

### 3 Theoretische Grundlagen

Dieser Abschnitt gibt einen Einblick in den aktuellen Stand der Forschung und beschreibt die in dieser Arbeit genutzten Begriffe und Theorienansätze. Er dient auch dazu, eine Abgrenzung von den Themen zu geben, die in dieser Arbeit nicht behandelt werden. Es wird mit einer Begriffsdefinition begonnen, bevor auf die Theorienansätze inkl. der Wirkungsdebatte und -forschung eingegangen wird. Anschließend wird der dieser Arbeit zugrundeliegende Forschungsansatz dargestellt.

#### 3.1 Begriffsdefinitionen und Abgrenzung

Kennzeichnend für die in dieser Arbeit verwendeten Begriffe ist, dass in der Literatur verschiedene Definitionen zu finden sind. Einige Begriffe werden synonym, andere konträr verwendet. Bei der Übersetzung der englischen Begriffe ins Deutsche sind ebenfalls keine eindeutigen Abgrenzungen definierbar. Daher werden in diesem Abschnitt Eingrenzungen festgelegt, die im weiteren Verlauf des Textes genutzt werden.

##### Wirkung/Wirksamkeit

In der Literatur wird der Begriff Wirkung häufig mit der im Duden dargelegten Definition „durch eine verursachende Kraft bewirkte Veränderung“ (Cornelsen Verlag GmbH 2023) beschrieben. Dabei entsteht der Fokus auf einem Kausalitätszusammenhang zwischen Ursache und Ergebnis. Dieses hat Konsequenzen für die Wirkungsdebatte und die Umsetzung einer Analyse (Ottmann und König 2018, S. 6). Auf diesen Zusammen-

hang wird in Kapitel 3.2.1 Wirkungsdebatte genauer eingegangen. Häufig wird der Begriff Wirkung von anderen Autor\*innen synonym mit dem Wort Wirksamkeit gesetzt. Manche Autor\*innen setzen die Begriffe zusätzlich mit Effektivität gleich.

Die DAC (Arbeitsgruppe Evaluierung der Entwicklungszusammenarbeit des OECD-Entwicklungsausschusses) beschreibt Wirksamkeit/Effektivität als das „Ausmaß, in dem die Ziele einer Entwicklungsmaßnahme unter Berücksichtigung ihrer relativen Bedeutung erreicht worden sind oder voraussichtlich erreicht werden.“ (2009, S. 27) Das Wort Wirkung wird von der DAC häufig verwendet, aber nicht näher erläutert.

Albrecht et al. (2013, S. 8) bezeichnen Wirkungen als „Veränderungen bei den Leistungsadressaten, deren Lebensumfeld und der Gesamtgesellschaft, die infolge der Leistungen erzielt werden bzw. auf diese zurückgeführt werden können“. Den Begriff Wirksamkeit verwenden sie selten und nicht synonym mit Wirkung.

Ottmann und König (2023, S. 18) differenzieren eindeutig zwischen Wirkung, welche sie auf die „individuelle Perspektive auf Einzelfall“ beziehen und Wirksamkeit, die sie als institutionelle Perspektive verorten.

Auch das Bundesamt für Gesundheit der Schweiz differenziert zwischen Wirkung/Effekt als „ursächlich auf eine Intervention rückführbare Differenz eines Zustands im Vergleich zu einem unbeeinflussten Zustand“ und Wirksamkeit/Effektivität als das „Ausmaß der durch das Programm, das Projekt oder einen anderen Gegenstand intendierten Wirkungen, die auch eintreten.“ (Widmer und Brunold, S. 5)

Gollwitzer et al. zitieren Hager und Hasselhorn (2000), welche

„eine konzeptuelle Trennung der Begriffe Wirksamkeit und Wirkung vor[schlagen]. [...]

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.“ (2023, S. 17)

Wirkungen werden von Grünhaus und Rauscher (2021, S. 7) als „jene positiven und/oder negativen Veränderungen, die an Begünstigten bzw. Betroffenen nach erbrachter Aktivität bzw. konsumierter Leistung [...] oder in der Umwelt festzustellen sind“, bezeichnet. Sie nutzen das Wort Wirksamkeit nicht synonym, erläutern ihr Verständnis davon aber nicht.

Baumgartner und Haunberger (2023, S. 331) setzen Wirksamkeit und Effektivität gleich. Sie meinen mit den Begriffen die Eigenschaft einer Maßnahme, „bestimmte Wirkungen auslösen zu können“. Sie nutzen dabei den Kausalzusammenhang und setzen voraus, dass ein Wirkungsnachweis oder eine -plausibilisierung vorliegt. Als Wirkungen beschreiben sie die auf eine Maßnahme zurückführbaren empirisch nachgewiesenen oder plausibilisierten Veränderungen.

In Zusammenhang mit einer CIE (counterfactual impact evaluations) beziehen Kluve et al. „Wirkungen auf die Nettoeffekte, die als Differenz zwischen den durchschnittlichen Ergebnissen nach Durchführung der Maßnahme und den durchschnittlichen kontrafaktischen Ergebnissen definiert werden“. Wirksamkeit beschreiben sie als „die „Erreichung von Zielstellungen“ und wird durch den Vergleich des Erreichten mit dem Geplanten (oder mit der Ausgangssituation) oder aber durch den Vergleich der Beobachtungswerte nach Stattfinden der Maßnahme mit dem, was geschehen wäre, wenn die Maßnahme nicht stattgefunden hätte (kontrafaktische Situation)“, bewertet. (Kluve et al. 2021, S. 121)

Der Begriff Wirkung wird im Bereich der Wirkungsanalysen nochmal in unterschiedliche Termini geteilt (z. B. Impact, Outcome, Effect). Einige Autor\*innen unterscheiden zwischen Nettoeffekt und Bruttoeffekt. „Der Nettoeffekt bzw. die Wirkung, ist das, was bei einem Vergleich zwischen dem Gesamteffekt (oder Bruttoeffekt) und dem, was ohne die Intervention erreicht worden wäre, verbleibt. Der Nettoeffekt kann auch einen negativen Wert annehmen, wenn die Intervention weniger wirksam ist als die Marktdynamik.“ (ebd., S. 6) Wirkungen, die auch ohne die Maßnahmen eingetreten wären, werden häufig als Deadweight betitelt (Grünhaus und Rauscher 2021, S. 7). Dazu mehr bei den folgenden Begriffsdefinitionen.

Bei allen Definitionen lässt sich als Kern die Veränderung eines Zustandes finden. Unterschiede gibt es in Bezug auf die Klient\*innen- und Organisationsebene. Es besteht Einigkeit darüber, dass es sich um „intendier-

te und nicht intendierte Wirkungen“ handeln kann, die sowohl bei der Zielgruppe als auch bei anderen Stakeholder\*innen auftreten können. Wirkungen können nützlich und schädlich bzw. negativ und positiv sein. (Kränzl-Nagl et al. 2019, 28f; Grünhaus und Rauscher 2021, S. 7; Ottmann und König 2023, S. 23)

In diesem Dokument werden die Begriffe Wirksamkeit und Wirkung wie folgt verwendet:

Wirkungen sind die intendierten und nicht-intendierten Veränderungen bei den Adressat\*innen und deren Umfeld, die auf die Intervention zurückzuführen sind. Sie können sowohl positiv als auch negativ sein. Wirksamkeit ist das Ergebnis der eingetretenen Wirkungen.

### Input

Bei dem Begriff „Input“ herrscht Konsens bei den zitierten Autor\*innen. Es ist damit der Mitteleinsatz gemeint. Dieser beinhaltet die finanziellen, personellen und materiellen Ressourcen. Größtenteils ist der Input als Kosten erfassbar (Prinz 2019, S. 293; Beywl und Niestroj 2009, 142f; Widmer und Brunold, S. 4).

### Income

Verschiedene Autor\*innen ergänzen den Begriff Input um den des Income. Damit werden die (persönlichen) Ressourcen beschrieben, die die Adressat\*innen mit einbringen (Kränzl-Nagl et al. 2019, 28f; Beywl und Niestroj 2009, 142f).

Beide Begriffe, Input und Income, finden in dieser Arbeit wie hier beschrieben Verwendung.

### Output

Der Begriff Output wird ebenfalls einheitlich beschrieben. Er gilt als die erbrachte Leistung bzw. zählbaren/messbaren Resultate, welche unmittelbar erlebbar sind (z. B. Anzahl Einzüge, Anzahl Workshops) (Halfar und IGC-AG 2010, S. 37; Ottmann und König 2018, S. 6; Albrecht et al. 2013, S. 8; Widmer und Brunold, S. 4; Kränzl-Nagl et al. 2019, 28f; Beywl und Niestroj 2009, 142f).

Output ergibt sich aus den Aktivitäten, welche die pädagogischen Maßnahmen sind. Beywl und Niestroj unterscheiden zwischen mehreren Arten von Aktivitäten: „(a) Interventionen, die sich auf die Zielerreichung richten (b) Hilfsaktivitäten, welche die Interventionen ermöglichen oder unterstützen, und (c) Zielgruppen-Aktivitäten, die eine aktive Beteiligung der Zielpersonen erfordern.“ (2009, 142f)

### Outcome

Uneinigkeit bzgl. der Definition besteht bei dem Begriff Outcome. Die International Group of Controlling um Bernd Halfar verwenden ihn für die gesellschaftliche Wirkung. Sie meinen damit die objektive kollektive Effektivität. (2010, S. 37)

Die meisten Autor\*innen beschreiben mit dem Wort die Wirkungen auf der Ebene der Zielgruppe. (Prinz 2019, S. 293; Albrecht et al. 2013, S. 29; Ottmann und König 2018, S. 6; Beywl und Niestroj 2009, S. 144)

Die DAC beschränkt sich bei dem Begriff auf keinen Bereich und betrachtet den Outcome als „voraussichtlich oder tatsächlich erreichte [...] kurz- und mittelfristige [...] Wirkungen der Outputs einer Maßnahme“. (DAC 2009, S. 37)

Als stakeholder\*innenspezifische Bruttowirkungen wird der Outcome von Grünhaus und Rauscher (2021, 6ff) definiert. Bruttowirkung beschreibt dabei alle intendierten und nicht-intendierten Wirkungen sowie die Wirkungen, die auch ohne Maßnahmen eingetreten wären.

In dieser Arbeit wird der Begriff, entsprechend der Mehrzahl der Autor\*innen, wie folgt genutzt: Mit Outcome werden die Wirkungen auf der Ebene der Zielgruppe beschrieben.

### Effect

Über den Begriff Effect besteht erneut ein einheitliches Verständnis. Es wird damit die „unmittelbare, objektiv ersichtliche und [empirisch valide] nachweisbare Wirkung für einzelne Stakeholder“ (Kränzl-Nagl et al. 2019, 28f), z. B. die Anzahl von Abonnent\*innen eines Newsletters, beschrieben. Dieses inkludiert die beabsichtigten oder unbeabsichtigten Veränderungen. (DAC 2009, S. 27)

## Impact

Der Impact wird, wie der Outcome, konträr beschrieben:

Grünhaus und Rauscher (2021, 6ff) bezeichnen den Impact als Nettowirkung, welcher sich aus der Bruttowirkung abzüglich des Deadweight (Wirkungen, die auch ohne die Maßnahme eingetreten wären) ergibt.

Halfar und die IGC (2010, S. 30) bezeichnen mit Impact die subjektive Effektivität, also die Wahrnehmung der Stakeholder\*innen als Reaktion auf Effect und Outcome (z. B. zufriedene Rückmeldung aus Nachbarschaft).

Prinz, Albrecht et al. und Ottmann/König verstehen unter Impact die gesellschaftliche Wirkung bzw. den Nutzen für die Gesellschaft (Prinz 2019, S. 293; Albrecht et al. 2013, S. 29; Ottmann und König 2023, S. 21).

Kränzl-Nagl et al. und Beywl und Niestroj beschreiben Impact als Wirkungen einer „Dienstleistung, die über die bei Zielgruppen auftretenden Outcomes hinausgehen.“ (2019, 28f; 2009, S. 144)

Diese letzten beiden Definitionen werden in dieser Arbeit als Grundlage genommen: Mit Impact werden die über die Zielgruppe hinausgehenden gesellschaftlichen Wirkungen bezeichnet.

## Performance Measurement

Auch der Begriff des Performance Measurements ist in der Literatur nicht einheitlich definiert. So wird darunter je nach Kontext Leistung, Ergebnis und/oder Erfolg verstanden. „Die Interessensvielfalt der Stakeholder sowie die Multidimensionalität der Ergebnisse“ (Bono 2010, S. 2) sind die Kernelemente aller Konzepte. Es handelt sich beim Performance Measurement um einen Optimierungsprozess, der zur Steigerung von Effizienz und Effektivität dient (ebd.). Aufgabe des Performance Measurements ist die Quantifizierung von Ergebnissen/Leistung mittels Kennzahlen und Indikatoren. Diese Messgrößen sollten sowohl monetärer als auch nicht-monetärer Art sein. (Kränzl-Nagl et al. 2019, S. 54–55)

Da der Fokus beim Performance Measurement auf dem Erfolg liegt, welcher Wirkungen umfasst aber auch darüber hinaus geht, findet der Begriff hier keine Verwendung.

### Social Impact Measurement

Der Begriff Social Impact Measurement wird in der Literatur uneindeutig verwendet. Im Profit Bereich beschreibt er die sozialen Maßnahmen der Unternehmen. Im NPO-Bereich wird er i. d. R. synonym mit Wirkungsmessung genutzt. (Costa 2021, S. 132) Er wird in dieser Arbeit mit letzterer Definition verwendet.

### Abgrenzung zum Nutzen

Da der Begriff Nutzen ähnlich wie das Wort Wirksamkeit eine Folge von Wirkungen beschreibt, ist es hilfreich, eine Abgrenzung zwischen den Begriffen Wirkung und Nutzen deutlich zu machen. Beim Begriff Wirkung wird das Ergebnis in der Regel aus der organisationalen Ebene bewertet. Der Nutzen ist in erster Linie die Betrachtungsweise aus der Adressat\*innenperspektive. Daher ist es wichtig, bei Wirkungsanalysen den Nutzen aus Sicht der Adressat\*innen miteinzubeziehen. (Ottmann und König 2023, S. 16; Grünhaus und Rauscher 2021, S. 7). Andere Autor\*innen sehen den Nutzen rein im Zusammenhang mit einer ökonomischen Einheit, welcher für eine monetäre Bewertung dient. (Baumgartner und Haunberger 2023, S. 44) In dieser Arbeit wird von dem Begriff Nutzen Abstand genommen.

## 3.2 Wissenschaftlicher Kontext

In diesem Kapitel erfolgt die Einbettung des Themas in den wissenschaftlichen Kontext. Das Thema der Wirkungsmessung wird im Folgenden dazu aus zwei Perspektiven beleuchtet. Zunächst wird auf die Inhalte der Wirkungsdebatte eingegangen. Dabei wird die zeitliche Entwicklung mit den jeweiligen (sozialpolitischen oder professionellen) Einflüssen dargestellt sowie die Kritik an der Debatte erläutert. Im darauffolgenden Abschnitt werden die Erkenntnisse und der aktuelle Stand der Wirkungsforschung beschrieben. Es werden die verschiedenen Wirkungsdimensionen, die Wirkungsebenen und die Forschungsansätze thematisiert.

### 3.2.1 Wirkungsdebatte

Grund für die Wirkungsdebatte ist das Umdenken von konsumtiven Mitteln zur Bedarfsdeckung hin zu dem Ansatz der sozialen Investitionen, die einen Ertrag für die Gesellschaft erzeugen müssen. Im Zuge der Wirtschaftskrise und der dadurch begrenzten Mittel geschah ein Umdenken von der makroökonomischen (Keynes) zur neoklassischen/-liberalen Sichtweise (Sesselmeier et al. 2017, 37ff). Die Kostenträger\*innen wünschen effizientere Leistungserbringung, bei der nicht mehr für erbrachte Leistung (Output), sondern für Ergebnisse (Impact) gezahlt wird. In Einzelfällen gibt es Kontrakte, bei denen, ausgerichtet an der Qualität, höher vergütet wird („pay for result“ oder success) (Burmester und Wohlfahrt 2018, 14f). Hierbei handelt es sich um einen Teil des fortlaufenden Prozesses der Ökonomisierung, welcher die Wirkungsdebatte der Sozialen Arbeit seit den 1970ern prägt. (ebd., S. 16) Zu diesem Zeitpunkt setzte sich die Qualitätsbeurteilung mit den drei Dimensionen Struktur-, Prozess- und Ergebnisqualität nach Donabedian im sozialen Bereich durch. Über die Jahrzehnte verschob sich der Fokus von der Struktur- und Prozessqualität zur Ergebnisqualität. Bis sich Mitte der 1990er die Diskussion zuspitzte, als es Konsens gab, dass Qualität alleine nicht ausreiche, um Wirkungen zu erzeugen (Merchel 2019, S. 129). Der Unterschied zwischen dem Qualitäts- und dem Wirkungsbegriff liegt in erster Linie auf der „Blickrichtung“. Qualität zielt hauptsächlich auf die Erstellung von Kriterien zur Bewertung der Leistung und der Absicherung, dass die Ziele und Ergebnisse durch sinnvolle Prozesse erreicht werden, ab. Bei der Wirkungsmessung stehen Auswirkungen für die Klient\*innen und die Gesellschaft im Fokus der Untersuchung. (Kehl et al. 2016, S. 17) Es ist somit ein Wandel in der Wirkungsdebatte von der Qualität der Angebote zum ‚gesellschaftlichen Mehrwert‘ zu beobachten. Sozialleistungen werden nicht mehr als Unterstützung, sondern als Investition in die Zukunft gesehen, dabei ist der Wirkungsnachweis Bedingung für die Finanzierung (Burmester und Wohlfahrt 2018, S. 7). In Folge dessen wurde der Ergebnisbegriff weiter differenziert und bis heute bestimmen die Begriffe Output, Outcome, Effect und Impact den aktuellen Diskurs (Boecker und Weber 2019, 229f). Geprägt wurde dieser Diskurs zusätzlich durch das New Public Management (NPM), in dem die Wirkung „als eine Zielgröße der Steue-



„rung“ (Burmester und Wohlfahrt 2018, 17f) in den Fokus kam. Das NPM kommt aus dem angelsächsischen Sprachraum und wurde in Deutschland als Neues Steuerungsmodell (NSM) in den Verwaltungen eingesetzt, um diese durch Orientierung an Zielgrößen effizienter zu machen. Durch das NSM wurden in der Verwaltung Managementkonzepte, wie z. B. Controlling, Kennzahlen und Audits eingesetzt. Dadurch wurde der Fokus der Arbeit auf eine Auseinandersetzung mit Zielen bzw. Zielorientierung gelegt. Diese Veränderung hatte Auswirkungen auf den sozialen Bereich. Beispielsweise entstand ein Wettbewerb mit Auftraggeber\*innen-Auftragnehmer\*innen-Beziehungen (ebd., S. 18) und die Forderungen, Wirkungen von Maßnahmen nachzuweisen (Otto 2007, S. 9). Aus sozialpolitischer und Verwaltungssicht geht es in erster Linie um Erfolg und Wirkung, um eine ‚gerechte‘ Verteilung öffentlicher Mittel durchzuführen. (Burmester und Wohlfahrt 2018, 22f; Kluve et al. 2021, S. 6). Infolge dieser Entwicklung entstand das Kontraktmanagement, bei dem Leistungs- und Zielvereinbarungen zwischen Leistungsträger\*innen und sozialen Organisationen per Vertrag festgelegt wurden. Dieses entspricht einer Art Mikro-Management, bei dem die Kostenträger\*innen Effizienzziele vorgeben, an denen die sozialen Einrichtungen ihre pädagogischen Maßnahmen ausrichten müssen (Burmester und Wohlfahrt 2018, 22f). Es handelt sich dabei um wirkungsorientierte Investitionslogiken, infolgedessen Wettbewerbsstrukturen entstehen, welche sich auf die Steuerungsmedien auswirken. Wenn der Fokus auf dem Sozialen Impact liegt, (z. B. Social Return of Investment (SROI)) wandeln sich zwangsläufig die Angebote und Konzepte der NPOs (Polutta 2022, 228ff).

Dass der Druck in Hinblick auf die Wirkungsmessung steigt, zeigt z. B. auch die Gründung einer Untergruppe der GECES („Groupe d’Experts de la Commission sur l’Entrepreneuriat Social“) im Jahr 2012, welche damit beauftragt wurde, eine Methodik zu entwickeln, die europaweit zum Einsatz kommen sollte, um die sozialen Auswirkungen der geförderten Programme nach EuSEF- und EaSI-Rechtsvorschriften zu messen. Ziel war die Erforschung und Gewährleistung des wirksamen und sparsamen Einsatzes der Mittel. (Clifford 2015, S. 90)

Soziale Arbeit ist aufgrund der Finanzierung durch den Sozialstaat abhängig von der Sozialpolitik, und somit von der Konjunktur und Kri-

sen. Aufgrund dieser Abhängigkeit und als staatliche Aufgabe, sowie durch das Leistungsdreieck ist soziale Dienstleistung nicht unmittelbar von der Nachfrage gesteuert und in der Regel nicht wirtschaftlich orientiert (Wohlfahrt 2017, S. 57). Sozialwirtschaft ist im Unterschied zu profitorientierten Unternehmen nicht im finanziellen Sinn wertschaffend bzw. erzeugt keinen finanziellen Mehrwert (ebd., S. 63). Als meritorische Güter wird soziale Dienstleistung unabhängig von der Zahlungsfähigkeit der Adressat\*innen angeboten, somit entsteht häufig eine Indifferenz gegenüber dem Preis. Die Wirksamkeit kann statt des Preises als Entscheidungsgrundlage für Nutzer\*innen dienen. (Boecker und Weber 2018, S. 7)

Die Wirkungsdebatte in der Sozialen Arbeit wird kontrovers diskutiert. Die Kritiker\*innen berufen sich u. a. auf die These des Technologiedefizites von Luhmann, welches besagt, dass es keine Ursachen-Wirkungs-Kausalität in der Sozialen Arbeit gibt und somit eine Wirkungsmessung nicht umsetzbar ist (Schulze-Krüdener 2017, S. 67). Diese Kritik wird durch die Koproduktionsthese ergänzt, welche die Mitwirkung der Klient\*innen an der sozialen Dienstleistung betont (Compliance). Dabei besteht Einigkeit, dass neben der Methode auch die\*der Adressat\*in sowie die Fachkraft am Ergebnis beteiligt sind (Eppler et al. 2011, S. 10). Zusätzlich ist die Ergebnisoffenheit ein wichtiges fachliches Prinzip der Sozialen Arbeit, was dem Ansatz festgelegter Zielgrößen widerspricht (Hagn 2017, S. 88). Auch die Partizipation, als wichtiger Bestandteil der Sozialen Arbeit, kann nicht uneingeschränkt umgesetzt werden, wenn die Ziele im Vorfeld vorgegeben sind (Ziegler 2012, S. 100; Albus und Ziegler 2013, S. 165). Ergänzend dazu wird das Attributionsproblem, welches die Zuordnung eines Ergebnisses zu einer Maßnahme beschreibt, und das Kontributionsproblem, welches den Beitrag einer Intervention an einer bestimmten Veränderung bezeichnet, als Kritik genannt. (Baumgartner und Haunberger 2023, S. 38) Weitere Kritikpunkte sind, dass die Komplexität der Sozialen Arbeit nicht angemessen abgebildet werden kann und es eine Technisierung der Qualitätsbeobachtung gibt, obwohl Ergebnisse nicht in Zahlen abbildbar sind. Diese Vorgehensweise lenkt von angemessenen Formen der Reflexion ab. Außerdem ist die Verknüpfung von Wirkungsanalysen mit einer finanziellen Steuerung kritisch zu betrachten. (Merchel 2003, S. 13)

In der Praxis entsteht häufig die Sorge, dass die Ökonomisierung das oberste Ziel der Ausrichtung auf Wirksamkeit ist. Wie oben bereits geschrieben, sollten die finanziellen Ziele als Grundlage für die Erreichung der inhaltlichen Wirkungen dienen. Fachliche Ziele können nur durch die entsprechende Fachkompetenz und nicht durch ökonomische Parameter erreicht werden. (Horak und Speckbacher 2022, S. 234)

Dennoch ist in den letzten Jahren deutlich geworden, dass die Soziale Arbeit dem „Wirkungsthema sozialpolitisch nicht ausweichen kann und darf und zum anderen die Wirkung in der Handlungslogik der Sozialen Arbeit den zentralen Bezugspunkt“ (Merchel 2019, S. 130) darstellt. Die Wirkungsmessung befindet sich somit im Spannungsfeld zwischen sozialpolitisch geforderten Interventionen und dem normativen Selbstverständnis der Sozialen Arbeit (Schulze-Krüdener 2017, S. 99). Aus der Sichtweise der Kritiker\*innen stehen soziale Unternehmen dauerhaft unter Effizienz- und Kostenersparnis-Druck. Dieses hat Folgen für die Profession, die Mitarbeiter\*innen und auch für die Klient\*innen. Freie Träger werden als Dienstleistende und nicht mehr als Professionelle mit Expert\*innenwissen und Parteilichkeit gesehen (Hagn 2017, S. 87). Erwerbswirtschaftliche Logiken haben in den letzten Jahren Einzug in die sozialen Dienste genommen. Gleichzeitig sind die Dienste nicht vollständig refinanziert. Dieser Widerspruch prägt die Vorgehensweisen bzw. Perspektive (Wohlfahrt 2017, S. 55). Es „zeigt sich, dass gerade für die nicht an Gewinnen orientierten NGOs die Möglichkeiten schwinden, am Wettbewerb teilzuhaben und zugleich politisch die Interessen der Bezugsgruppen zu repräsentieren.“ (Deller und Brake 2014, S. 246)

Die zunehmende Bedeutung von wirtschaftlichen Aspekten im NPO-Management sollte unter anderem auch kritisch betrachtet werden, da sich bereits in der Vergangenheit Ansätze zur Steigerung der Qualität als Kosteneinsparungsmaßnahmen herausgestellt haben. Weitere Kritikpunkte sind, dass durch Wirkungsmessung ein „Titanic-Syndrom“ entstehen kann, welches starre Zielplanung statt situativen Managements und fehlende Auseinandersetzung mit Risiken und dem Umfeld bezeichnet. Darüber hinaus werden die bereits hohe Bürokratie und die aufwendigen Dokumentationsvorgaben in NPOs kritisiert, welche durch Wirkungsorientierung noch gesteigert werden könnte. Auch kann es zu einer zu hohen Erwartungs-

haltung in der Annahme kommen, dass durch Wirkungsorientierung „alles besser“ wird. Dieses kann zu Enttäuschung führen, da mit vertretbarem Aufwand ein kausal abgeleiteter Wirkungsnachweis von sozialen Maßnahmen nicht möglich ist. Außerdem können zu starre Leitplanken Situationspotenziale der professionell handelnden Personen einschränken. (2016, S. 121) Zusätzlich kann es zu Differenzen zwischen der Sicht der NPO und der gesellschaftlichen Perspektive kommen, wenn die Analyse unterschiedlich differenziert und mit verschiedenen Blickrichtungen durchgeführt wurde (Grünhaus und Rauscher 2022, S. 511). Wenn Ergebnisse von Wirkungsevaluationen aufzeigen, dass Maßnahmen vermeintlich keine Wirkung erzielen, ist mit großer Gegenwehr in den Organisationen und Kritik an der Vorgehensweise zu rechnen. Somit ist es wichtig, eine möglichst valide Messung zu erzielen. (Baumgartner und Haunberger 2023, S. 26)

Die Befürworter\*innen halten einen Wirkungsnachweis von Effektivität und Effizienz aufgrund der Knappheit der Mittel für unumgänglich (Sesselmeier et al. 2017, S. 31). Dabei wird davon ausgegangen, dass Stakeholder\*innen und Geldgeber\*innen Kriterien zur Bewertung der Wirksamkeit einfordern, die belegen, dass die finanziellen Mittel ‚sinnvoll‘ eingesetzt werden. Neben der Output-Messung besteht darüber hinaus Bedarf an Nachweisen von subjektiven, klient\*innenbezogenen und gesellschaftlichen Wirkungen (Eppler et al. 2011, S. 9). Die Haltung bzgl. sozialer Dienste und ehrenamtlichem Engagement hat sich verändert. Es geht nicht mehr um den Nachweis des Einsatzes von Ressourcen, sondern um die soziale Investition. (Then & Kehl 2012; Then & Kehl 2015a zitiert nach Kehl et al. 2016, S. 10) Der reine Verweis auf Prozesse reicht für einen hohen Professionalitätsanspruch nicht aus. Es bedarf glaubhafter und systematischer Darlegung von Effekten und Nebeneffekten, die durch Prozessgestaltung erzeugt werden und belegen „welche Wirkungen das Handeln in der Sozialen Arbeit nach sich zieht“. (Merchel 2019, S. 130) Aufgrund der einfachen Vergleichbarkeit durch die leicht verfügbaren Medien ist es nötig, den Stakeholder\*innen und der Öffentlichkeit umfangreiche Wirkungsnachweise darzulegen. Auch wenn es z. B. dafür keine Vorgaben durch die Leistungsträger\*innen gibt. (Kehl et al. 2012 zitiert nach Kehl et al. 2016, S. 10)

Die Bundesarbeitsgemeinschaft der Freien Wohlfahrtspflege e. V. (BAGFW) geht davon aus, dass soziale Organisationen ein intrinsisches Interesse haben, sich mit Hilfe der Darstellung von Wirkungen zu rechtfertigen (Burmester und Wohlfahrt 2018, S. 26). Bereits im Jahr 2015 hat die BAGFW das Centrum für soziale Investitionen und Innovationen (CSI) beauftragt, ein Gutachten<sup>5</sup> über die Einsatzmöglichkeit verschiedener Instrumente für die Wirkungsmessung zu analysieren. Das Interesse der Organisationen an einer Auseinandersetzung mit dem Thema stieg in den letzten Jahren zunehmend an. In der Praxis wird deutlich, dass es kein einheitliches Verständnis über den Begriff und die Auswirkungen für die operative Umsetzung gibt. (Kehl et al. 2023, S. 5) Dabei ist festzustellen, dass „Zielerreichung vielfach mit Wirkung gleichgesetzt“ (Burmester und Wohlfahrt 2018, S. 22) wird.

Die oben beschriebene Haltung, dass Wirkungsforschung als Grundlage für eine gerechte Mittelverteilung dient, ist konträr zu der Auffassung im englischsprachigen Bereich. Dort wird durch die Wirkungsforschung Einfluss auf die Sozialpolitik genommen, um soziale Dienstleistungen gut zu planen, zu steuern und umzusetzen (Otto 2007, 21f). Im deutschsprachigen Raum hat sich in der Forschung der Eindruck entwickelt, dass nur politisch opportune Wirkungen erwünscht sind. (Albus und Ziegler 2013, 163f)

Ein weiterer Diskussionspunkt in der Wirkungsdebatte ist die Frage, wer über die Wirkungen entscheidet. Dieser Punkt geht in die Richtung des machttheoretischen Ansatzes in der Sozialen Arbeit. Es macht einen Unterschied, ob die Verantwortlichen in den Organisationen, die Adressat\*innen, die Politik, die Geldgeber\*innen oder andere Stakeholder\*innen Kriterien für die Wirkungsmessung definieren. Indikatoren werden aus den unterschiedlichen Perspektiven entsprechend unterschiedlich

---

5 Dieses Gutachten wird u. a. von Burmester und Wohlfahrt kritisiert, da sie die genutzte Definition der sozialen Investition anders interpretieren. Laut ihrem Ansatz unterscheidet die EU zwischen sozialen Investitionen und Sozialschutzmaßnahmen. Kehl et al. machen diese Unterscheidung in dem Gutachten nicht. Darüber hinaus kritisieren sie den Fokus auf rein monetäre Ansätze. Auch sind sie der Meinung, dass das „Gutachten des CSI [...] eher für den Verfall der Sitten im Sozialsektor als für mehr Transparenz [steht].“ (Burmester und Wohlfahrt, 280f).

festgelegt und interpretiert (Burmester 2020, S. 41). In der Literatur wird in der Regel davon ausgegangen, dass die Leistungsträger\*innen und die Politik an Einsparungen (kurz-, mittel-, oder langfristig) interessiert sind. Dieses impliziert, dass diese Ziele nicht (immer) den individuellen Wünschen der Klient\*innen entsprechen und ggfs. auch den strukturellen Maßnahmen der Leistungserbringenden widerspricht. (Albus und Ziegler 2013, S. 166)

“Those in need have a right to know that the help they receive is likely to help their problem rather than harm, an government and others will only pay for interventions that are likely to make positive difference” (Otto 2007, S. 46 zit. nach Buchanan 2005, 2)

Dieses Zitat macht die Haltung der Befürworter\*innen der Wirkungsforschung deutlich. Damit sind alle potenziellen Stakeholder\*innen miteingeschlossen, die ein Interesse an den Ergebnissen der Wirkungsforschung haben. Das Ziel ist somit eine „systematische Untersuchung der Überprüfung der Wirksamkeit von zielgerechten Maßnahmen, Leistungen und Programmen und deren anschließende Bewertung“ (Schulze-Krüdener 2017, S. 44).

Es gibt dadurch einen tragfähigen Diskurs über Qualität und Qualitätsentwicklung durch eine empirische und/oder analytische Grundlegung. (Merchel 2003, S. 13) Dabei ist eine „konsequente Fokussierung auf die Adressatinnen“ (Gössler 2016, S. 121) ein potentiell positives Resultate von Wirkungsorientierung. Dieser Ansatz kann dem Vorwurf gegenüber Sozialunternehmen entgegenwirken, dass sie nur noch am Selbsterhalt statt am Nutzen für die Kund\*innen interessiert sind und dabei die Kosten explodieren. Durch die Aufstellung von klaren Effektivitätskriterien kann Wirkungsorientierung als Korrektiv fungieren. Außerdem werden im Rahmen der Wirkungsmessung integrative Steuerungsmodelle erstellt, welche die Chance der Entwicklung von zielführenden qualitativen und quantitativen Kriterien mit sich bringt. Diese Kriterien entstehen im Optimalfall im Zusammenwirken von Leistungsträger\*innen, NPOs und Vertreter\*innen der Zielgruppe. Selbstwertstärkung ist ein möglicher weiterer positiver Effekt. Sowohl im Gehalt als auch in der Gesellschaft spiegelt

sich häufig eine Abwertung gegenüber sozialen Berufen wider. Mitarbeitende in sozialen Bereichen geraten dadurch unter Rechtfertigungsdruck anstatt ihre Arbeit als wertvoll zu beschreiben. Wirkungsmessung kann dabei helfen, die Erfolge der Arbeit darzustellen und somit in ein positives Licht zu rücken. (ebd., S. 123) Die Fokussierung auf Ziele führt im Idealfall zu einer strukturierten und planvollen Vorgehensweise statt zu (blindem) Aktionismus. (ebd., S. 121) Gössler benennt auch den professionellen Fassadenbau als einen positiven Aspekt. Er zielt damit darauf ab, dass Managementkonzepte ein Zugehörigkeitsgefühl erzeugen und Unsicherheiten beseitigen. Zusätzlich wirken sie positiv auf Externe. (ebd.) Weiterhin kann mit einer proaktiven Vorgehensweise entgegengewirkt werden, dass Geldgeber\*innen oder Leistungsträger\*innen eigene Instrumente zur Wirkungsmessung entwickeln und die NPOs somit erneut in Abhängigkeit geraten. Denn auch Spender\*innen möchten eine größtmögliche Wirkung mit ihren Spenden erreichen. In dem Zusammenhang wird der Begriff effektiver Altruismus verwendet. (Grünhaus und Rauscher 2022, S. 508–509; Schober und Rauscher 2014, S. 262)

### 3.2.2 Wirkungsforschung

Ähnlich wie die Wirkungsdebatte finden sich auch in der Wirkungsforschung unterschiedliche Ansätze. Grundsätzlich werden die Vorgehensweisen in Dimensionen, Ebenen und Methoden differenziert.

Die Dimensionen beziehen sich auf sechs verschiedene Formen von Wirkung. Diese sind Folgende:

- „• Psychologische Wirkung (z. B. Linderung von Belastungserleben oder Depressionen, Ermöglichung von Selbstwirksamkeitserfahrungen, Stärkung des Selbstbewusstseins);
- Physische Wirkung (z. B. Vermeidung von körperlichen Belastungen, Beiträge zur Aufrechterhaltung oder Wiedererlangung körperlicher Gesundheit);

- Ökonomische Wirkung (z. B. zusätzliche volkswirtschaftliche Wertschöpfung, Kosteneinsparungen öffentlicher Haushalte und der Sozialversicherungssysteme);
- Soziale Wirkung (z. B. Etablierung und Stärkung von sozialen Beziehungen in Familien, Peer Groups und weiteren sozialen Netzwerken);
- Kulturelle Wirkung (z. B. Vermittlung von spezifischen Werten, Normen und Haltungen);
- Politische Wirkung (z. B. politischer Einfluss durch Anwaltschaft, Stellungnahmen, Lobbying)“ (Kehl et al. 2023, S. 8)

Studien belegen, dass Fach- und Leitungskräfte den Schwerpunkt auf die sozialen, kulturellen und psychologischen Perspektiven des Begriffs legen. Für sie sind die Klient\*innen die primären Stakeholder\*innen. In der Literatur wird hingegen besonders die ökonomische Seite hervorgehoben und die Zielgruppen sind Politik und Gesellschaft. (ebd., S. 4)

Neben den Wirkungsdimensionen werden zusätzlich drei verschiedene Ebenen unterschieden. Es wird von der Mikro-Ebene, mit welcher die individuelle Ebene der Klient\*innen gemeint ist, gesprochen. Die Meso-Ebene bezieht sich auf die Organisationen und die Makro-Ebene stellt die gesellschaftliche Dimension dar. (ebd., S. 8; Baumgartner und Haunberger 2023, S. 84) Sowohl in der Forschung als auch in der Debatte und den politischen Diskussionen finden sich alle Ebenen wieder und es kommt zu Vermischungen.

Zusätzlich zu der Unterscheidung zwischen Dimensionen und Ebenen bedarf es einer Klärung der Zielsetzung, wofür die Wirkungsmessung dient und in welchem Ausmaß sie angelegt werden soll. Je nach verfügbaren Ressourcen und Zielen gibt es unterschiedliche Stufen von Wirkungsmessung:



- „1) Modellbildung und anschauliches Monitoring
- 2) Modellbildung und informierte Schätzung aus öffentlich verfügbaren oder geschätzten Daten [(Wirkungsplausibilisierung)] [...]
- 3) Modellbildung und wissenschaftlich stringente (empirische) Messung“ (Kehl et al. 2016, S. 64–65)

Neben den unterschiedlichen Stufen bzw. Intensität der Wirkungsmessung gibt es verschiedene Forschungssettings, mittels derer die Stufen belegt werden können. Die Forschungsansätze differieren dabei hinsichtlich der Datengrundlage und der Umsetzung. Neben drei Kategorien, die von Dollinger (2018, 245f) benannt werden, die evidenzbasierte Forschung, die realistische Evaluation und die performative Folgenforschung, beschreiben Kluge et al. (2021) den Ansatz der counterfactual impact evaluations (CIEs). Bei diesen Ansätzen besteht kein Konsens über den Wirkungsbegriff, die Art, wie Wirkung zu messen ist und was dieses für Fachkräfte bedeutet. Dabei wird auf zwei Aspekte verwiesen. Erstens gibt es, angelehnt an den Ansatz von Luhmann, keine einheitliche und eindeutige Definition des Kausalitätsansatzes und zweitens ist ein Kausalitätsvergleich auf den Bezug einer bestimmten Handlung zu ihren Besonderheiten angewiesen (Dollinger 2018, S. 247).

Der empirische Nachweis von Kausalzusammenhängen ist nur über Blackbox-Verfahren (z. B. RCTs) möglich. Sie beantworten, ob die Maßnahme eine Wirkung erzeugt hat. Sie zählen zu den quantitativen Ansätzen. Soll ermittelt werden, wie eine Wirkung entstanden ist, kommen Whitebox-Verfahren (z. B. realistische Evaluation) zum Einsatz. Sie beinhalten quantitative und qualitative Methoden. (Ottmann und König 2023, S. 119)

Die unterschiedlichen Vorgehensweisen bzw. Paradigmen werden im Folgenden näher beschrieben.

### Evidenzbasierte Forschung:

Die evidenzbasierte Forschung ist in der Literatur und Praxis, die am häufigsten erwähnte und genutzte Methode. Dabei wird der Begriff Evidenzbasierung unterschiedlich interpretiert. Das Kernprinzip ist die Bereitstellung unstrittiger Evidenz, es dürfen nur gesicherte Ergebnisse genutzt werden. Programme müssen Unilinearität aufweisen sowie apersonale und dekontextualisierte Wirkungsannahmen treffen. Die Messung erfolgt durch streng kontrollierte und kausaltheoretisch angelegte Studien, z. B. mit Hilfe von randomized controlled trials (RCTs). Dabei ist nachweisbar, dass etwas funktioniert, aber häufig kann nicht belegt werden, warum (Dollinger 2018, 248f).

Es gibt zusätzlich unterschiedliche Auslegungen des Ansatzes im englisch- und deutschsprachigen Raum. Ursprünglich kommt der Begriff „evidenzbasiert“ aus der Medizin. Im englischsprachigen Raum ist die evidence-base practice bereits weit fortgeschritten. Aufgrund der engeren Verknüpfung zum Gesundheitswesen gibt es eine größere Offenheit für evidenzbasierte Soziale Arbeit. Wie oben bereits erwähnt, prägen dort die Ergebnisse der Forschung sozialpolitische Entscheidungen, wohingegen in Deutschland die Wirkungsforschung bislang in erster Linie zur Verteilung der Gelder genutzt wird. Dabei wird der Ansatz verfolgt, der im oben genannten Zitat von Buchanan angesprochen wird, dass die Stakeholder\*innen ein Anrecht haben, über Wirkungen informiert zu sein. (Otto 2007, 12ff)

Die Kritik an der evidenzbasierten Forschung ist in der Literatur weit verbreitet und es gibt Konsens, dass empirische Belege in der Sozialen Arbeit weiterhin schwierig sind, da es an Kontrollgruppen mangelt, weil es aus ethischer Sicht nicht praktikabel ist. Wenn Kontrollgruppen aus ethischen Gründen nicht für die Forschung in Frage kommen, besteht die Möglichkeit, quasi-experimentelle Verfahren durchzuführen. Hierbei werden bereits bestehende Gruppen als Vergleichsgruppe herangezogen. Dieses setzt voraus, dass eine solche Gruppe verfügbar ist. In der Praxis kann der Vergleich z. B. bei zwei Wohngruppen in der gleichen Einrichtung durchgeführt werden, wobei bei einer Gruppe ein neues Wohnkonzept erprobt wird und bei einer nicht. (Ottmann und König 2023, S. 126) Zusätzlich müssen empirische Studien in einem geschützten Raum statt-

finden, damit keine anderen Einflüsse die Ergebnisse beeinflussen. Häufig werden daher Programme oder Methoden und nicht die Wirkungen evaluiert. (Schulze-Krüdener 2017, 52f) In dem Zusammenhang wird das fehlende Passungsverhältnis kritisiert, da bezweifelt wird, dass Studien unter Laborbedingungen auf reale Lebenssituationen angepasst werden können (Albus und Ziegler 2013, S. 168). Eine weitere Kritik ist, dass die evidenzbasierte Wissenschaft das professionelle Urteilsvermögen ablöst. Es wird zusätzlich in Frage gestellt, inwieweit dieser Ansatz zum Berufsethos der Sozialen Arbeit passt, da die Individualität der Persönlichkeit der Adressat\*innen und der Fachkräfte sowie der Kontext und die Lebensumstände ausgeblendet werden. (Dollinger 2018, S. 250f)

Darüber hinaus sind aufgrund des Mangels an erforderlichen Forschungsbedingungen und Ressourcen bestehende Studien häufig nicht valide bzw. plausibel. Weiter wird argumentiert, dass der Hathrone Effect, welcher bedeutet, dass neue Maßnahmen Routinen brechen und daher wirksam sind, berücksichtigt werden muss (Schulze-Krüdener 2017, 62f).

#### Realistische Evaluation:

Eine andere Form der Forschung ist die realistische Evaluation. Die Kernprinzipien dieses Ansatzes sind kontextspezifisch wirkende Mechanismen. Anders als in der empirischen Forschung wird bei dieser Vorgehensweise das Umfeld und der Kontext mit einbezogen. Es wird davon ausgegangen, dass Kausalitäten nicht immer wahrnehmbar sind und dass sie von anderen Faktoren überlagert werden können. Generative Mechanismen, bei denen es Abläufe gibt, die abhängig vom jeweiligen Kontext wirken oder nicht, bilden die Grundlage. Wirkung wird als kontingent angesehen. Vergleiche und Variationen sind somit wichtig. Über die Evaluation von Maßnahmen und Erfahrungen in verschiedenen Kontexten werden „sets of ideas“ entwickelt, welche als Basis bzw. Orientierungsrahmen genutzt werden. Es handelt sich dabei um einen emanzipatorischen Ansatz, weil die Handlungsoptionen der Adressat\*innen erweitert werden. (Dollinger 2018, S. 250ff) Pawson und Tilley sind zwei Vertreter dieses Ansatzes. Sie haben einen Evaluationszyklus entworfen, der darauf abzielt, Reflexionswissen und Spezifizierung für Programme zu entwickeln, um auf dieser Basis Entscheidungen für das Handeln zu rechtfertigen. Bestandteile die-

des vierschrittigen Zyklus sind die Theoriebildung mit Blick auf Mechanismen, Kontexte und Outcomes; die Entwicklung von Hypothesen, darüber, was wirken könnte; die Beobachtung inklusive einer Datensammlung mit qualitativen und quantitativen Methoden sowie die Programmoptimierung (s. Abb. 3). (Schulze-Krüdener 2017, 108ff; Ottmann und König 2023, S. 141).

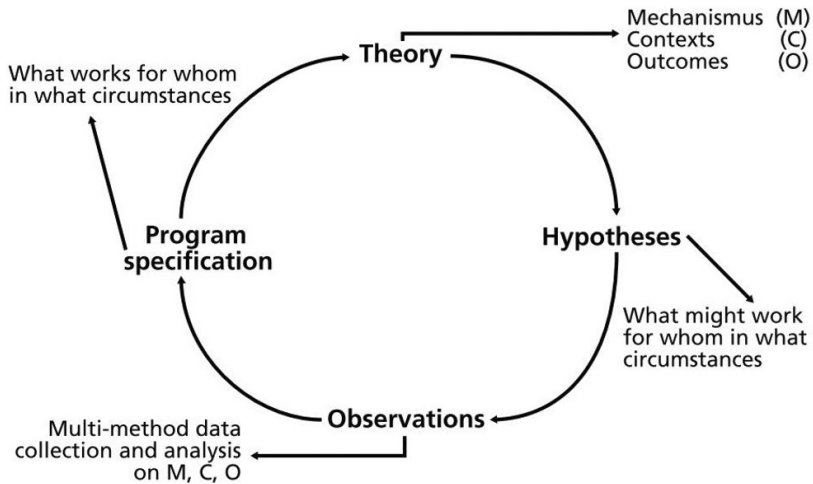


Abbildung 3: Realistic Evaluation Cycle: Ottmann nach Pawson, R. & Tilley, N. (1997) (Ottmann und König 2023, S. 141)

Es handelt sich bei der realistischen Evaluation um eine quasi-experimentelle Vorgehensweise, die nicht nach „what works?“, sondern „why does it work?“ fragt. Dieses bedeutet, dass der Fokus nicht darauf gerichtet ist, ob eine Intervention erfolgreich ist oder nicht, sondern unter welchen Bedingungen welche Maßnahmen für welche Zielgruppen welche Wirkungen erzielen. Wie bereits erwähnt, ist die Einbeziehung des Kontextes elementar. Der Vorteil dieser Vorgehensweise ist, dass Rückschlüsse gezogen werden können, ob Maßnahmen auch in anderen Kontexten erfolgsversprechend sein könnten. Ein Nachteil ist, dass Einbußen bei der Vorhersage von Wirkungswahrscheinlichkeiten gemacht werden müssen. (Micheel 2013, S. 190)

Durch ihre theoriebasierte Vorgehensweise eignet sich die realistische Evaluation besonders für die Verbesserung von Maßnahmen. Sie ist

besonders hilfreich für die Beurteilung von Projekten, die erfolgreich verlaufen, aber noch nicht deutlich wurde, warum und für wen. Außerdem kann sie für die Vorbereitung der Ausweitung von Maßnahmen genutzt werden, die auf neue Kontexte angepasst werden müssen. Sie kann auch zu Erkenntnissen führen, wenn Maßnahmen bisher unterschiedlich gute Ergebnisse geliefert haben, aber nicht zu erklären ist, wieso die Differenzen aufgetreten sind. (Baumgartner und Haunberger 2023, 198f)

#### Performative Folgenforschung:

Der dritte genannte Ansatz ist die performative Folgenforschung. Sie ist bisher wissenschaftlich nicht etabliert und wird z. B. von Ottmann nicht als eine eigene Form anerkannt. Das Kernprinzip ist die empirisch-realistische Orientierung. Performativ bedeutet in dem Zusammenhang, dass „Soziale Maßnahmen [...] in diesem Sinne Interventionen in komplexe soziale Zusammenhänge[n] [sind], die stets aktiv reproduziert werden“ (Dollinger 2018, S. 253). Der Fokus in dieser Vorgehensweise liegt auf Maßnahmen und deren Deutungen bezogen auf den Kontext. Adressat\*innen selbst geben den Maßnahmen einen spezifischen Sinn. Die empirische Ausrichtung wird geteilt, aber auf Grundlage faktisch realisierter Praxen beleuchtet. Dieser Ansatz wird im englischen Sprachraum „practice-based evidence“ genannt. Bei der performativen Kausalität handelt es sich um eine nicht weit verbreitete Position, welche aufgrund ihres spezifischen Kausalitätsverständnisses dennoch zu erwähnen ist. Teil des Selbstverständnisses dieses Ansatzes ist die Erzeugung von Wirkungen (ebd., 253ff).

#### Kontrafaktische Wirkungsanalysen (counterfactual impact evaluations, CIEs):

Den kontrafaktischen Wirkungsanalysen liegt eine Annahme von kausalen Zusammenhängen zwischen Intervention und Ergebnis zugrunde. Sie schließen alternative Erklärungen aus, indem sie auf die Frage antworten, was geschehen wäre, wenn die Intervention nicht durchgeführt worden wäre. Methodisch wird z. B. auf die Theorie des Wandels oder den Logikmodell-Ansatz zurückgegriffen. Es werden im Rahmen dessen experimentelle und quasi-experimentelle Ansätze genutzt. Der experimentelle Ansatz erfolgt als RTC oder im Rahmen von „sozialer Erprobung“. Für die quasi-

experimentelle Vorgehensweise gibt es verschiedene Methoden, z. B. das Propensity-Score-Matching, bei dem ein „statistischer Zwilling“ genutzt wird. Die CIEs werden z. B. für den Beleg der Wirkungen von europäischen ESF+-geförderten Maßnahmen genutzt. Sie dienen in erster Linie dazu, darzulegen, in welchem Umfang es Wirkungen gab und ob diese negativ oder positiv sind. Sie helfen nicht dabei zu ergründen, warum eine Maßnahme gewirkt hat, bzw. warum nicht. Das Umfeld wird bei diesem Ansatz nicht explizit mit einbezogen. Stakeholder\*innenperspektiven werden bedingt in Betracht gezogen. (Kluve et al. 2021, 12ff)

Auch hier ist kritisch zu beurteilen, wie die Messungen erfolgen, da bei dem experimentellen Ansatz eine Messung zum gleichen Zeitpunkt, bei den gleichen Personen mit und ohne der Maßnahme durchgeführt werden müsste. (Ottmann und König 2023, S. 121) Dieses ist in der Praxis nicht möglich, weshalb auf Kontrollgruppen zurückgegriffen wird. Die Herausforderungen dabei sind bereits unter dem Punkt „Evidenzbasierte Forschung“ thematisiert worden. Von manchen Autor\*innen wird dieser Ansatz als heuristische Vorgehensweise eingestuft, da die Hypothesen über mögliche Ereignisabläufe und deren Folgen nicht belegbar sind. (Baumgartner und Haunberger 2023, S. 38)

Es gibt Grenzen der Wirkungsforschung, welche im Folgenden benannt sind. Wenn keine empirische Belegung durch eine Kontrollgruppe durch Vorher-/Nachher-Studien besteht, kann nur von Wirkungsplausibilisierung gesprochen werden. Es sollte kommuniziert werden, warum ein empirischer Beleg nicht möglich ist. Für Vorher-/Nachher-Vergleiche ist eine frühzeitige Planung vor Maßnahmenbeginn wichtig. Dieses ist bei laufenden Projekten nicht mehr umsetzbar. Für neue Projekte sollten diese Vergleiche im Projektplanungsprozess inkludiert werden.

Darüber hinaus sind empirische Studien im Design mit sowohl interner als auch externer Validität nicht umsetzbar. Mit externer Validität ist die Generalisierung der Ergebnisse gemeint. Wenn externe Einflüsse ausgeschlossen sind, kann keine Übertragung auf andere Kontexte erfolgen. Werden Studien mittels Klient\*innen- und Fachkraftbefragungen durchgeführt, ist die Objektivität zu hinterfragen, da Personen ggfs. nicht wahrheitsgemäß antworten, wenn sie Konsequenzen befürchten oder Personen

nicht schlecht darstellen wollen. Dieses Problem kann aufgelöst werden, indem Kontrollbefragungen von weiteren Gruppen im Sinne einer Triangulation durchgeführt werden. (Ottmann und König 2018, 35f)

Zusätzlich zu der Kritik an der Validität von empirischen Studien in der Sozialen Arbeit kommt der Kostenfaktor hinzu. Sozialunternehmen können sich i. d. R. keine kostenintensiven Langzeitstudien leisten. Wirkmodelle ersetzen diese und reduzieren die Komplexität von Wirkungszusammenhängen. Sie dienen dadurch als Grundlage zur Entscheidungsfindung und Steuerung von Sozialunternehmen. (Kränzl-Nagl et al. 2019, S. 34)

Die Subjektivität der Fragestellung und die begrenzte Aussagekraft der jeweiligen Forschungsmethode sind allgemeine Kritikpunkte an der Forschung in der Sozialen Arbeit (Schulze-Krüdener 2017, S. 35). Es wird davon ausgegangen, dass es keine Wenn-Dann-Kausalitäten gibt, sondern Wirkungszusammenhänge, die individuell überprüft werden müssen (ebd., S. 67). Zusätzlich ist die empirische Forschung mit einem hohen Einsatz von Ressourcen verbunden. Dementsprechend werden die Ansätze der realistischen Evaluation und der performativen Folgenforschung am praxistauglichsten angesehen.

König hat für die Praxisforschung in der Sozialen Arbeit Gütekriterien aufgestellt. Diese sind folgende: Realisierbarkeit mit Blick auf Bedingungen und Ressourcen, Angemessenheit bei der Auswahl der Methoden und der Datenquellen, Gültigkeit (Korrespondenz zwischen Gegenstand und Indikatoren), Regelgeleitetheit (Prozessstreue und Transparenz), Verwertbarkeit (Anschlussfähigkeit der Ergebnisse). (2016, 86f) Diese Kriterien sollten auch bei der Umsetzung von Wirkungsmessung berücksichtigt werden.

In dem Transparenzgutachten der CSI, welches von der BAGWF beauftragt wurde, wird deutlich, dass es aktuell kaum Ansätze gibt, die Wirkungen vollumfänglich messen. Der Großteil der Ansätze legt den Fokus darüber hinaus auf die Outputs statt den Outcome oder den Impact. Wirkungskausalitäten werden weitestgehend nicht vollständig nachvollzogen. Somit können Wirkungen, die ohnehin eingetreten wären (Deadweight), nicht identifiziert werden. Die meisten Ansätze beschränken sich nur auf eine Ebene, z. B. Mikro- oder Makroebene, ohne die Erkenntnisse zu verknüpfen. I. d. R. liegt der Fokus auf der ökonomischen Dimen-

sion. Die kulturelle und politische Perspektive wird nahezu in keinem Ansatz behandelt. Aus einigen Ansätzen können Methoden entnommen werden, die in Kombination vielversprechend erscheinen. Darauf wird u. a. in Kapitel 4.4 eingegangen. Die untersuchten Modelle geben bisher noch keine Auskunft über „sozialintegrative (Teilhabe-) und wertebildende/-festigende Effekte wohlfahrtsverbandlicher Arbeit sowie themenanalytische Beiträge zur sozialpolitischen Diskussion und Entscheidungsfindung“. (Kehl et al. 2016, 54f)

### 3.2.3 Einordnung des Themas in den wissenschaftlichen Diskurs

Die Diskussion über die Verwendung von Managementansätzen im sozialen Bereich ist ein weiterer Schritt in der langjährigen professionalisierungstheoretischen Auseinandersetzung. Die Debatte über Qualität und Wirkung lenkt den Fokus weg von den ehemaligen Themen in Bezug auf die Verberuflichung der Sozialen Arbeit hin zu den Strukturproblemen sozialpädagogischen Handelns. (Dewe und Otto 2018, S. 1198)

Laut Luhmann ist die heutige professionelle Hilfe mehr als die reine „gute Tat“ und setzt daher voraus, dass Organisationen den Hilfezustand als Fall verstehen und proaktiv Programme entwickeln, die dort ansetzen. U.a. durch das Subsidiaritätsprinzip und die zahlreichen Anbieter am Markt haben sich die Maßnahmen qualitativ unterschiedlich gut entwickelt. Aktuell nimmt der Staat eine passive Rolle ein und es besteht keine Vergleichbar- und Messbarkeit bzgl. der Wirkung und Qualität der Maßnahmen. Zeitgleich schmücken sich die Organisationen als „Wohltäter\*innen“ durch ihr Marketing, u. a. um Mittel zu akquirieren, während der Status der Sozialarbeiter\*innen zunehmend sinkt. Die einschränkende Vorgehensweise der Organisationen gegenüber der Professionalität der Mitarbeitenden verstärkt den Statusverlust. (Erath und Balkow 2017, 114ff) Leistungsorientierte Debatten überlagern die Professionalität (Dewe und Otto 2018, S. 1198). In dem Zusammenhang wird auf Managementkonzepte wie z. B. Wirkungsmessung verwiesen.

Die Debatte über den Einsatz von Managementinstrumenten in der Sozialen Arbeit wird von manchen Autor\*innen als „Reformulierung der Ordnung des Sozialen“ (Otto und Ziegler 2006, 96ff) gesehen. Kritisch



wird vermerkt, wenn der Fokus nur auf dem Faktor Geld liegt. Wichtig ist dabei, die „zentrale[n] Werte wie „Effizienz, Flexibilität, Qualität, Wettbewerb, Effektivität, Kundenorientierung und ‚value for money‘“ (Gibbs 2005: 229)“ (Otto und Ziegler 2006, S. 96) nicht aus dem Blick zu verlieren. Dabei wird das new public management als Misstrauen gegenüber professionellem Handeln interpretiert. Organisationswissen überlagere Fachwissen und die Rahmenbedingungen für Professionelle würden so eng geschnürt, dass fachliche Ermessensentscheidungen nicht mehr zum Tragen kommen können. Otto und Ziegler sprechen von einer „evidenzbasierten De-Professionalisierung“ (ebd., S. 105), da nicht mehr der Fall an sich im Fokus ist, sondern Risikokategorien bearbeitet werden. Wenn Wirkungsmessung als strenge Zielvorgabe interpretiert wird, widerspricht sie der Autonomie der Adressat\*innen<sup>6</sup>. Was ggfs. zu evidenzbasierten Erfolgen führt, steht einer erfolgreichen sozialpädagogischen Praxis gegenüber. (ebd., S. 96) Wird dahingegen z. B. der Capability-Ansatz von Amartya Sen und Martha Nussbaum herangezogen und Wirkungsmessung so ausgelegt, dass das individuelle Ziel der Adressat\*innen die Zielvorgabe ist, dann steht die Wirkungsmessung zu der pädagogischen Fachlichkeit nicht im Widerspruch. Wirkungsforschung kann nur erfolgreich sein, wenn sie in reflexiven Verwendungszusammenhängen zum Einsatz kommt. (Albus et al. 2018, S. 1828)

Wenn anzunehmen ist, dass professionelles Handeln immer auf einen bestimmten Zweck ausgerichtet ist, kann davon ausgegangen werden, dass „Wirksamkeit „... die zentrale Leitorientierung aller Professionen und damit die *conditio sine qua non* jeglicher Professionalität schlechthin“ (Baumgartner und Sommerfeld 2012, S. 1163, zitiert nach Baumgartner und Haunberger 2023, S. 25) ist.<sup>7</sup> Wird dieses vorausgesetzt, dann sind Maßnahmen per se wirkungsorientiert. Sofern zwei Bedingungen erfüllt sind: die Profession begründet fachlich, was Wirkung bedeutet und sie begründet dieses empirisch. (ebd.)

6 Wobei Soziale Arbeit an sich bereits Eingriffe in das Leben anderer durchführt. Dieses ist aus der ethischen Perspektive zu beleuchten und wird im Rahmen dieser Arbeit nicht thematisiert.

7 Diese Behauptung ist aus Sicht der Autorin dieser Arbeit kritisch zu betrachten. In der Theorie ist zwar davon auszugehen, dass Soziale Arbeit diesen Anspruch hat, es bedeutet aber nicht zwangsläufig, dass es in der Praxis so umgesetzt wird.

Optimalerweise bleibt die Soziale Arbeit anschlussfähig an die Wissenschaft und gleichzeitig offen für professionelles praktisches Handeln. Daheim kategorisiert drei Gruppen von Wissen: wissenschaftliches Wissen, berufliches Wissen und Alltagswissen. (1992, 28f zitiert nach Deller 2009, S. 98) Stichweh weist auf die „Verbindung zwischen wissenschaftlich generiertem Wissen und der Professionspraxis“ hin (ebd., 97f). Oevermann erweitert diese Ansicht, indem er Profession als die Stelle der „Vermittlung von Theorie und Praxis unter Bedingungen der verwissenschaftlichten Rationalität“ (1996, 80 zitiert nach ebd., S. 99) beschreibt. Wissenschaft und Praxis werden von Dewe/Ferchhoff/Radtke als gleichwertige Anteile der Profession bewertet (ebd.). Die Generierung von wissenschaftlichem Wissen kann bspw. über Methoden-, Vergleichs- oder Effektstudien sowie Metaanalysen erfolgen. Die Wirkungsmessung ist im Bereich der Effektstudien angesiedelt. (Erath und Balkow 2017, 114ff)

### 3.3 Theoriegrundlage der Arbeit

Da die performative Folgenforschung nicht allgemein als Wirkungsforschungsansatz akzeptiert ist und die evidenzbasierte Forschung für die Soziale Arbeit kritisch zu betrachten ist, wird in dieser Arbeit die realistische Evaluation als Grundlage genommen.

Nach diesem Ansatz gibt es keine „wirksamen“ oder „unwirksamen“ Maßnahmen. Die Stärke der Wirksamkeit wird in einem Zusammenspiel von fachlichen und normativen Entscheidungen beurteilt. Somit gibt es eine hohe oder weniger starke Wirksamkeit einer Maßnahme, im Vergleich zu einer anderen oder keiner Maßnahme. (Ziegler 2012, S. 95)

Aufgrund der fehlenden Möglichkeit in der Sozialen Arbeit lineare Ursachen-Wirkungsbeziehung nachzuweisen, ist es elementar, den Kontext mit einzubeziehen. Dabei wird davon ausgegangen, dass sich Wirkungen durch ein Zusammenspiel von Angeboten und Kontext ergeben (Ottmann und König 2019, S. 69). Grundlage hierfür ist das von Pawson und Tilley aufgestellte KOM-Modell, welches in dem Zusammenhang Kontext, Outcome und Mechanismus bedeutet. (Ottmann und König 2023, 49f).

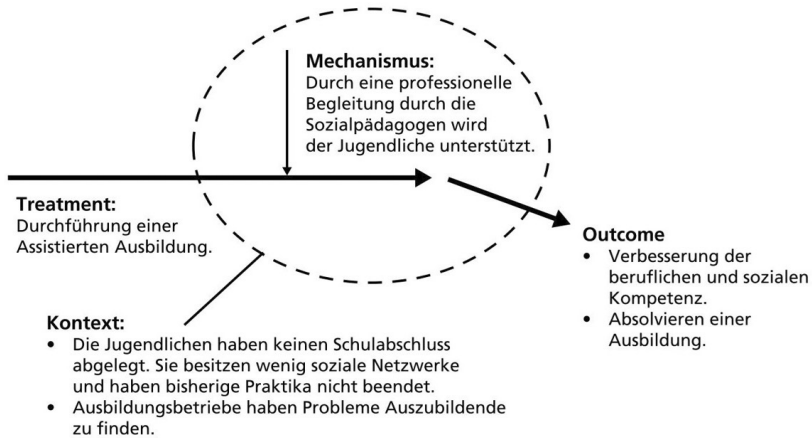


Abbildung 4: KOM: Einbeziehung des Kontextes in Wirkmodell (ebd., S. 50)

Wie in der Abb. 4 zu sehen ist, wird die sozialpädagogische Intervention in Bezug zum Kontext gesetzt, dieser hat Einfluss auf den Outcome.

Wenn Ergebnisse nicht empirisch belegt sind und somit kein kausaler Ursache-Wirkungs-Zusammenhang vorliegt, wird von einer Wirkungsplausibilisierung gesprochen. Diese erfolgt durch eine Ergebnisinterpretation, welche fachlich begründet, welchen Anteil die Maßnahme und welchen äußere Einflüsse an den Veränderungen haben (Ottmann und König 2019, S. 67; Merchel 2019, S. 135; Baumgartner und Haunberger 2023, S. 42).

In Kapitel 3.2.2 (Wirkungsforschung) wurde bereits erläutert, dass es keinen idealen Ansatz der Wirkungsmessung gibt. Deshalb bedarf es aktuell der Verknüpfung von verschiedenen Methoden aus den unterschiedlichen Ansätzen, um ein umfassendes und mehrdimensionales Ergebnis zu erhalten. (Kehl et al. 2016, S. 56) In Kapitel 4 werden daher einige Modelle vorgestellt, um anschließend eine für das Praxisfeld taugliche Variante zu kombinieren. Im nächsten Kapitel wird dafür der Arbeitsbereich näher beschrieben.

### 3.4 Vorstellung des Praxisfeldes

Der Leistungsbereich Beruf und Arbeit (LB BuA) ist eine von fünf Abteilungen des Geschäftsfeldes Integration, welches neben dem Geschäftsfeld Teilhabe und dem Geschäftsfeld Alter und Pflege die operativen Bereiche des Caritasverbandes für die Stadt Köln e. V. beinhaltet. Der Verband umfasst zusätzlich die Abteilung Zentrale Services sowie das Innovationsmanagementteam und Stabsstellen. Der hauptamtliche Vorstand des Vereines besteht aus zwei Personen und wird durch eine besondere Vertretung des Vorstandes unterstützt. „Der Caritasverband für die Stadt Köln ist Träger von 80 Diensten und Einrichtungen der Sozialen Arbeit und Pflege im Kölner Stadtgebiet.“ (Bertke) Aktuell arbeiten ca. 2000 Menschen hauptamtlich im Verband. Weitere 1.630 Personen sind ehrenamtlich tätig. Im Geschäftsfeld Integration arbeiten ca. 330 Menschen. Dem LB BuA gehören 70 Personen an. Er ist auf vier Einrichtungen aufgeteilt. Diese sind die Schuldner-, Insolvenz- und Sozialberatung; die Arbeitsmarktintegration; das Jugendbüro und die offene Kinder- und Jugendarbeit (OKJA). Weitere Details sind den Organigrammen des Verbandes und des LBs im Anhang (S. 119f) zu entnehmen.

In den Einrichtungen gibt es unterschiedliche Projekte und Maßnahmen, die von vielen verschiedenen Leistungsträger\*innen finanziert werden. Einige der Geldgeber\*innen sind z. B. das Jobcenter, die Kommune, die Kirche, das Land NRW, die EU, Stiftungen und Vereine<sup>8</sup>. Außer einem Teil der OKJA handelt es sich um keine Pflichtleistungen. Nahezu keines der Projekte bzw. Maßnahmen ist auskömmlich refinanziert, da Eigenanteile bei Projektmitteln gefordert sind und/oder Personalkostensteigerungen und Overheadkosten nicht geltend gemacht werden können. Es besteht der Druck, regelmäßige Ko-Finanzierungen zu organisieren und/oder Maßnahmen zu beenden.

Die Anforderungen der Leistungsträger\*innen bzgl. der Zielerreichung sind hoch und teilweise aus pädagogischer Sicht nicht immer realistisch umsetzbar.

8 Diese Aufzählung ist nicht abschließend.

Es besteht der Bedarf, Strategien zur Steuerung zu entwickeln, um effizient und effektiv zu arbeiten. Zusätzlich werden Kriterien benötigt, auf Grundlage derer Entscheidungen zur Weiterführung, -entwicklung oder Schließung von Angeboten getroffen werden können.

