

WORKING PAPER NO. 22

Vom „Fassadenmanagement“ zur evidenzbasierten Steuerung?

Reflexionsfragen zur Umsetzung von Wirkungsevaluation sozialer Interventionen

Hemma Mayrhofer

© IRKS

www.uibk.ac.at/irks

Wien, Juni 2022

ISSN 1994-490 X

Mayrhofer, H. (2022): Vom „Fassadenmanagement“ zur evidenzbasierten Steuerung? Reflexionsfragen zur Umsetzung von Wirkungsevaluation sozialer Interventionen, IRKS Working Paper, 22

Inhaltsverzeichnis

1. Einleitung: Fragestellung und theoretischer Rahmen	3
2. Begriffsver- und -entwurrungen	6
3. Zentrale Reflexionsfragen für eine gute Praxis der Wirkungserfassung bzw. -evaluation	8
3.1. Funktionen: Wozu will ich Erkenntnisse über die Wirkungen bestimmter Interventionen, Projekte bzw. Programme?.....	10
3.2. Akteurskonstellationen: Wer führt die Wirkungsevaluation in wessen Auftrag und über wen durch?	11
3.3. Zieldefinitionen und -präferenzen: Welche Wirkungen werden mit den zu untersuchenden Interventionen angestrebt und welche davon untersuche ich?.....	13
3.4. Erfolgskriterien bzw. Bewertungsmaßstäbe: Was definiere ich als Kriterium und Hinweis für Wirkungen?.....	15
3.5. Validität der Wirkungserfassung: Mit welchen Methoden bzw. Instrumenten komme ich zu belastbaren empirischen Wirkerkenntnissen?.....	16
3.6. Reaktivität der Instrumente: Inwiefern wirken die Methoden bzw. Instrumente auf den evaluierten Gegenstand zurück?.....	18
3.7. Ergebnistransfer: Wie fließen die Wirkerkenntnisse in die untersuchte Praxis zurück – und was hat das für (Aus-)Wirkungen?	18
4. Zum Schluss: Chancen elaborierter Wirkungsevaluation.....	20
<i>Literatur</i>	21

Vorbemerkungen: Die in diesem Working Paper zusammengefassten Reflexionen zur wissenschaftlichen Erfassung und Bewertung von Wirkungen sozialer Interventionen beruhen auf meiner langjährigen theoretischen und empirischen Auseinandersetzung mit sowie Umsetzung von Wirkungsevaluationen im Feld der Sozialen Arbeit, aber auch im Kontext der Wirksamkeit rechtlicher Regelungen. Der Text nahm seinen Ausgangspunkt in einem Vortrag, den ich Ende 2017 im Rahmen des wissenschaftlichen Beiratstreffens des IRKS halten durfte, und entwickelte sich über die nächsten zwei bis drei Jahre aus unterschiedlichen Anlässen weiter. Da sich der auf diese Weise entstandene Beitrag nur bedingt in den formellen Rahmen wissenschaftlicher Zeitschriften einfügt, es mir aber widerstrebt, seine Struktur zu verändern und seine Länge zu kürzen, wähle ich nun das freizügigere Format des Working Papers. Genau das ist der Text nämlich: ein Arbeitspapier, das work in progress-Überlegungen zusammenfasst. Ich formuliere und diskutiere zentrale Reflexionsfragen für Wissenschaft, Politik und Praxis, die dabei unterstützen können, eine gute Praxis der Wirkungsevaluation – bzw. generell der sozialwissenschaftlichen Erfassung von Wirkungen von sozialen Interventionen, Maßnahmen, Projekten oder Programmen – zu realisieren. Die Überlegungen erheben keinen Anspruch auf Vollständigkeit und sind nicht als standardisierte Checklisten bzw. kochbuchartige Handlungsanleitungen zu verstehen, sondern bieten Anregungen an, die eigene Evaluationspraxis zu reflektieren. Sie sollen aber auch für die Reflexion von Management-Instrumenten zum Wirkungscontrolling bzw. von wirkungsbezogenen Vorgaben seitens der Fördergeber:innen an Soziale Dienste gewinnbringende Denkanstöße geben.

1. Einleitung: Fragestellung und theoretischer Rahmen

Die Forderung nach Wirkungsorientierung in der Sozialen Arbeit lässt sich aus einer neoinstitutionalistischen Theorieperspektive als neu institutionalisierte Umwelterwartung an Träger, Einrichtungen und Fachkräfte der Sozialen Arbeit betrachten. Die Praxis kann sich dem Ansinnen relevanter Umwelten, die Wirksamkeit ihrer Interventionen und damit zugleich die Legitimität des Ressourceneinsatzes nachzuweisen, zumeist nur begrenzt oder auch gar nicht entziehen, selbst wenn solche Forderungen teils als Risiko für die autonome Fachlichkeit problematisiert werden. An wissenschaftliche Wirkungsevaluation werden in diesem Zusammenhang oft große Erwartungen gerichtet, dabei handelt es sich – so die hier vertretene und näher auszuführende These – durchaus auch um ein riskantes und voraussetzungsreiches Vorhaben.

Die empirische Sozialforschung bietet sich Politik und Praxis als Partnerin beim Nachweis von Wirkungen der Sozialen Arbeit an und befindet sich zugleich selbst mitten im Aushandlungsprozess über adäquate Forschungsdesigns und -methoden (u.a. Albus & Ziegler, 2013; Borrmann & Thiessen, 2016; Haunberger & Baumgartner, 2017; May, 2011; Menold, 2007; Micheel, 2013; Müller, 2017; Otto, 2007; Schneider, 2011). Und sie agiert dabei oft in einem komplexen Interessens- und Kräftefeld zwischen öffentlichen oder privaten Auftraggeber:innen der Studien, evaluierter professioneller Praxis und Nutzer:innen bzw. Adressat:innen der Sozialen Dienste.

Der Titel des Working Papers verweist bereits auf gewisse Ambivalenzen der skizzierten Entwicklungen: So deutet der Begriff „Fassadenmanagement“ Oberflächlichkeiten an und der Weg zur evidenzbasierten Steuerung ist mit einem Fragezeichen versehen. Mit beiden Aspekten sind wichtige Grundannahmen oder Rahmungen für die nachfolgend aufgeworfenen und diskutierten Reflexionsfragen zur Umsetzung von Wirkungsevaluation in der Sozialen Arbeit benannt. Sie sollen deshalb einleitend kurz umrissen werden.

Die Vorstellung des „Fassadenmanagements“, das Organisationen betreiben, lehnt sich an die neoinstitutionalistische Organisationstheorie an (Meyer & Rowan, 1991 [1977], S. 57ff., auch Drepper, 2010, S. 153f.), die sich hierbei wiederum von Goffman (2003 [1959]) inspirieren ließ. Der Begriff bezieht sich auf die „Schauseite“ (Luhmann, 1964, S. 116) der Organisation nach außen gegenüber relevanten Umwelten. Dabei ist „... die Differenzierung in Schauseite und Innenansicht (...) nicht reiner Schwindel, sondern eine strategische Konzeption, die eine sinnvolle Anpassung an spezifischen Systemgrenzen mit ihren je besonderen Anforderungen ermöglicht“ (Luhmann, 1964, S. 116). Schutzmechanismen gegen Inkonsistenzen zwischen diesen Anforderungen stellen erstens die Entkopplung von Formalstruktur und Aktivitätsebene sowie eine lediglich lose Kopplung verschiedener Organisationsteile wie Abteilungen oder Einrichtungen, zweitens die Logik des Vertrauens und guten Glaubens und drittens das Mittel zeremonieller Inspektionen und Evaluationen dar (Meyer & Rowan, 1991 [1977], S. 57ff.).

Gerade Soziale Dienste, aber auch andere Organisationen (z.B. im politischen Feld etc.) sind für ihr Weiterbestehen in hohem Ausmaß darauf angewiesen, dass sie ihre Existenz gegenüber verschiedenen relevanten Umwelten laufend legitimieren. Diese Umwelten richten nicht selten unterschiedliche, teils widersprüchliche Zielerwartungen an die Sozialen Dienste, die in der Regel nicht einfach zurückgewiesen werden können. Somit ist – so die hier vertretene Hypothese – ein gewisses Ausmaß an Fassadenmanagement für solche Organisationen kaum vermeidbar. Und zugleich kann elaborierte, nach außen transparente Wirkungsevaluation das Fassadenmanagement empfindlich stören, geht es dann doch nicht mehr nur um zeremonielle Inspektion und Evaluation. Dadurch können Zielwidersprüche (d.h. differierende und nicht miteinander zu vereinbarende Wirkungsziele

sowie deren Erreichung oder auch die Abweichung davon) offen zutage treten und für die unterschiedlichen sozialen Umwelten explizit gemacht werden.

Zur evidenzbasierten Steuerung sei einleitend angemerkt, dass sich sowohl beim Schlagwort der Evidenzbasierung als auch in Bezug auf Steuerung riesige Diskussionsfelder auf-tun, die hier lediglich angerissen, keinesfalls aber erschöpfend behandelt werden können. In Zeiten des postheroischen Managements (Baecker, 1994) gewinnt die Einsicht an Gewicht, dass nicht-triviale Systeme nicht direkt und linear steuerbar sind und es vermehrt über Kontextsteuerung zu arbeiten gilt (Willke, 2001, S. 132). Und Soziale Arbeit ist durch und durch solch ein nicht-triviales System und hat mit ebensolchen Systemen zu tun: Mit Individuen (seien es nun Klient:innen, Adressat:innen, Mitarbeiter:innen etc.), Gruppen, Teams, Organisationen (etwa Fördergeber:innen, andere Soziale Dienste, Behörden etc.), dem Gemeinwesen u.v.m.

Die Forderung nach einer evidenzbasierten Praxis (EBP) wiederum kommt von fachlicher Seite, es geht dabei um die Etablierung einer Praxis, die sich systematisch auf wissenschaftlich fundierte Erkenntnisse stützt (Mayrhofer, 2014; Otto, 2007; Otto et al., 2010; Sommerfeld, 2016, kritisch zur deutschen Diskussion: James, 2016). Hierzu findet eine vielschichtige und lebhaft Diskussions statt, und zwar vor allem im deutschsprachigen Raum. Unter anderem wird Kritik daran geübt, dass diesem Ansatz zu einfache Vorstellungen davon zugrunde liegen, wie sich wissenschaftliche Erkenntnisse in die Praxis transferieren lassen, dass EBP von einer linearen, hierarchischen Transfervorstellung von Wissen ausgehe, welches es dann in der Praxis kochbuchartig anzuwenden gelte (Hüttemann, 2011, 2016). Die Diskussion weist noch zahlreiche andere Facetten auf, einzelne – etwa die Frage der Methodenhierarchie – werden in den nachfolgenden Ausführungen aufgegriffen.

Eingebettet in diesen theoretischen Rahmen und auf Basis eigener umfassender Erfahrungen in der Evaluationsforschung (u.a. Mayrhofer, 2016; 2017) werden im vorliegenden Beitrag zentrale Reflexionsfragen für Wissenschaft, Politik und Praxis aufgeworfen und diskutiert, die dabei unterstützen sollen, eine gute Praxis der Wirkungsevaluation – bzw. generell der Erfassung von Wirkungen – im Feld der Sozialen Arbeit zu realisieren. Die Fragen beziehen sich auf sieben verschiedene Phasen bzw. Dimensionen in der Planung und Umsetzung von Wirkungsevaluation: Funktionen der Evaluation, Akteurskonstellationen, Zieldefinitionen, Erfolgskriterien bzw. Indikatoren, Validität der Wirkungserfassung, Methodenreaktivität und Ergebnistransfer. Zunächst aber soll ein Versuch unternommen werden, ein wenig Ordnung in das Begriffs-Wirrwarr im Feld der Wirkungsorientierung und -evaluierung zu bringen.

2. Begriffsver- und -entwirrungen

Bevor Reflexionsfragen benannt werden können, bedarf es der kurzen Erläuterung zentraler Begriffe im Feld der Wirkungsorientierung und -evaluation. Dies ist umso wichtiger, als die Auseinandersetzung mit Wirkungen von uneinheitlichen Begriffsverwendungen gekennzeichnet ist: Mit ein und demselben Begriff kann sehr Unterschiedliches, manchmal auch Gegensätzliches gemeint sein. Zunächst gilt es zwischen Wirkungsorientierung, Wirkungsmessung bzw. -erfassung und Wirkungsevaluation zu unterscheiden. Danach werden ausgewählte Kernbegriffe der Wirkungsdebatte (Input, Output, Outcome, Impact,...) erläutert, um abschließend einen Vorschlag für eine eindeutigere Begriffsverwendung zu unterbreiten.

Doch was ist überhaupt eine Wirkung? Unter Wirkung soll hier eine spezifische Form von Veränderung eines Zustandes oder einer Entwicklung verstanden werden, bei der diese Veränderung einer oder mehreren Ursachen zugerechnet werden kann. Damit ist als erstes zentrales Merkmal die Veränderung benannt, d.h. eine Vorher-Nachher-Differenz. Unter Umständen kann auch der Erhalt eines ansonsten gefährdeten Zustandes bzw. die Vermeidung einer erwartbaren Entwicklung bewirkt werden (=Prävention), doch auch hier muss die Intervention bzw. Maßnahme eine Veränderung gegenüber der befürchteten Entwicklung bewirken, d.h. einen Unterschied machen, um von Wirkung sprechen zu können. Dass solch präventive Wirkungen nur schwer nachzuweisen sind, sei an dieser Stelle nur kurz angedeutet. Das zweite essenzielle Merkmal einer Wirkung bezieht sich darauf, dass die Veränderung auf einen oder mehrere ursächliche Impulse zurückgeführt werden kann. Es braucht somit ausreichend bestimmbare Ursache-Wirkungs-Zusammenhänge, um von einer Wirkung – und nicht bloß von Veränderung – sprechen zu können. Genau dieser kausale Aspekt, der dem Wirkungsbegriff innewohnt, der ihn wesentlich ausmacht, bereitet für die Wirkungserfassung eine enorme Herausforderung.

Grundsätzlich ist mit dem vorgeschlagenen Wirkungsbegriff nicht zwangsläufig Zielorientierung bzw. absichtsvolle Intervention verbunden, es sind selbstverständlich auch ungeplante Wirkungen möglich. Mit Wirkungsorientierung hingegen schon: Der Begriff bezieht sich auf einen umfassenderen Prozess und es kommt das Element der beabsichtigten bzw. angestrebten Wirkungen hinzu, d.h. es existieren Wirkungsziele, deren Erreichen absichtsvoll herbeigeführt werden soll. Der Begriff steht zugleich für die im Rahmen der Modernisierung der Verwaltung (Stichwort: New Public Management) eingeführten neuen Steuerungsmodelle, die auf vereinbarte und erbrachte Leistungen sowie deren geplante und erzielte Wirkungen fokussieren. So ist seit 2013 der Grundsatz der wirkungsorientierten Haushaltsführung im österreichischen Bundeshaushaltsgesetz verankert, auf Länderebene existieren teils vergleichbare Orientierungen (Bundesministerium für Finanzen, n. d.; Dachverband Wiener Sozialeinrichtungen, 2013).

Theoretisch könnte man annehmen, dass damit eine starke Außenorientierung einhergeht, da sich die Wirkungen an den Adressat:innen der Maßnahmen bzw. Dienstleistungen entfalten (sollen). Manche Akteur:innen der Sozialen Arbeit erhoff(t)en sich auch durch zunehmende Transparenz der erbrachten Leistung und erreichten Wirkungen eine Stärkung von Adressat:innenrechten, Wirkungsorientierung wird mitunter als anschlussfähig an emanzipatorische Prämissen in der Sozialen Arbeit beobachtet. Die Realität scheint diesen Erwartungen allerdings nicht immer gerecht zu werden (Albus & Ziegler, 2013, S. 164f.).

Der Nachweis von Wirkungen wird meist als Wirkungsmessung bezeichnet. Ich schlage vor, das neutralere Wort der Wirkungs-ERFASSUNG im Sinne einer Erfassung von kausalen Ursache-Wirkungs-Zusammenhängen zu verwenden. Der Begriff Messung impliziert quantifizierendes Erfassen von Wirkungen; darum geht es zwar manchmal, aber nicht immer und schon gar nicht ausschließlich.

Bei der Wirkungserfassung lassen sich zumindest zwei Zugänge unterscheiden:

- Management-Instrumente wie Monitoring-Systeme bzw. Controlling-Instrumente haben nicht nur Kontrollfunktionen, sondern häufig zugleich mehr oder weniger direkte Steuerungsfunktionen. Sie arbeiten i.d.R. mit Audits, formalisierten Richtlinien und standardisierten Instrumenten, um den Prozess der Leistungserbringung zu vereinheitlichen und zu kontrollieren (z.B. Steuerung über Kennzahlen, mithilfe derer der Grad der Zielerreichung kontrolliert werden soll, Wirkungscontrolling, wirkungsorientierte Folgenabschätzung – Otto, 2007, S. 52).
- Wissenschaftliche Wirkungsevaluation ist eine spezifische Form der Wirkungsforschung und strebt mit Mitteln der empirischen Sozialforschung die Bewertung von Maßnahmen, Interventionen, Projekten oder Programmen an. Sie will auf wissenschaftlicher Basis Ursache-Wirkungs-Zusammenhänge erfassen und analysieren.

Beide Zugänge sollen hier nicht in eine vergleichende Wertung gebracht werden, es geht eher um ein Ergänzungs- denn ein Konkurrenzverhältnis. Sie haben unterschiedliche Vor- und Nachteile, die nachfolgend teilweise angedeutet werden, auch wenn der Schwerpunkt auf Wirkungsevaluation liegt.

Besondere Begriffswirren lassen sich um die beiden aus dem Englischen übernommenen Bezeichnungen Outcome & Impact beobachten. Input und Output hingegen werden meist einheitlich verwendet, sie bezeichnen einerseits Mitteleinsatz und andererseits Produkte bzw. erbrachte Leistungen oder Ergebnisse. Wenn man die Literatur zu Wirkungsorien-

tierung und Wirkungsforschung sichtet, stößt man beispielsweise auf folgende Definitionen von Outcome und Impact – es ist davon auszugehen, dass damit noch nicht die gesamte Vielfalt an Begriffsdefinitionen abgebildet ist:

- Als Outcome gelten teilweise Wirkungen, die ursächlich auf die Intervention zurückführbar sind (Reade-Soh & Stockmann, 2009), teilweise nur die kurz- und mittelfristigen Auswirkungen (Caspari & Barbu, 2008), vereinzelt auch – von den anderen Definitionen völlig abweichend – die Bruttowirkungen (d.h. alle Veränderungen, die nach Setzen der Intervention zu beobachten sind, ohne dass eine kausale Zuordnung zur Intervention möglich ist – Rauscher, Schober & Millner, 2012, S. 6) oder gesellschaftliche Wirkungen (Halfar & Hegenauer, 2010).
- Der Begriff Impact wird manchmal für Bruttowirkungen im eben beschriebenen Sinn verwendet (Reade-Soh & Stockmann, 2009), teils ist er für Langzeit-Effekte (Caspari & Barbu, 2008) oder auch Wirkungen auf System- bzw. gesellschaftlicher Ebene (Beywl, Speer & Kehr, 2004, S. 37) reserviert. Er wird vereinzelt auch als subjektive Wirkungseinschätzung der Personen definiert, die von einer Maßnahme betroffen sind (Halfar & Hegenauer, 2010), oder auch – wieder deutlich konträr zu anderen Begriffsverwendungen – als Netto-Wirkungen (Rauscher et al., 2012, S. 7).

Daraus lässt sich das Fazit ziehen, dass es empfehlenswert ist, stets nach dem zugrundeliegenden Begriffsverständnis zu fragen, wenn von Outcome oder Impact gesprochen wird. Um den Begriffsverwirrungen zu entgehen, wird hier lediglich die sehr wichtige Unterscheidung zwischen Produkt bzw. Ergebnis (Output) und Wirkung vorgeschlagen (Produkte bzw. erbrachte Dienstleistungen müssen noch nicht zwangsläufig Wirkungen zeitigen). Zugleich erweist es sich als nützlich, zwischen drei verschiedenen Wirkungsebenen zu unterscheiden (Mayrhofer, 2017): Mikroebene (Individuum, Kleingruppen), Mesoebene (Organisationen oder Sozialraum/Stadtteil) und Makroebene (Gesellschaft). Darüber hinaus können zeitliche Wirkaspekte, d.h. kurz-, mittel- und langfristige Wirkungen, differenziert werden.

3. Zentrale Reflexionsfragen für eine gute Praxis der Wirkungserfassung bzw. -evaluation

Im Folgenden sollen auf Basis eigener Evaluationserfahrungen zentrale Reflexionsfragen aufgeworfen und diskutiert werden, die dabei unterstützen können, eine gute Praxis der Wirkungsevaluation – bzw. generell der Erfassung von Wirkungen – zu realisieren oder sich ihr so gut wie möglich anzunähern. Obwohl wissenschaftliche Wirkungsevaluation

dabei den Hauptbezugspunkt bildet, wird teilweise auch auf Management-Instrumente zur Wirkungserfassung eingegangen. Die Fragen können – in unterschiedlicher Weise bzw. mit verschiedenen Schwerpunkten – sowohl für Wissenschaft, Politik, Fördergeber:innen und Praxis Sozialer Arbeit relevant sein.

1. Wozu will ich Erkenntnisse über die Wirkungen bestimmter Interventionen/Maßnahmen/Programme? (= intendierte Funktionen)
2. Wer führt die Wirkungsevaluation in wessen Auftrag und über wen durch? (= Akteurskonstellationen)
3. Welche Wirkungen werden mit den zu untersuchenden Interventionen angestrebt und welche davon untersuche ich? (= Zieldefinitionen, Bestimmung und Eingrenzung der zu erfassenden Wirkungsziele)
4. Was definiere ich als Kriterien und Hinweise für Wirkungen? (= Erfolgskriterien und Wirkungsindikatoren)
5. Mit welchen Methoden bzw. Instrumenten komme ich zu belastbaren empirischen Ergebnissen, ob und wie bestimmte Interventionen wirken? (= Validität der Wirkungserfassung)
6. Inwiefern wirken die Methoden bzw. Instrumente selbst auf den evaluierten Gegenstand zurück? (= Reaktivität der Erfassungsinstrumente/Methodenreaktivität)
7. Wie fließen die Wirkerkenntnisse in die untersuchte Praxis zurück und was hat das für (Aus-)Wirkungen? (= Ergebnistransfer)

Der „Reflexionsreigen“, der hiermit eröffnet wird, lässt sich grafisch wie folgt darstellen:



Abbildung 1: „Reflexionsreigen“ Wirkungsnachweis bzw. Wirkungsevaluation

3.1. Funktionen: Wozu will ich Erkenntnisse über die Wirkungen bestimmter Interventionen, Projekte bzw. Programme?

Die Frage danach, welche Funktion die Wirkungserfassung bzw. -evaluation erfüllen soll, steht am Anfang des Entscheidungs- und Umsetzungsprozesses. In der Praxis dürfte dieser Schritt allerdings nicht immer explizit vollzogen werden, sondern die Funktionen manchmal latent und möglicherweise auch mehrdeutig dem Vorhaben zugrunde liegen. Nach Stockmann (2007, S. 36ff.) lassen sich vier verschiedene Funktionen bzw. Ziele von Evaluationen (der Kategorisierungsvorschlag bezieht sich nicht nur auf Wirkungsevaluation) unterscheiden. Sie werden im Folgenden mit Fokus auf die Erfassung und Bewertung von Wirkungen reformuliert:

- Die *Erkenntnisfunktion* kann in Bezug auf Wirkungserfassung ganz allgemein so zusammengefasst werden, dass damit mehr Wissen über die Erreichung der Wirkungsziele erlangt werden soll. Es geht im Kern um die Frage, *ob* etwas wirkt. In der Folge können aus diesem Wissen Steuerungsentscheidungen abgeleitet werden.

- Wenn die Evaluation vorrangig *Kontrollfunktionen* erfüllt, dann soll vor allem *sichergestellt* werden, dass die angestrebten Wirkungsziele auch erreicht werden. Stockmann (2007, S. 38) ist beizupflichten, dass jede ausreichend aussagekräftige Evaluation direkt oder indirekt auch eine Form von Kontrolle mit ermöglicht, selbst wenn dies nicht dezidiert Evaluationsziel ist. Das macht Evaluationen grundsätzlich zu einer sensiblen Angelegenheit.
- Soll die Wirkungsevaluation *Lern- und Weiterentwicklungsfunktionen* im Sinne der Optimierung und Professionalisierung der sozialen Interventionen wahrnehmen, dann benötigt man vor allem ein höheres Verständnis von den Wirkmechanismen und -zusammenhängen. Die Evaluation muss Wissen bzw. Erkenntnisse darüber bereitstellen können, *wie* etwas wirkt.
- *Legitimationsfunktionen* spielen im Zusammenhang mit Wirkungsnachweisen meist eine Rolle, es geht häufig auch darum, mit diesen Nachweisen gegenüber unterschiedlichen relevanten Umwelten die erhaltenen und eingesetzten Ressourcen, aber auch die angewandten Interventionsmethoden und erreichten Ziele zu legitimieren. Wirkungsnachweise können als Strategie in der Konkurrenz um knappe öffentliche Gelder eingesetzt werden – oder von der anderen Seite betrachtet: Wirkungsnachweise stellen eine Entscheidungsgrundlage öffentlicher Stellen für die Vergabe knapper Mittel dar. Dies macht deutlich, dass es bei Wirkungserfassung bzw. -evaluation faktisch in der Regel auch nicht einfach nur um Effektivität im Sinne von Wirksamkeit geht, sondern i.d.R. auch um Effizienz, d.h. um das Verhältnis des Ressourceneinsatzes zu den erzielten Leistungen und Wirkungen.

Oft werden der Evaluation mehrere Funktionen zugrunde liegen, nicht selten kommt aber bestimmten Funktionen eine größere Bedeutung bei. Für die nachfolgenden Reflexionsfragen soll festgehalten werden, dass je nachdem, welche Funktionen im Vordergrund stehen, auch unterschiedliche Instrumente der Wirkungserfassung empfehlenswert erscheinen (vgl. Punkt 3.5).

3.2. Akteurskonstellationen: Wer führt die Wirkungsevaluation in wessen Auftrag und über wen durch?

Bereits eingangs wurde darauf verwiesen, dass sich Evaluationsforscher:innen in komplexen Interessens- und Kräftefeldern bewegen. Die jeweilige Konstellation kann Auswirkungen auf den Evaluationsprozess, die eingesetzten Instrumente der Wirkungserfassung und deren Ergebnisse haben. Folgende fünf Akteursebenen sollen hier unterschieden werden, wobei sie teilweise auch deckungsgleich sein können (s.u.): erstens Auftrag- bzw. Fördergeber:in der evaluierten Einheit, zweitens evaluierte Einheit (Einrichtung, Projekt,

Programm etc. und deren Träger), drittens Adressat:innen bzw. Nutzer:innen der evaluierten Einheit, viertens Auftraggeber:in der Evaluation sowie fünftens Evaluators:innen.

Grundsätzlich lässt sich zwischen interner und externer Evaluation unterscheiden. Von internen Evaluationen spricht man dann, „wenn sie von der gleichen Organisation vorgenommen werden, die auch das Programm oder das Projekt durchführt“ (Stockmann, 2010, S. 80). Sind die evaluierenden Personen zugleich die das Projekt, Programm bzw. Angebot umsetzenden Personen (und nicht z.B. Kolleg:innen aus einer anderen Organisationseinheit), dann handelt es sich um eine Selbstevaluation. Bei externen Evaluationen hingegen gehören die Evaluators:innen nicht der durchführenden Organisation an. Dadurch lässt sich u.a. eine größere Unabhängigkeit und persönliche Distanz zum Evaluationsgegenstand als auch fundierte Methodenkompetenz realisieren. Die folgenden Ausführungen beziehen sich auf externe Evaluationen, konkret auf wissenschaftlich-empirische Evaluationsforschung zu Wirkungen sozialer Interventionen, d.h. auf Konstellationen, in denen externe Evaluationsforscher:innen die Evaluation durchführen.

Die Akteurskonstellationen können sehr verschieden sein, folgende drei Möglichkeiten, mit denen ich selbst Erfahrung habe, sollen hier beschreiben werden:

- *Dyadische Auftragsbeziehung*: Der bzw. die Auftraggeber:in der Evaluation ist zugleich für die Umsetzung des Evaluationsgegenstandes verantwortlich, d.h. ein Sozialer Dienst will eine selbst durchgeführte Maßnahme, ein eigenes Programm etc. evaluiert haben. Dies kann einerseits eine hohe Abhängigkeit der Forscher:innen vom evaluierten Dienst mit sich führen, und zwar vor allem dann, wenn die Evaluation vorrangig Legitimationsfunktion nach außen hat. Bei ausreichender Vertrauensbasis, wechselseitiger Rollenakzeptanz und starker Lern- und Entwicklungsorientierung auf Seiten des Auftraggebers bzw. der Auftraggeberin lassen sich aber meiner Erfahrung nach in solch einer Konstellation fruchtbare Kooperationen realisieren.
- *Auftragsdreieck*: Der bzw. die Auftraggeber:in der Evaluation und die evaluierte soziale Dienstleistung bzw. Einrichtung etc. sind unterschiedliche Akteur:innen, erstere:r ist zugleich Fördergeber:in der evaluierten Einheit, die Forschung steht zwischen beiden. Handlungsmöglichkeiten und Evaluationsprozess sind in dieser Konstellation stark abhängig von der Beziehung zwischen Auftraggeber:in und evaluiertem Dienstleister sowie von den Funktionen, die die Wirkungsevaluation hat. Steht Kontrolle bzw. Legitimation im Vordergrund und sind in der Folge die Evaluationsergebnisse eine zentrale Entscheidungsgrundlage für weitere Mittelgewährung, dann ist von einem prekären Vertrauensverhältnis auszugehen und die Umsetzung der Wirkungsevaluation von zahlreichen Spannungsfeldern geprägt.

- „Viererkonstellation“ mit externer Evaluationsfinanzierung: Wenn die Wirkungsevaluation von einer unabhängigen Stelle finanziert wird, z.B. einer Forschungsförderungsstelle (Antragsforschung), die zu keinem der direkt beteiligten und von der Evaluation betroffenen Akteur:innen in einem Nahe- oder Abhängigkeitsverhältnis steht, kann dies in inhaltlicher, zeitlicher und ressourcenmäßiger Hinsicht große Freiräume eröffnen. Es bestehen auch besonders gute Voraussetzungen für eine Kooperation auf Augenhöhe zwischen den beteiligten Akteursebenen (Mayrhofer, 2017). Allerdings ist davon auszugehen, dass solch ein Setting vergleichsweise schwer zu realisieren ist, da Antragsforschung meist mit langwierigen, zeitintensiven und ergebnisunsicheren Antragsverfahren einhergeht und es zudem eine für die meist anwendungsorientierten Forschungsziele geeignete Förderschiene braucht.

Es ist kein Zufall, dass die Adressat:innen der Interventionen, Maßnahmen bzw. Programme in diesen Konstellationsbeschreibungen noch nicht vorkamen. Sie haben faktisch – so ist (selbst-)kritisch anzumerken – bei der Konzeption und Beauftragung von Wirkungsevaluation zumeist keine "Parteienstellung". Die Frage, inwieweit Wirkungsevaluation auch als partizipatives Forschungsprojekt (hierzu u.a. Mayrhofer & Schachner, 2013) umsetzbar ist, kann an dieser Stelle nicht vertiefend behandelt werden. Es ist aber darauf zu verweisen, dass man sich damit nicht nur Vorteile, sondern auch beträchtliche Herausforderungen und Risiken einhandeln dürfte. So werden etwa die Fragen virulent, wie sich die methodisch oft sehr anspruchsvollen Verfahren umsetzen lassen, wer tatsächlich eingebunden werden kann und will und wer nicht und wie mit den Doppelrollen (Betroffene:r bzw. Adressat:in und zugleich partizipative:r Forscher:in) sowie den damit verbundenen Spannungsfeldern umzugehen ist.

3.3. Zieldefinitionen und -präferenzen: Welche Wirkungen werden mit den zu untersuchenden Interventionen angestrebt und welche davon untersuche ich?

Die Frage nach den Zielen der zu evaluierenden Interventionen, Projekte oder Programme ist zugleich die Frage nach den intendierten Wirkungen – d.h. sie ist essenziell in der Wirkungsevaluation sowie generell in der Wirkungsorientierung und -erfassung. Es gilt zu klären, auf welche Ziele durch die implementierten Instrumente des Wirkungsnachweises fokussiert wird und welche damit weniger oder u.U. gar nicht erfasst werden. Das impliziert die Frage danach, wessen Ziele dies sind, d.h. welche internen und externen Akteur:innen bzw. Stakeholder diese Ziele einbringen und anstreben. Und wie wird darüber entschieden, welche Ziele bei der Wirkungsevaluation im Mittelpunkt stehen, und welche Folgen haben die Präferenzsetzungen gegebenenfalls? So lassen sich beispielsweise im

Bereich des Strafvollzugs und der Straffälligenhilfe Resozialisierung, Schutz der Gesellschaft (Spezial- und Generalprävention) oder auch Sanktion bzw. Bestrafung als Wert für sich als Zielsetzungen identifizieren. Je nach Fokussierung wird etwas völlig anderes evaluiert.

Diese aufgeworfenen Fragen verweisen bereits darauf, dass die Ziele Sozialer Dienste vielfältig sein können. Die Ziele setzen sich die Organisationen nicht nur selbstgewählt, sie werden durch die relevanten Umwelten mitdefiniert. Solche Umwelten können etwa öffentliche Auftraggeber:innen bzw. fördernde/finanzierende Stellen (die meist ihrerseits in hohem Ausmaß Legitimationserfordernisse ihren relevanten Umwelten gegenüber haben), andere Soziale Dienste, die Standards setzen, die professionelle Bezugsebene, verschiedene politische Player, unterschiedliche Sozialraum-Akteur:innen, eine medial repräsentierte Öffentlichkeit u.a.m. sein. Hinzu kommen die Klient:innen bzw. Nutzer:innen, deren besondere Relevanz einerseits unumstritten ist. Andererseits wird aber auch lebhaft diskutiert, wie sich diese Relevanz faktisch zu entfalten hat, inwieweit und in welchem Umfang etwa wirklich mit ihren Zielen gearbeitet wird und werden kann und welche ihrer Ziele eventuell auch zurückgewiesen werden.

Die unterschiedlichen relevanten Umwelten richten nicht nur teils verschiedene, sondern manchmal auch widersprüchliche Erwartungen an die Sozialen Dienste, die oft nicht zurückgewiesen werden können und zwischen denen es zu vermitteln gilt. Hinzu kommt, dass die Ziele nicht immer eindeutig und konkret sind – bzw. sind sie insgesamt oft nur eingeschränkt S.M.A.R.T. (spezifisch, messbar, akzeptiert, realistisch, terminisierbar), und das hat mitunter gute Gründe. So besteht eine häufige Lösungsstrategie für widersprüchliche Zielsetzungen darin, die Organisationsziele abstrakt und mehrdeutig zu formulieren. Das ermöglicht es, die Differenzen und Widersprüche latent zu halten: Man kann auf allgemeiner Ebene einen Konsens herstellen und erlaubt zugleich jeweils unterschiedliche Interpretationen und Umsetzungen. Abstrakte Zielformulierungen können in manchen Handlungsfeldern durchaus funktional sein! Zielwidersprüche und -konflikte können mit dieser Strategie jedoch nur so lange latent gehalten werden, wie keine Notwendigkeit besteht, Zielerreichungen konkret zu überprüfen (Mayrhofer, 2014, S. 285).

Zugleich lassen sich manche Ziele vorweg und überindividuell schwer konkretisieren, vor allem bei einer stark flexiblen und Adressat:innen-orientierten Arbeitsweise. So braucht es beispielsweise in der Offenen Jugendarbeit eine ganzheitliche Orientierung an den Lebenslagen und Bedürfnissen von jungen Menschen (Mayrhofer, 2017). Aber auch im Bereich des Erwachsenenschutzes bzw. der Erwachsenenvertretung kann das Ziel, Selbstbestimmung zu ermöglichen und zu fördern, flexiblen Konkretisierungsbedarf zeitigen. Was heißt das etwa, wenn sich eine Person selbst stark rechtliche Stellvertretung wünscht, weil es das Leben einfacher macht?

Diese Beschaffenheit der zu evaluierenden Wirkungsziele macht es auch schwer, die intendierten Wirkungen zu konkretisieren bzw. sich darauf zu einigen, was als Erfolg gelten kann. Konsensfindungen über Ziele sind übrigens Thomas Klatetzki zufolge dann besonders schwierig, „wenn die Klienten der Organisation als abweichend definiert werden und wenn die Organisation versucht, ein großes Spektrum der Problemlagen von Klienten zu bearbeiten“ (Klatetzki, 2010, S. 14) – sprich: wenn sie sich in ganzheitlicher und niederschwelliger Weise um Probleme und Unterstützungsbedarfe kümmert (Mayrhofer, 2018).

3.4. Erfolgskriterien bzw. Bewertungsmaßstäbe: Was definiere ich als Kriterium und Hinweis für Wirkungen?

Auf dem Weg hin zur empirischen Erfassung von Wirkungen muss die essenzielle Frage geklärt werden, woran man Zielerreichungen grundsätzlich erkennen könnte. Es geht um die Ableitung von Indikatoren, d.h. „empirisch (quantitativ oder qualitativ) zu erfassende(r) Kenngrößen, die einen Soll-Ist-Vergleich bezüglich der Zielsetzungen von Projekten oder Programmen ermöglichen sollen“ (Meyer, 2007, S. 197). Wichtig ist dabei, die Funktion des Indikators „als Anzeige eines nicht direkt messbaren theoretischen Konstrukts“ (Meyer, 2007, S. 197) zu berücksichtigen. Es gilt theoretisch abzuleiten oder empirisch-explorativ zu ergründen, wie Wirkung auf welchen Wirkungsebenen (auf individueller, sozialräumlicher bzw. organisatorischer oder gesellschaftlicher Ebene) grundsätzlich zum Ausdruck kommen könnten (hierzu exemplarisch Mayrhofer, 2017). Dies impliziert auch die Frage danach, welche und wessen Bewertungsmaßstäbe dabei zur Anwendung kommen sowie welche alternativen Bewertungskriterien noch denkbar wären – und welchen Unterschied sie machen würden.

Erschwert wird dieser Schritt bei Wirkungserfassungen im Feld der Sozialen Arbeit dadurch, dass nicht nur Ziele teilweise überindividuell schwer konkretisierbar sind, sondern sich auch die Indikatoren für Zielerreichungen mitunter überindividuell schwer bestimmen lassen (Mayrhofer, 2017). Damit geht auch ein bedeutsames Risiko einher, nämlich dass leicht zu erfassende Faktoren (das bedeutet oft: quantitativ messbare Faktoren) tendenziell gegenüber den schwerer erfassbaren Aspekten bevorzugt werden. Eine gewisse Pragmatik hierbei ist in der Praxis auch naheliegend, geht es doch u.a. auch um Legitimation nach außen. Kannonier-Finster und Ziegler (1998) verweisen in Bezug auf psychosoziale Betreuung auf die Gefahr, „... daß diese leicht meßbaren Faktoren in der Folge auch als *relevante* Faktoren anerkannt und betont werden“ (S. 42, Hervorhebung im Original). Bei der Implementierung von Wirkungsnachweisen gilt es deshalb zu prüfen, inwieweit durch die Instrumente nicht die leicht messbaren Faktoren vorschnell mit den relevanten Faktoren gleichgesetzt werden.

Generell zeigen solche Indikatoren, die etwa in Kennzahlen ihren Niederschlag finden können, bei differenzierter Analyse sehr schnell die Grenzen ihrer Aussagekraft. Wenn man beispielsweise die Zieldimension „Wohnfähigkeit fördern“ in der Wohnungslosenhilfe, die u.a. den Indikator „Fähigkeit zu selbstständiger Haushaltsführung“ haben könnte (das Beispiel ist nicht rein fiktiv, ohne hier konkreter zitieren zu wollen), an der wohnenden Bevölkerung anwenden würde, müsste man vermutlich bei nicht wenigen Personen eine Delogierung überlegen. Oder wenn die Zufriedenheit mit der Betreuung in psychosozialen Einrichtungen als ein Kriterium für erfolgreiche, personenorientierte Betreuung betrachtet wird, gilt es mit zu beachten, dass die Befragten ihre Erwartungen und Bedürfnisse auch der eigenen Lebens- und Lerngeschichte und den aktuellen Kontexten anpassen: Wer lange in institutionellen Betreuungssettings gelebt hat, hat möglicherweise gelernt, die Erwartungen herunterzuschrauben – oder gar nie ausreichend gelernt, eigene Bedürfnisse zu haben und zu artikulieren. Mit diesen Beispielen soll nicht die Sinnlosigkeit des Suchens nach Indikatoren demonstriert, sondern lediglich nochmals verdeutlicht werden, wie sorgfältig solche Indikatoren, die nicht direkt messbare Sachverhalte zu erfassen versuchen, auf ihre Gegenstandsangemessenheit und Aussagekraft hin zu prüfen sind.

3.5. Validität der Wirkungserfassung: Mit welchen Methoden bzw. Instrumenten komme ich zu belastbaren empirischen Wirkerkenntnissen?

Dieser Fragenkomplex bezieht sich zunächst auf die Gültigkeit der erfassten Wirkungen, d.h. die interne Validität, die etwas über die Güte kausaler Schlussfolgerungen aussagt (Bortz & Döring, 2003, S. 530; Müller, 2017, S. 211). Die zentrale Frage hierbei lautet: Wie verlässlich können die beobachteten Veränderungen den gesetzten Interventionen kausal zugerechnet werden bzw. wie zuverlässig kann ausgeschlossen werden, dass die festgestellten Wirkungen nicht durch andere Faktoren (sog. „Störfaktoren“ bzw. „Drittvariablen“) verursacht werden? Das impliziert zugleich die Frage danach, wie gut die Instrumente des Wirkungsnachweises die Komplexität der i.d.R. nichtlinearen, pfadabhängigen und multikausalen Wirkzusammenhänge zu erfassen vermögen. Genau durch diese Merkmale der Wirkweisen vieler sozialer Interventionen wird die interne Validität von Wirkungsevaluationen geschwächt. In der Regel wird es deshalb um eine bestmögliche Annäherung an kausale Ursache-Wirkungs-Zusammenhänge gehen – es gilt, die Grenzen der Wirkungsevaluation oder auch des Wirkungscontrollings stets mit zu reflektieren.

Die Frage nach der Validität der Wirkungserfassung zielt zweitens auf den Geltungsbereich von Wirkerkenntnissen ab, d.h. auf deren externe Validität: Welchen Gültigkeitsbereich haben die Ergebnisse, in welchem Ausmaß sind sie verallgemeinerbar? Wie zuverlässig sind die empirisch beobachteten Ursache-Wirkungs-Zusammenhänge auf ähnliche

Interventionen in mehr oder weniger abweichenden Situationen und Kontexten übertragbar? – Und die Situationen und Kontextbedingungen für bestimmte Interventionen gleichen sich in Arbeitskontexten der Sozialen Arbeit niemals völlig, Abweichungen sind immer gegeben. Selbst wenn eine Maßnahme prinzipiell wiederholbar wäre, würde sie immer auf andere äußere Situationen und veränderte Rahmenbedingungen stoßen (Otto, 2007, S. 58ff.; Albus & Ziegler, 2013, S. 167f.; Menold, 2007, S. 38).

Dies wirft viele Fragen auf, die sich auf adäquate Untersuchungsdesigns für Wirkungsevaluation beziehen. Hierzu findet auf wissenschaftlicher Ebene eine lebhaft Diskussions statt, insbesondere auch im Zusammenhang mit einer Evidenzbasierung der Praxis, die hier nicht wiedergegeben werden kann. Für eine einfache Übertragung der in der Evidenzbasierten Praxis (EBP) geltenden Methodenhierarchie bzw. methodischen Standards (der „Levels of Evidence“ bzw. der sog. „Goldstandards“) auf Wirkungsevaluation in der Sozialen Arbeit bestehen wesentliche Einschränkungen – etwa aufgrund einer hohen individuellen Varianz auf Seiten der Klient:innen und deren lebensweltlicher Kontexte.

Aus den komplexen Methodendiskussionen und Evaluationserfahrungen lässt sich zunächst schlussfolgern, dass es vorweg ein angemessen komplexes Verständnis vom zu evaluierenden Arbeitsfeld braucht. Es sind solide Kenntnisse über die Charakteristika der Wirkmöglichkeiten der dort gesetzten Interventionen und über die damit beabsichtigten und erwartbaren Wirkmechanismen und -zusammenhänge notwendig. Darauf aufbauend können gegenstandsadäquate Forschungsmethoden der Wirkungsevaluation identifiziert und gegebenenfalls (weiter-)entwickelt bzw. adaptiert werden, die etwa auch qualitativ-rekonstruktive Verfahren bzw. generell qualitative Forschungsansätze umfassen und insbesondere komplexe Wirkzusammenhänge mit Kontextfaktoren zu erfassen vermögen. Die Instrumente sollten beispielsweise auch nicht-intendierte Wirkungen erfassen können, wofür es ein exploratives Vorgehen und nicht nur hypothesenprüfende Verfahren benötigt.

Generell stellen Zahlen eine massive Reduktion von Komplexität dar – und das ist zugleich eine Stärke und Schwäche. Sie suggerieren starke und schnelle Aussagekraft, mit der es aber oft nicht weit her ist, wenn man die dahinterliegenden komplexen Zusammenhänge rekonstruiert – zumindest wenn man alleine auf die Zahlen zurückgreift und keine vertiefenden, oft qualitativen, Einblicke in die Wirkungsweisen hat, also zu wenig darüber weiß, *wie* etwas genau wirkt. Dies unterstreicht die große Bedeutung eines komplexen Forschungsdesigns mit einem passenden Methodenmix, die erst in ihrer Kombination ausreichend valide Erkenntnisse zu den Wirkungen Sozialer Arbeit gewähren. Es braucht dabei sowohl methodische Ansätze, die eine ausreichend belastbare Annäherung an die Frage erlauben, *ob* eine Intervention, eine Maßnahme, ein Programm etc. wirkt, als auch solche, die valide Erkenntnisse darüber erlauben, *wie* die Wirkung zustande kommt, d.h. die zu einem tieferen Verständnis der Wirkweisen beitragen. Erstere Ansätze erfüllen eher

Kontroll- und Legitimationsfunktionen, es handelt sich aber oft um sogenannte „Black-Box-Evaluationen“ (hierzu auch Haunberger & Baumgartner, 2017, S. 122), zweitens können deutlich besser Lern- und Entwicklungsfunktionen erfüllen. Die Realisierung dieses Anspruchs ist allerdings immer auch eine Frage der Forschungskompetenzen, denn diese Forschungsdesigns stellen hohe methodische Ansprüche an die Forscher:innen. Und es ist insbesondere auch eine Frage der zeitlichen und finanziellen Ressourcen, die für die Wirkungsevaluation zur Verfügung stehen.

3.6. Reaktivität der Instrumente: Inwiefern wirken die Methoden bzw. Instrumente auf den evaluierten Gegenstand zurück?

Wie bereits ausgeführt, manifestieren sich sowohl in Management-Instrumenten der Wirkungssteuerung und -kontrolle als auch in der methodischen Umsetzung von Wirkungsforschung und -evaluation Zielpräferenzen und Bewertungsmaßstäbe (Albus & Ziegler, 2013). Das hat zunächst Auswirkungen auf die Wirkerkenntnisse. Zusätzlich ist die Frage danach zu stellen, inwieweit sich der evaluierte Gegenstand dem Bewertungsinstrument unbemerkt (und damit unkontrolliert) anpasst. Inwiefern wirken also die Modelle zur Erfassung von Wirkungen – seien es nun managerielle oder auch wissenschaftliche – unbeabsichtigt auf den Gegenstand (die zu erbringende soziale Dienstleistung) zurück? Solche Rückkopplungseffekte können insbesondere dann negative Auswirkungen haben, wenn die Instrumente dem Handlungsfeld nicht ausreichend gerecht werden, wenn sie dessen Arbeitsweise und Wirkzusammenhänge nur ungenügend abbilden.

Es ist beispielsweise davon auszugehen, dass die Vorgabe von Wirkungsindikatoren und die Erfassung von Wirkweisen in definierten Kennzahlen auf die tatsächliche Aufgabenerfüllung abfärben. Die Aktivitäten orientieren sich zumindest in gewissem Ausmaß daran, es werden damit Aufmerksamkeiten spezifisch gelenkt. Das kann unter Umständen die Adressat:innen-Orientierung empfindlich einschränken. Managerielle Modelle der Wirkungssteuerung führen oft zur stärkeren Formalisierung von Strukturen und Prozessen, was zur Lähmung der „Kerngeschäfte“ der Organisation beitragen und die dort notwendige Flexibilität reduzieren kann. Wissenschaftliche Wirkungsevaluation ist vermutlich meist gegenstandssensibler, man kann sich mit ihr aber andere Folgeprobleme einhandeln, wie nun im letzten Punkt näher thematisiert wird.

3.7. Ergebnistransfer: Wie fließen die Wirkerkenntnisse in die untersuchte Praxis zurück – und was hat das für (Aus-)Wirkungen?

Die Frage danach, wie die Wirkerkenntnisse an die Praxis rückgebunden werden können, ist von Anfang an mitzureflectieren, Formen und Zeitpunkte der Ergebnistrückkopplung sind bereits bei der Ausarbeitung des Evaluationskonzepts mit einzuplanen. Zu klären gilt

es u.a., inwieweit bzw. in welcher Weise eine ausreichende Vertrauensbasis für Lernprozesse geschaffen und garantiert werden kann, wer die Adressat:innen der Evaluationsergebnisse sind und in welchem Zusammenhang die Ergebnisse zu deren Interessen stehen. Die Herausforderungen des Wissenschafts-Praxis-Transfers lassen sich in folgenden Fragen zusammenfassen: Wie können die Erkenntnisse der Wirkungsevaluation zu einer Form von Wissen transformiert werden, das an die Praxis anschlussfähig ist? Wie können sie in das Handlungswissen der Praxis einfließen und so eine evidenzbasierte Praxis mit ermöglichen?

Eine ‚Gretchenfrage‘ der Evaluationsforschung allgemein (nicht nur der Wirkungsevaluation) ist die nach der Ergebnistransparenz: Dürfen und sollen die Evaluationsergebnisse veröffentlicht werden? Wem sollen und müssen sie zugänglich gemacht werden? Bei einem grundsätzlichen Bekenntnis zum Ziel, Forschungsergebnisse möglichst allgemein zugänglich zu machen, gilt es zu beachten, dass Wirkungsevaluation auch riskante Transparenz schaffen kann, wie einleitend dargestellt wurde. Sie stört insbesondere dann, wenn sie aussagekräftige und valide Ergebnisse zu generieren vermag, potenziell die Schutzmechanismen gegen Inkonsistenzen zwischen unterschiedlichen Umwelтанforderungen und -erwartungen, die von institutionalisierten Organisationen (im neoinstitutionalistischen Begriffsverständnis – siehe Einleitung) oft nicht einfach zurückgewiesen werden können. Inwieweit kann man im entsprechenden Evaluationsfeld als Evaluator:in ausreichend interessenssensibel agieren, in welchem Ausmaß ist dies zugleich zulässig, ohne essenzielle Grundsätze wissenschaftlichen Arbeitens und forschungsethische Standards zu verletzen? Diese Fragen sind nicht nur innerhalb des Forschungsteams abzuwägen, sondern auch möglichst bald (genau genommen von Beginn an) mit Auftraggeber:innen der Wirkungsevaluation und den Evaluierten zu erörtern und klären.

Veröffentlichen bedeutet immer, Ergebnisse ein großes Stück weit aus der Hand zu geben – sie mitunter auch in eine politische Kampfarena zu ‚werfen‘. Auch wenn Forscher:innen mitreflektieren müssen, was dann mit den Ergebnissen gemacht wird, und Risiken einer eventuellen Instrumentalisierung – von welcher Seite auch immer – teilweise ein Stück weit vorgebeugt werden können, lassen sich die Verwertungsprozesse nur begrenzt steuern. Bei Entscheidungen für oder gegen eine Veröffentlichung braucht es deshalb Antworten auf folgende Fragen: Inwieweit ist es in bestimmten Konstellationen vertretbar oder gar erforderlich, den Evaluierten geschützte Bereiche für Lern- und Entwicklungsprozesse zu ermöglichen? Wie gut lässt sich dieses – teilweise sehr nachvollziehbare – Interesse an einem geschützten Rahmen mit einem allgemeinen Interesse an Transparenz, an einer bestmöglichen Leistungserbringung und auch einem verantwortungsvollen Ressourceneinsatz vereinbaren? Welche anderen wichtigen Interessen könnten durch eine Nicht-Veröffentlichung unter Umständen verletzt werden?

4. Zum Schluss: Chancen elaborierter Wirkungsevaluation

In einem letzten Schritt lässt sich der Bogen hin zum ersten Fragenfeld schlagen und prüfen, inwieweit mit den erzielten Wirkungserkenntnissen die damit verfolgten Funktionen erfüllt werden können. Zugleich ist darauf zu verweisen, dass sich die zahlreichen aufgeworfenen Fragen des „Reflexionsreigens“ abstrakt nur ungenügend beantworten lassen, sie können nur auf den konkreten Fall bezogen geklärt und entschieden werden. Die zahlreichen Herausforderungen und Spannungsfelder, die mit Wirkungserfassung und -evaluation einhergehen und oben aufgezeigt wurden, sollen zudem nicht davon abhalten, sich mit den Wirkungen Sozialer Arbeit vertiefend auseinanderzusetzen und aussagekräftige Wirkungsevaluationen anzustreben. Denn, so ist abschließend mit Nachdruck zu betonen, elaborierte (und zumeist methodenplurale) Wirkungsevaluation führt nicht nur Risiken mit sich, sie eröffnet auch beachtliche Chancen.

Wirkungsevaluation bietet im besten Fall die Möglichkeit, gegenüber Dritten nachdrücklich zu vermitteln, worin die spezifische Beschaffenheit der jeweiligen Maßnahmen bzw. Interventionen besteht, d.h. was die ihnen eigene Qualität ist und warum diese Qualität wichtig ist (Mayrhofer, 2018). Dies wird immer bedeutsamer, je mehr Zielerreichungen nach außen in fundierter Weise nachgewiesen werden müssen – und nicht nur beispielsweise eher zeremoniell über Selbstbeschreibungen im Jahresbericht. Je weniger es also gelingt, die benötigten Freiräume für das professionelle Wirken durch Intransparenz nach außen sicherzustellen, desto mehr ist es notwendig, die eigenen fachlichen Kriterien und die eigentliche professionelle Leistung überzeugend zu vermitteln. Komplexe Wirkungsevaluation kann die Chance zur Kommunikation der eigenen Leistung nach außen bieten und im besten Fall zu einem besseren Verständnis und zu höherer Akzeptanz für diese Leistung in den relevanten Umwelten beitragen.

Literatur

- Albus, S. & Ziegler, H. (2013). Wirkungsforschung. In G. Graßhoff (Hrsg.), Adressaten, Nutzer, Agency. Akteursbezogene Forschungsperspektiven in der Sozialen Arbeit (S. 163-180). Wiesbaden: Springer VS.
- Baecker, D. (1994). Postheroisches Management: Ein Vademecum. Berlin: Merve.
- Beywl, W., Speer, S. & Kehr, J. (2004). Wirkungsorientierte Evaluation im Rahmen der Armuts- und Reichtumsberichterstattung. Perspektivstudie. (Forschungsbericht Bundesministerium für Arbeit und Soziales). Köln. Univation – Institut für Evaluation Dr. Beywl & Associates GmbH; Bundesministerium für Gesundheit und Soziale Sicherung. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-316312> (letzter Zugriff: 12.06.2022).
- Borrmann, S. & Thiessen, B. (Hrsg.). (2016). Wirkungen Sozialer Arbeit. Potentiale und Grenzen für Evidenzbasierung für Profession und Disziplin. Opladen, Berlin, Toronto: Barbara Budrich.
- Bortz, J. & Döring, N. (2003). Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler (3. Aufl.). Berlin, Heidelberg: Springer.
- Bundesministerium für Finanzen. (n. d.). 2. Etappe der Haushaltsrechtsreform. Verfügbar unter <https://www.bmf.gv.at/themen/budget/haushaltsrechtsreform/zweite-etappe-haushaltsrechtsreform.html> (letzter Zugriff: 12.06.2022).
- 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.
- Dachverband Wiener Sozialeinrichtungen (2013). Rahmenrichtlinie zur Qualitätssicherung für die vom Fonds Soziales Wien anerkannten und geförderten Einrichtungen der Wiener Wohnungslosenhilfe. Verfügbar unter http://wohnen.fsw.at/downloads/dokumente/Rahmenrichtlinie_Qualitaetssicherung_WWH.pdf (letzter Zugriff: 05.09.2014)
- Drepper, T. (2010). Soziale personenbezogene Dienstleistungsorganisationen aus neoinstitutionalistischer Perspektive. In T. Klatetzki (Hrsg.), Soziale personenbezogene Dienstleistungsorganisationen. Soziologische Perspektiven (S. 129-165). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Goffman, E. (2003 [1959]). Wir alle spielen Theater: Die Selbstdarstellung im Alltag. München: Piper.
- Halfar, B., & Hegenauer, T. (2010). Wirkungsorientiertes Non Profit Organisation-Controlling. Controller-Magazin, 3, 87-92. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-330292> (letzter Zugriff: 12.06.2022).
- Haunberger, S. & Baumgartner, E. (2017). Wirkungsevaluierungen in der Sozialen Arbeit. Zeitschrift für Evaluation, 16, 121-145.
- Hüttemann, M. (2011). Effekthascherei oder wirklicher Fortschritt? Ein Kommentar zur Wirkungsorientierung in Forschung und Praxis der Sozialen Arbeit. In N. Eppler, I. Miethe & A. Schneider (Hrsg.), Qualitative und Quantitative Wirkungsforschung. Ansätze, Beispiele, Perspektiven (S. 53-67). Opladen, Berlin, Farmington Hills, MI: Barbara Budrich.
- Hüttemann, M. (2016). Wissensproduktion und Wissensverwendung in Projekten – kooperative Wissensbildung als Alternative zu evidenzbasierter Praxis? In S. Borrmann & B. Thiessen (Hrsg.), Wirkungen Sozialer Arbeit. Potentiale und Grenzen der Evidenzbasierung für Profession und Disziplin (S. 42-56). Opladen, Berlin, Toronto: Barbara Budrich.
- James, S. (2016). Inside the belly of the beast. Möglichkeiten und Grenzen der evidenzbasierten Praxis. In S. Borrmann & B. Thiessen (Hrsg.), Wirkungen Sozialer Arbeit. Potentiale und Grenzen der Evidenzbasierung für Profession und Disziplin (S. 143-160). Opladen, Berlin, Toronto: Barbara Budrich.
- Kannonier-Finster, W. & Ziegler, M. (1998). Methodologische Überlegungen zur Evaluierung sozialer Dienste. Störfaktor, 42, 37-57. Verfügbar unter http://bidok.uibk.ac.at/library/finster_ziegler-evaluierung.html (letzter Zugriff: 12.06.2022).

- Klatetzki, T. (2010). Zur Einführung: Soziale personenbezogene Dienstleistungsorganisation als Typus. In T. Klatetzki (Hrsg.), Soziale personenbezogene Dienstleistungsorganisationen. Soziologische Perspektiven (S. 7-24). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Luhmann, N. (1964). Funktionen und Folgen formaler Organisation. Berlin: Duncker & Humblot.
- May, M. (2011). Wirkung und Qualität in den verschiedenen Ansätzen quantitativer und qualitativer Evaluationsforschung. In N. Eppler, I. Miethe & A. Schneider (Hrsg.), Qualitative und Quantitative Wirkungsforschung. Ansätze, Beispiele, Perspektiven (S. 33-53). Opladen, Berlin, Farmington Hills, MI: Barbara Budrich.
- Mayrhofer, H. & Schachner, A. (2013). Partizipative Fotobefragung in der Evaluationspraxis: Möglichkeiten und Grenzen des Arbeitens mit anspruchsvollen qualitativen Verfahren am Beispiel einer Projektevaluierung im Behindertenbereich [122 Absätze]. Forum Qualitative Sozialforschung / Forum: Qualitative Social Research, 14, Art. 9. Verfügbar unter <http://nbn-resolving.de/urn:nbn:de:0114-fqs130296> (letzter Zugriff: 12.06.2022).
- Mayrhofer, H. (2014). Krise des Widerspruchsmanagements: Wirkungsorientierung in der Sozialen Arbeit aus neoinstitutionalistischer Theorieperspektive. Österreichische Zeitschrift für Soziologie, 3, 281-305.
- Mayrhofer, H. (2016). Biographische Fallrekonstruktionen in der Wirkungsevaluation. Erfahrungen aus einem Forschungsprojekt zu den Wirkweisen Mobiler Jugendarbeit. In S. Borrmann & B. Thiessen (Hrsg.), Wirkungen Sozialer Arbeit. Potentiale und Grenzen der Evidenzbasierung für Profession und Disziplin (S. 185-200). Opladen, Berlin, Toronto: Barbara Budrich.
- Mayrhofer, H. (2018). Wirkungsevaluation als Chance. Methodische Zugänge und Wirkerkenntnisse zur mobilen Jugendarbeit. Sozial Extra, 42, 50-54.
- Mayrhofer, H. (Hrsg.). (2017). Wirkungsevaluation mobiler Jugendarbeit. Methodische Zugänge und empirische Ergebnisse. Opladen, Berlin, Toronto: Barbara Budrich.
- Menold, N. (2007). Methodische und methodologische Aspekte der Wirkungsmessung. In P. Sommerfeld & M. Hüttemann (Hrsg.), Evidenzbasierte Soziale Arbeit. Nutzung von Forschung in der Praxis (S. 26-39). Baltmannsweiler: Schneider.
- Meyer, J. W. & Rowan, B. (1991 [1977]). Institutionalized Organizations: Formal Structure as Myth and Ceremony. In W.W. Powell & P. J. DiMaggio (Hrsg.), The New Institutionalism in Organizational Analysis (S. 41-62). Chicago: University of Chicago Press.
- Meyer, W. (2007). Messen: Indikatoren – Skalen – Indizes – Interpretationen. In R. Stockmann (Hrsg.) Handbuch zur Evaluation: Eine praktische Handlungsanleitung (S. 195-222). Münster: Waxmann.
- Micheel, H-G. (2013). Methodische Aspekte der Wirkungsforschung. In G. Graßhoff (Hrsg.), Adressaten, Nutzer, Agency. Akteursbezogene Forschungsperspektiven in der Sozialen Arbeit (S. 181-193). Wiesbaden: Springer VS.
- Müller, C. E. (2017). Kausale Wirkungsevaluation zwischen methodischem Anspruch und empirischer Praxis. In R. Stockmann & W. Meyer (Hrsg.), Die Zukunft der Evaluation. Trends, Herausforderungen, Perspektiven (S. 205-222). Münster, New York: Waxmann.
- Otto, H-U., Polutta, A. & Ziegler, H. (2010). Zum Diskurs um evidenzbasierte Soziale Arbeit. In H-U. Otto, A. Polutta & H. Ziegler (Hrsg.), What Works - Welches Wissen braucht die Soziale Arbeit. Zum Konzept evidenzbasierter Praxis (S. 7-25). Opladen, Farmington Hills, MI: Barbara Budrich.
- Otto, H-U. (2007). What works. Zum aktuellen Diskurs um Ergebnisse und Wirkungen im Feld der Sozialpädagogik und Sozialarbeit - Literaturvergleich nationaler und internationaler Diskussion. Berlin: Arbeitsgemeinschaft für Jugendhilfe.
- Rauscher, O., Schober, C. & Millner, R. (2012). Social Impact Measurement und Social Return on Investment (SROI) – Analyse. Wirkungsmessung neu? (Working Paper). Wien: NPO-Kompetenzzentrum. Verfügbar unter https://www.wu.ac.at/fileadmin/wu/d/cc/npocompetence/working_paper_social_impact_measurement_vs_sroi-analyse.pdf (letzter Zugriff: 12.06.2022).
- Reade-Soh, N. & Stockmann, R. (2009). Wirkungsorientierung und Evaluierung in der Entwicklungszusammenarbeit. Berlin: Berlin Institut für Bevölkerung und Entwicklung. Verfügbar unter http://www.forschungsnetzwerk.at/downloadpub/2009_Reade_Stockmann_EvaluationEZ.pdf (letzter Zugriff: 12.06.2022).

- Schneider, A. (2011). Professionelle Wirkung zwischen Standardisierung und Fallverstehen: Zum Stand der Wirkungsforschung. In N. Eppler, I. Miethe & A. Schneider (Hrsg.), *Qualitative und Quantitative Wirkungsforschung. Ansätze, Beispiele, Perspektiven* (S. 13-32). Opladen, Berlin, Farmington Hills, MI: Barbara Budrich.
- Sommerfeld, P. (2016). Evidenzbasierung als ein Beitrag zum Aufbau eines professionellen Wissenskorpas in der Sozialen Arbeit. In S. Borrman & B. Thiessen (Hrsg.), *Wirkungen Sozialer Arbeit. Potentiale und Grenzen der Evidenzbasierung für Profession und Disziplin* (S. 21-41). Opladen, Berlin, Toronto: Barbara Budrich.
- Stockmann, R. (2007). *Handbuch zur Evaluation – eine praktische Handlungsanleitung*. Münster: Waxmann.
- Willke, H. (2001). *Systemtheorie III: Steuerungstheorie* (3. Aufl.). Stuttgart: Lucius & Lucius.