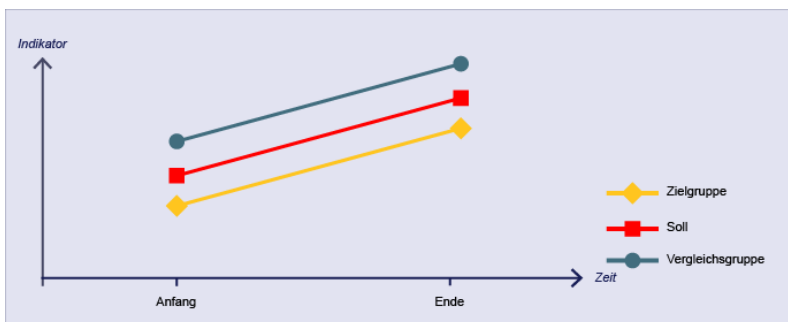


Wirkungsmessung für NPO

Zewo-Leitfaden für Dienstleistungen und Projekte im Inland



Vorher-Nachher und Soll-Ist Vergleich mit Vergleichsgruppe



Charakteristik

Die Entwicklung der Zielgruppe wird mit jener der Vergleichsgruppe verglichen und zur Zielsetzung in Bezug gesetzt wird.

Beispiel

25% der Schulabbrecher in der Region x haben innerhalb von 6 Monaten eine Arbeitsstelle gefunden. Zu Beginn des Projektes waren es nur 15%. Die Verbesserung um 10 %-Punkte ist gleich gut wie in der Vergleichsgruppe in der Region y, bei der sich der Prozentsatz von 25% auf 40% verbessert hat. Das Ziel von 35 % konnte nicht erreicht werden.

Vorteile

Es wird gezeigt, was sich bei der Zielgruppe durch die Intervention verändert hat.

Es wird festgestellt, ob die Ziele erreicht worden sind.

Der Ansatz ist methodisch fundiert.

Die Zuordnung der Wirkung zur Intervention ist möglich.

Nachteile

Die Datenerhebung ist aufwändig und erfordert mindestens 4 Messungen.

Das Vorgehen ist methodisch anspruchsvoll.

Es ist keine Aussage zur Nachhaltigkeit möglich.

Eignung

Eignet sich zum Lernen, Lenken und Legitimieren. Wird bei Zielvereinbarungen angewandt und insbesondere dann, eingesetzt, wenn die Wirkung eindeutig nachgewiesen werden soll und es keine ethischen Bedenken zum Vergleich mit Vergleichsgruppen gibt. Geeignete Vergleichsgruppen müssen bestimmt oder konstruiert werden können. Wenn die Ausgangslage der Vergleichsgruppe nicht mit jener der Zielgruppe

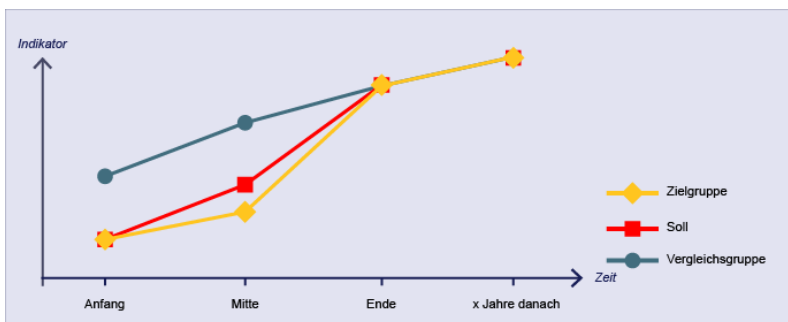
identisch ist, muss die Differenz zwischen der Ziel- und der Vergleichsgruppe am Anfang und am Ende der analysierten Periode des Dienstleistungsangebotes oder des Projektes bestimmt werden. Die Aussage zur Wirkung hängt in diesem Fall davon ab, ob sich die Differenz vergrößert oder verringert hat. Dieses Verfahren wird difference-in-difference genannt.

Wirkungsmessung für NPO

Zewo-Leitfaden für Dienstleistungen und Projekte im Inland



Ex-post und Soll-Ist Vergleich mit Vergleichsgruppe



Charakteristik

Die Entwicklung der Zielgruppe wird mit jener der Vergleichsgruppe über die analysierte Periode eines Dienstleistungsangebotes oder über das Projektende hinaus verglichen und mit den Zielsetzungen in Bezug gebracht.

Beispiel

10% der Schulabbrecher haben in der Region x hat im vergangenen Jahr innerhalb von 3 Monaten eine Arbeitsstelle gefunden. Das sind 5%-Punkte mehr als zu vor dem Projekt. In der gleichen Zeit hat sich die Situation in der Vergleichsgruppe nur um 3%-Punkte verbessert. Nach Beendigung des Projekts verläuft die Entwicklung der Ziel- und Vergleichsgruppe gleich. Das Projekt hat die Soll-Ziele erreicht.

Vorteile

Es kann wissenschaftlich fundiert nachgewiesen werden, ob eine Intervention nachhaltig wirksam war und ob die Ziele langfristig erreicht wurden.

Nachteile

Die Datenerhebung ist sehr aufwändig (mind. 6 Messungen).

Das Vorgehen ist methodisch anspruchsvoll.

Eignung

Eignet sich für die Überprüfung der langfristigen Entwicklungspolitik der Organisation und der strategischen Ausrichtung der Organisation. Geeignet für die Legitimation der nachhaltigen Wirkung gegenüber Dritten. Wegen der zeitlichen Verzögerung ist der Ansatz weniger geeignet für das unmittelbare Lenken und Lernen. Wird dann, eingesetzt, wenn eine nachhaltige Wirkung

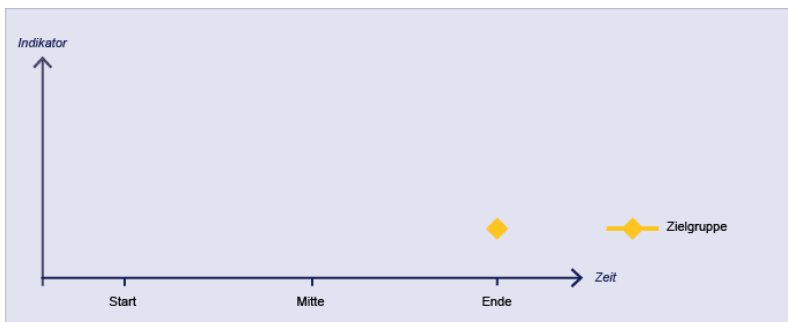
eindeutig nachgewiesen werden soll und es keine Bedenken zum Vergleich mit Vergleichsgruppen gibt. Geeignete Vergleichsgruppen müssen bestimmt oder konstruiert werden können. Wenn die Ausgangslage der Vergleichsgruppe nicht mit jener der Zielgruppe identisch ist, muss die Differenz zwischen der Ziel- und der Vergleichsgruppe verfolgt werden. Die Aussage zur Wirkung hängt in diesem Fall davon ab, ob sich die Differenz vergrößert oder verringert.

Wirkungsmessung für NPO

Zewo-Leitfaden für Dienstleistungen und Projekte im Inland



Beschreibung



Charakteristik

Es handelt sich um eine punktuelle Betrachtung, bei der zu einem bestimmten Zeitpunkt die Situation der Zielgruppe beschrieben wird.

Beispiele

30% der Alleinerziehenden in der Region x nutzen die Kultur-Legi, welche diverse Vergünstigungen für kulturelle Veranstaltungen bietet.

Vorteile

Der Aufwand für die Datenerhebung ist gering.

Das methodische Know-how ist vorhanden.

Nachteile

Es gibt keinen Bezug zu den gesetzten Zielen: Man weiss nicht, ob und inwieweit die Organisation ihre Ziele erreicht hat.

Es gibt keinen Bezug zur Ausgangslage: Man weiss nicht, ob und wie sich die Situation der Zielgruppe seit dem Beginn der Intervention verändert hat.

Man weiss nicht, ob und wie sich die Situation der Zielgruppe verändert hätte, ohne die Dienstleistung oder ohne das Projekt.

Es ist keine Aussage zur Wirkung möglich.

Es ist keine Aussage zur Nachhaltigkeit möglich.

Eignung

Durch ihre geringe Aussagekraft, ist die punktuelle Beschreibung wenig geeignet für das Lernen oder Lenken einer Organisation. In der Praxis wird

insbesondere dann auf die punktuelle Beschreibung zurückgegriffen, wenn keine Ziele formuliert wurden, Daten zur Ausgangslage und Vergleichsgruppen fehlen oder diese nur mit unverhältnismässig grossem Aufwand konstruiert werden können. Zur Wirkungsmessung ist die reine Beschreibung eines Zustandes jedoch nicht geeignet.



Wirkungsmessung für NPO

Zewo-Leitfaden für Dienstleistungen und Projekte im Inland



Planung und Eignung

Der Zeitpunkt der Messungen muss geplant werden. Die Häufigkeit der Messungen erhöht die Genauigkeit der Resultate, aber auch den Aufwand. Die folgende Tabelle zeigt, wann es welche Messungen braucht, wofür sie geeignet sind und wie weit sie in der Praxis verbreitet sind.

Planung von Zeitpunkt und Häufigkeit der Messung

Zeitpunkt	Vor Aufnahme der Tätigkeit (Ex-ante)	Während der Tätigkeit	Am Ende einer Periode / eines Projekts	Nach Abschluss (Ex-post)
Notwendigkeit	Nötig, wenn erforderliche Daten zur Ausgangslage fehlen	Nötig, wenn die Verläufe betrachtet werden sollen	Nötig, wenn der Erfolg beurteilt werden soll	Nötig, wenn die nachhaltige Wirkung der Intervention beurteilt werden soll
Eignung	Geeignet, um über die Durchführung zu entscheiden	Geeignet, zur Überwachung und Steuerung der Umsetzung	Geeignet zur Steuerung und Rechenschaftsablage	Geeignet zur Überprüfung der Strategie und Politik
Verbreitung	Wird eher selten und wenn, dann vor grossen Programmen angewandt	Wird häufig nur auf Output-Ebene (Monitoring) angewandt	Wird auch auf Outcome-Ebene angewandt	Wird eher selten und wenn, dann auf Impact-Ebene angewandt

Die reine Beschreibung eines Zustandes sagt ebenso wenig über die Wirkung aus wie ein herkömmlicher Soll-Ist Vergleich. Damit Aussagen über die Wirkung möglich werden, muss der erreichte Zustand der Zielgruppe zu deren Ausgangslage oder zu einer Vergleichsgruppe in Bezug gesetzt werden. Die Kombination von beidem wird gelegentlich auch als «gold standard» der Wirkungsmessung bezeichnet. Die nachfolgende Tabelle gibt einen Überblick, welche Vergleiche möglich sind und wozu sie sich eignen.

Eignung der Vergleichsarten

Aussagekraft	Zielerreichung	Wirkung	Plausibilisierung (Contribution)	Eindeutige Zuordnung (Attribution)	Nachhaltigkeit	Minimale Anzahl Messungen	Dauer
Basis Vergleichsmodelle							

Soll-Ist Vergleich	o.k.					1	punktuell
Vorher-Nachher Vergleich		o.k.	o.k.			2	3 bis 8 Jahre
Vergleich mit Vergleichsgruppe		o.k.	o.k.			2	punktuell
Vorher-Nachher Vergleich mit Vergleichsgruppe		o.k.	o.k.	o.k.		4	3 bis 8 Jahre
Komplexere Modelle und Kombinationen							
Vorher-Nachher und Soll-Ist Vergleich	o.k.	o.k.	o.k.			2	3 bis 8 Jahre
Ex-post Vergleich		o.k.	o.k.	o.k.		3	6 bis 12 Jahre
Ex-post und Soll-Ist Vergleich	o.k.	o.k.	o.k.	o.k.		3	6 bis 12 Jahre
Ex-post Vergleich mit Vergleichsgruppe		o.k.	o.k.	o.k.	o.k.	6	6 bis 12 Jahre
Soll-Ist Vergleich mit Vergleichsgruppe	o.k.	o.k.	o.k.			2	punktuell
Vorher-Nachher und Soll-Ist Vergleich mit Vergleichsgruppe	o.k.	o.k.	o.k.	o.k.		4	3 bis 8 Jahre
Ex-post und Soll-Ist Vergleich mit Vergleichsgruppe	o.k.	o.k.	o.k.	o.k.	o.k.	6	6 bis 12 Jahre
Kein Vergleich							
Beschreibung						1	punktuell



Stiftung Zewo
Schweizerische Zertifizierungsstelle für gemeinnützige,
Spenden sammelnde Organisationen

Stiftung Zewo
Lägernstrasse 27
8037 Zürich
Telefon +41 (0)44 366 99 55

Wirkungsmessung für NPO

Zewo-Leitfaden für Dienstleistungen und Projekte im Inland



Einbezug der Zielgruppen bei der Wirkungsmessung



Grundsätzlich ist es zu empfehlen die Zielgruppen als eigentliche Empfänger von Dienstleistungen oder Massnahmen in die Wirkungsmessung einzubeziehen. Aus inhaltlichen oder methodischen Gründen ist dies jedoch nicht immer möglich. Es gilt daher von Fall zu Fall zu entscheiden, mit welchem Aufwand die Zielgruppen in die Wirkungsmessung einbezogen werden können.

In der Praxis geht das von ein paar Interviews mit Vertreterinnen und Vertretern der Zielgruppe bis hin zu umfangreichen standardisierten Befragungen. Können Vertreterinnen und Vertreter der Zielgruppen nicht direkt befragt werden, gibt es manchmal die Möglichkeit, dass Mittler beziehungsweise im unmittelbaren Umfeld der Zielgruppen tätige Akteure für eine Einschätzung der Wirkungen bei den Zielgruppen beigezogen werden können.



Stiftung Zewo

Schweizerische Zertifizierungsstelle für gemeinnützige,
Spenden sammelnde Organisationen

Stiftung Zewo
Lägernstrasse 27
8037 Zürich
Telefon +41 (0)44 366 99 55

Wirkungsmessung für NPO

Zewo-Leitfaden für Dienstleistungen und Projekte im Inland



Spezialfälle



Die Wirkungsmessung befasst sich mit konkreten Dienstleistungen oder Projekten. In der Praxis müssen deshalb spezielle Gegebenheiten, Eigenheiten oder Umstände einer Dienstleistung oder eines Projekts berücksichtigt werden. Aus diesem Grund müssen in der Planung der Wirkungsmessung adäquate Lösungen für solche Besonderheiten gefunden werden. In der Folge werden zwei Spezialfälle kurz diskutiert: Dienstleistungen mit Einmalkontakten und Case Management. Zum Case Management wird zudem noch ein kurzes Beispiel erläutert, wie die Wirkungsmessung im Rahmen des Case Managements geplant werden könnte (Ressourcenmodell).



Stiftung Zewo
Schweizerische Zertifizierungsstelle für gemeinnützige,
Spenden sammelnde Organisationen

Stiftung Zewo
Lägernstrasse 27
8037 Zürich
Telefon +41 (0)44 366 99 55

Wirkungsmessung für NPO

Zewo-Leitfaden für Dienstleistungen und Projekte im Inland



Dienstleistungen mit Einmalkontakten

Bei Dienstleistungen, bei denen häufig nur einmalige Kontakte mit der Zielgruppe stattfinden (z.B. Sorgentelefon), besteht die Schwierigkeit die Wirkung dieser Tätigkeiten auf eine geeignete Art zu messen. Je nach Thematik der Dienstleistung müssen direkte Befragungen der Zielgruppe als nicht geeignet eingestuft werden (z.B. bei Anrufen von suizidgefährdeten Personen).

In solchen Fällen gilt es Wege zu finden, wie eine Wirkungsmessung über alternative Ansätze erfolgen kann. Folgende Punkte können helfen eine alternative und finanzierbare Betrachtungsweise zu bestimmen:

Was für alternative Methoden gibt es?

- Die Wirkung kann über breiter abgestützte, repräsentative Befragungen der potentiellen Zielgruppe analysiert werden
- Die leistungserbringende Person im Feld selbst sammelt Daten für die Wirkungsmessung

Mit wem können grössere Untersuchungsprojekte durchgeführt werden?

- Öffentlich finanzierte Leistungen können evtl. in andere Untersuchungen der öffentlichen Hand integriert werden
- Organisationen mit ähnlichen Angeboten können gemeinsam eine breiter angelegte Studie zur Wirkungsmessung finanzieren
- Universitäten, Fachhochschulen oder andere Institutionen können angefragt werden
- Unter Umständen existieren bereits bestehende Untersuchungen oder Forschungsprojekte, in welche die Analyse eingeschlossen werden könnte



Stiftung Zewo

Schweizerische Zertifizierungsstelle für gemeinnützige,
Spenden sammelnde Organisationen

Stiftung Zewo
Lägernstrasse 27
8037 Zürich
Telefon +41 (0)44 366 99 55

Wirkungsmessung für NPO

Zewo-Leitfaden für Dienstleistungen und Projekte im Inland



Case Management

Case Management ist ein spezifisches Verfahren zur koordinierten Bearbeitung komplexer Fragestellungen im Sozial-, Gesundheits- und Versicherungsbereich. In einem systematisch geführten, kooperativen Prozess wird eine auf den individuellen Bedarf abgestimmte Dienstleistung erbracht bzw. unterstützt, um gemeinsam vereinbarte Ziele und Wirkungen mit hoher Qualität effizient zu erreichen. In den meisten Case Management Konzepten werden verschiedene Prozess-Schritte auf der Klientenebene beschrieben, welche von der Aufnahme, über das Assessment, zur Zielvereinbarung, Durchführung von Massnahmen bis hin zur Evaluation reichen. Ausgehend von den Informationen und Daten des Case Management sind durchaus Aussagen zu den Wirkung der Massnahmen auf individueller Ebene möglich. Können diese Informationen in geeigneter Form aggregiert werden, ermöglichen sie umfassende Aussagen zu Dienstleistungen oder Projekten.

Beim Case Management werden Einzelfälle betrachtet. Der Fokus liegt dabei auf einzelnen Klienten. In der Regel wird dabei qualitativ mit persönlichen Befragungen gearbeitet.

Für die Wirkungsmessung muss überlegt werden, wie diese Datenerhebungen genutzt werden können. Insbesondere gilt es zu klären, wie die Daten für die Wirkungsmessung ausgewertet werden können. Zu beachten gibt es folgende Fragen:

- Handelt es sich um objektive oder um subjektive Daten?
- Können Kontrollvariablen erfasst werden?
- Ist es möglich Vergleichsgruppen zu erfassen?
- Können die Daten quantifiziert werden?



Stiftung Zewo
Schweizerische Zertifizierungsstelle für gemeinnützige,
Spenden sammelnde Organisationen

Stiftung Zewo
Lägernstrasse 27
8037 Zürich
Telefon +41 (0)44 366 99 55

Wirkungsmessung für NPO

Zewo-Leitfaden für Dienstleistungen und Projekte im Inland



Beispiel: Ressourcenmodell und Wirkungsinstrument aus dem Bereich der Suchtarbeit

Anhand eines standardisierten Interviews zu Beginn der Arbeit mit einem Klienten werden die im Vorfeld definierten Schlüsselressourcen erfragt. Diese werden dann über den Zeitraum der Betreuung beobachtet und immer wieder erfasst. So kann festgestellt werden, wie sich die Ressourcen in verschiedenen Bereichen entwickelt haben.

Die 16 Schlüsselressourcen sind wie folgt definiert:

Primärgruppe

- Selbstbezug
- Beziehungs- und Konfliktverhalten
- Wohnfähigkeit
- Selbstadministration

Bildung

- Schul- und Ausbildung
- Fort- und Weiterbildung
- Lernbereitschaft und –fähigkeit
- Kriterienfähigkeit

Arbeit

- Arbeitstätigkeit
- Einkommen/Vermögen
- Rollenhandeln
- Status

Freizeit

- Freie Beziehungen
- Animationspotenzial
- Sozialkompetenz
- Verwurzelte Sinnorientierung

Bei der Beurteilung werden drei Levels unterschieden:

1. Mittelschicht
2. Unterschicht
3. Desintegration, Randexistenz

Quelle: *Denkwerkstatt sie&m GmbH*

Wichtig

Analog zu diesem Beispiel aus dem Suchtbereich können im Case Management auch in anderen thematischen Bereichen Schlüsselressourcen definiert werden. Dies ist eine gute Hilfe um individuelle Entwicklungsprozesse systematisch zu erfassen.



Wirkungsmessung für NPO

Zewo-Leitfaden für Dienstleistungen und Projekte im Inland



Praxisbeispiel zum Case Management

Zur Förderung der sozialen Integration von Menschen mit Behinderung bietet eine gemeinnützige Non-Profit-Organisation (PI) eine breite Palette unterschiedlicher Dienstleistungen an. Eine davon ist die Sozialberatung, die eine umfassende Stärkung der Ressourcen von Klientinnen und Klienten vorsieht, indem deren Situation ganzheitlich erfasst und bei Bedarf mehrschichtige Interventionen eingeleitet werden.

In diesem Zusammenhang stellte Pro Infirmis in den vergangenen Jahren eine Zunahme der Komplexität der Fälle fest (Walcher 2009). In der Beratungspraxis Oftstellen sich den Klientinnen und Klienten oft gleichzeitig dringliche Fragen zu in verschiedenen Lebensbereichen (Wohnen, Arbeit, Finanzen, Gesundheit u.a.) und Daraus kann sich ein Bedarf an Leistungen mehrerer Institutionen des Sozial- und Gesundheitsbereichs ergeben. Mit der Zunahme der beteiligten Leistungserbringer steigt auch die Komplexität des Unterstützungsprozesses. Um in solchen Situationen die zielorientierte Kooperation der Leistungserbringer im Sinne der Klientinnen und Klienten zu gewährleisten, hat die gemeinnützige Non-Profit-Organisation das Case Management eingeführt, welches die folgende Ziele verfolgt:

- die Handlungsmöglichkeiten der Klientinnen und Klienten sollen erweitert werden
- interne und externe Ressourcen sollen genutzt und erschlossen werden
- zwischen Institutionen den beteiligten Institutionen soll eine aktive und verbindliche Zusammenarbeit zwischen Institutionen herbeigeführt werden

Als teils privat teils aus der öffentlichen Hand finanzierte Organisation setzt sich die gemeinnützige Non-Profit-Organisation immer auch mit der Frage auseinander, ob die vorhandenen Mittel optimal eingesetzt werden und ob die bestehenden Angebote und Dienstleistungen tatsächlich die erwünschten Wirkungen erzielen. Deshalb hat die Organisation vor der Lancierung des Case Management-Angebotes unter wissenschaftlicher Begleitung einer Fachhochschule, ein Konzept zur Wirkungsmessung entwickelt.

Zur Wirkungsmessung wurde ein multiperspektivischer Ansatz gewählt, der einerseits aus einer Wirkungsprüfung durch Klientinnen und Klienten und Case-Management-Fachkräfte besteht und andererseits eine Erhebung im Längsschnitt beinhaltet, die entlang des CM-Prozesses erfolgt ist und Aufschluss über die Veränderung der Lebensqualität gibt.

Zentrale Fragestellungen auf der Ebene Wirkungen waren dabei:

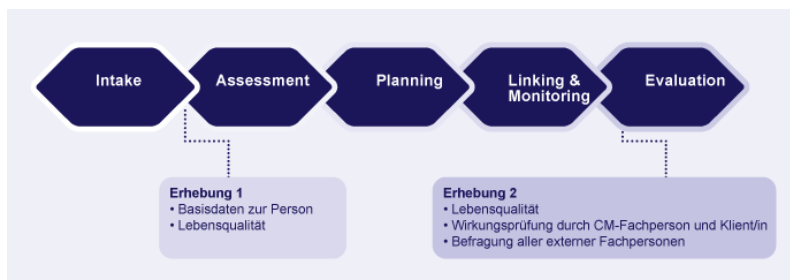
- Hat sich die Lebensqualität bzw. die Selbstwirksamkeit der Klientinnen und Klienten nach Abschluss des Case Managements verbessert?
- In welchen Lebensbereichen erreichen die Klientinnen und Klienten

Verbesserungen? Welche Bereiche bleiben stabil? Gibt es Bereiche mit Verschlechterungen?

- Besteht ein Zusammenhang zwischen der Ausgangslage der Klientinnen und Klienten und den beobachteten Veränderungen zur Lebensqualität?
- Besteht ein Zusammenhang zwischen der Interventionsdichte der Beratungen und der Veränderung der Lebensqualität?
- In welchen Lebensbereichen wurden Rahmenziele für das Case Management entwickelt? Konnten die Ziele aus Sicht der Case Managerinnen und Manager erreicht werden? Konnten die Ziele aus Sicht der Klientinnen und Klienten bzw. Angehörigen erreicht werden? Unterscheidet sich die Sicht der Case Managerinnen und Manger von der Sicht der Klientinnen und Klienten?
- Wie schätzen die externen Fachkräfte die mit dem Case Management intendierten Wirkungen ein?

Um Wirkungen messen zu können, wurden die Erhebungen im Längsschnitt angelegt. Dabei orientierten sich die Erhebungen am Verfahrensablauf des Case Managements und erfolgten zu zwei unterschiedlichen Zeitpunkten: Eine erste Erhebung erfolgte in der Anfangsphase der Unterstützung. Weitere Befragungen haben mit Abschluss des Case Managements stattgefunden.

Damit kann die Situation zu Beginn des Case Management-Prozesses mit jener beim Abschluss verglichen werden.



Wie aus der Abbildung hervorgeht, sind im Rahmen des Case Management Prozesses die folgenden Erhebungen geplant:

1. **Basiserhebung:** Zu sozio-demografischen Merkmalen und zur Behinderungsart
2. **Erhebung Lebensqualität:** Die Messung der Lebensqualität erfolgt mittels eines Erhebungsinstrumentes der Weltgesundheitsorganisation (Angermeyer et al. 2002). Dieses wird mit Fragen zur Selbstwirksamkeit ergänzt. Der Fragebogen umfasst 30 Items und kann von den Klientinnen und Klienten selbst oder stellvertretend von nahen Angehörigen ausgefüllt werden. Die Klientinnen und Klienten füllen zu Beginn des Unterstützungsprozesses einen Fragebogen aus und einen bei Abschluss des Case Managements.
3. **Erhebung Wirkungsüberprüfung im Case Management:** Case-Management-Fachperson sowie Klientinnen und Klienten nehmen eine Einschätzung der Wirkung vor, indem sie bei Fallabschluss auf die vereinbarten Ziele zurückblicken. Diese Einschätzungen werden von der Case-Management-Fachperson und der Klientin oder dem Klienten je separat vorgenommen. Der Fragebogen umfasst die Lebensbereiche Gesundheit, Hilfsmittel und Therapie, Lebensplan/-sinn, Soziale Kontakte und Familie, Tagesstruktur, Arbeit/Ausbildung, Freizeit, Finanzen sowie Wohnen. Darüber hinaus werden die Klientinnen und Klienten um eine allgemeine Einschätzung des Case Managements gebeten.

4. **Befragung der am Fall beteiligten externen Fachpersonen:** Bei Fallabschluss werden die Case Managerinnen und Manager alle relevanten Fachpersonen um eine Einschätzung des Unterstützungsprozesses beten. Der Fragebogen thematisiert die im Case Management intendierten Wirkungen auf der Systemebene.

(Quelle: „Case Management Pro Infirmis – Zwischenbericht zur Wirkungsmessung der Periode Juli 2009 bis August 2012“, Prof. Dr. Haller, BFH/Fachbereich Soziale Arbeit, 2013)



Stiftung Zewo

Schweizerische Zertifizierungsstelle für gemeinnützige,
Spenden sammelnde Organisationen

Stiftung Zewo
Lägernstrasse 27
8037 Zürich
Telefon +41 (0)44 366 99 55

Wirkungsmessung für NPO

Zewo-Leitfaden für Dienstleistungen und Projekte im Inland



Durchführung

Die Durchführung der Wirkungsmessung kann in verschiedener Weise erfolgen. Es ist zu planen, durch wen die Daten erhoben und ausgewertet werden. Die Wirkungsmessung kann grundsätzlich von den Dienstleistungs- oder den Projektverantwortlichen selber (Selbstevaluation) oder von unabhängigen Dritten (Fremdevaluation) durchgeführt werden. Dazwischen gibt es verschiedene Mischformen (Hybridevaluation).

Dabei ist Folgendes zu beachten:

- **Unabhängigkeit**
Externe Experten oder auch Dienstleistungs- oder den Projektverantwortliche müssen, je nach Verwendungszweck der Wirkungsmessung (z.B. fürs Legitimieren oder Lenken), die nötige Unabhängigkeit haben, um von den Nutzerinnen und Nutzern der Ergebnisse als genügend unparteiisch und unvoreingenommen wahrgenommen zu werden. Für andere, interne Zwecke gibt es wichtigere Aspekte als die Unabhängigkeit (z.B. die Fachkompetenz beim Lernen).
- **Glaubwürdigkeit**
Das Team muss die nötige fachliche und methodische Kompetenz haben, um die Evaluation fachgerecht durchführen zu können.
- **Akzeptanz**
Das Team muss das nötige Einfühlungsvermögen und die nötige Erfahrung haben, um von den Beteiligten und den Nutzerinnen und Nutzern der Ergebnisse akzeptiert zu werden.



Wirkungsmessung für NPO

Zewo-Leitfaden für Dienstleistungen und Projekte im Inland



Selbstevaluation

Die Selbstevaluation ist ein Verfahren, bei dem die praxisgestaltenden Fachleute identisch sind mit den Evaluatorinnen und Evaluatoren. Dies bedeutet, dass die Akteure ihre eigene Tätigkeit überprüfen: Sie sind also gleichzeitig praxisverantwortlich und evaluationsverantwortlich.¹ Inhaltlich unterscheiden sich die Fragestellungen der Selbstevaluation nicht von denjenigen einer externen Evaluation. Beurteilt werden insbesondere die Relevanz, die Wirksamkeit und die Wirtschaftlichkeit einer Dienstleistung oder eines Projekts. Bei einer unterstützten Selbstevaluation berät und unterstützt ein anerkanntes Institut oder eine anerkannte Fachperson die Dienstleistungs- oder Projektdurchführenden in der Planung, Durchführung und Berichterstattung einer Selbstevaluation.

Vorteile

- Motivation der Evaluierten ist grösser
- Insiderwissen kann eingesetzt werden
- Evaluatoren sind mit dem Thema vertraut
- Ownership für Ergebnisse führt zu rascher Umsetzung
- Akzeptanz der Ergebnisse in der Organisation grösser
- Organisation ist weniger aufwändig

Nachteile

- Weniger Distanz
- Grundsatzfragen werden weniger gestellt
- Evaluationswissen fehlt
- Weniger Legitimationskraft gegen aussen

Eignung

Die Selbstevaluation eignet sich besonders für Evaluationen, die gemacht werden, um daraus zu lernen. Sie eignet sich gut dafür Prozesse zu analysieren und schrittweise Verbesserungen zu erzielen. Selbstevaluationen können auch bei knappen finanziellen Ressourcen durchgeführt werden.

WICHTIG

Auch bei Selbstevaluationen ist darauf zu achten, dass dafür genügend Zeit und Ressourcen eingesetzt werden, dass das nötige Know-how vorhanden ist oder bereit gestellt wird und Verbindlichkeiten klar festgelegt sind. Selbstevaluationen scheitern sonst gern an der Methodenkomplexität und mangelnden Ressourcen.

¹ Quelle: Hildegard Müller-Kohlenberg, Wolfgang Beywl (2003): Standards der Selbstevaluation, Begründung und aktueller Diskussionsstand, Zeitschrift für Evaluation 1/2003, Köln.



Stiftung Zewo
Schweizerische Zertifizierungsstelle für gemeinnützige,
Spenden sammelnde Organisationen

Stiftung Zewo
Lägernstrasse 27
8037 Zürich
Telefon +41 (0)44 366 99 55

Wirkungsmessung für NPO

Zewo-Leitfaden für Dienstleistungen und Projekte im Inland



Fremdevaluation

Die Verantwortung für die Durchführung liegt bei Personen, welche nicht in die Umsetzung einer Massnahme einbezogen sind. Expertinnen und Experten verschaffen sich Informationen über den Evaluationsgegenstand und geben den Verantwortlichen Rückmeldungen darüber, wie sie diesen einschätzen.

Vorteile

Keine «Betriebsblindheit»
Unbefangenheit

Methodenkompetenz

Legitimationskraft gegen aussen ist
höher

Nachteile

Geringe Einbindung in Arbeitsfeld
Kontextbezug und Sachkompetenz
fehlen

Akzeptanz der Ergebnisse in der
Organisation kleiner

Eignung

Die Fremdevaluation eignet sich besonders für Evaluationen, die gemacht werden, um eine Tätigkeit zu legitimieren. Fremdevaluationen kommen zum Einsatz, wenn die eigenen personellen Ressourcen knapp sind, wenn wenig internes Evaluations-Know-how vorhanden ist oder wenn die Finanzierung (und die Planung) einer Tätigkeit ausserhalb der ausführenden Organisation erfolgt ist.

WICHTIG

Die Zielsetzung und die Fragestellung der Evaluation müssen klar festgelegt werden. Die Auswahl geeigneter externer Evaluatoren ist entscheidend für das Gelingen der Evaluation. Evaluatoren müssen natürlich die nötigen Sachkenntnisse mitbringen, aber auch über die nötige Sensibilität im Umgang mit verschiedenen Anspruchsgruppen verfügen und von diesen akzeptiert werden. Schliesslich sind realistische Zeitpläne und Budgets bereitzustellen.

Partner

Als potentielle Partner für Fremdevaluationen kommen entweder private Evaluationsunternehmen oder Bildungseinrichtungen wie Universitäten oder Fachhochschulen in Frage. Welche Partner sich am besten eignen, hängt vom Themenfeld ab.

Schweizerische Evaluationsgesellschaft Seval

Liste von anerkannten Schweizer Hochschulen



Stiftung Zewo
Schweizerische Zertifizierungsstelle für gemeinnützige,
Spenden sammelnde Organisationen

Stiftung Zewo
Lägernstrasse 27
8037 Zürich
Telefon +41 (0)44 366 99 55

Wirkungsmessung für NPO

Zewo-Leitfaden für Dienstleistungen und Projekte im Inland



Hybridevaluation

Die Kombination von Selbst- und Fremdevaluation kann als Hybridevaluation bezeichnet werden. Dabei werden die Vorteile von Selbst- und Fremdevaluation kombiniert. Im Zentrum dieses Ansatzes steht ein systematischer, durch ein unabhängiges Evaluationsteam unterstützter Prozess einer Selbstevaluation. Das Evaluationsteam unterstützt die Dienstleistungs- oder Projektverantwortlichen bei der Selbstevaluation und der kritischen Beurteilung des Projekts. Daten und Informationen, welche der Beurteilung zugrunde liegen, werden dabei vom Evaluationsteam verifiziert, zusätzliche Fragen werden aufgeworfen und alternative Hypothesen der Interpretation der Daten zur Diskussion gestellt. Dieses Vorgehen stellt sicher, dass die Evaluation eng auf die Bedürfnisse der Leistungserbringenden oder des Projektteams abgestimmt und der distanzierte Blick dennoch gewährleistet ist.

Vorteile

Informationen können rechtzeitig ins Projekt einfließen

Lernorientierte Atmosphäre

Nachteile

Rollen sind nicht immer klar

Braucht Zeit, bis alle ihre Rolle kennen

Ressourceneinsatz insgesamt hoch

Eignung

Hybridevaluationen eignen sich sowohl fürs Lernen als auch fürs Legitimieren. Bei knappen eigenen personellen Ressourcen oder wenn wenig internes Evaluations-Know-how vorhanden ist, aber trotzdem nicht auf die Vorteile einer Selbstevaluation verzichtet werden soll, ist die Hybridevaluation das geeignete Mittel.

WICHTIG

Die Zielsetzung und die Fragestellung der Evaluation müssen klar festgelegt werden. Die Auswahl geeigneter externer Evaluatoren ist entscheidend für das Gelingen der Evaluation. Der Verteilung der Rollen und der Festlegung der Verantwortlichkeiten sollte besondere Aufmerksamkeit geschenkt werden.



Wirkungsmessung für NPO

Zewo-Leitfaden für Dienstleistungen und Projekte im Inland



Indikatoren

Damit Wirkungen gemessen werden können und das Erreichen von Wirkungszielen überprüft werden kann, müssen diese in objektiv messbaren Grössen konkretisiert werden. Eine solche objektiv messbare Grösse wird Indikator genannt. Ein Indikator gibt Antwort auf die Frage: «Wie werden wir wissen, ob das, was wir geplant haben, auch tatsächlich eingetroffen ist und dass wir unser Ziel erreicht haben?» Ein Indikator für ein Wirkungsziel gibt also an, wie wir die Wirkung messen und wann wir das Wirkungsziel als erreicht betrachten:

Ziel	Indikator	Soll
Verbesserte Integration von Schulabbrechern in den Arbeitsmarkt	Prozentsatz der Schulabbrecher, welche innerhalb von 6 Monaten eine Arbeitsstelle finden	80%

WICHTIG

Ein guter Indikator sollte **SMART** sein:

- Specific: Der Indikator muss eindeutig und klar sein.
- Measurable: Der Indikator muss messbar sein, auch mit angemessenen Kosten.
- Achievable: Der vom Indikator vorgegebene Sollwert muss erreichbar sein.
- Relevant: Die Information die der Indikator liefert soll relevant für Projektmanager sein.
- Time-bound: Der Indikator muss angeben, bis wann das Ziel erreicht werden soll.

Quelle: *European Commission, PCM Guidelines*

Gute Indikatoren zu finden und zu wählen ist entscheidend dafür, dass eine Wirkungsmessung nützliche Informationen liefert, aber keineswegs einfach. Eine hohe Bedeutung kommt auch hier der partizipativen Erarbeitung zu: Ein guter Indikator wird insbesondere auch von der Zielgruppe akzeptiert und als sinnvoll empfunden.

Es ist oft nötig, mehrere Indikatoren für ein Ziel zu definieren. In der Praxis werden quantitative und qualitative Indikatoren häufig kombiniert. Grundsätzlich gilt aber, dass man sich auf ein Minimum beschränken sollte, um keine unnötigen Datenmengen zu produzieren.

Zu einem Indikator gehört immer auch die Angabe, aus welcher Quelle die

Daten zum Indikator stammen und mit welcher Methode diese erhoben werden. Dadurch wird sichergestellt, dass der Indikator messbar ist. Zeigt es sich dabei, dass die Daten zum Indikator nicht oder nur mit unverhältnismässigem Aufwand erhoben werden können, muss der Indikator durch einen einfacheren ersetzt werden. Zu berücksichtigen ist auch die Möglichkeit, auf bestehende Quellen zurückzugreifen, z.B. auf nationale Statistiken oder auf Daten von Partnerorganisationen.

ACHTUNG

In der Praxis wird den Quellen oft zu wenig Aufmerksamkeit gewidmet und die Daten sind später gar nicht verfügbar, was den Indikator nutzlos macht. Ein Indikator ohne echte Datenquelle ist nicht messbar und kann damit nicht für eine Wirkungsmessung verwendet werden.

In der Praxis gibt es verschiedene Varianten Ziel und Indikator zu definieren bzw. voneinander zu unterscheiden. Diese sind gleichberechtigt. Man sollte sich jedoch immer bewusst sein, welche Definition gilt und diese konsistent anwenden. Diskussionen darüber sind fast unvermeidbar. Zu beachten ist, dass ein Indikator im hier verwendeten Sinne (der Indikator gibt an, ob das Ziel erreicht wurde oder nicht) immer auch eine Zielsetzung und damit die Wahl eines Vergleiches enthält.

Indikatoren sollen sowohl auf den Wirkungsebenen (insbesondere Outcome), wie auch auf der Leistungsebene (Output) festgelegt werden.

Beispiele

- **Projekt Schuldenberatungskurs**

Ziel: Menschen mit Schulden aufgrund von zu hohen Ausgaben für Konsum sollen ihre finanzielle Lage in den Griff bekommen.

Indikator: Prozentsatz x der 250 Kursteilnehmer geben ein Jahr nach dem Kursbesuch an, dass sich ihre finanzielle Lage stabilisiert hat.

Soll: 70%

- **Projekt Förderung landwirtschaftliche Infrastruktur im Berggebiet**

Ziel: Die ökonomische Situation der Bauern in den Berggebieten soll verbessert werden.

Indikator: Ein Prozentsatz x der Begünstigten geben an, dass die realisierten Infrastrukturprojekte einen positiven Effekt auf die ökonomische Situation gezeigt haben.

Soll: 80%

Wirkungsmessung für NPO

Zewo-Leitfaden für Dienstleistungen und Projekte im Inland



Beispiel Schuldenberatung

Durch das Beratungsangebot für Menschen mit Schuldenproblemen soll die Anzahl von verschuldeten Menschen gesenkt und damit ein Beitrag zur Reduzierung der Armut geleistet werden.



© Urs Siegenthaler / Caritas Zürich

	Ziel	Indikator	Soll	Messung
Wirkungsziel (W1)	Zielgruppe kann Schulden reduzieren	Prozentsatz x der Personen, welche ein Angebot der Schuldenberatung in Anspruch genommen haben, konnten innerhalb von 24 Montaten nach dem Erstkontakt ihre Schulden halbieren	>60%	Befragung
<i>Wirkungshypothese W1</i>	<i>Die Schuldenberatung trägt dazu bei, dass die Zielgruppe ihre Schulden reduzieren kann.</i>			
Leistungsziel (L1)	Steigerung der individuellen Beratungskontakte	Die Anzahl der Kontakte mit betroffenen Menschen	Steigerung um 20%	Erfassung der Kontakte
<i>Wirkungshypothese L1</i>	<i>Je grösser die Anzahl der Beratungskontakte, desto mehr verschuldete oder von der Verschuldung gefährdete Haushalte können erreicht werden.</i>			
Leistungsziel (L2)	Kursangebot ausbauen	Anzahl Teilnehmer, die einen Kurs zur Bewältigung von finanziellen	250 Teilnehmer im kommenden Jahr	Teilnehmerlisten

Notlagen
besuchen

Wirkungshypothese *Je grösser die Anzahl der Teilnehmenden, desto grösser ist die*
L2 Wirkung auf verschuldete oder von der Verschuldung gefährdete
Haushalte.



Wirkungsmessung für NPO

Zewo-Leitfaden für Dienstleistungen und Projekte im Inland



Beispiel Wolf

Durch verschiedene Massnahmen sollen Bedingungen geschaffen werden, dass sich die Population von Wölfen in den Alpen wieder hergestellt wird und Wölfe in der freien Wildbahn leben können.



© Chris Martin Bahr / WWF-Canon

	Ziel	Indikator	Soll	Messung
Wirkungsziel (W1)	Ein erheblicher Anteil des potentiellen Lebensraums von Wölfen in den Alpen stehen unter Schutz oder nachhaltiger Nutzung	Anzahl Hektare des Habitats in den Gebieten a, b und c der Alpen unter Schutz und nachhaltiger Nutzung	Steigerung des Gebietes unter Schutz oder mit nachhaltiger Nutzung bis zum Jahr 2018 um 15% im Vergleich zu 2005	Öffentliche Daten des Bundes
Wirkungshypothese W1	<i>Wenn der Lebensraum unter Schutz steht, kann sich die Wolfpopulation in der Wildnis entwickeln</i>			
Leistungsziel (L1)	Politisches Lobbying und öffentlicher Druck führen dazu, dass der Schutzstatus der Grossraubtiere in der Schweiz verbessert wird und Abschüsse nicht zulässig sind.	Status des Gesetzes	Erreichung der politischen Etappenziele A und B bis 2015	Beschlüsse der Politik
Wirkungshypothese L1	<i>Ein verbesserter gesetzlicher Schutz ist ein erster Schritt zur Vergrösserung der Population des Wolfes</i>			
Wirkungsziel (W2)	Ein verbesserter Herdenschutz reduziert das Konfliktpotential	Anzahl eingesetzte Herdenschutzhunde	Jährliche Steigerung eingesetzter Herdenschutzhunde	Umfrage

	zwischen Wolf und Mensch		um x%	
Wirkungshypothese W2	<i>Je weniger Schafe gerissen werden, desto grösser die Akzeptanz bei den Bauern in den Berggebieten</i>			
Leistungsziel (L2)	Ein erhöhtes Angebot von Herdenschutzkursen wird von Landwirten in den Berggebieten genutzt.	Anzahl Kurse und Anzahl Teilnehmer	Steigerung der Anzahl Kurse und Teilnehmer im Vergleich zum Vorjahr	Teilnehmerlisten
Wirkungshypothese L2	<i>Je mehr Kurse angeboten werden, desto mehr Bauern in den Berggebieten können erreicht werden</i>			
Wirkungsziel (W3)	Der Tourismus im Berggebiet profitiert ökonomisch oder kulturell vom Wolfschutz	Anzahl positive Presseberichte zum Wolf	5 positive Berichte in regionalen Zeitungen	Mediendienste
Wirkungshypothese W3	<i>Je mehr Menschen direkt oder indirekt vom Wolfschutz profitieren, desto eher/stärker werden Schutzmassnahmen unterstützt</i>			
Leistungsziel (L3)	Die Erlebniswege zum Thema Wolf und Bär erreichen viele Menschen	Anzahl Menschen erreicht	5'000 wurden im Jahr x erreicht	Schätzung
Wirkungshypothese L3	<i>Je mehr Menschen über positive Erlebnisse über das mögliche Miteinander von Wolf und Mensch aufgeklärt werden, desto höher ist die Akzeptanz des Wolfes in den Alpen</i>			



Wirkungsmessung für NPO

Zewo-Leitfaden für Dienstleistungen und Projekte im Inland

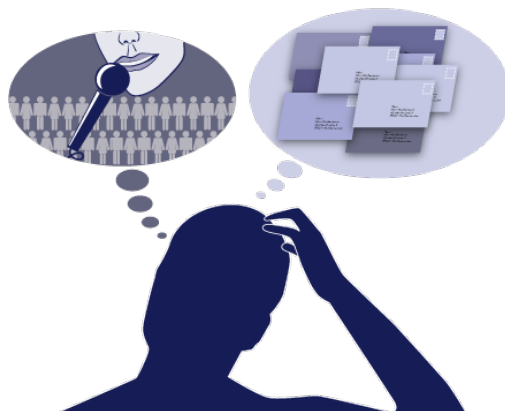


Erhebungsmethoden

Im Rahmen der Wirkungsmessung können verschiedene Erhebungsmethoden angewendet werden. Grundsätzlich kann zwischen qualitativen und quantitativen Erhebungsmethoden unterschieden werden. Neben der Wahl der Erhebungsmethode gilt es, in einem zweiten Schritt zu überlegen, ob alle oder nur ein Teil der betroffenen Einheiten, Zielgruppen oder Fälle in der Erhebung zur Wirkungsmessung berücksichtigt werden sollen (Untersuchungseinheit). Weiter gilt es, die Qualität der erhobenen Daten kritisch zu prüfen (Datenqualität).

Methodenmix

Im heutigen Forschungsalltag ist es üblich, qualitative und quantitative Methoden kombiniert einzusetzen und so von den Vorteilen beider Methoden zu profitieren. Man spricht in diesem Fall von einem Methodenmix. Der Nutzen der Kombination von qualitativen und quantitativen Methoden gilt als unbestritten und ist zum Standard aussagekräftiger Wirkungsmessungen geworden. So kann z.B. bei der Messung der Effektivität eines Programms einerseits ein standardisierter Fragebogen an die Zielgruppen verteilt und andererseits ein Interview mit Mitarbeitenden oder eine Gruppendiskussion mit Expertinnen und Experten durchgeführt werden. In Abhängigkeit der gewählten Erhebungsmethode haben die erhobenen Daten eine bestimmte Form (Interviewprotokolle, Protokolle von Expertengesprächen, Prozentzahlen aus einer Befragung, Häufigkeiten einer Beobachtung usw.) und müssen daher mit geeigneten Auswertungsverfahren analysiert werden.





Wirkungsmessung für NPO

Zewo-Leitfaden für Dienstleistungen und Projekte im Inland



Quantitative Methoden

Bei der quantitativen Methodik geht es darum, Verhalten oder Veränderungen in Form von zahlenmässigen Ausprägungen möglichst genau zu beschreiben und zu erfassen.

Anzahl der Untersuchungseinheiten:	Viele
Annahmen:	Klare Vorstellung über relevante Zusammenhänge
Ausgangspunkt:	Überprüfen von Vorstellungen
Fokus:	Wissen der Forschenden steht im Zentrum
Absicht:	Theorieevaluation

Eignung

Quantitative Verfahren eignen sich mit ihrer standardisierten Befragungs- und Beobachtungsform für die Untersuchung grosser Stichproben und für die Anwendung statistischer Auswertungsverfahren zur objektiven Messung und Quantifizierung von Sachverhalten. Sie sind ideal, um objektive Daten über die Zeit zu vergleichen und daraus Entwicklungen abzulesen. Quantitative Datenerhebungsmethoden erlauben es, eine grosse Menge von Informationen über zuvor definierte Verfahren zu untersuchen. Die gewonnenen Informationen können mit Hilfe statistischer Verfahren und Auswertungstechniken analysiert und verglichen werden.

Erhebung

Quantitative Daten werden mit folgenden Verfahren erhoben:

- Strukturierte Beobachtung, Messung, Zählung
- Analyse von Sekundärdaten (Statistiken, Prozessdaten)
- Verschiedene Formen von Befragungen und Experimenten

Stichprobengrösse

Die zu wählende Stichprobengrösse ist davon abhängig, wie genau die aus der Befragung resultierenden Ergebnisse sein müssen. Am einfachsten ist es daher, wenn alle Untersuchungseinheiten befragt werden können. Man spricht dann von einer Vollerhebung. Bei einer Vollerhebung erübrigen sich statistische Tests zur Signifikanz von Unterschieden, weil die Daten nicht auf einer Sichtprobe basieren, welche auf die Grundgesamtheit hochgerechnet werden. Aus der folgenden Tabelle lässt sich ableiten, dass bei Untersuchungseinheiten, die weniger als 300 Fälle umfassen am besten eine Vollerhebung durchgeführt wird. Ersichtlich wird auch, dass mit 300 befragten

Einheiten relativ verlässliche Aussagen zu grossen Grundgesamtheiten gemacht werden können.

N =Grösse der Grundgesamtheit

Grösse der Grundgesamtheit	n = minimale Stichprobengrösse bei einem Fehler von +/-3 Prozentpunkten	n = minimale Stichprobengrösse bei einem Fehler von +/-5 Prozentpunkten
100	92	80
200	169	132
300	234	169
400	291	196
500	341	217
1000	516	278
5000	880	357
10 000	964	370
100 000	1056	383
1 000 000	1066	384

Non-response

Non-response (Nichtbeantwortung) kann bei Befragungen zu einem Problem führen, weil durch fehlende Antworten die Stichprobe verkleinern und damit die Aussagekraft der Daten schmälern. Das non-response-Problem kann daher die Repräsentativität einer Stichprobe gravierend beeinträchtigen, weil eine Nichtbeantwortung meist mit den Befragungsgegenständen verknüpft ist. Mögliche Lösungen der non-response-Problematik bilden Nachfassaktionen beziehungsweise gezielte Nachfragen bei Nicht-Teilnehmenden (sofern diese bekannt sind).

Auswertung

Bei quantitativen Erhebungsverfahren erfolgt die Auswertung mittels verschiedener statistischer Verfahren und Kennzahlen wie Häufigkeiten, prozentuale Anteile oder Mittelwerte sowie mit komplexeren statistischen Verfahren.

Vorteile

Exakt quantifizierbare Ergebnisse

Ermittlung von statistischen Zusammenhängen möglich

Möglichkeit, eine grosse Stichprobe zu untersuchen und damit repräsentative Ergebnisse zu erhalten

Nachteile

Keine Flexibilität während der Untersuchung durch die Standardisierung der Untersuchungssituation. Die Fragen sind schon vorher festgelegt, es ist kein individuelles Eingehen auf die Testpersonen möglich.

Man ermittelt nicht die Ursachen für einen Befund oder eine Einstellung wie beispielsweise Unzufriedenheit. Zur Verringerung dieses Problems empfiehlt sich der Einsatz offener Fragen.

Man erhält keine Verbesserungsvorschläge. Dieser Nachteil kann ebenfalls durch die Integration offener Fragen verringert werden.

Hohe externe Validität durch grosse
Stichprobe

Grössere Objektivität und
Vergleichbarkeit der Ergebnisse



Wirkungsmessung für NPO

Zewo-Leitfaden für Dienstleistungen und Projekte im Inland



Qualitative Methoden

Bei der qualitativen Methodik geht es um das Beschreiben, Interpretieren und Verstehen von Zusammenhängen.

Anzahl der Untersuchungseinheiten:	Wenige
Annahmen:	Wenig gefestigtes Wissen über Wirkungszusammenhänge
Ausgangspunkt:	Bedarf an vertieften Informationen
Fokus:	Wissen der Akteure/Akteurinnen steht im Zentrum
Absicht:	Theoriekonstruktion

Eignung

Qualitative Datenerhebungsmethoden erlauben es, einen bestimmten Untersuchungsgegenstand im Detail und in der Tiefe zu untersuchen. Dabei können auch neue und nicht erwartete Informationen hervorgebracht werden. Dies kann zu einem tiefen Verständnis des Untersuchungsgegenstandes führen, erschwert aber Verallgemeinerungen über den konkret betrachteten Gegenstand hinaus. Die qualitative Befragung bzw. Beobachtung zeichnet sich durch eine unverzerrte und umfassende Informationen liefernde Herangehensweise aus und ist hierdurch überall dort geeignet, wo man eine differenzierte und ausführliche Beschreibung individueller Meinungen und Eindrücke benötigt. Insbesondere zur Sammlung von detaillierten Verbesserungsvorschlägen und zur Erkundung von Ursachen (für Sachverhalte wie beispielsweise Unzufriedenheit) sind qualitative Methoden ideal.

Erhebung

Qualitative Daten werden mit folgenden Verfahren erhoben:

- Verschiedene Formen von Interviews (Einzelgespräche, Gruppeninterviews, Fokusgruppen)
- Dokumentenanalyse

Stichprobengrösse

Über die Anzahl der Gespräche, die geführt werden sollen, gibt es in der Literatur keine einheitliche Meinung. Die Vorstellung von einer geeigneten Stichprobengrösse bewegt sich zwischen einigen wenigen Gesprächen und rund 200 Personen, wobei in Abhängigkeit der untersuchten Fragestellung ab einer gewissen Anzahl Gespräche eine theoretische Sättigung eintritt. Das heisst, dass sich der Erkenntnisgewinn durch zusätzliche Gespräche nicht

weiter steigern lässt. Die erforderliche Stichprobengröße ist im Allgemeinen deutlich geringer als bei quantitativen Verfahren. Bei der Zusammensetzung der Stichprobe gelten die Grundsätze des theoretical sampling, das heisst, die Stichprobe sollte den theoretischen Überlegungen und der Fragestellung angepasst werden, heterogen zusammengesetzt sein und möglichst typische Vertreter enthalten.

Auswertung

Bei qualitativen Erhebungsverfahren erfolgt die Auswertung anhand verschiedener Formen der Inhaltsanalyse. Diese Formen gehen von der Zusammenfassung und der schrittweisen Reduktion des Datenmaterials aus. Wichtig: Auch qualitativ gewonnene Daten können quantitativ ausgewertet werden.

Vorteile

Flexible Anwendung der Methode, die Methode passt sich an den Untersuchungsgegenstand an und nicht umgekehrt.

Die Offenheit des Vorgehens ermöglicht es, neue, bisher unbekannt Sachverhalte zu entdecken.

Da die Teilnehmenden keinerlei Vorgaben haben, erhält man eher wahre und vollständige Informationen über die subjektive Sicht der Gesprächspartner.

Der Fokus wird von den Teilnehmenden selbst bestimmt, dadurch liegt er vor allem auf den für sie relevanten Sachverhalten.

Durch die persönliche Interaktion gibt es die Möglichkeit, Hintergründe zu erfragen und Unklarheiten zu beseitigen.

Hohe inhaltliche Validität durch nicht prädestinierte Vorgehensweise

Tiefgründiger Informationsgehalt durch offene Befragung

Größere Subjektivität der Ergebnisse

Nachteile

Die Anforderungen an die Qualifikation der beobachtenden oder interviewenden Person sind recht hoch, die Qualität der Daten ist zu einem gewissen Teil auch davon abhängig.

Die Auswertung ist vor allem im Vergleich zu den quantitativen Methoden relativ aufwändig.

Aus qualitativen Daten kann man keine zahlenmässigen Mengenangaben ableiten.

Häufig wird die Robustheit der Daten aufgrund der subjektiven Perspektive bei der Dateninterpretation angezweifelt.



Wirkungsmessung für NPO

Zewo-Leitfaden für Dienstleistungen und Projekte im Inland



Untersuchungseinheit

Die Auswahl der für die Wirkungsmessung zu untersuchenden Einheiten, Zielgruppen oder Fälle hängt massgebend vom Design bzw. den Vergleichen ab.

Idealerweise werden alle betroffenen Einheiten, Zielgruppen oder Fälle in der Erhebung zur Wirkungsmessung berücksichtigt. Man spricht in diesem Fall von einer Vollerhebung. In der Praxis sind Vollerhebungen aus spezifischen Gründen oder wegen den Kosten nicht immer möglich. Daher muss entschieden werden, welche Untersuchungseinheiten bei der Wirkungsmessung berücksichtigt werden sollen. Bei quantitativen Verfahren spricht man in solchen Fällen von einer Stichprobe.

Stichproben haben gegenüber den Vollerhebungen den Nachteil, dass Informationen immer nur für einen Teil aller interessierenden Beobachtungen erhoben wird. Entsprechend muss man sich überlegen, ob die Ergebnisse, welche auf einer Stichprobenauswahl basieren, auch für die gesamte Untersuchungseinheit gelten. Ist dies nicht der Fall, dann wurde die Stichprobe nicht richtig gezogen bzw. die zu analysierenden Einheiten nicht korrekt ausgewählt.

Auswahlkriterien bei qualitativen Verfahren

Nicht nur bei quantitativen Verfahren stellt sich die Frage nach der Auswahl der Stichprobe. Auch bei qualitativen Verfahren muss man sich überlegen, welche Einheiten für die Untersuchung berücksichtigt werden sollen. Aufgrund der Auswahlkriterien ergibt sich meistens automatisch die Anzahl der zu berücksichtigenden Einheiten. Oft möchte man pro Auswahlkriterium mindestens eine oder zwei Einheiten berücksichtigen. Aus theoretischer Sicht ist die Anzahl der Untersuchungen genügend, wenn das Prinzip der Sättigung eintritt. Als saturiert gilt eine Auswahl oder Stichprobe, wenn zusätzlich Erhebungen keine neuen Informationen mehr erbringen und mit dem erhobenen Material der Erkenntnisgewinn gesättigt ist. Zur Bestimmung der richtigen Auswahl der Analyseeinheiten kann ein Vorgehen in drei Schritten angewandt werden:

- In einem ersten Schritt empfiehlt es sich, das inhaltliche Interesse an bestimmten Gruppen zu präzisieren.
- In einem zweiten Schritt kann darauf geachtet werden, ob alle möglichen Ausprägungen und Merkmale der zu untersuchenden Einheit bei der Auswahl berücksichtigt worden sind.
- Im dritten Schritt ist nach der Datenerhebung nochmals zu prüfen, welche Konstellationen und Merkmale nicht in der bisherigen Erhebung vorkommen. Gegebenenfalls sind zusätzliche Untersuchungseinheiten zu berücksichtigen. Das bedeutet auch, dass die Auswahl der Untersuchungseinheiten nicht wie bei den quantitativen Verfahren anhand eines einzigen Schrittes durchgeführt werden muss.

Bei sehr knappen Ressourcen besteht auch die Möglichkeit, dass man den Fokus auf extreme Ausprägungen legt. Unter diesen Bedingungen beschränkt

man sich auf eine Auswahl von Fällen, bei denen besonders positive bzw. besonders negative Resultate erwartet werden. Dieses Verfahren setzt folglich eine subjektive Auswahl voraus, was nicht dem optimalen Vorgehen entspricht.

Auswahlkriterien bei quantitativen Verfahren

Um Fehler oder Verzerrungen aufgrund einer falschen Auswahl von Fällen zu vermeiden, sollte genau geklärt werden wer oder was zur Population gehört, die untersucht werden soll: Es ist insbesondere darauf zu achten, dass auch schlecht erreichbare (z.B. räumlich) und marginalisierte Gruppen wie ethnische oder religiöse Minderheiten oder Frauen bei der Stichprobenziehung angemessen berücksichtigt werden. Weiter muss die Grösse der Stichprobe bestimmt werden: Zentrales Kriterium ist dabei, wie genau die Ergebnisse sein müssen. Die Grösse der Population hat, zumindest bei eher grossen Populationen, wenig Einfluss auf die Mindestgrösse der Stichprobe (siehe dazu auch quantitative Methoden). Natürlich spielen in der Praxis dabei auch die verfügbare Zeit und die Kosten eine wichtige Rolle.

Zur Auswahl von Stichproben bei quantitativen Verfahren gibt es verschiedene Möglichkeiten. Grundsätzlich wird zwischen «Zufallsstichproben» und «Nicht-Zufallsstichproben», welche nach bestimmten Kriterien zusammengestellt werden, unterschieden. Wenn die Stichprobe nach dem Zufallsprinzip zusammengesetzt sein soll, hat jeder, welcher der Grundgesamtheit angehört, dieselbe Wahrscheinlichkeit als Stichprobenmitglied «gezogen» zu werden. Im Folgenden werden einige der wichtigsten Auswahlverfahren kurz beschrieben.

Zufallsstichproben

- **Einfache Zufallsstichprobe**

Es gelangt jede Einheit der Grundgesamtheit mit der selben Wahrscheinlichkeit in die Auswahl (z.B. Namen aus einer Urne ziehen oder jede x-te Person).

- **Geschichtete Zufallsstichprobe**

Die Untersuchungseinheiten werden nach einem bestimmten Merkmal in Gruppen (Schichten) unterteilt (z.B. Gemeinden, Kurse). Aus diesen Teilgesamtheiten werden dann nach dem Zufallsprinzip Stichproben entnommen.

- **Gestufte Zufallsstichprobe**

Es werden zuerst die Stufungskriterien bestimmt (z.B. Regionen A bis D). Dann wird die Grundgesamtheit nach diesen aufgeteilt und eine zufällige Auswahl getroffen (z.B. Region B und D) und auf eine bestimmte Anzahl von Primäreinheiten begrenzt, die man untersucht (z.B. 5 Kurse je Region). Die restlichen Teilgesamtheiten werden ignoriert. Aus den zufällig ausgewählten Primäreinheiten (z.B. 5 Kurse) ermittelt man nun die Zufallsstichprobe der Merkmalsträger (je 20 Kursteilnehmer pro Kurs). In den beiden Regionen werden je 100 Kursteilnehmer befragt, die dann zu einer Gesamtstichprobe zusammengeführt werden.

Nicht-Zufallsstichproben

- **Quoten-Stichprobe**

Zuerst werden die Elemente der Grundgesamtheit in Gruppen eingeteilt. Die Stichprobe ist nun so zu ziehen, dass dieses Gruppenverhältnis in der Stichprobe möglichst genau so aussieht wie in der Grundgesamtheit. Es wird also versucht, eine gewünschte Populationsstruktur in der Stichprobe nachzuahmen. Dazu werden den Interviewern Vorgaben gemacht, welche Eigenschaften die zu Befragenden haben müssen. Wen er konkret auswählt, bleibt dabei dem Interviewer überlassen.

- **Homogene und heterogene Fallauswahl**

Die Beobachtungen werden so für die Stichprobe ausgewählt, dass sie möglichst ähnliche/unähnliche Eigenschaften aufweisen. In Fallstudien

(welche per Definition keine Zufallsstichproben darstellen) werden beispielsweise häufig zwei Beobachtungen mit möglichst gegensätzlichen Eigenschaften untersucht.

- **Die Auswahl von typischen Fällen**

Hier werden diejenigen Beobachtungen für die Stichprobe ausgewählt, von denen man weiss oder annimmt, dass sie typische, durchschnittliche und eben keine extremen Eigenschaften aufweisen.

- **Die Auswahl von kritischen Fällen**

Es werden insbesondere oder ausschliesslich diejenigen Beobachtungen ausgewählt, von denen man weiss, dass ihr Einschluss in die Stichprobe für die Glaubwürdigkeit oder Akzeptanz einer Studie ausschlaggebend ist.



Wirkungsmessung für NPO

Zewo-Leitfaden für Dienstleistungen und Projekte im Inland



Datenqualität

Die Informations- oder Datenqualität bezeichnet die Relevanz und Korrektheit von Informationen. Sie gibt Hinweise dazu, wie gut Daten die Realität oder die tatsächlichen Situationen beschreiben. Die Qualität der erhobenen Daten ist von zentraler Bedeutung dafür, dass eine Wirkungsmessung exakte Resultate liefern kann. In der wissenschaftlichen Methodik der Datenerhebung gibt es zwei Gütekriterien, welche die erhobenen Daten erfüllen sollten:

- **Reliabilität** (Zuverlässigkeit)
Der Begriff Reliabilität bezieht sich auf die Zuverlässigkeit der Datenerhebung. Ein Erhebungsverfahren ist dann zuverlässig, wenn unter den gleichen Bedingungen eine Wiederholung der Erhebung oder Messung zu den gleichen Ergebnissen führt.
- **Validität** (Gültigkeit)
Der Begriff Validität bezieht sich auf die Gültigkeit der Datenerhebung. Eine Datenerhebung ist dann gültig, wenn sie das misst, was man messen wollte. Eine Messung oder Erhebung ist valide, wenn die erhobenen Werte geeignete Kennzahlen für die zu untersuchende Fragestellung liefern.

Daten und Datenquellen prüfen

Daten und Datenquellen sollten bezüglich der Kriterien Reliabilität und Validität überprüft werden. Das ist insbesondere dann nötig, wenn es sich um fremde Datenquellen handelt oder wenn neue Datenquellen verwendet werden. Neue Erhebungs- und Verarbeitungsmethoden sollten ebenfalls getestet werden. Es kann sich also lohnen, die Datenerhebung vorgängig in einem Testlauf zu prüfen. Zusätzlich sollte geprüft werden, ob die Erhebungen die gewünschten Informationen liefern.

Fehlerquellen identifizieren und minimieren

Bei der Erhebung und der Erfassung von qualitativen und quantitativen Daten gibt es verschiedene Fehlerquellen, die vermieden werden sollten. Wird eine eigene Datenerhebung mittels qualitativer oder quantitativer Erhebungsinstrumente durchgeführt, dann sollten die Instrumente (Fragebogen, Gesprächsleitfaden usw.) nach Möglichkeit mit einem Pretest geprüft werden. Dazu wird das Erhebungsinstrument an Testpersonen oder Testfällen erprobt. Diese sollten, wenn möglich, der Zielgruppe der Befragung respektive den zu analysierenden Fällen ähnlich sein. Zudem sollte der Pretest unter möglichst ähnlichen Bedingungen durchgeführt werden wie die geplante Erhebung. In Abhängigkeit der Ergebnisse des Pretests ist eventuell eine Überarbeitung oder Anpassung der Erhebungsinstrumente notwendig. Es ist daher wichtig, dass die dazu benötigte Zeit bei der Planung berücksichtigt wird. Häufige Fehlerquellen liegen bei der Auswahl der Untersuchungseinheit.



Wirkungsmessung für NPO

Zewo-Leitfaden für Dienstleistungen und Projekte im Inland



Auftrag formulieren

Wenn die Wirkung durch ein internes Team gemessen und beurteilt wird, besonders aber wenn damit ein externes oder gemischtes Team beauftragt wird, sollte dieser Auftrag und die Planung der Wirkungsmessung schriftlich festgehalten werden.

Ein Auftrag für eine Wirkungsmessung kann der folgenden Struktur für Evaluationen folgen:

Muster zur Struktur für Evaluationen

Zweck und Nutzen (rationale, purpose)

Weshalb wird die Evaluation durchgeführt und wozu werden die Ergebnisse genutzt?

Zum Beispiel:

- Prozesse oder Verfahren optimieren
- Strategie oder Politik der Organisation weiterentwickeln
- Über die Weiterführung eines Projekts oder Programms entscheiden
- Rechenschaft gegenüber Geldgebern oder der Öffentlichkeit ablegen

Ziel (objectives)

Worüber soll die Evaluation etwas aussagen?

Zum Beispiel:

- Bestätigen, dass eine bestimmte Leistung (Output) erbracht, eine bestimmte Wirkung bei der Zielgruppe (Outcome) erreicht oder ein Beitrag zum übergeordneten Ziel geleistet wurde.
- Beurteilen, ob eine bestimmte Intervention effizient, effektiv und relevant war.
- Darlegen von Feststellungen, Schlussfolgerungen und Empfehlungen zu einer bestimmten Dienstleistung oder zu einem bestimmten Projekt .

Umfang (scope)

Der Umfang der Evaluation soll klar definiert werden, indem die wichtigsten Eckpunkte klar umschrieben werden.

Zum Beispiel:

- Zu behandelnde Themen
- Zu betrachtende Zeitperiode
- Zu betrachtende Aktivitäten
- Bereits verwendete Mittel
- Geografische Ausdehnung
- Zielgruppen

Beteiligte und Betroffene

Wer ist an der Evaluation beteiligt oder von ihr betroffen?

Zum Beispiel:

- Dienstleistungs- oder Projektverantwortliche Mitarbeitende
- Mittlerorganisationen
- Zielgruppe
- Partnerorganisationen
- Regierung

Welche Interessen und Bedürfnisse haben diese Personen(gruppen)? Wie werden diese berücksichtigt?

Berichterstattung

Wie wird über die Ergebnisse berichtet? Sind neben dem klassischen Bericht auch weitere «Produkte» zu liefern?

Zum Beispiel:

- Klassischer Bericht
- Workshop mit den Beteiligten
- Debriefing mit Projektverantwortlichen
- Präsentationen für Linienverantwortliche
- Schriftliche «Lessons learnt»

Budget

- Kosten im Verhältnis zur Komplexität der Fragestellung und zum Wert der erwünschten Information verhältnismässig?

Was sollte schriftlich vereinbart werden?

- Ziel, Zweck, Blickwinkel der Evaluation festhalten
- Evaluationsfragen formulieren
- Termine und Budget vereinbaren
- Einem solchen Auftrag können Qualitätsstandards zugrunde gelegt werden (siehe [SEVAL](#))
- Berichtsformat definieren



Stiftung Zewo
Schweizerische Zertifizierungsstelle für gemeinnützige,
Spenden sammelnde Organisationen

Stiftung Zewo
Lägernstrasse 27
8037 Zürich
Telefon +41 (0)44 366 99 55

Wirkungsmessung für NPO

Zewo-Leitfaden für Dienstleistungen und Projekte im Inland



SEVAL Evaluationsstandards

Die Schweizerische Evaluationsgesellschaft (SEVAL) ist eine multidisziplinäre Organisation, die sich für die Qualität und Verbreitung von Evaluationen einsetzt. Sie bildet ein Forum für den Austausch über Evaluationen zwischen Politik, Verwaltung, Hochschulen, NGOs und der Beratung. Die Evaluationsstandards der SEVAL sollen einen Beitrag zur Professionalisierung der Evaluation in der Schweiz leisten. Sie definieren die Qualitätsanforderungen, die an Evaluationen gestellt werden und richten sich insbesondere an Evaluatorinnen und Evaluatoren sowie Auftraggeberinnen und Auftraggeber. Die Leitlinien sind für Evaluationen aller Art ausgelegt, mit Ausnahme von Personalevaluationen. Sie sind in vier Themenbereiche gegliedert: Die Nützlichkeitsstandards stellen sicher, dass sich eine Evaluation an den Informationsbedürfnissen der vorgesehenen Nutzerinnen und Nutzer ausrichtet. Die Durchführbarkeitsstandards stellen sicher, dass eine Evaluation realistisch, gut durchdacht, diplomatisch und kostenbewusst ausgeführt wird. Die Korrektheitsstandards stellen sicher, dass eine Evaluation rechtlich und ethisch korrekt durchgeführt wird und dem Wohlergehen der Beteiligten und Betroffenen gebührende Aufmerksamkeit widmet. Die Genauigkeitsstandards stellen schliesslich sicher, dass eine Evaluation gültige und verwendbare Informationen hervorbringt und vermittelt.

Download

SEVAL Standards, Evaluationsstandards der Schweizerischen Evaluationsgesellschaft, 2000, vollständiger Text

Link

SEVAL