

Wirkungsanalysen und ihre Forschungsdesigns: Kritische Gegenüberstellung ihrer konzeptionellen und methodologischen Herangehensweise im Hinblick auf mögliche Komplementaritäten und Anwendungen für die Beratungswissenschaften*

Bernhard Plé^{1,2,3}✉

Zusammenfassung

Der Artikel gibt zunächst einen Überblick über die wichtigsten Ansätze zur Messung und Analyse von sozialen Wirkungen. Ziel ist es herauszustellen, welche Herausforderungen konzeptioneller und methodologischer Art sich verschiedenen Forschungsdesigns stellen: dem experimentellen Design, quasi-experimentellen Design, Ex-post-facto-Design, Regression Discontinuity Design, Pretest-Posttest-Design und den qualitativen Designs. Eine vergleichende Gegenüberstellung derselben zeigt auf, welche Komplementaritäten zwischen den verschiedenen Ansätzen möglich sind.

Abstract

This paper begins with giving an overview of the main approaches to social impact measurement and analysis. Its purpose is to point out the conceptual and methodological challenges posed to the different types of research-designs: experimental designs, quasi-experimental designs, ex post facto design, regression discontinuity design, pretest-posttest design, and qualitative designs. A comparison between these different designs shows that there are important points of complementary between qualitative and quantitative approaches to social impact analysis.

Keywords: Beratungswissenschaften, Evaluationsforschung, Forschungsdesigns, social impact analysis, Wirkungsanalysen

1. Einleitung

Seit Ende der 1990er Jahre werden Wirkungsanalysen verstärkt thematisiert. Anlässe dazu gaben verschiedene Reflexionen über ihre Notwendigkeit sowie

Durchführbarkeit. Leitend waren damals drei Gesichtspunkte, die noch heute die Diskussion bestimmen. Erstens vermehrten sich die Wahrnehmungen von Defiziten bei Wirkungskontrollen, die im Rahmen von Evaluationen des Verwaltungshandelns, der staatlichen Verwaltungs- und Institutionenmodernisierung (zum Beginn vgl. Wollmann, 2000, S. 200-202) sowie der staatlichen Entwicklungspolitik (zum Beginn vgl. Stockmann, 2000, S. 382) durchgeführt wurden. Als zentrales Problem der Wirkungsanalysen im Bereich staatlichen Handelns gilt nach wie vor der Nachweis, dass bestimmte Veränderungen – oder auch nur der Erhalt eines Ausgangszustandes – kausal mit einer gesetzten Intervention zusammenhängen. Zu diesem Ergebnis gelangen sowohl der *Arbeitskreis Evaluation von Entwicklungspolitik* (2009) als auch der im Auftrag des *Austrian Research Centers* erstellte Report mit dem Titel „Neuere Entwicklungen im Bereich der Wirkungsanalyse und

* Überarbeitung der Vorträge mit dem Titel: Wirkungsforschung Teil I & II – mögliche Analyserahmen für den Nachweis von Wirkungen von Beratung?, im Wissenschaftssalon der ARGE Bildungsmanagement am 20. Jänner 2014 und 20. April 2015

¹ FH Joanneum, August-Eichhorn-Institut für Soziale Arbeit, Eggenberger Allee 11, 8010 Graz

² Institut ARGE Bildungsmanagement am Department Psychologie der Sigmund Freud Privatuniversität

³ Universität Bayreuth, Bayern, Deutschland

✉ Korrespondenz über diesen Artikel ist zu richten an FH-Prof. apl. Prof. DR. Bernhard Plé, Österreich. E-Mail: Bernhard.Ple@fh-joanneum.at

Lizenzbedingungen:



-abschätzung“, der auf den Bereich der Forschungs-Technologie- und Innovationspolitik eingeht (vgl. Rohmberg, Steindl & Weber, 2006). Seit dem Jahr 2006 sind allein im Bereich staatlichen Handelns mindestens 19 erprobte Verfahren zur Wirkungsanalyse im Einsatz, von denen knapp die Hälfte sektorspezifisch angewendet wird, um die direkten, also geplanten Wirkungen z.B. in der Bildung, im Gesundheitswesen oder in der Zivilgesellschaft, zu erfassen. Einen Überblick über diese verschiedenen Verfahren gibt der *Arbeitskreis Evaluation von Entwicklungspolitik* (2009).

Anlass zur Thematisierung von Wirkungsanalysen gab zweitens der im Bereich von Schulungsmaßnahmen zunehmende Bedarf nach Erfolgsnachweisen. Sowohl bei Personalverantwortlichen als auch bei durchführenden Organisationen wurde das Erfordernis wahrgenommen, Erfolgsnachweise zu erbringen oder auch zu fordern, damit Nutzen und Rentabilität des Aufwandes messbar werden (Hoffman & Luisser, 2009). Infolge der zunehmenden Bedeutung von Coaching ist die Forschung über dessen Wirkungen und Erfolgsfaktoren in den letzten Jahren stark angestiegen. Überblicke über die Wirkungsforschung im Bereich Coaching geben die Studien von Jansen, Mäthner & Bachmann (2004, 2005), Künzli (2005, 2009), Greif (2008) und Kühl (2014).

Wirkungsanalysen wurden Ende der 1990er Jahren auch in einer dritten Hinsicht thematisiert. Anlass dazu gab eine in den USA aufgekommene Diskussion über die praktische Relevanz der in sozialpsychologischen oder Feldexperimenten ermittelten Wirkungen von Coachings, massenmedialer Stimuli oder Förderprogrammen. Seit dem Jahr 2000 finden sich auch in deutschsprachigen Ländern Ansätze einer diesbezüglichen Methodendiskussion. Sie vereinen sich mit neuen Konzeptionen von Medienwirkungen (Jäckel, 2011) und sie bestimmen auch konkrete Wirkungsanalysen von Programmen in einzelnen Handlungsfeldern wie Familienhilfe, Kinder- und Jugendhilfe (Haubrich, 2009). Dabei finden sich auch allgemeine Überlegungen zum Einsatz qualitativer Methoden in der Sozialen Arbeit (Bohnsack & Nentwig-Gesemann, 2010), insbesondere in der Jugendhilfe (Kelle & Erzberger, 2006), sowie handlungsfeldübergreifende Einführungen in die qualitative Wirkungsanalyse (Kuckartz et. al., 2007; Bohnsack, 2010).

Welche Herausforderungen konzeptioneller und methodologischer Art sich der sozialwissenschaftlichen Wirkungsanalyse heute stellen, soll in diesem Beitrag behandelt werden. Dazu sollen leitende Ver-

ständnisse von Wirkungen, Wirkungsebenen und Wirkungserhebungen dargestellt und daraufhin überprüft werden, welche Stärken und Schwächen sie jeweils aufweisen. Anhand einer kritischen Gegenüberstellung derselben soll abschließend aufgezeigt werden, welche Komplementaritäten in den verschiedenen Ansätzen möglich sind.

2. Herausforderungen an die Wirkungsanalyse

Nach wissenschaftlichem Verständnis sind Wirkungen alle Veränderungen, die nach einer Intervention (Maßnahme) auftreten. Hierzu gehören auch nicht erwartete und unerwünschte Wirkungen. Wirkungen umfassen somit die intendierten Veränderungen und die nicht-intendierten Veränderungen, wobei zu den Letzteren die positiven sowie die negativen Veränderungen zählen. Zudem kann zwischen internen, bei der Durchführungsorganisation ausgelösten Wirkungen und externen, bei den Zielgruppen sowie in deren Umfeld hervorgerufenen Wirkungen unterschieden werden. Zunehmend findet Beachtung, dass die Komponenten „Ressourcen“ und „Vernetzung der Durchführungsorganisationen“ nicht nur für die unmittelbare Steuerung der Intervention, sondern auch für die Diffusion der Wirkungen bei den Zielgruppen von zentraler Bedeutung sind. Daher sollte auch die Leistungsfähigkeit der Organisationen, die mit der Durchführung von Interventionen betraut sind, im Fokus der Wirkungsanalyse stehen (Stockmann, 2006).

Grundlage der Interpretation kausaler Zusammenhänge ist die Feststellung von Veränderungen bei denjenigen, die an einem Programm oder an einer Maßnahme teilgenommen haben. Dieser zunächst trivial klingende Schritt bedeutet einen Vorher-Nachher-Vergleich, wozu mindestens zwei Erhebungszeitpunkte vorgesehen sein müssen. Dabei gilt insbesondere für Programme und Maßnahmen, die dauerhafte Veränderungen im Verhalten, im Bewusstsein, in Strukturen und/oder Prozessen zum Ziel haben, dass solche Effekte erst nach Abschluss der Teilnahme an einer Maßnahme, einem Programm oder einer Beratung erhoben werden können. Dazu muss entschieden werden, wann die richtigen Erhebungszeitpunkte sein sollen. Zudem muss in der Praxis die Frage entschieden werden, wie mit Klienten und Klientinnen in der Wirkungsanalyse verfahren werden soll, die das Programm vorzeitig verlassen haben oder die für die Nachuntersuchung nicht mehr

zur Verfügung stehen (Groenemeyer & Schmidt, 2011).

Es stellen sich also bei der Wirkungsanalyse bestimmte Herausforderungen, die im Folgenden thematisiert werden sollen. Sie beginnen bereits an der Stelle, wo die Vorstellungen in der Zielgruppe über die mit einer Intervention zu erreichenden Ziele sehr weit auseinander liegen können. So können die Akteure z.B. der Öffentlichen Verwaltung oder einer Organisation mit der Verabschiedung und Finanzierung eines Programms andere Ziele verfolgen als die durchführenden Organisationen in ihren Programmen festlegen; wieder andere Vorstellungen entwickeln dann die Professionellen in ihrer alltäglichen Arbeit mit der Klientel. Die Festlegung der Zielüberprüfung kann auch über einen Aushandlungsprozess der an den Maßnahmen Beteiligten verlaufen, was nicht nur zeitaufwändig und unter Umständen konfliktrichtig ist, sondern auch keine Gewähr für eindeutige Festlegungen gibt. Verkompliziert wird diese Festlegung, wenn zudem der Anspruch erhoben wird, auch die nicht-intendierten Folgen oder Nebenwirkungen in den Blick zu nehmen. Ähnliches gilt auch für die Übersetzung der Zielvorstellungen in messbare Indikatoren. Sehr häufig werden auf dieser Grundlage Erfolgsfaktoren berechnet, d. h. diejenigen, bei denen sich Veränderungen in der angestrebten Richtung zeigen, werden als Erfolge der Maßnahme, diejenigen, bei denen keine Veränderungen gemessen werden kann oder deren Situation sich in Bezug auf das Ziel verschlechtert hat, als Misserfolge gezählt. Es stellt sich auch das Problem, dass sich Individuen, Organisationen oder soziale Verhältnisse auch ohne Teilnahme an der Maßnahme entwickeln oder sie während und nach der Maßnahme sehr unterschiedlichen anderen Einflüssen ausgesetzt sind, und dies umso mehr, je weiter der letzte Erhebungszeitpunkt von der Beendigung der Maßnahme entfernt ist. Als Kausalanalyse steht die Wirkungsforschung vor dem Problem, gemessene Veränderungen ursächlich auf die Maßnahme zurückzuführen.

Als Vorbilder zur Lösung dieser Probleme galten bis in die 1980er Jahre die experimentelle Methode in Kleingruppen und der quasi-experimentelle Versuchsplan im Feld (Wollmann, 2000). Diese Methoden bestehen bekanntlich darin, dass mindestens eine Untersuchungsgruppe, die an der Maßnahme teilnimmt, mit mindestens einer Kontrollgruppe, die nicht an der Maßnahme teilnimmt, verglichen wird. Dabei soll die Zufallszuweisung zu beiden Gruppen sicherstellen, dass alle "Störvariablen", also unter-

schiedliche individuelle Voraussetzungen und situative Umwelteigenschaften, die die Wirkungen der Maßnahme beeinflussen können, in beiden Gruppen zufällig verteilt sind und von daher das Ergebnis nicht "verfälschen" können. Nach Abschluss der Maßnahme werden dann Unterschiede zwischen beiden Gruppen auf die ursächliche Wirkung der Maßnahme zurückgeführt. Diese Methoden bilden Bestandteil der quantitativen Wirkungsanalysen, die es im Folgenden genauer zu betrachten gilt.

3. Quantitative Wirkungsanalyse

Quantitative Wirkungsanalysen zielen darauf ab, mit größtmöglicher Zuverlässigkeit zu *messen*, ob eine Intervention die intendierten Wirkungen auslöst und welche nicht-intendierten Wirkungen ebenfalls entstanden sind. Voraussetzung zur Wirkungsmessung ist, dass differenzierte Ursache- Wirkungsbeziehungen aufgestellt werden (Stockmann, 2006) und ein sogenannter *theoriebasierter Ansatz* zur Wirkungsmessung angewendet wird.

Welche Wirkungen im Fokus der Wirkungsmessung stehen, wird im Folgenden näher betrachtet. Die wissenschaftliche Diskussion zur Wirkungsmessung bezeichnet *alle* infolge einer Intervention auftretenden Veränderungen eines Zustands als Wirkungen und subsumiert diese unter dem Begriff „Impact“. Innerhalb des Impacts wird noch zusätzlich der Outcome gekennzeichnet, er umfasst jenen Teilbereich der Wirkungen, der kausal auf die Interventionsmaßnahme zurückzuführen ist. Aktuell zeigen sich zwei Herangehensweisen an die so konzipierten Wirkungen. Das breite Wirkungsverständnis zielt auf das Messen aller auf der Mikro-, Meso- und Makroebene auftretenden Wirkungen, das enge zielt hingegen (nur) auf die Mittel- und Langzeit-Wirkungen auf jenen drei Ebenen.

Im Prinzip handelt es sich bei der quantitativen Wirkungsforschung um ein hypothesentestendes Verfahren, nämlich um die Prüfung der Frage, ob und in welchem Ausmaß durch eine Maßnahme, durch ein Programm oder durch eine Beratung Effekte erzielt werden, bzw. ob die gemessenen Effekte kausal auf die Maßnahme zurückgeführt werden können. Hierfür muss ein explizit theoretisches Modell über die erwarteten Ursache-Wirkungs-Zusammenhänge aufgestellt werden (Stockmann, 2006). Andernfalls hinterlässt der empirische Nachweis von Wirkungen eine "black box" (Caspari & Barbu, 2008; Bloom, 2006; White & Barbu, 2006; Ravallion, 2005). Zu berücksichtigen

sind ebenfalls externe Faktoren, die ein mögliches Risiko für die konstatierten Ursache-Wirkungs-Zusammenhänge darstellen (Caspari & Barbu, 2008). Wirkungshypothesen müssen für jede der drei Ebenen des Wirkungsmodells formuliert werden.

Die Erstellung eines Wirkungsmodells ist eine äußerst komplexe Aufgabe. Die bei der Erstellung eines Wirkungsmodells zugrunde gelegten Hypothesen müssen von den Forschenden rekonstruiert und meist überarbeitet werden (Caspari & Barbu, 2008). Die Methodenliteratur empfiehlt dazu, dass in einem ersten Schritt auf den aktuellen Forschungsstand zu dem betreffenden Bereich zurückgegriffen wird, um in einem weiteren Schritt mit der Zielgruppe und den Stakeholdern über das Wirkungsmodell zu reflektieren, wozu sich Fokusgruppeninterviews anbieten. Ansonsten bleibt die Wahrscheinlichkeit groß, dass indirekte und vor allem nicht intendierte Wirkungen unerkannt bleiben, da dann nur die in dem Wirkungsmodell spezifizierten Wirkungen überprüft werden würden.

Der theoriebasierte Ansatz zur Wirkungsmessung muss auch die Frage berücksichtigen, was ohne Intervention geschehen wäre, d. h. das Kontrafaktische muss bei der Wirkungsmessung berücksichtigt werden (Caspari & Barbu, 2008). Hierfür ist die Verwendung spezieller Datenerhebungsdesigns notwendig, die festlegen, wie, wann, wo, und wie oft Daten zu Wirkungen erfasst werden. Im Folgenden soll auf die wichtigsten Designs eingegangen werden.

Das gewählte Forschungsdesign ist entscheidend für den Grad der Gewissheit, mit dem die Frage nach dem Zusammenhang zwischen Ursache und Wirkung beantwortet werden kann. Hierzu gibt es aus der empirischen Sozial-, Verwaltungs- und Wirtschaftsforschung eine Reihe von Empfehlungen, die hier darzustellen sind (Stockmann, 2006; Wollmann, 2000; Baker, 2000; ADB, 2006).

Zunächst ist festzuhalten: Allen Designs für Wirkungsanalysen ist gemeinsam, dass sie auf Vergleichen beruhen (Stockmann, 2006). Im Idealfall werden zwei Gruppen miteinander verglichen, eine, bei der die geplante Intervention stattfindet (Zielgruppe), und eine zweite, bei der keine Intervention vorgenommen wird. Diese zweite Gruppe heißt *Kontrollgruppe* oder *Vergleichsgruppe*, je nachdem, ob die miteinander zu vergleichenden Gruppen per Zufallszuteilung oder nicht per Zufallszuteilung gebildet werden. Aus dem Vergleich der Ziel- und Kontroll- bzw. Vergleichsgruppe kann dann auf die Wirkung geschlossen werden. Eine Betrachtung nur der Zielgruppe zu nur einem Zeitpunkt, nämlich nach einer Intervention, ermöglicht

keinerlei Aussagen über Veränderungen aufgrund einer Intervention, weil das Kontrafaktische hierbei nicht berücksichtigt wird.

Die Vorgehensweise zur Analyse von Wirkungen ist im Idealfall die sogenannte *double-difference* Methode (Caspari & Barbu, 2008).

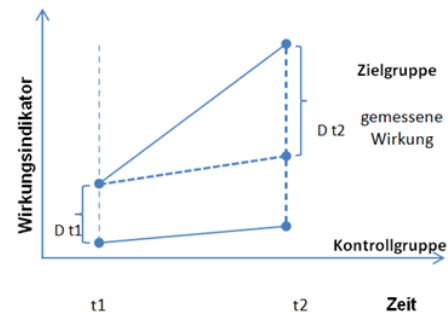


Abbildung 1. Double-difference Methode.

Um optimale Aussagen zu bekommen, soll die Wirkungsmessung nach dieser Methode möglichst kurz vor der Maßnahme zum Zeitpunkt t_1 stattfinden, um den Ausgangswert oder Referenzwert der Wirkung festzustellen, und spätestens kurz vor oder kurz nach Beendigung der Maßnahme zum Zeitpunkt t_2 erfolgen, um die Ex-post Betrachtung von Wirkungen zu ermöglichen. Die Ermittlung der Wirkung besteht aus einem Vergleich der Werte zum Zeitpunkt t_2 mit den Ausgangswerten gemessen in t_1 . Die Wirkung einer Intervention ergibt sich aus dem Unterschied zwischen Zielgruppe und Vergleichsgruppe kurz nach der Maßnahme (D_{t_2}) minus dem Unterschied zwischen Zielgruppe und Vergleichsgruppe kurz vor der Maßnahme (D_{t_1}) (Caspari & Barbu, 2008; Baker, 2000; ADB, 2006).

Die im Folgenden angeführten Datenerhebungsdesigns sind in der quantitativen Wirkungsanalyse dazu bestimmt, mit je verschiedenem Grad der Gewissheit die Ursache-Wirkungs-Zusammenhänge zu belegen.

1. Den Königsweg in der Kleingruppenforschung bilden – wie schon dargestellt – die *experimentellen Designs*. Die Störvariablen werden hier durch Randomisierung ausgeschlossen, in dem die Zuteilung von Personen in Ziel- und Kontrollgruppe per Zufall und vor Beginn einer Intervention erfolgt. Die *double-difference* Methode kann hier uneingeschränkt angewendet werden. Experimentelle Designs sind – so

bleibt hier kritisch anzumerken – zwar ideal zur Überprüfung und Identifikation von Kausalzusammenhängen zwischen den Leistungen einer Intervention (verstanden als unabhängige Variable) und den erfassten Veränderungen (abhängige Variable), doch sind sie im institutionellen Kontext der zu erbringenden Leistungen in Form von Beratungen, Entwicklungs- und Steuerungsmaßnahmen oder auch publikumsbezogenen Veranstaltungen selten anwendbar. Eine Kontrollgruppe müsste bereits im Rahmen der Durchführung einer Intervention per zufallsgesteuertem Auswahlprozess gebildet werden. Dies nachträglich durchzuführen, ist aber technisch unmöglich.

2. Wenn das experimentelle Design aus technischen Gründen nicht oder nicht mehr eingesetzt werden kann, wird häufig das *Quasi-Experiment* durchgeführt. Der Unterschied zum Experiment besteht darin, dass anstelle einer Randomisierung, also einer per Zufallsauswahl erfolgende Aufteilung der Ziel- und Kontrollgruppe, eine statistisch kontrollierte Vergleichsgruppe konstruiert wird. Dies kann auf zwei Wegen erfolgen. Der erste hat die Bezeichnung *Matching on Observables*: So können Personen oder auch Institutionen, die im Hinblick auf bestimmte Merkmalsausprägungen der Zielgruppe gleichen, der Vergleichsgruppe zugeordnet werden. Angewandt werden kann dieses Design auch auf wenige Vergleichspaare. Der zweite Weg hat die Bezeichnung *Propensity Score Matching* und hat die Generierung zahlenmäßig großer Vergleichsgruppen zum Ziel. Ähnlich wie beim ersten Weg werden der Vergleichsgruppe bestimmte Einheiten (Personen, Institutionen) zugeordnet, so dass sich Ziel- und Vergleichsgruppen „im Hinblick auf relevante pre-treatment Merkmale möglichst ähnlich sind“ (Legewie, 2012, S. 132). Die statistische Konstruktion großer Vergleichsgruppe wird auf der Grundlage existierender Census-Daten, repräsentativer Haushaltsbefragungen und sonstiger Datensätze aus Bevölkerungsumfragen möglich. Aufgrund der fehlenden Randomisierung ist jedoch nicht mehr sicher, ob eventuelle Störvariablen neutralisiert werden können.

4. Wenn die Bedingungen für ein Quasi-Experiment nicht gegeben sind, dann kann auf mehreren Wegen versucht werden, die Wirkungen zu erheben.

1. Im „Ex-post-facto-Design“ wird mit retrospektiven Fragen versucht, Informationen über frühere Zeit-

abschnitte zu erhalten. Dies erfolgt in Querschnittsuntersuchungen, in denen alle zu messenden Variablen nur zu einem Zeitpunkt erhoben werden (Baker, 2000; White & Barbu, 2006). Solche Querschnittsuntersuchungen sind das in der Entwicklungspolitik und Organisationsentwicklung am häufigsten vorkommende Untersuchungsdesign. Nach Caspari und Barbu (2008, S. 8) ist die zentrale Grundannahme, dass die Ausgangssituation der Ziel- und Vergleichsgruppe vor der Maßnahme identisch ist. Der gefundene Unterschied zwischen der Ziel- und Vergleichsgruppe wird somit allein der Maßnahme zugeschrieben (Caspari & Barbu, 2008). Diese Zuschreibung erfolgt nach der *single-difference Methode*, also ganz im Gegensatz zur hier erläuterten *double difference Methode*, und hat die Konsequenz, dass die auf Interventionen folgenden Veränderungen über- oder unterbewertet werden, weil die Veränderungen allein der Maßnahme zugeschrieben werden.

2. Ein weiteres Erhebungsdesign ist die Konstruktion von Vergleichsgruppen mittels der „Regression Discontinuity“ Methode (Caspari & Barbu, 2008). Ist z.B. die Teilnahme an einer Maßnahme an eine bestimmte Voraussetzung mit gesetztem Schwellenwert gebunden – d. h. wird z. B. für die Teilnahme an einer Schulungsmaßnahme ein Einkommen von weniger als der Betrag X vorgesehen – kann eine Vergleichsgruppe konstruiert werden. Der leitende Gedanke ist, dass als Vergleichsgruppe diejenigen ausgewählt werden können, die den Schwellenwert nur knapp nicht erreicht haben, somit aber ähnliche Charakteristika wie die Teilnehmenden aufweisen (Baker, 2000; Bamberger, 2006). Bei Interventionen, wo die Zulassungsvoraussetzungen überprüft werden, liegen Daten zum Zeitpunkt t_1 für Personen oder Institutionen vor, die letztendlich in die Maßnahme aufgenommen wurden, es liegen aber auch Daten für solche vor, die abgelehnt wurden, weil sie zum Zeitpunkt t_1 den Schwellenwert leicht unter- bzw. überschritten hatten.

3. Vergleichsgruppen werden auch ohne vorausgehende Konstruktion verwendet, wenn Interventionseffekte bei der Zielgruppe mit typischen Veränderungen in der Gesamtpopulation verglichen werden. Grundlage für den Vergleich bildet die Methode der „generischen Kontrolle“: Kennwerte wie z.B. Sterbe- und Fruchtbarkeitsziffern, Krankenstände und weitere Indikatoren zur Charakterisierung der Erwerbsbevölkerung werden herangezogen, um abzuschätzen, was sich ohne die Intervention (z.B. betriebliche Ge-

sundheitsvorsorge) ereignet hätte. Unterschiede zwischen den Messwerten einer großen Population und einer kleineren Zielgruppe werden als Wirkung der Intervention zugeschrieben.

4. Besonders häufig kommt in Wirkungsanalysen der einfache Vorher-Nachher-Vergleich zur Anwendung. Er wird auch „Pretest-Posttest-Design“ genannt, weil der Vergleich nur innerhalb der Zielgruppe stattfindet. Dabei werden die zum ersten Zeitpunkt gewonnenen Daten wie Daten aus einer Vorher-Messung verwendet. Falls die Intervention noch nicht allzu lange läuft, werden die Veränderungen zu mehreren Zeiten nach der Intervention gemessen, was in der Form von Längsschnittstudien erfolgt. Die Differenz der Messwerte soll Aufschluss über die Wirkungen einer Intervention geben. Dabei wird davon ausgegangen, dass die Messwerte bei Pretest und Posttest gleich ausgefallen wären, wenn es keine Intervention gegeben hätte. Kritisch bleibt hier festzustellen, dass andere, externe Faktoren wie z.B. Wirtschaftskrisen, Krankheitsfälle oder auch allgemeine Veränderungsprozesse die Wirkung einer Intervention schwächen, verstärken oder neutralisieren können. Solche Faktoren, die teilweise oder ganz für die gemessene Veränderung mit ausschlaggebend sein können, bleiben bei einem reinen Vorher-Nachher-Vergleich der Zielgruppe unberücksichtigt (Caspari & Barbu, 2008). Das Kontrafaktische kann hier nicht berücksichtigt werden, weil keine Kontroll- oder Vergleichsgruppe gegeben ist. Für Wirkungsmessungen ist dieses Erhebungsdesign daher wenig geeignet. Es kann lediglich die Veränderungen der Zielgruppe in Zeitreihen erheben, weshalb der Grad der Gewissheit, mit dem die Frage nach dem Zusammenhang zwischen Ursache und Wirkung beantwortet werden kann, sehr gering ist (Caspari & Barbu, 2008).

5. Eine Variante des einfachen Vorher-Nachher-Vergleichs bildet die Panelanalyse. Hier werden mehrere Messungen bei denselben Personen oder denselben Organisationen zu verschiedenen Zeitpunkten durchgeführt, um innerhalb der einzelnen Intervalle Vergleiche vornehmen und schließlich Vergleichsreihen bilden zu können. Der solche Vergleichsreihen leitende Gedanke besteht vor allem darin, dass die als unabhängig und abhängig angenommenen Variablen in verschiedenen Zeitabständen aufeinander bezogen werden können. Trotz dieser Vergleichsreihen mit variierbaren Zeitabständen bleibt auch hier

kritisch festzuhalten, dass das Kontrafaktische damit noch nicht berücksichtigt werden kann.

Aus der Kritik an der quantitativen Wirkungsforschung haben sich neue, qualitative Analyseansätze entwickelt. Nicht nur wurde bezweifelt, dass die quasi-experimentellen Designs die Komplexität jener Umfelder, in welchen die Interventionen durchgeführt werden, würdigen können. Auch im Hinblick auf die anderen hier dargestellten Designs (siehe oben die Punkte 3.1 bis 3.5) wurde kritisch festgestellt, dass die Ergebnisse der einschlägigen Studien bestenfalls nur die Frage beantworten können, ob und in welchem Ausmaß eine Intervention Wirkungen zeitigen kann. Ihr Informationsgehalt erweist sich bis heute gegenüber derselben Kritik als sehr beschränkt, insofern als die Frage, wie diese Wirkungen erzeugt worden sind, unbeantwortet bleibt.

Ausgangspunkt der qualitativen Wirkungsforschung ist nicht die Frage, wie möglichst viele "Störvariablen" mittels eines Forschungsdesign ausgeschlossen oder kontrolliert werden können, sondern vielmehr die Frage, wie möglichst viele Bedingungen in ihren Wechselwirkungen mit der Intervention in die Wirkungsanalyse einbezogen werden können.

5. Qualitative Ansätze der Wirkungsanalyse

Qualitative Ansätze der Wirkungsanalyse richten ihr Hauptaugenmerk sowohl auf die subjektiven Interpretationen der an Interventionen beteiligten Akteure als auch deren Interaktionsprozesse. Ihr Anspruch ist es, den Prozesscharakter der auf Interventionen folgenden Veränderungen sowohl bei der Zielgruppe als auch bei den Durchführenden zu rekonstruieren, wobei die subjektiven Interpretationen, die das Handeln der Beteiligten leiten, sichtbar und verstehbar gemacht werden sollen. Anders als in der quantitativen Wirkungsanalyse wird die Zielgruppe nicht als quasi passives Objekt betrachtet, vielmehr wird mit hermeneutischen Modellen operiert, in denen die Akteure ihre je eigenen Vorstellungen von Ursache-Wirkungsnetzen haben und mit ihrem Handeln aktiv in diese eingreifen. Die Stärke qualitativer Verfahren besteht in der Wirkungsforschung auch darin, dass sie bislang unbekannte und nicht antizipierte Wirkungen von Interventionen in den Blick nehmen können (Kelle, 2006; Caspari & Barbu, 2008; Groenemeyer & Schmidt, 2011).

Die Stärke qualitativer Verfahren kann auf drei Ebenen zum Tragen kommen, wie im Folgenden zu zeigen ist.

1. Qualitative Methoden können dazu eingesetzt werden, die *outcomes* von Interventionen aus der Sicht der Zielgruppe zu identifizieren. Wie Udo Kelle (2006) aufgezeigt hat, setzt die quantitative Kausalanalyse oft ein, ohne im Voraus hinreichend zu klären, „welche sinnvolle Indikatoren für den Erfolg einer Intervention sein können“ (Kelle, 2006, S. 127). Die routinemäßige Verwendung standardisierter Erhebungsinstrumente birgt das Risiko, dass die gemessenen Variablen zu allgemein sind, um die Veränderungen, die auf eine Intervention folgen, in ihrer Aspekthaltigkeit für die Zielgruppe erfassen zu können. So kann z.B. ein Modellprojekt zur Einführung eines Betreuungssystems in der häuslichen Altenpflege als nur mäßig erfolgreich evaluiert werden, wenn der *outcome* in Form einer standardisierten Messung der Lebenszufriedenheit erhoben wird. Anstatt auf vorhandene standardisierte Messinstrumente zurückzugreifen, müsste hier in einem ersten Schritt in offenen, qualitativen Interviews mit verschiedenen Gruppen von Beteiligten – also mit Angehörigen, ÄrztInnen, Pflegekräften, SozialarbeiterInnen – dialogisch geklärt werden, was als *outcome* bzw. Erfolg der Intervention aus der Sicht von Betroffenen und deren Bezugspersonen gelten kann. Auf diesen „Gesprächscharakter“ einer qualitativen Wirkungsanalyse hat auch Ralf Bohnsack (2010, S. 45) aufmerksam gemacht, wenn er den Nutzen der kommunikativen Einbindung sowohl der Zielgruppe wie auch der relevanten Stakeholder darin herausstellt, dass sich die Chance eröffnet, deren analytische und interpretative Kompetenzen im Hinblick auf die intendierten und nicht intendierten Veränderungen zu identifizieren und „zur Artikulation zu bringen“ (Bohnsack, 2010, S. 50). In diesem Vorgehen soll (und kann) sich das Verhältnis zur Praxis verändern: Anstelle der Gegenüberstellung von Wissenschaft und Praxis soll die Anerkennung der in der Praxis implizierten „Klugheit“ (Bohnsack, 2010, S. 56) treten.
2. Qualitative Methoden können auch dazu eingesetzt werden, um die Wirkungsweise und -verläufe bestimmter Maßnahmen zu beschreiben (Kelle, 2006). Wie und durch welche Pfade bestimmte Interventionen wirken, kann auf der Mikro-, Meso- und sogar Makroebene beschrieben werden. So lassen sich jene Effekte einer Intervention empirisch identifizieren und in ihrem Verlauf beschreiben, über die es zu Beginn der Intervention keine oder nur ungenaue Vorstellungen gab. Vor allem ist die Identifizierung unbeabsichtigter Nebenwirkungen unerlässlich, weil diese die geplanten Effekte überlagern und neutralisieren können oder sogar die Gestaltungsabsicht in Gegenteiliges verwandeln. Gezeigt werden kann dies z.B. anhand von Beschreibungen sogenannter Verlaufskurven, in denen Professionelle ihre Beziehungen zu Mitgliedern ihrer Zielgruppe gestalten, indem sie ihr Wissen über relevante Einflussgrößen zunächst in routinemäßiges Handeln übersetzen, nach und nach aber ihre Aufmerksamkeitsfelder an konkrete Situationen anpassen und in mehrfachen Interventionsschleifen eine Rekursivität entwickeln, aus der ein Interaktionsmodus mit nachfolgenden Veränderungen sowohl in der Zielgruppe als auch in ihnen selbst entsteht (Brüsemeister, 2003). Anhand von Verlaufskurven können jene „kausalen Pfade“ aufgezeigt werden, durch die eine bestimmte Intervention erst wirken kann. Kausale Pfade zu erkennen, wird umso wichtiger, desto weniger die Technologisierbarkeit von Interventionen gegeben ist.
3. Qualitative Methoden können dazu eingesetzt werden, die Mitwirkung von Seiten der Zielgruppe als besondere Wirkungsweise in je besonderen kausalen Pfaden zu identifizieren (Compliance). Ob und wie eine Intervention in der Beziehungsarbeit nämlich wirkt, hängt von Prozessen ab, die auch von der Zielgruppe mitbestimmt werden. Compliance ist eine besondere, nicht nur in verschiedenen Beratungsformaten vorkommende Wirkungsweise, sie ist auch in der Erbringung sozialer Dienstleistungen und in der Implementierung von Technologien eine notwendige Voraussetzung dafür, dass die mit der Intervention verfolgte Wirkungsentention sich als erfolgreich zeigt. Mit der Compliance eröffnen sich somit zwei weitere kausale Pfade. Zum einen kann die Zielgruppe auch Koproduzentin der mit der Intervention intendierten Wirkungen werden. Zum anderen können in Fällen, wo das gemeinsame Produzieren nicht ohne weiteres erfolgt, die Abwehrhaltungen und entsprechenden Handlungen identifiziert werden. Diesen zweiten Wirkungspfad hat schon August Aichhorn in den 1920er Jahren aufgezeigt, indem

er belegte, dass selbst in der Verwehrlosung lebende Hilfsbedürftige Abwehrhaltungen mitbringen und dann mühsam zur Annahme von Hilfe und ihrer Umsetzung angeleitet werden müssen (Aichhorn, 2011). Dieser Wirkungspfad ist in der Psychoanalyse und insbesondere von Udo Rauchfleisch weiter untersucht worden. Wie er gezeitigt hat, gehören zum Spektrum möglicher Abwehrhaltungen und daraus folgender Handlungen erstens die bis zur Machtdemonstration reichende und im fehlenden Urvertrauen gründende Entwertung und Ablehnung von Hilfe, zweitens der Abbruch der eigenen Mitwirkung auf Grund eines narzisstischen Beziehungsmusters und drittens das angstvolle Ausweichen vor Zuwendung und Unterstützung als Folge eines Festklammerns an der eigenen negativen Identität (Rauchfleisch, 1996). Eine ergänzende, in der Sozialpädagogik entwickelte Heuristik betrifft die positiven Orientierungsmuster gegenüber psychosozialen Dienstleistungen, womit der andere Wirkungspfad ins Sichtfeld kommt. Wie Klaus Mollenhauer exemplarisch nachgewiesen hat, gehören zum Spektrum positiver Orientierungsmuster vier Motive samt daraus folgender Verhaltensweisen in der Annahme und Umsetzung von Hilfe. Dazu zählen erstens die mit Passivität vereinte Fürsorgeerwartung, in der die Bereitschaft zur Übernahme von Verantwortung für den Erfolg der Hilfe gering ist; zweitens die den Gegenpol dazu bildende egoistische, nach individuellem Nutzen strebende Orientierung an Hilfe, die das Eingehen von Verpflichtungen ablehnt; drittens die durch das Ziel der Selbsthilfe bestimmte Leistungsorientierung, in der Verantwortung für den Erfolg von Hilfe übernommen wird. Zu diesem Spektrum der für die Umsetzung von Hilfe möglichen Orientierungsmuster gehört viertens das Motiv, an der Verwirklichung eines moralischen, nicht schon durch den eigenen Nutzen erfüllten Zwecks mitzuwirken, was z.B. dadurch erfolgen kann, dass die eigene Mitwirkung am intendierten Ergebnis der Einzelfallhilfe als Beitrag zur Menschenwürde verstanden wird, dem Zweck des Kindeswohls verpflichtet ist, der Gottwohlgefälligkeit dienen soll, im Zeichen der Generationengerechtigkeit verstanden wird, als Beitrag zur Lebensqualität im eigenen Wohnumfeld geleistet oder auch als Chance zur Stärkung der Eigenverantwortung interpretiert wird.

6. Exkurs: Exemplarische Wirkungsanalysen

Die hier behandelten Forschungsdesigns kommen in Reinform oder Mischformen zur Anwendung. Ihre Anwendung soll im Folgenden in drei Studien identifiziert werden, wobei es darauf ankommt, kritisch die zuvor behandelten Stärken und Schwächen aufzuzeigen.

6.1. Langfristige Wirkungsmessungen von AMS-Maßnahmen – Ein Beispiel für die Kombination quantitativer Forschungsdesigns

Bekanntlich sollen die Qualifizierungsmaßnahmen des Arbeitsmarktservice die Anstrengungen von arbeitslosen Personen darin unterstützen, ihre Beschäftigungsintegration wieder herzustellen. Das Arbeitsmarktservice überprüft regelmäßig, ob die von ihm gesetzten Qualifizierungsmaßnahmen die intendierte Wirkung entfalten. In der vorliegenden Studie „Die langfristige Wirkung von AMS-Maßnahmen“ (AMS, 2009) soll sowohl die kurzfristige als auch langfristige Wirkung (5 Jahre nach der Maßnahme) erhoben und gemessen werden. Zu diesem Zweck werden als Zielgruppe ehemalige KlientInnen ausgewählt, die im Jahr 2003 an Qualifizierungsmaßnahmen teilgenommen haben.

Die Erstellung der statistischen Gruppe, an der die Wirkungen festgestellt werden sollen, soll zum einen Vorher-Nachher-Vergleiche und zum anderen Vergleiche mit der erwerbstätigen Bevölkerung ermöglichen. Es handelt sich also hier um eine Kombination von zwei Designs, des „Pretest-Posttest-Design“ (s. oben das quantitative Design Nr. 4.4) und des *Propensity Score Matching* (s. oben das quantitative Design Nr. 2).

Damit die Wirkungsmessung auch mittels Vergleich mit der erwerbstätigen Bevölkerung im Fünfjahreszeitraum möglich wird, muss, – so die Studie – die statistische Gruppe auf jenen Personenkreis beschränkt werden, für den „sinnvollerweise eine solche Analyse vorgenommen werden kann“ (AMS, 2009, S. 11). In diesem Sinn sind von den 116.200 Mitgliedern der Zielgruppe – also den ehemaligen TeilnehmerInnen an den Maßnahmen – folgende Personenkreise nicht weiter untersucht bzw. aus den Vergleichen ausgeschlossen worden, wobei die von mir mit „!“ markierten Gruppen – so sei hier kritisch angemerkt – die Erfolgsquoten erheblich verschlechtert hätten:

- Personen, die in dem Fünfjahreszeitraum verstorben sind (!)

- Personen, die innerhalb des Fünfjahreszeitraums in die Erwerbspension übergetreten sind (!);
- Personen, die nicht in jedem einzelnen Jahr sozialversicherungsrechtlich dokumentiert sind (!)
- Frauen mit einer Geburt im Fünfjahreszeitraum (!);
- Personen außerhalb des Haupterwerbsalters (jünger als 25 und älter als 44) (!)
- Personen, die im Fünfjahreszeitraum an weiteren AMS-Maßnahmen teilgenommen haben, damit der Effekt der Maßnahmen des Jahres 2003 bestimmt werden kann;

Die Wirkungsanalyse der AMS-Qualifizierungsmaßnahmen konzentriert sich somit auf 18.780 Personen (43% Frauen, 57% Männer).

Die Wirkung wird zunächst im Hinblick auf vier Dimensionen der Arbeitsmarktintegration gemessen: die Beschäftigungs- und Erwerbstätigkeitstage pro Jahr und die Bruttobeschäftigungseinkommen, unterschieden nach Jahreseinkommen und Stundeneinkommen. Die Indikatoren für die Langzeitwirkung bestehen aus Quoten in jeder dieser vier Dimensionen. Die Quoten erfassen zum einen, wie viel Prozent der ehemaligen TeilnehmerInnen fünf Jahre nach der Maßnahme mehr als die Hälfte der vollen Erwerbstätigkeitstage und Beschäftigungstage nachweisen, und sie erfassen zum anderen, wie viel Prozent der ehemaligen TeilnehmerInnen fünf Jahre nach der Maßnahme ein bestimmtes Bruttojahresbeschäftigungseinkommen und Bruttotageseinkommen überschreiten (s. Abbildung 2).

»Nachher«: langfristige Perspektive - Arbeitsmarktpositionierung
(Maßnahmenteilnehmender Personenkreis »langfristige Wirkung«¹)
Im Jahr 2008, fünf Jahre nach der Maßnahmenteilnahme

	Frauen	Männer
Personenkreis »langfristige Wirkung«¹		
Erwerbsaktivitätstage²		
0 Tage	0,0%	0,0%
1-182 Tage	1,6%	2,1%
183-365 Tage	98,4%	97,9%
Beschäftigungstage³		
0 Tage	2,4%	3,5%
1-182 Tage	3,5%	5,0%
183-365 Tage	94,1%	91,6%
Jahresbeschäftigungseinkommen⁴ in EUR		
0,-	2,2%	3,2%
1,- bis 4.000,-	3,8%	3,5%
4001,- bis 8.000,-	9,0%	6,2%
Mehr als 8.000,-	85,0%	87,2%
Tageseinkommen⁵ in EUR		
0,-	2,5%	3,6%
1,- bis 30,-	16,5%	8,8%
31,- bis 60,-	44,7%	18,8%
Mehr als 60,-	36,4%	68,9%

Abbildung 2. Wirkungsdimensionen und Messeinheiten für die Arbeitsmarktintegration (Quelle: AMS, 2009; S. 27).

Die Wirkungen werden auf diese Weise Jahr für Jahr gemessen. Dabei wird festgestellt, ob sich die vierdimensionale Arbeitsmarktintegration nach dem zweiten, dritten und weiteren Jahr erheblich verbessert hat.¹ Die Langzeitwirkungen werden auf einer

weiteren Betrachtungsebene gemessen. Während die erste Wirkungsmessung keine Vergleichsgruppe heranzieht, geht die zweite Messung gemäß dem *Propensity Score Matching* vor. Die Vergleichsgruppe besteht aus allen erwerbstätigen Personen in Österreich in demselben Jahr, in dem die Maßnahme gesetzt wurde. Der Vergleichsmaßstab wird aus den gleichen vier Dimensionen der Arbeitsmarktintegration

¹ Ein kritischer Vergleich der Ergebnisse meinerseits ergibt – so sei hier angemerkt –, dass sich die Arbeitsmarktpositionierung im fünften Jahr nicht mehr erheblich gegenüber dem ersten Jahr verändert hat (vgl. ebd., S. 26-27 die Tabellen S. 4 und 5).

gewonnen, die beim internen Vorher-Nachher-Vergleich angewandt werden (s. Abbildung 3).

Die Wirkung wird mittels eines Vergleichs der Medianwerte gemessen, welche die zwei Gruppen in den vier Dimensionen der Arbeitsmarktintegration aufweisen. Bei den ehemaligen SchulungsteilnehmerInnen wird ermittelt, wie hoch die Medianwerte im 5. Jahr nach Besuch der Maßnahme in jeder der vier Dimensionen der Arbeitsmarktpositionierung sind, um sie mit den Medianwerten aller Erwerbstätigen zu vergleichen, die diese ein Jahr vor Beginn der Schulungs-

maßnahme aufwiesen. Was hier nicht mehr begründet wird und in methodischer sowie legitimatorischer Hinsicht problematisch ist, ist die Relationierung der Zeiträume 2002 und 2008. So stellt sich die Langzeitwirkung der AMS-Maßnahmen darin dar, dass die vier Medianwerte, welche die Gesamtheit der damals Erwerbstätigen im Jahr 2002 aufweist, zu 65 bis 89 % von der wieder beschäftigten Zielgruppe im Jahr 2008 erreicht werden!

Maßnahmenteilnehmender Personenkreis »langfristige Wirkung«¹ und ihre relative Positionierung zu allen erwerbsaktiven Personen in Österreich
Jahreswerte 2002 (Jahr vor der Maßnahmenteilnahme)

	<i>Alle erwerbsaktiven Personen im Haupterwerbsalter³</i>	<i>Maßnahmenteilnehmender Personenkreis (»langfristige Wirkung«)</i>	<i>Maßnahmenteilnehmende in % von allen</i>
Alle Personen			
Erwerbsaktivitätstage ²	365	365	100,0%
Beschäftigungstage ²	365	259	71,0%
Jahresbeschäftigungseinkommen ²	18.482,-	12.032,-	65,1%
Tageeinkommen ²	56,-	50,-	88,9%
Alle Frauen			
Erwerbsaktivitätstage ²	365	365	100,0%
Beschäftigungstage ²	365	246	67,4%
Jahresbeschäftigungseinkommen ²	13.061,-	8.388,-	64,2%
Tageeinkommen ²	41,-	35,-	85,7%
Alle Männer			
Erwerbsaktivitätstage ²	365	365	100,0%
Beschäftigungstage ²	365	266	72,9%
Jahresbeschäftigungseinkommen ²	23.632,-	15.694,-	66,4%
Tageeinkommen ²	69,-	60,-	87,8%

Abbildung 3 Wirkungsmessung nach dem Design des Propensity Score Matching (Quelle: AMS, 2009, S. 24).

6.2. Langfristige Wirkungsmessungen von Substitutionstherapien – Ein Beispiel für das „Pretest-Posttest-Design“

Wie die Wirkungen von langfristigen Substitutionstherapien gemessen werden, soll hier am Beispiel einer aufwändigen Studie mit der Kurzbezeichnung PREMOS verdeutlicht werden (Wittichen, Bühringer & Rehm, 2010). Die im Auftrag des deutschen Bundesministeriums für Gesundheit durchgeführte PREMOS-Studie hatte das Ziel, Langzeitwirkungen bei einer großen Population opioidabhängiger PatientInnen aufzuzeigen, die über einen Zeitraum von sechs Jahren in Substitutionseinrichtungen behandelt werden.

Die Zielgruppe bildet hier eine nach Einrichtungstyp geschichtete Zufallsstichprobe. Die Stichprobe wurde geschichtet, um eine ausreichend hohe Zahl von PatientInnen verschiedener Einrichtungen zu erhalten. Die Zufallsstichprobe wurde aus 86 Hausarztordinationen, 101 mittleren Einrichtungen und 36 großen Substitutionszentren gezogen. Ihre anfängliche Größe von rund 2700 Patienten und Patientinnen musste im Laufe der Jahre naturgemäß reduziert werden, weil ein Teil der Ärzte und Ärztinnen aus verschiedenen Gründen ihre Ordinationen oder die Substitutionsbehandlung aufgaben. Auch hier stellte sich somit das Problem, wie mit den Patienten und Patientinnen verfahren werden sollte, die aus der anfänglichen Stichprobe ausgeschieden sind. Ähnlich

wie bei der oben dargestellten AMS-Wirkungsstudie meinte die Forschergruppe im Verlaufe dieser Ausfälle, dass sie die statistische Gruppe, an der die Wirkungen gemessen werden sollten, zusätzlich reduzieren müsse. Herausgenommen aus der Langzeitanalyse wurde schließlich auch ein Teil der PatientInnen, der in den betrachteten Zeitintervallen aus disziplinarischen Gründen, meist wegen erheblichen Beigebrauchs und längerer Inhaftierung, von der Behandlung ausgeschlossen werden musste!

Die aufwändige und an Exaktheit höchst anspruchsvolle Erhebung erfolgte in vier Messzeitpunkten von den Jahren 2004 bis 2009. Um die längerfristigen Effekte zu messen, wurden u.a. die folgenden *Outcome*-Variablen bestimmt: Erfolg oder Misserfolg einer langfristigen Substitution sollte anhand von Indikatorenmesswerten für „Haltequote“, „Unterbrechungen“, „stabile Abstinenz“, „Beikonsum“ und „Erwerbstätigkeit“ bestimmt werden. Alle aufgefundenen PatientInnen wurden standardisiert befragt und zusätzlich wurden medizinische Befunde wie standardisierte Urin-Drogenscreenings und ärztliche Abstinenzassessments ausgewertet.

Die Ergebnisse, zu welchen die Wirkungsanalyse anhand von Vorher-Nachher-Vergleichen gelangt, sind zwar außergewöhnlich aufwändig, aber nur mittels der *single-difference* Methode erzielt worden. Einige dieser Ergebnisse seien hier in aller Kürze angeführt, weil ihre Konsequenzträchtigkeit auch in gesundheitspolitischer Hinsicht offensichtlich wird:

1. Haltequote: Für die Mehrzahl aller PatientInnen ergibt sich hier – wie die Forschergruppe herausstellt – „ein positives Bild“: 70% aller (in der Stichprobe verbleibenden!) PatientInnen konnten in dem 6-jährigen Beobachtungszeitraum in der Substitution gehalten werden.

2. Unterbrechungen: Für den gesamten Zeitraum zeigt sich, dass eine vollständig "stabile" Substitution, d. h. eine Substitution ohne jegliche Unterbrechung und ohne bedeutsame Veränderungen von Dosierung, nur für 30% der (in der Stichprobe verbleibenden!) PatientInnen erreicht wurde. Typisch für den Verlauf der meisten (in der Stichprobe verbleibenden!) PatientInnen sind zumindest einmalige, oft aber auch mehrmalige Unterbrechungen bzw. die Wiederaufnahme der Substitution bei dem/der gleichen oder einem/einer anderen Arzt/Ärztin.

3. Hinsichtlich der Abstinenz zeigt sich, dass die meisten Fälle als temporär erfolgreiche Abstinenz einzuordnen sind. 85% der zum ersten Messzeitpunkt als abstinent klassifizierten PatientInnen waren beim letzten Messzeitpunkt, also 5 Jahre danach, wieder in Substitution.

4. Beikonsum: Es ergibt sich aufgrund der Urinproben ein gewisser Rückgang nahezu aller konkomitant gebrauchten Substanzen. Aller Beikonsum reduziert sich von rund 59% auf rund 41% (in der verbleibenden Stichprobe!). Diesen geringen Rückgang interpretiert die Forschergruppe dahingehend, dass ein nicht unerheblicher Teil der PatientInnen Beikonsum betrieben habe, um die möglicherweise unzureichende Substitutionsdosis zu kompensieren.

5. Als positiv wird die Verbesserung der beruflichen Situation in dem 6-jährigen Beobachtungszeitraum bewertet, indem zwei Indikatoren miteinander verglichen werden: So stieg der Anteil der berufstätigen PatientInnen von 24% auf 34% (in der verbleibenden Stichprobe!) und der Anteil derer, die in berufsqualifizierenden Maßnahmen waren, erhöhte sich von 7,5% auf 19% (in der verbleibenden Stichprobe!).

Auch an dieser Studie ist kritisch festzustellen, dass das Design dieser aufwändigen Wirkungsanalyse es nicht leisten kann, das Kontrafaktische zu berücksichtigen, also die Frage zu beantworten, was ohne Substitutionsmaßnahme geschehen wäre. Womöglich handelt es sich hier um eine Zielgruppe, zu der sich keine Vergleichsgruppe bilden lässt oder es sich verbietet, eine solche für einen längeren Zeitraum zu bilden.

6.3. Identifikation sowohl von outcomes als auch von kausalen Pfaden im Coaching – eine Masterarbeit als Beispiel für die qualitative Wirkungsanalyse

Wie qualitative Verfahren in der Wirkungsanalyse eingesetzt werden können, wird auch in Masterarbeiten erprobt, die bei der ARGE Bildungsmanagement betreut werden. Eine davon soll hier beispielhaft für die zum Tragen kommenden Stärken vorgestellt werden.

Um die Veränderungen durch Coaching zu erheben, werden in der Masterthesis von Matthias Wilhelm (2014) ausgewählte Coachees nach den kurz- und langfristigen Auswirkungen sowohl im beruflichen

Bereich als auch im privaten Bereich befragt. Erhoben werden zum einen die Veränderungen als Effekte, also die outcomes, die dem Coaching subjektiv attribuiert wurden, – erhoben werden zum anderen die Prozesse, durch welche aus der Sicht der Coachees diese Effekte entstanden sind. Die Besonderheit des Samplings besteht darin, dass alle Interviewten nach Abschluss des Coachings beim gleichen Institut eine Ausbildung zum Coach absolviert haben und somit in den Wahrnehmungen der Prozesse, aus denen die Effekte hervorgegangen sein könnten, gleichermaßen geschult sind. Um die Offenheit für eine Varietät von Wirkungspfaden sicherzustellen, sind die Interviewpersonen aus unterschiedlichen Branchen gewählt worden – aus dem Ingenieurwesen, Sozialwesen, Telekommunikation, der Elektronik, Bildung und Akustik.

Effekte und Wirkungspfade werden mittels eines Leitfadenterviews erhoben. Der Leitfaden ist so aufgebaut, dass einerseits die positiven Veränderungen und ihre jeweiligen Attribuierungen identifiziert werden können und andererseits die negativen Veränderungen und ausbleibenden Veränderungen in den Blick kommen und ihre jeweiligen Attribuierungen erfasst werden können.

Die Ergebnisse dieser Masterarbeit bestätigen den besonderen Erkenntniswert qualitativer Wirkungsanalysen und seien hier in Kürze angeführt.

- Ein langfristiger Effekt wird in der Erhöhung der Selbstreflexion festgestellt. Ein weiterer Langzeiteffekt wird darin wahrgenommen, dass sich der eigene Blickwinkel sowohl im beruflichen als auch im privaten Bereich verändert hat. Weitere Langzeiteffekte werden in der Erweiterung der Handlungsspielräume identifiziert, ferner im gesteigerten Selbstbewusstsein, in der Ausgeglichenheit und im planvollen Erarbeiten Können von Zielen.
- Auch negative unbeabsichtigte Effekte können festgestellt werden. Sie sind hauptsächlich kurzfristig und können in zwei Kategorien aufgeteilt werden. Zum einen handelt es sich um emotionale Belastungen als Folge von Coaching, die beispielsweise wie folgt beschrieben werden: *„Und da hat es auch negative Auswirkungen geben, dass ich dann mehr Streit oder Diskussionen gehabt habe, was natürlich für mich negativ war in dem Sinne, wer streitet schon gerne mit seinem Umfeld“* (zit. nach Wilhelmy, 2014, S. 69). Bei den negativen Effekten handelt es sich zum anderen um

körperliche Belastungen, die man nach Abschluss der Coaching-Maßnahme selbst erzeugen kann. Typisch dafür sind Wahrnehmungen wie die folgende: *„Da habe ich komplette Zusammenbrüche gehabt. Also ich weiß nicht, ob man Nervenzusammenbruch dazu sagen kann. Die waren zwar die Konsequenzen aus Einzelcoachings, ist aber nicht in der Coaching-Session passiert.“* (zit. nach Wilhelmy, 2014, S. 69)

- Nach den Faktoren und Prozessen befragt, welche die positiven Effekte herbeigeführt haben, geben die Interviewpersonen die fachliche Kompetenz des Coaches an. Der geschützte Rahmen, in dem das Coaching stattfindet, wird als weiterer Faktor für die positiven Effekte identifiziert. Als entscheidend für die Erzeugung positiver Effekte werden Vertrauen und Zuhören Können festgestellt, was in folgender Aussage wie der folgenden formuliert wird: *„Es hat schon viel mit Vertrauen und Zuhören zu tun oder mit der Wahrnehmung, dass man jemandem gegenüber sitzt, der sich wirklich darauf einlassen möchte“* (zit. nach Wilhelmy, 2014, S. 72).
- Neben den Faktoren, die förderlich für die positiven Effekte sind, können auch Faktoren für die negativen ermittelt werden. An erster Stelle werden fachliche Fehler dafür verantwortlich gemacht, dass es zu negativen Effekten kommt. Fachliche Fehler werden im Erteilen von Ratschlägen identifiziert, ebenso in der zeitlichen Planung und in den falschen Erwartungen der Coachees bezüglich des Coachings.

Die Vorteile der qualitativen Wirkungsanalyse werden hier exemplarisch deutlich. Ihre Stärke besteht zum einen darin, dass sie bislang unbekannte und nicht antizipierte Wirkungen von Interventionen in den Blick nehmen können. Zum anderen kann mit Hilfe qualitativer Methoden beschrieben werden, wie und durch welche kausale Pfade bestimmte Maßnahmen wirken.

7. Schlussbetrachtung

Wie die bisherigen Ausführungen gezeigt haben, sind Wirkungsanalysen Attributionen identifizierter Veränderungen zu Maßnahmen. Zur ersten Identifizierung möglicher Wirkungen müssen sowohl quantitative und theoriebasierte Ansätze als auch qualitative Ansätze angewendet werden. Hinzu kommen auch partizipati-

ve Ansätze, welche die Perspektive der Nutzer und Nutzerinnen in die Feststellung möglicher Wirkungspfade einbezieht. Eine Kombination solcher Ansätze ist Voraussetzung dafür, dass keine „black box“ zwischen Maßnahmen und Wirkungen übrig bleibt.

Dabei ist darauf zu achten, dass neben den Daten zu Veränderungen bei der Ziel- und Vergleichsgruppe auch Daten zu den Interventionen selbst und ihres Kontextes erhoben werden müssen. Insbesondere bei Beratungsleistungen sollten für die Komponenten ihrer Aktivitäten sinnvolle, also dem Gütekriterium der Validität entsprechende, Indikatoren gebildet und entsprechende Daten erhoben werden. Dies ist im Kontext von Beratungsleistungen deutlich schwieriger zu realisieren und bedarf eines tiefergehenden Verständnisses der sozialen Prozesse, die Wirkungen erst möglich machen oder verhindern können. Zur Untersuchung dieser Aspekte können qualitative Ansätze im besonderen Maße beitragen und quantitative Verfahren sinnvoll ergänzen.

Qualitative Methoden können dabei angewandt werden, um Wirkungshypothesen zu erstellen und Wirkungspfade vertiefend zu untersuchen. Qualitative Methoden können im Rahmen der Wirkungsanalyse auch die Kausalitätsfrage aus der Perspektive der Zielgruppe beantworten, indem sie die Interpretationen der gefundenen Ergebnisse seitens der Adressaten berücksichtigt und dabei auch mögliche nicht-intendierte Wirkungen identifiziert.

Insbesondere bleibt als Desiderat für weitere Forschungen und damit auch für empirisch ausgerichtete Masterarbeiten eine Wirkungsdimension, welche die Durchführenden selbst und den institutionellen Rahmen ihrer Maßnahmen betrifft. Neben den externen Wirkungen, welche bei den Zielgruppen hervorgerufen werden, sollten auch die internen, bei der Durchführungsorganisation und den Durchführenden ausgelösten Wirkungen in den Blick genommen werden. Das Entstehen von internen Wirkungen beim Durchführen von Maßnahmen verdient mehr Beachtung.

Literatur

Aichhorn, Th. (2011). „...nicht Anwalt der Gesellschaft, sondern Anwalt der Verwehrlosten zu sein!“ Beiträge zur Biographie und zum Werk August Aichhorns. In Th. Aichhorn (Hg.), *August Aichhorn. Pionier der psychoanalytischen Sozialarbeit* (7–54). Wien: Löcker.

AMS (2009). *Die langfristige Wirkung von AMS-Maßnahmen*. Erstellt im Auftrag des AMS. Für den Inhalt verantwortlich Synthesis Forschung Gesellschaft m.b.H. Wien: Arbeitsmarktservice Österreich.

Asian Development Bank (2006). *Impact Evaluation. Methodological and Operational Issues*. Manila: Asian Development Bank.

Arbeitskreis Evaluation von Entwicklungspolitik (2009). *Verfahren der Wirkungsanalyse. Eine Landkarte für die entwicklungspolitische Praxis*. Freiburg i. Br.: Arnold-Bergstraesser-Institut.

Baker, J. L. (2000). *Evaluating the Impact of Development Projects on Poverty: A Handbook for Practitioners*. Washington, Washington D.C.: World Bank.

Bloom, H. S. (2006). *The Core Analytics of Randomized Experiments for Social Research*. MDRC Working Papers on Research Methodology. http://www.mdrc.org/sites/default/files/full_533.pdf. Abrufdatum: 15.04.2015.

Bohnsack, R. (2010). Qualitative Evaluationsforschung und dokumentarische Methode. In R. Bohnsack & I. Nentwig-Gesemann, (Hg.), *Dokumentarische Evaluationsforschung. Theoretische Grundlagen und Beispiele aus der Praxis* (23–62). Leverkusen: Barbara Budrich.

Bohnsack, R. & Nentwig-Gesemann, I. (Hg.) (2010). *Dokumentarische Evaluationsforschung. Theoretische Grundlagen und Beispiele aus der Praxis*. Leverkusen: Barbara Budrich.

Brüsemeister, Th. (2003). Flugbahn und Verlaufskurve – Pierre Bourdieu und Fritz Schütze zur Transintentionalität von Biographien. In: R. Greshoff, G. Kneer & U. Schimank (Hg.), *Die Transintentionalität des Sozialen. Eine vergleichende Betrachtung klassischer und moderner Sozialtheorien* (192–221). Opladen: Westdeutscher Verlag.

Caspari, A. & Barbu, R. (2008). *Wirkungsevaluierungen: Zum Stand der internationalen Diskussion und dessen Relevanz für Evaluierungen der deutschen Entwicklungszusammenarbeit*. Evaluation Working Papers. Bonn: Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung.

Greif, S. (2008). *Coaching und ergebnisorientierte Selbstreflexion*. Göttingen: Hogrefe.

Groenemeyer, A. & Schmidt, H. (2011). Evaluation und Evaluationsforschung. In H.-U. Otto & H. Thiersch (Hg.), *Handbuch Soziale Arbeit* (4. Auflage) (366–378). München/Basel: Ernst Reinhardt Verlag.

Haubrich, K. & Loidl-Keil, R. & Drilling, M. (2009). Evaluation in der Sozialen Arbeit im Ländervergleich. In: Widmer, W. Beywl, C. Fabian (Hg.), *Evaluation: Ein systematisches Handbuch* (469–474). Wiesbaden: VS Verlag.

Hoffmann, K. & Luisser, P. (2009). *Wirkungen lösungsfo-kussierter Trainings auf Mitarbeiterführung und Produktivität*. München/Mering: Rainer Hampp Verlag.

Jansen, A., Mäthner, E. & Bachmann, T. (2004). *Erfolgreiches Coaching. Wirkfaktoren im Einzel-Coaching*. Kröning: Asanger.

Jansen, A., Mäthner, E. & Bachmann, T. (2005). Wirksamkeit und Wirkfaktoren von Coaching. In Rauen, C. (Hg.), *Handbuch Coaching* (3. Auflage) (55–75). Göttingen/Bern/Wien/Toronto/Seattle/Oxford/Prag: Hogrefe.

- Jäckel, M. (2011). *Medienwirkungen: Ein Studienbuch zur Einführung* (5. Auflage). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Kelle, U. (2006). Qualitative Evaluationsforschung und das Kausalitätsproblem. In U. Flick (Hg.), *Qualitative Evaluationsforschung: Konzepte – Methoden – Umsetzung* (117–134). Reinbek bei Hamburg: Rowohlt Taschenbuch Verlag.
- Kelle, U. & Erzberger, C. (2006). Stärken und Probleme qualitativer Evaluationsstudien – ein empirisches Beispiel aus der Jugendhilfeforschung. In: Flick, U. (Hg.): *Qualitative Evaluationsforschung: Konzepte - Methoden – Umsetzung* (284–300). Reinbek: Rowohlt.
- Kühl, W. (2014). Wirkung von Führungskräfte-Coachings in der sozialen Arbeit. *Organisationsberatung – Supervision – Coaching*, 21 (1), 39–53.
- Künzli, H. (2005). Wirksamkeitsforschung im Führungskräfte-Coaching. *Organisationsberatung – Supervision – Coaching*, 12 (3), 231–243.
- Künzli, H. (2009). Wirksamkeitsforschung im Führungskräfte-Coaching. *Organisationsberatung–Supervision – Coaching*, 16 (1), 1–15.
- Kuckartz, U., Dresing, T., Rädiker, S. & Stefer, C. (2007). *Qualitative Evaluation: Der Einstieg in die Praxis*. Wiesbaden: VS Verlag.
- Rohmberg, W., Steindl, C., Weber, M. (2006). *Neue Entwicklungen im Bereich der Wirkungsanalyse und -abschätzung FTI-politischer Maßnahmen*. Wien: Austrian Research Centers.
- Legewie, J. (2012). Die Schätzung von kausalen Effekten: Überlegungen zu Methoden der Kausalanalyse anhand von Kontexteffekten in der Schule. *Kölner Zeitschrift für Soziologie*, 64, 123–153.
- Ravallion, M. (2005). *Evaluating Anti-Poverty Programs*. Washington, D.C.: World Bank.
- Stockmann, R. (2000). Evaluation staatlicher Entwicklungspolitik. In Reinhard Stockmann (Hg.), *Evaluationsforschung: Grundlagen und ausgewählte Forschungsfelder* (375–408). Opladen: Leske + Budrich.
- Stockmann, R. (2006). *Evaluation und Qualitätsentwicklung: eine Grundlage für wirkungsorientiertes Qualitätsmanagement*. Münster: Waxmann.
- White, H. & Barbu, A. (2006). *Impact Evaluation: The Experience of the Independent Evaluation Group of the World Bank*. Washington, D.C.: World Bank.
- Wilhelmy, M. (2014). *Auswirkungen von Einzelcoaching nach der Resonanzmethode bei Führungskräften aus der Sicht der Coachees*. Unveröffentlichte Masterarbeit, Department für Psychologie der Sigmund Freud Privat Universität Wien, Institut ARGE Bildungsmanagement.
- Wittichen, H.-U., Bühringer, G. & Rehm, J. (2011). *Predictors, Moderators and Outcome of Substitution Treatments – Effekte der langfristigen Substitution Opioidabhängiger*. Schlussbericht an das Bundesministerium für Gesundheit. [jektbericht_PREMOS_-
_Langfristige_Substitution_Opiatabhaengiger.pdf.
Abrufdatum: 14.04.2015.](http://www.bundesgesundheitsministerium.de/fileadmin/da teien/Publikationen/Drogen_Sucht/Forschungsberichte/Pr</p>
</div>
<div data-bbox=)

Wollmann, H. (2000). Evaluierung und Evaluationsforschung von Verwaltungspolitik und –modernisierung – zwischen Analysepotential und –defizit. In Reinhard Stockmann (Hg.), *Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder* (195–232). Opladen: Leske + Budrich.

Eingegangen: 21. April 2015

Peer Review: 27. Mai 2015

Angenommen: 3. Juni 2015

Autor

FH-Prof. apl. Prof. DDr. Bernhard Plé; Professor für Soziologie an der FH Joanneum, Graz und Universität Bayreuth; Mitglied der wissenschaftlichen Leitung der ARGE Bildungsmanagement, Universitätsinstitut für Beratungs- und Managementwissenschaften am Department Psychologie der Sigmund Freud Privatuniversität.

Diesen Artikel zitieren als: Plé, B. (2015). Wirkungsanalysen und ihre Forschungsdesigns: Kritische Gegenüberstellung ihrer konzeptionellen und methodologischen Herangehensweise im Hinblick auf mögliche Komplementaritäten und Anwendungen für die Beratungswissenschaften. *Zeitschrift für Beratungs- und Managementwissenschaften*, 2, 25–38.

Reichen Sie Ihr Manuskript beim Journal der ARGE Bildungsmanagement, Universitätsinstitut für Beratungs- und Managementwissenschaften am Department Psychologie der Sigmund Freud Privatuniversität ein und profitieren Sie von:

- Peer-reviewed
- Bequemer Online-Einreichung
- Keine Platzbeschränkungen
- Veröffentlichung nach Aufnahmeverfahren
- Ihre Arbeit ist öffentlich zugänglich

Senden Sie Ihr Manuskript an:

forschungsjournal@bildungsmanagement.ac.at

