

Leseprobe aus Gollwitzer • Pohl • Jäger, Evaluation, ISBN 978-3-621-28886-6  
© 2023 Programm PVU Psychologie Verlags Union in der Verlagsgruppe Beltz, Weinheim Basel  
<http://www.beltz.de/de/nc/verlagsgruppe-beltz/gesamtprogramm.html?isbn=978-3-621-28886-6>

# Inhalt

|         |    |
|---------|----|
| Vorwort | 10 |
|---------|----|

## I »Grundlagen« Was ist Evaluation und wie funktioniert sie?

|   |    |
|---|----|
| <b>1 Was ist Evaluation?</b>  | 14 |
| 1.1 Wirksamkeit und Wirkung von Maßnahmen   | 16 |
| 1.2 Weitere Evaluationsgegenstände und -kriterien                                     | 23 |
| 1.3 Historische Strömungen und theoretische Positionen<br>in der Evaluationsforschung | 25 |
| 1.3.1 Meilensteine  | 25 |
| 1.3.2 Theoretische Positionen   | 26 |
| 1.4 Wichtige Begriffe der Evaluationsforschung  | 27 |
| Zusammenfassung   | 30 |
| Übungsaufgaben  | 31 |
| <b>2 Evaluationsprojekte: Strukturen, Abläufe und Qualitätsstandards</b>              | 33 |
| 2.1 Struktur und Ablauf eines Evaluationsprojekts                                     | 33 |
| 2.1.1 Das Modell von Balzer (2005)  | 34 |
| 2.1.2 Das Modell von Posavac und Carey (1980)   | 36 |
| 2.2 Standards und Gütekriterien   | 40 |
| 2.2.1 Allgemeine Evaluationsstandards   | 40 |
| 2.2.2 Methodische Standards zur Sicherung der Validität                               | 43 |
| 2.2.3 Psychometrische Gütekriterien   | 45 |
| Zusammenfassung   | 47 |
| Übungsaufgaben  | 48 |

## II »Fragestellungen« Welche Arten von Evaluationsprojekten gibt es?

|  |    |
|--|----|
| <b>3 Prospektive Evaluation und Maßnahmenplanung</b> | 52 |
| 3.1 Beschreibung und Bewertung des Ist-Zustands      | 54 |
| 3.1.1 Quantitative Methoden zur Zustandsbeschreibung | 55 |
| 3.1.2 Qualitative Methoden zur Zustandsbeschreibung  | 63 |
| 3.1.3 Multimethodale und multimodale Diagnostik      | 66 |
| 3.1.4 Bewertung eines Ist-Zustands                   | 68 |
| 3.2 Bestimmung der Zielgruppe                        | 72 |
| 3.3 Definition des Soll-Zustands                     | 76 |

|          |  |            |
|----------|--|------------|
| 3.4      | Programmauswahl und Konzeptionsanalyse   | 86         |
| 3.4.1    | Programmauswahl  | 86         |
| 3.4.2    | Konzeptionsanalyse   | 89         |
|          | Zusammenfassung  | 91         |
|          | Übungsaufgaben   | 93         |
| <b>4</b> | <b>Prozessevaluation: Programmoptimierung, Outcome-Monitoring und Implementationskontrolle</b> | <b>95</b>  |
| 4.1      | Programmoptimierung  | 96         |
| 4.2      | Outcome-Monitoring   | 101        |
| 4.3      | Implementationskontrolle   | 105        |
|          | Zusammenfassung  | 109        |
|          | Übungsaufgaben   | 110        |
| <b>5</b> | <b>Wirksamkeitsevaluation</b>  | <b>112</b> |
| 5.1      | Variation unterschiedlicher Bedingungen in einer Evaluationsstudie                             | 114        |
| 5.2      | Interne Validität  | 118        |
| 5.2.1    | Gefahren für die interne Validität   | 118        |
| 5.2.2    | Strategien zur Erhöhung der internen Validität   | 125        |
| 5.3      | Persistenz und Transfer  | 134        |
| 5.3.1    | Persistenz   | 134        |
| 5.3.2    | Transfer   | 136        |
| 5.4      | Externe Validität (Generalisierbarkeit)  | 138        |
|          | Zusammenfassung  | 140        |
|          | Übungsaufgaben   | 141        |
| <b>6</b> | <b>Evaluation des Nutzens und der Kosten einer Maßnahme</b>                                    | <b>143</b> |
| 6.1      | Quantifizierung von Kosten und Nutzen  | 145        |
| 6.1.1    | Kostenarten  | 145        |
| 6.1.2    | Wirksamkeit und Nutzen   | 146        |
| 6.2      | Modelle der Effizienzanalyse   | 152        |
| 6.2.1    | Kosten-Nutzen-Analyse  | 153        |
| 6.2.2    | Kosten-Effektivitäts-Analyse   | 154        |
| 6.3      | Bestimmung der Nutzenschwelle  | 155        |
|          | Zusammenfassung  | 157        |
|          | Übungsaufgaben   | 158        |

### III »Methoden« Wie werden Evaluationsstudien geplant und ausgewertet?

|          |  |     |
|----------|--|-----|
| <b>7</b> | <b>Designfragen: Planung von Evaluationsuntersuchungen</b> | 162 |
| 7.1      | Arten von Designs  | 163 |
| 7.2      | Einfaktorielle Untersuchungsdesigns                        | 165 |
| 7.3      | Designs mit Zuweisung auf Gruppen- oder Kohortenebene      | 171 |
| 7.4      | Mehrfaktorielle Untersuchungsdesigns                       | 174 |
|          | Zusammenfassung  | 177 |
|          | Übungsaufgaben   | 179 |
| <br>     |  |     |
| <b>8</b> | <b>Auswertungsverfahren</b>                                | 181 |
| 8.1      | Theorie kausaler Effekte                                   | 182 |
| 8.1.1    | Definition kausaler Effekte                                | 186 |
| 8.1.2    | Kausale Interpretierbarkeit von Effekten in der Anwendung  | 188 |
| 8.2      | Statistische Kontrolle von Selektionseffekten              | 191 |
| 8.2.1    | Generalisierte Kovarianzanalyse (ANCOVA)                   | 193 |
| 8.2.2    | Kovariaten-Matching  | 196 |
| 8.2.3    | Propensity-Score-Methoden                                  | 200 |
| 8.2.4    | Praktische Tipps zur Wahl und Anwendung einer Methode      | 206 |
| 8.2.5    | Regressions-Diskontinuitäts-Analyse                        | 211 |
| 8.3      | Umgang mit Non-Compliance                                  | 213 |
| 8.3.1    | As-Treated- und Per-Protocol-Analysen                      | 213 |
| 8.3.2    | Intention-To-Treat   | 214 |
| 8.3.3    | Instrumenten-Variablen-Ansatz                              | 214 |
| 8.3.4    | Wahl einer Methode   | 216 |
| 8.4      | Umgang mit fehlenden Werten                                | 220 |
| 8.4.1    | Mechanismen fehlender Werte                                | 220 |
| 8.4.2    | Methoden zum Umgang mit fehlenden Werten                   | 222 |
| 8.4.3    | Praktische Tipps   | 225 |
|          | Zusammenfassung  | 226 |
|          | Übungsaufgaben   | 228 |
| <br>     |  |     |
|          | <b>Anhang</b>  |     |
|          | Literatur  | 232 |
|          | Glossar  | 238 |
|          | Sachwortverzeichnis  | 252 |
|          | Abkürzungsverzeichnis                                      | 255 |
|          | Hinweise zum Online-Material                               | 256 |

# 1 Was ist Evaluation?

## ■ Was Sie in diesem Kapitel erwartet

In diesem ersten Kapitel werden wir uns mit der Frage befassen, was Evaluation genau ist. Wir werden sehen, dass Evaluation sehr viele unterschiedliche Tätigkeiten umfassen kann, die sich mit Begriffen wie Überprüfung, Testung, Kontrolle, Bewertung, Abschätzung, Bestimmung, Stärkung, Unterstützung etc. beschreiben lassen. Das Entscheidende dabei ist, dass professionelle Evaluation immer auf wissenschaftlichen Standards basieren (also »evidenzbasiert« sein) sollte; hierin unterscheidet sich die »Evaluationsforschung« von der alltäglichen Evaluation außerhalb eines professionellen Kontexts. Wir werden am Beispiel der Wirksamkeitsevaluation einer Maßnahme veranschaulichen, wieso Evidenzbasierung wichtig ist. Wir geben in diesem Kapitel einen Überblick über Meilensteine der Evaluationsforschung, zentrale theoretische Positionen und wichtige Grundbegriffe.

Am Ende des Kapitels

- ▶ kennen Sie den Unterschied zwischen Evaluation und Evaluationsforschung,
- ▶ haben Sie einen Überblick über Evaluationsfragestellungen und -tätigkeiten,
- ▶ haben Sie wichtige Meilensteine und Epochen der Evaluationsforschung kennengelernt, und
- ▶ kennen Sie die wichtigsten Begriffe der Evaluationsforschung.

## Drei Beispiele

Bevor wir erörtern, was man unter Evaluation versteht, stellen wir zunächst drei Beispiele vor. Diese Beispiele betreffen unterschiedliche Zielgruppen und zugleich unterschiedliche Themenbereiche. An ihnen werden typische Fragestellungen von Evaluationsvorhaben deutlich gemacht.

**Impfkampagne im Zuge der COVID-19-Pandemie.** Nachdem zu Beginn des Jahres 2021 Impfstoffe gegen das Coronavirus (SARS-CoV-2) zur Verfügung standen, hat die Politik unterschiedliche Versuche unternommen, die deutsche Bevölkerung zu einer Impfung zu motivieren. Dazu gehörten Kommunikationskampagnen des Bundesgesundheitsministeriums (im Zuge der Aktion »Zusammen gegen Corona«; Bundesministerium für Gesundheit, 2022), lokale Aktionen in deutschen Städten und Gemeinden (wie etwa der »Impfmarathon« der Stadt München im Dezember 2021), Werbekampagnen von Seiten der Privatwirtschaft, aber auch verhaltensbeeinflussende politische Maßnahmen wie etwa die »2G-Regel« (der zufolge nur Geimpfte oder Genesene Zutritt zu einer Veranstaltung hatten) oder die Rücknahme des Angebots, sich auf Kosten des Staates auf COVID-19 testen zu lassen. Evaluationsrelevante Fragestellungen hierbei sind:

- ▶ Wie wirkt sich die Impfkampagne auf die Impfquote aus?
- ▶ Wie unterscheiden sich die unterschiedlichen Maßnahmen hinsichtlich ihrer Akzeptanz in der Bevölkerung?
- ▶ Wird die »2G-Regel« vor Ort (z. B. in Restaurants, Kinos oder Theatern) überhaupt so implementiert, wie sie vom Gesetzgeber intendiert war?

**Reform des Psychotherapeutengesetzes.** Seit dem 15.11.2019 gibt es ein neues Gesetz über den Beruf des/der Psychotherapeut:in (»Psychotherapeutengesetz«, kurz PsychThG), das regelt, wer unter welchen Bedingungen und mit welcher Qualifikation in Deutschland psychotherapeutisch tätig werden darf. Es ist seit September 2020 in Kraft. Ein entscheidender Unterschied zum »alten« PsychThG, das noch aus dem Jahr 1998 stammte, ist, dass die Approbation, also die staatliche Zulassung, die zur Ausübung des Berufs des/der Psychotherapeut:in erforderlich ist, nicht mehr im Anschluss an eine postgraduale Weiterbildung, sondern jetzt direkt im Anschluss an ein fünfjähriges Studium der Psychologie mit psychotherapeutischer Ausrichtung erworben wird. Evaluationsrelevante Fragestellungen hierbei sind:

- ▶ Wird die Gesetzesreform dazu führen, Engpässe bei der psychotherapeutischen Versorgung in Deutschland zu reduzieren?
- ▶ Unterscheiden sich Psychotherapeut:innen, die ihre Approbation nach dem »alten« bzw. nach dem »neuen« System erworben haben, hinsichtlich ihrer Fertigkeiten, ihrer Kenntnisse oder der Qualität ihrer Berufsausübung?
- ▶ Geht die Gesetzesreform mit einer Kostenersparnis für das Gesundheitssystem (oder sogar mit Mehrkosten) einher?
- ▶ Geht die Gesetzesreform mit einer geringeren Schuldenlast für die Approbationskandidat:innen einher, weil es nun keine prekären Beschäftigungsverhältnisse mehr in der Ausbildungsphase gibt?

**Online-Therapie.** Seit einiger Zeit wird vermehrt das Potenzial webbasierter oder -unterstützter Psychotherapie in den Blick genommen. So können nicht nur Menschen in Krisenregionen (bspw. Soldat:innen in Kriegsgebieten) oder Menschen in unterversorgten Gebieten (bspw. in infrastrukturschwachen ländlichen Gebieten) erreicht werden, sondern auch Wartezeiten auf einen Therapieplatz überbrückt werden. Aber auch in der Regelversorgung spielt Online-Therapie nicht nur wegen der Covid-19-Pandemie eine zunehmend größere Rolle. Fragestellungen sind hierbei:

- ▶ Wie werden solche webbasierten Angebote genutzt? Wie hoch ist die Akzeptanz durch die Nutzer:innen?
- ▶ Kann eine webbasierte Psychotherapie ähnlich gute Therapieerfolge erbringen wie eine reguläre Psychotherapie vor Ort?
- ▶ Wie ist das Kosten-Nutzen-Verhältnis webbasierter Psychotherapieangebote im Vergleich zu konventionellen Angeboten zu bewerten?

### **Evaluation und Evaluationsforschung**

Fragestellungen wie die oben genannten sind Gegenstand der Evaluation und der Evaluationsforschung. Der Unterschied zwischen beiden Begriffen wird in den folgenden Definitionen dargestellt.

## ■ Definition

**Evaluation** bedeutet wörtlich übersetzt *Bewertung* und bezieht sich auf den Bewertungsprozess und das Ergebnis der Beurteilung des Wertes eines Produkts, eines Prozesses oder eines Programms. Man kann den Begriff Evaluation eher eng oder eher weit definieren. In der weiten Definition umfasst Evaluation jede Art von Wertbeurteilung – auch dann, wenn diese subjektiv, unsystematisch und nicht nach wissenschaftlichen Standards vorgenommen wird (z. B. ist dann auch die Aussage »Das Essen schmeckt mir gut« eine Evaluation). In der engen Definition ist Evaluation eine Wertbeurteilung, die auf der Basis wissenschaftlicher Prozesse und Befunde vorgenommen wird (Suchman, 1967). In der engen Definition entspricht Evaluation dem Begriff Evaluationsforschung.

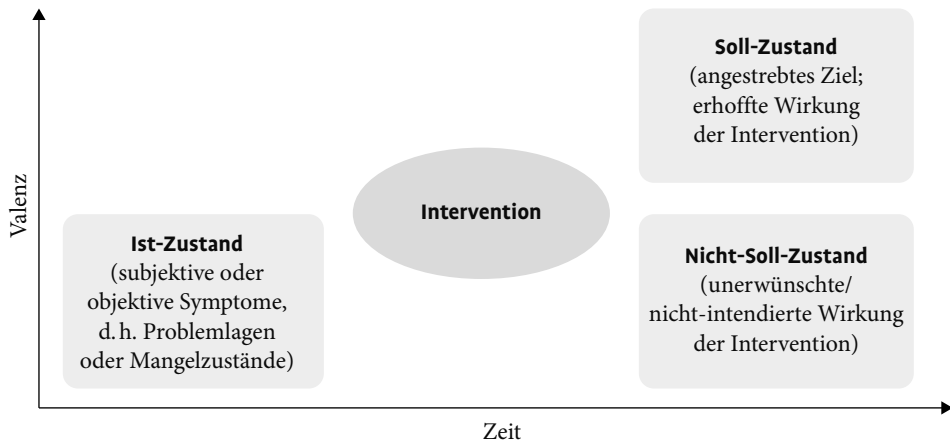
**Evaluationsforschung** wird als ein Prozess verstanden, »... bei dem nach zuvor festgelegten Zielen und explizit auf den Sachverhalt bezogenen und begründeten Kriterien ein Evaluationsgegenstand bewertet wird. Dies geschieht unter Zuhilfenahme sozialwissenschaftlicher Methoden durch Personen, welche hierfür besonders qualifiziert sind. Das Produkt eines Evaluationsprozesses besteht in der Rückmeldung verwertbarer Ergebnisse in Form von empirisch fundierten Beschreibungen, begründeten Interpretationen und Empfehlungen an möglichst viele Beteiligte und Betroffene, um den Evaluationsgegenstand zu optimieren und zukünftiges Handeln zu unterstützen« (Balzer, 2005, S. 16).

Die Evaluationsforschung verlangt also eine besondere Kompetenz der Evaluator:innen sowie die Anwendung sozialwissenschaftlicher Methoden bzw. expliziter Regeln bei der Bewertung von Evaluationsgegenständen (Suchman, 1967). Wir verwenden in diesem Lehrbuch den Begriff Evaluation in der engen Definition und damit synonym zum Begriff Evaluationsforschung.

## 1.1 Wirksamkeit und Wirkung von Maßnahmen

Bei Evaluationsprojekten geht es in vielen Fällen um die Wirksamkeit bzw. die Wirkung einer Maßnahme, also z. B. die erhofften (gesellschaftlichen und individuellen) Auswirkungen einer Impfkampagne, einer Gesetzesreform oder eines Online-Therapieangebots. Solche Maßnahmen oder Interventionen (wörtlich: »Eingriffe«) gehen meist mit der Erwartung einer Veränderung einher: Ein als »problematisch« oder verbesserungswürdig bewerteter »vorher«-Zustand (Ist-Zustand) soll mit Hilfe einer Intervention in einen quantitativ oder qualitativ besseren »nachher«-Zustand (Soll-Zustand) überführt werden. Das Prinzip ist in Abbildung 1.1 veranschaulicht. Ob diese Veränderung tatsächlich eingetreten ist und wie stark sie ist, muss empirisch ermittelt werden. Dies gehört zum Aufgabenbereich der Wirksamkeitsevaluation.

Man kann dieses Prinzip auf die Ebene gesellschaftlicher Veränderungen anwenden, aber auch auf die Ebene (intra-)individueller Veränderungen: Nehmen wir an, Sie haben



**Abbildung 1.1** Die Wirksamkeitsevaluation überprüft die Annahme, dass mit Hilfe einer Intervention aus einem als problematisch erlebten Ist-Zustand (niedrige Valenz bedeutet eine Bewertung des Zustands als schlecht und verbesserungsbedürftig) über die Zeit hinweg ein positiv bewerteter Soll-Zustand wird, wobei eine Intervention auch unerwünschte oder nicht-intendierte Wirkungen haben kann (Nicht-Soll-Zustand)

Kopfschmerzen (problematischer Ist-Zustand). Dann wäre eine denkbare Intervention, ein Analgetikum (ein Schmerzmittel) einzunehmen. Ihre Hoffnung ist, dass mit der Zeit die Kopfschmerzen verschwinden und sich Ihr Wohlergehen bessert (Soll-Zustand). Allerdings könnte sich diese Hoffnung als falsch erweisen: Es wäre möglich, dass das Schmerzmittel nicht wirkt (also der Soll-Zustand nicht erreicht wird) oder sogar Nebenwirkungen wie etwa Magenschmerzen auftreten (Nicht-Soll-Zustand). Dass eine Intervention also die gewünschten Wirkungen hat, ist nicht garantiert. Gerade deshalb ist es wichtig, die Wirksamkeit einer Intervention umfassend zu evaluieren.

### Wirksamkeit und Wirkung

Hager und Hasselhorn (2000) schlagen eine konzeptuelle Trennung der Begriffe Wirksamkeit und Wirkung vor.

#### Definition

- ▶ Mit **Wirksamkeit** sind die durch eine Maßnahme verursachten beobachteten Konsequenzen (»Effekte«) gemeint.
- ▶ Mit **Wirkung** sind die modelltheoretisch angenommenen Mechanismen (»Prozesse«) gemeint, die zu den beobachtbaren Effekten geführt haben.

Ein Nachweis der Wirksamkeit beschränkt sich also auf die Feststellung, *dass* eine Maßnahme etwas bewirkt hat. Eine Analyse der Wirkung einer Maßnahme fragt danach, *wie* eine Maßnahme gewirkt hat, d. h. welche Prozesse zur beobachteten Wirksamkeit einer Maßnahme geführt haben.



Man mag nun einwenden, dass eine Evaluation der Wirksamkeit im Vordergrund stehen müsse und eine Analyse des genauen Wirkmechanismus für die meisten Anwendungsfälle doch eigentlich irrelevant und eher eine Aufgabe der Grundlagenforschung sei. Dies ist aber aus drei Gründen zu kurz gedacht:

- (1) Erstens birgt nur eine genaue Erforschung des Wirkmechanismus die Chance, die Wirksamkeit einer Maßnahme zu optimieren, d. h. ihre Hauptwirkung zu maximieren und unerwünschte Neben- und Folgewirkungen einzudämmen.
- (2) Zweitens kann man nur auf der Basis einer Wirkungsanalyse herausfinden, welche Bestandteile und Durchführungsbedingungen einer Maßnahme wirklich nötig sind und auf welche man hingehen ohne Wirksamkeitseinbußen verzichten kann.
- (3) Drittens lassen sich durch eine genaue Kenntnis des Wirkmechanismus Kosten reduzieren. Nur wenn man den »Hauptwirkstoff« (das »Verum«) einer Maßnahme kennt, kann man zielgerichtet nach effizienteren Wegen suchen, diesen herzustellen bzw. zu verabreichen.

Die Analyse von Wirkmechanismen ist nicht nur etwas für die Grundlagenforschung, sondern hat eine unmittelbare praktische Bedeutsamkeit, etwa für die Optimierung von Maßnahmen oder die Verringerung von Kosten.

### ■ Beispiel • Wie wirkt eigentlich Aspirin?

Seit mehr als 120 Jahren wird Acetylsalicylsäure (ASS), eine Verbindung aus Essig- und Salicylsäure, als Mittel gegen Fieber und Schmerzen eingesetzt, seit 1897 wird sie als »Aspirin« künstlich synthetisiert. Dass Aspirin wirksam ist, wusste man schon lange, aber der genaue Wirkmechanismus blieb bis Anfang der 1970er-Jahre ungeklärt. Erst John Robert Vane (1927–2004), ein britischer Biochemiker, fand heraus, dass ASS über eine Hemmung entsprechender Enzyme die Prostaglandinproduktion reduziert. Prostaglandine sind Gewebshormone, die unter anderem die Schmerzempfindlichkeit und die Körpertemperatur erhöhen. Ihre Hemmung wirkt sich also fiebersenkend und schmerzreduzierend aus. Aber ASS hat auch Nebenwirkungen: Prostaglandine fördern auch die Produktion schützenden Schleims (etwa im Magen). Ihre Hemmung führt dazu, dass die Magenwand in Kontakt mit den Magensäuren kommt; Entzündungen und Blutungen sind die Folge.

Seitdem der genaue Wirkmechanismus von ASS bekannt ist, sucht man nach Möglichkeiten, die Nebenwirkungen zu minimieren, ohne die Wirkung des Hauptwirkstoffs zu verringern. Erfolgversprechend scheint die Suche nach Präparaten, die speziell jene Enzyme hemmen, die für Entzündungen zuständig sind, und gleichzeitig jene Enzyme unbeeinflusst lassen, die für den Schutz der Magenschleimhaut zuständig sind (sog. selektive COX-2-Hemmer). Zwar gibt es bereits entsprechende Präparate, allerdings haben diese offenbar wieder spezifische unerwünschte Nebenwirkungen. Die Suche geht also weiter – aber erst durch die Entdeckung des Wirkmechanismus von ASS kann sie systematisch betrieben werden.

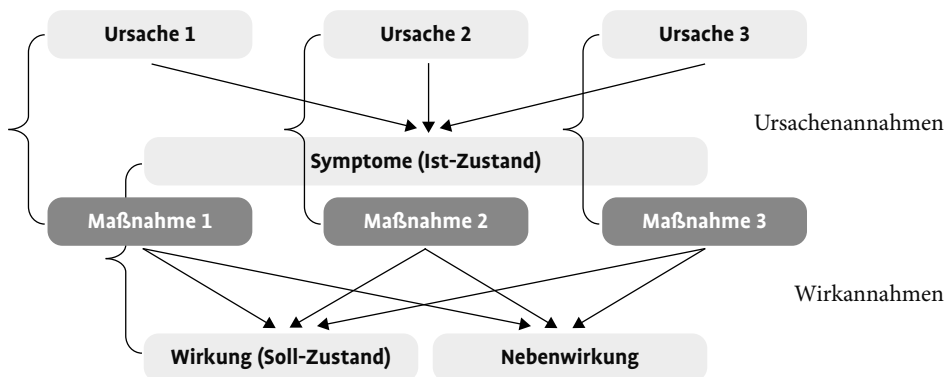
## Ursachen- und Wirkannahmen

Hinter jeder Intervention verbergen sich Annahmen. Wenn Sie Kopfschmerzen haben, treffen Sie – explizit oder implizit – Annahmen über die vermeintlichen Gründe für den problematischen Ist-Zustand: zu wenig frische Luft, eine ungünstige Körperhaltung beim Sitzen, alpines Föhnwetter oder der Rotwein am gestrigen Abend. Aber auch in Bezug auf die Intervention treffen Sie Annahmen: Von der Einnahme eines Schmerzmittels erwarten Sie, dass dies Ihre Kopfschmerzen lindert. Wenn Ihre Kopfschmerzen zwei Stunden nach der Einnahme noch nicht verschwunden sind, dann nehmen Sie an, dass das Medikament unwirksam war usw. Tatsächlich basiert jede Intervention – egal ob es sich um eine »private« individuelle Maßnahme handelt oder eine politische Reform, welche die ganze Gesellschaft betrifft – auf solchen Ursachen- und Wirkannahmen. Aufgabe der Evaluationsforschung ist es, diese Annahmen zu überprüfen.

- ▶ **Ursachenannahmen** beziehen sich darauf, welche Faktoren als ursächlich für einen problematischen Ist-Zustand in Frage kommen.
- ▶ **Wirkannahmen** beziehen sich darauf, welche Maßnahmen vermutlich wirken.

Dabei sollte zum einen zwischen einer (vermuteten) Ursache und einer entsprechenden Maßnahme eine Passung (eine »funktionale Äquivalenz«) bestehen: Wenn Sauerstoffmangel der Grund für Ihre Kopfschmerzen wäre, dann sollten Sie einen Spaziergang machen. Wenn eine übermäßige Anspannung der Nackenmuskulatur die Ursache wäre, sollten Sie einen Massagetermin buchen etc. Zum anderen sollte funktionale Äquivalenz zwischen Ist- und Soll-Zustand bestehen: Um herauszufinden, ob eine Intervention wirklich wirksam war, sollte der Soll-Zustand sich auf die gleichen Merkmale beziehen wie der Ist-Zustand (also bspw. die Stärke des Schmerzes).

Abbildung 1.2 veranschaulicht die Ursachen- und Wirkannahmen (dargestellt in Pfeilen) sowie die dazugehörigen funktionalen Äquivalenzen (dargestellt in geschwungenen Klammern) grafisch.



**Abbildung 1.2** Grafische Darstellung von **Ursachenannahmen** (bzgl. der kausalen Wirkung einer oder mehrerer Ursachen auf den Ist-Zustand), **Wirkannahmen** (bzgl. der kausalen Wirkung einer oder mehrerer Maßnahmen auf den Soll-Zustand, einschl. Nebenwirkungen), **funktionale Äquivalenzen** zwischen Ursache und Maßnahme sowie zwischen Ist- und Soll-Zustand

## Wirkmodelle

Eine Wirkannahme – also die Annahme, dass durch eine Maßnahme eine Veränderung vom Ist-Zustand hin zum Soll-Zustand stattfindet – zieht immer drei Fragen nach sich:

- ▶ **Wie** soll die Maßnahme wirken? (Frage nach den Prozessen der Wirkung)
- ▶ **Wieso** soll die Maßnahme wirken? (Frage nach der Begründung für die Wirkannahme)
- ▶ **Wann** bzw. unter welchen Umständen sollte die Maßnahme wirken? (Frage nach den Randbedingungen, von denen die Wirkung abhängen sollte)

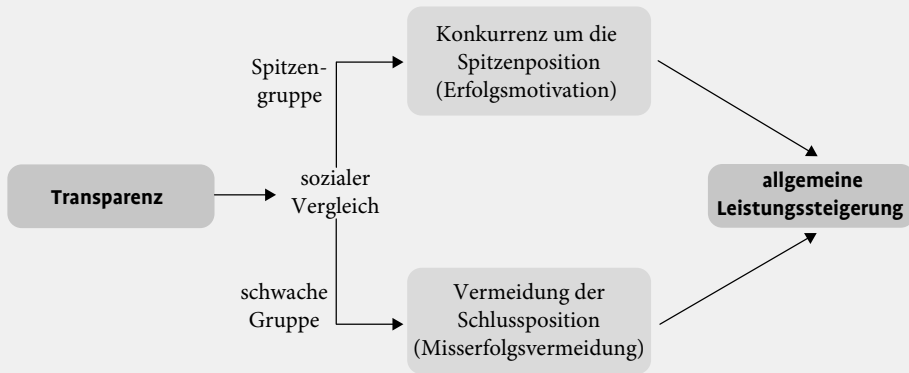
Wirkmodelle versuchen, all diese Fragen umfassend zu beantworten. Ob die Antworten dann wirklich stimmen oder nicht, soll im Rahmen einer Evaluation herausgefunden werden. Das bedeutet, um eine Maßnahme wirklich umfassend zu evaluieren, ist es erforderlich, vorab das Wirkmodell zu spezifizieren. Aus dem Wirkmodell lassen sich dann empirische Hypothesen ableiten, welche im Rahmen von Wirksamkeitsevaluationen überprüft werden können. Dies wird im Folgenden an zwei Beispielen veranschaulicht.

### ■ Beispiel • Beispiel 1: Sanfte Organisationsentwicklung

Ein Firmenchef möchte die Leistungsorientierung seiner Mitarbeitenden erhöhen. Insbesondere möchte er eine Stärkung der intrinsischen, also der von äußeren Anreizen unabhängigen, Leistungsmotivation erreichen. Um das Konfliktpotenzial, das üblicherweise mit Veränderungen in Institutionen einhergeht, möglichst gering zu halten, wählt er eine sanfte Form der organisationalen Strukturveränderung: Er kündigt ein Transparenzprogramm an. Ab sofort sollen die Leistungen aller Mitarbeitenden offengelegt und allen anderen zugänglich gemacht werden.

**Prozesse.** Mit dieser Maßnahme verknüpft der Firmenchef ein spezifisches Wirkmodell: Die Transparenz (alle sehen die Leistungen aller anderen) sorgt für eine Dynamik sozialer Vergleichsprozesse, bei der alle Mitarbeitenden motiviert sind, gut abzuschneiden. Die Personen im oberen Leistungsbereich (Spitzengruppe) erleben Stolz und sind motiviert weiterhin die Spitzenpositionen zu besetzen bzw. um die Topposition zu konkurrieren (Erfolgsmotivation). Die Personen im unteren Leistungsbereich (schwache Gruppe) fürchten um ihr Ansehen und wollen nicht an letzter Stelle stehen (Misserfolgsvermeidung).

Abbildung 1.3 veranschaulicht dieses Wirkmodell grafisch: Die Intervention löst einen sozialen Vergleich aus, der – in Abhängigkeit von der jeweiligen Zugehörigkeit zur Spitzen- oder schwachen Gruppe – unterschiedliche motivationale Prozesse (Erfolgsmotivation bzw. Misserfolgsvermeidung) evoziert und dadurch zu einer allgemeinen Leistungssteigerung führt.



**Abbildung 1.3** Wirkmodell der Leistungssteigerung in Form eines Transparenzprogramms. Transparenz führt zu einem sozialen Vergleichsprozess, der alle Mitarbeitenden motiviert, gut abzuschneiden. Die Personen im oberen Leistungsbereich (Spitzen-Gruppe) konkurrieren um die Topposition (Erfolgsmotivation). Die Personen im unteren Leistungsbereich (schwache Gruppe) wollen verhindern, an letzter Stelle zu stehen (Misserfolgsvermeidung). Dies soll zu einer allgemeinen Leistungssteigerung führen

**Begründung.** Die postulierten Wirkprozesse basieren auf Hypothesen, die ihrerseits theoretisch begründet werden müssen:

- ▶ Hypothese 1: Eine Offenlegung von Leistungsdaten evoziert soziale Vergleichsprozesse.
- ▶ Hypothese 2: Personen, die sich im oberen Leistungsspektrum befinden, vergleichen sich mit anderen Personen innerhalb dieser Teilgruppe.
- ▶ Hypothese 3: Personen, die sich im unteren Leistungsspektrum befinden, vergleichen sich mit anderen Personen innerhalb dieser Teilgruppe.
- ▶ Hypothese 4: Der Vergleich der Spitzenpersonen mit anderen Spitzenpersonen löst die Motivation aus, noch besser als die anderen zu sein; dies führt zu einer Leistungssteigerung unter allen Spitzenpersonen.
- ▶ Hypothese 5: Der Vergleich der Schwachen mit anderen Schwachen löst die Motivation aus, nicht schlechter als die anderen zu sein; dies führt zu einer Leistungssteigerung unter allen Schwachen.

Mit solchen Hypothesen sind wir in der (psychologischen) Grundlagenforschung angelangt. Die Begriffe soziale Vergleichsprozesse, Erfolgsmotivation, Misserfolgsvermeidung etc. sind Konzepte aus der Sozial- und Motivationspsychologie. Entsprechend müssten sich in diesen Disziplinen Hinweise darauf finden lassen, ob die angenommenen Wirkmechanismen theoretisch plausibel und empirisch gestützt sind. Dies scheint in der Tat der Fall zu sein: Die Theorie sozialer Vergleichsprozesse von Festinger (1954) besagt explizit, dass soziale Vergleichsprozesse spontan entstehen (Hypothese 1) und dass sich Personen eher mit jenen vergleichen, die ihnen ähnlich sind (Hypothesen 2 & 3). Die Annahmen bezüglich Erfolgsmotivation und Misserfolgsvermeidung können hingegen mit Motivations- und Selbstregulations-theorien begründet werden (z. B. Carver & Scheier, 1998; Lockwood, 2002).

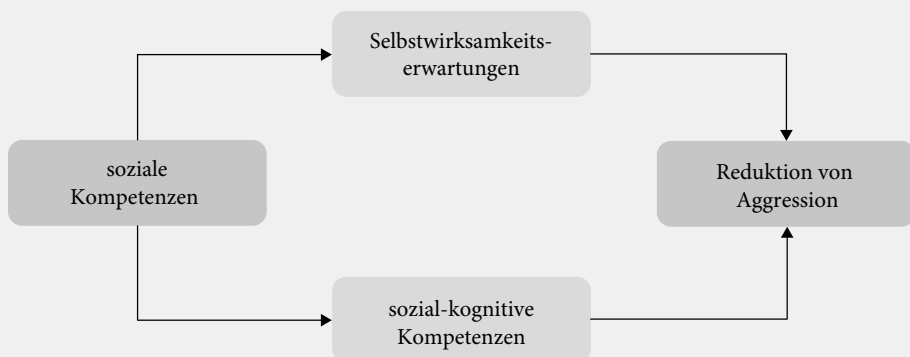
**Randbedingungen.** Die Sozial- und Motivationspsychologie definiert gleichzeitig Randbedingungen, unter denen der erhoffte Effekt des Transparenzprogramms mehr oder weniger wahrscheinlich wird. So hat sich etwa gezeigt, dass Personen weniger erfolgsmotiviert sind, wenn sie sich bedroht fühlen (Taylor & Lobel, 1989) – beispielsweise durch die Befürchtung, schlechte Leistungen gingen ab sofort mit Lohnkürzungen einher. Zudem spielen stabile interindividuelle Unterschiede im Regulationsfokus eine Rolle: Möglicherweise hat man es bei Personen im oberen Leistungsspektrum eher mit misserfolgsmotivierten Personen zu tun (und umgekehrt). Solche Persönlichkeitsunterschiede können Interventionseffekte also moderieren (Lockwood et al., 2002).

### Beispiel • Beispiel 2: Soziale Kompetenzen als Schutzfaktoren aggressiver Tendenzen

Pädagogisch-psychologische Trainings zur Steigerung und Stärkung sozialer Kompetenzen gehören zu den wichtigsten Maßnahmen der Aggressionsprävention (Gollwitzer et al., 2007; Karing & Beelmann, 2016).

**Prozesse.** Mit dem Aufbau sozialer Kompetenzen (z. B. differenzierte soziale Wahrnehmung, Kommunikationsfertigkeiten, Antizipation von Handlungskonsequenzen) werden zwei Wirkungen verknüpft, die sich in folgende Hypothesen überführen lassen (Petermann et al., 2012; s. a. Abb. 1.4):

- ▶ Hypothese 1: Wer über soziale Kompetenzen verfügt, fühlt sich im sozialen Umgang sicherer und vertraut stärker auf die intendierten Wirkungen seines Handelns (Selbstwirksamkeitshypothese).
- ▶ Hypothese 2: Wer über eine differenzierte soziale Wahrnehmung sowie die Fähigkeit zur Antizipation eigener Handlungskonsequenzen verfügt, wird in einer Konfliktsituation weniger aggressiv handeln (Sozial-kognitive Kompetenzhypothese).



**Abbildung 1.4** Wirkmodell zur Aggressionsreduktion: Soziale Kompetenzen führen zu erhöhten Selbstwirksamkeitserwartungen und zu sozial-kognitiven Kompetenzen, die beide wiederum zur Reduktion von Aggression führen

**Begründung.** Die Selbstwirksamkeitshypothese basiert auf der sozial-kognitiven Lerntheorie von Bandura (1979). Die Sozial-kognitive Kompetenzhypothese basiert auf einer Umkehrung der Vorhersagen des »Sozialen Informationsverarbeitungsmodells der Aggression« (Crick & Dodge, 1994). Eine der Annahmen dieses Modells ist, dass aggressive Kinder und Jugendliche über weniger Reaktionsoptionen in Konfliktsituationen verfügen als nicht-aggressive Kinder und Jugendliche. Die Umkehrung besteht nun in der Hypothese, dass aggressives Verhalten weniger wahrscheinlich auftritt, wenn es gelingt, die Verfügbarkeit eines breiteren Reaktionsinventars zu schulen.

**Randbedingungen.** Die Wahrscheinlichkeit, dass eine solche Verfügbarkeitssteigerung alternativer, nicht-aggressiver Handlungstendenzen tatsächlich aggressionsreduzierend wirkt, hängt von einer Reihe von Randbedingungen ab. Die Fähigkeit, in der heißen Phase eines Konflikts zunächst einmal zu überlegen, welche Reaktionsmöglichkeiten zur Verfügung stehen, wie diese zu bewerten sind und inwiefern sie zu positiven oder negativen Konsequenzen führen könnten, erfordert kognitive Ressourcen sowie zusätzlich die Fähigkeit zur Impulskontrolle. Verfügt ein Kind nicht über diese Eigenschaften, wird die Schulung sozialer Kompetenzen wahrscheinlich nicht zum gewünschten Erfolg führen. Kognitive Ressourcen sowie die Fähigkeit zur Impulskontrolle sind demnach notwendige individuelle Voraussetzungen für einen Zusammenhang zwischen dem zu schulenden Merkmal (Verfügbarkeitssteigerung alternativer, nicht-aggressiver Handlungstendenzen) und seiner vermuteten Wirkung (Reduktion aggressiver Verhaltenstendenzen).

Wirksamkeitsevaluation ist also weitaus mehr als der Nachweis, dass infolge einer Intervention eine Veränderung stattgefunden hat. Wirksamkeitsevaluation kann und sollte vielmehr verstanden werden als die systematische und kritische Überprüfung von Ursachen- und Wirkannahmen bzw. Hypothesen, die aus sogenannten Wirkmodellen abgeleitet werden können.

## 1.2 Weitere Evaluationsgegenstände und -kriterien

Evaluation ist mehr als nur Wirksamkeitsevaluation. Zum einen bezieht sich Evaluation nicht unbedingt immer nur auf Interventionen (wie eine Therapie, ein Medikament, ein Gesetz, eine Reform, s. a. die Beispiele zu Beginn dieses Kapitels), sondern oft auch auf ganz andere Gegenstände, z. B. einen Zustand (Ist- oder Soll-Zustand), eine Veränderung (ohne jegliche Intervention dazwischen), die Eigenschaften eines Menschen, das Klima einer Organisation, die ästhetischen Merkmale eines Kunstwerks etc.

Jeder dieser Evaluationsgegenstände kann im Hinblick auf unterschiedliche Eigenschaften (»Evaluationskriterien«) bewertet werden. Meist müssen diese Eigenschaften zunächst spezifiziert werden; selten sind sie von vornherein explizit. Wenn Sie beispielsweise beauftragt werden, das Teamklima einer Organisation zu evaluieren, dann müssen Sie zunächst definieren, was unter dem Begriff »Teamklima« eigentlich genau verstanden

wird. In Abschnitt 3.1 werden wir diesen Aspekt detaillierter aufgreifen und Möglichkeiten diskutieren, Evaluationskriterien gemeinsam mit mehreren Beteiligten zu spezifizieren und in empirisch messbare Größen zu überführen (d. h. zu »operationalisieren«).

Aber auch wenn der Evaluationsgegenstand eine Intervention ist, kann es mehrere Evaluationskriterien – und damit auch mehrere mögliche Evaluationsfragestellungen – geben. Tabelle 1.1 macht dies deutlich. Auf der rechten Seite der Tabelle finden sich Fragestellungen, die an das Evaluationsteam von Seiten der auftraggebenden Partei (einer Firma, einer staatlichen Institution, einer Schule etc.) herangetragen werden könnten. Auf der linken Seite ist das entsprechende Evaluationskriterium genannt. In diesem Buch werden wir auf viele dieser Kriterien genauer eingehen, daher sind in der rechten Spalte auch Verweise auf die jeweiligen Buchkapitel eingefügt.

**Tabelle 1.1** Beispiele für Evaluationskriterien und dazugehörigen Fragestellungen

| <b>Evaluationskriterium</b>                       | <b>Beispiele für spezifische Fragestellungen (und Verweise auf entsprechende Kapitel in diesem Lehrbuch)</b>   |
|---|--|
| <b>Interventionsbedarf</b>                        | Wird die Gewaltbelastung an einer Schule von allen Beteiligten als so schlimm erlebt, dass gegengesteuert werden muss? (Kap. 3)  |
| <b>Qualität des Interventionskonzepts</b>         | Ist eine psychotherapeutische Intervention in Bezug auf die ihr zugrundeliegenden theoretischen Annahmen plausibel, im Einklang mit grundlegenden klinisch-psychologischen Theorien und entsprechenden empirischen Befunden? (Abschn. 3.4.2)                                 |
| <b>Qualität der Einführung und Implementation</b> | Sind die Rahmenbedingungen für eine erfolgreiche Umstellung universitärer Lehre auf webbasierte Formate (Online-Lehre) gegeben? Wird die webbasierte Lehre wie vorgesehen umgesetzt? (Abschn. 4.3)   |
| <b>Akzeptanz</b>                                  | Wird das Angebot sozialtherapeutischer Maßnahmen von Straftäter:innen angenommen? (Abschn. 4.3)  |
| <b>Wirksamkeit (Effektivität)</b>                 | Welche Maßnahme trägt stärker zur Senkung der Rückfallquote bei Straftäter:innen bei: soziale Kompetenztrainings oder härtere Strafen? Gibt es unerwartete Neben- oder Folgewirkungen aufgrund dieser Maßnahmen? (Kap. 5)  |
| <b>Nachhaltigkeit</b>                             | Hält die Wirksamkeit einer Psychotherapie auch nach der letzten gemeinsamen Sitzung mit dem / der Psychotherapeut:in an? (Abschn. 5.3.1)   |
| <b>Transfererfolg</b>                             | Können Studierende ihre in der Vorlesung erworbenen Kenntnisse auch außerhalb des Universitätskontextes abrufen (Situationstransfer)? Können Studierende ihr Wissen über Prinzipien der Kausalität auch auf Alltagsprobleme anwenden (Anforderungstransfer)? (Abschn. 5.3.2) |
| <b>Kosten-Nutzen-Bilanz (Effizienz)</b>           | Wie hoch ist der betriebs- oder volkswirtschaftliche Nutzen sozialtherapeutischer Maßnahmen im Strafvollzug? Übersteigt der Nutzen der Maßnahme dessen Kosten? (Kap. 6)  |