

Ist der experimentelle Goldstandard wirklich Goldwert für eine Evidenzbasierung der Praxis Früher Hilfen?

Frühe Hilfen und die Forderung nach Wirkungsnachweisen

Der aktuellen Begriffsbestimmung Früher Hilfen, wie sie das „Nationale Zentrum Frühe Hilfen (NZFH)“ vorlegt, zufolge, sollten Frühe Hilfen „lokale und regionale Unterstützungssysteme mit koordinierten Hilfsangeboten für Eltern und Kinder ab Beginn der Schwangerschaft und in den ersten Lebensjahren mit einem Schwerpunkt auf der Altersgruppe der 0- bis 3-Jährigen“ sein. Ziel dieser Hilfen sei es, „Entwicklungsmöglichkeiten von Kindern und Eltern in Familie und Gesellschaft frühzeitig und nachhaltig zu verbessern [... und dabei neben einer] alltagspraktische[n] Unterstützung [...] einen Beitrag zur Förderung der Beziehungs- und Erziehungskompetenz von (werdenden) Müttern und Vätern [zu] leisten [...] zum gesunden Aufwachsen von Kindern [...] beizutragen und] deren Rechte auf Schutz, Förderung und Teilhabe [zu sichern]“ [1]. Obwohl im Bereich der Frühen Hilfen verschiedene Organisationen aus dem Gesundheits- und Sozialwesen (zusammen)arbeiten, ist die Kinder- und Jugendhilfe empirisch die Hauptakteurin in der Praxis Früher Hilfen [2].

Gerade die Frühen Hilfen stellen ein Handlungsfeld dar, in dem die Kinder- und Jugendhilfe verstärkt mit der Forderung konfrontiert ist, die Effektivität und Effizienz ihrer Leistungen unter Beweis zu stellen. Zumal Frühe Hilfen Eingriffe in Familienkonstellationen und die Lebensführung von Eltern darstellen, ist ihre Legitimation über den Nachweis ih-

rer Wirksamkeit angebracht. Zwar ist in diesem fachlich und politisch umstrittenen Feld nicht jede Maßnahme, die gewünschte Wirkungen entfaltet, automatisch angemessen. Umgekehrt gilt jedoch allemal, dass Maßnahmen, die keine Nutzwerte im Sinne positiver Effekte auf die Zustände und Lebensführungen ihrer AdressatInnen zeigen, nicht zu rechtfertigen sind.

Zumindest in der Kinder- und Jugendhilfe ist die Wirkungsorientierung dennoch nicht nur das Ergebnis einer fachlichen Debatte. Sie ist auch mit dem verwaltungstechnischen, politischen und moralischen Programm einer neuen wettbewerbsorientierten Steuerung öffentlicher Leistungen verknüpft. Mit der sogenannten Wirkungsorientierung richtet sich diese Steuerung über eine Rationalisierung und Optimierung von Verwaltungsabläufen hinaus auch auf die pädagogischen Prozesse der unmittelbaren Erbringung sozialer Dienste, die bislang weitgehend im Ermessen und der Entscheidungskompetenz von Wohlfahrtsprofessionellen lagen.

Während die gegenwärtigen Versuche einer neuen Steuerung sozialer Dienste in einem hohen Maße durch den Zweifel an der Effektivität und Effizienz professioneller Handlungs- und Entscheidungsautonomie gekennzeichnet sind, wird von Wirkungsevaluationen erwartet, auf wissenschaftlicher Evidenz basierende Ergebnisse bereitzustellen, die es erlauben, drängende praktische Probleme zu lösen und eine effektive Praxis anzuleiten. Unausgesprochen scheint dabei eine manualisierte Erbringung von Leistungen – das

heißt eine spezifische, formalisierte Festlegung und Prozessanweisung für die einzelnen Schritte der Leistungserbringung –, wie sie etwa in der evidenzbasierten Humanmedizin an Bedeutung gewinnt, als Vorbild zu dienen.

Die Frage, ob und inwiefern eine Manualisierung der Praxis der Kinder- und Jugendhilfe fachlich angemessen und zielführend ist, ist jedoch hoch umstritten [3, 4]. Die Debatte um den methodischen Goldstandard „Wirkungsforschung“ [5], das heißt um die randomisierte Experimentalforschung, ist mit dieser Frage verbunden. Während sich im medizinischen Feld seit Längerem eine Diskussion um die Aussagekraft der Ergebnisse von Kontrollexperimenten mit Blick auf die reale Lebenswelt beziehungsweise um naturalistische klinische Settings findet¹ und es einige – methodisch umstrittene [9] – Versuche gibt, die Ergebnisse der randomisierten Kontrollforschung mit naturalistischen Studien zusammenzuführen, ist

¹ So wollen etwa David Sackett et al. [6] unter der evidenzbasierten Medizin die Verknüpfung der besten verfügbaren externen Evidenz, das heißt der Ergebnisse von Experimentalforschungen mit individueller klinischer Expertise, verstanden wissen. Methodisch ist diese Forderung mit dem Problem verknüpft, dass damit genau jene interne Validität der Wirkungsaussagen infrage gestellt wird (siehe unten), deren Sicherstellung gerade die zentrale Stärke der randomisierten Experimentalforschung ist [7]. In der sozialwissenschaftlichen Debatte findet sich diesbezüglich häufig die eher pessimistische Position, dass das Verhältnis von „fidelity to the evidence base versus innovation“ eine nicht auflösbare „tensions inherent in evidence-based programmes“ darstelle ([8], S. 5).

die Debatte um die Angemessenheit, den Nutzen und die Grenzen der Experimentalforschung in der Kinder- und Jugendhilfe noch relativ neu.

Im folgenden Beitrag werden Zweifel an der Angemessenheit einer manualisierten Form evidenzbasierter Früher Hilfen erhoben. Zwar können die Aussagekraft und die Zuverlässigkeit methodisch anspruchsvoller Experimentalforschungen per se nicht ernsthaft bestritten werden, wohl ist aber zu bezweifeln, dass die Wirkungsbeschreibungen der Experimentalforschung eine hinreichende Grundlage sind, um die sozialarbeiterische Praxis in den Frühen Hilfen „anzuleiten“. Um die Zweifel zu begründen, wird zunächst ein Blick auf die grundlegende Frage geworfen, was die experimentelle Wirkungsmessung denn eigentlich misst.

Was misst die Wirkungsmessung?

Während es mittlerweile eine ganze Reihe von Studien gibt, die behaupten, Wirkungen der Kinder- und Jugendhilfe zu messen, findet sich kaum eine (deutschsprachige) Studie, die dieses Versprechen in einer methodisch validen Weise einlöst. Randomisierte, kontrollierte Experimentalstudien (RCTs) gehören demgegenüber zu jenen Studiendesigns, die das anspruchsvolle Versprechen halten können, belastbare Aussagen über die Wirkung solcher Maßnahmen zu formulieren. Dies ist auch dann kaum zu bestreiten, wenn man dem Einwand des Sachverständigenrat für die konzertierte Aktion im Gesundheitswesen folgt, dass „[r]andomisierte, kontrollierte, klinische Studien, die lediglich die Effektivität (und gegebenenfalls Effizienz) einer Maßnahme unter artifizielle Studienbedingungen (efficacy) beschreiben, [...] in vielen Fällen überbewertet [werden]“² ([10], S. 87).

RCTs basieren darauf, Ergebnisse von AdressatInnengruppen, die eine bestimmte Maßnahme erhalten haben, mit

den Ergebnissen von Gruppen zu vergleichen, die diese Maßnahme nicht erhalten. Sofern methodisch sichergestellt ist, dass die zwischen diesen Gruppen beobachteten Unterschiede nicht auf andere Einflüsse als durch die Maßnahme zurückzuführen sind, gelten diese als Wirkung der Maßnahme. Die experimentelle Wirkungsforschung überprüft also die Wahrscheinlichkeit, mit der auf ein Ereignis „A“ (zum Beispiel eine Frühe Hilfe) unter definierten Bedingungen ein Ereignis „B“ (die Wirkung dieser Hilfe) folgt. Die Wirkungsüberprüfungen von Maßnahmen auf Basis von RCTs stellen drei grundlegende Kriterien sicher: Erstens stellen sie sicher, dass das ursächliche Ereignis A der Wirkung (Ereignis B) zeitlich vorausgeht.³ Zweitens ermitteln RCTs, ob „A“ einen messbaren Einfluss auf „B“ hat. Das heißt, ob die konditionale Wahrscheinlichkeit von „B“ in Anbetracht von „A“ höher ist als die unbedingte Wahrscheinlichkeit von „B“ (Granger-Kausalität). Diese beiden Kriterien können nicht nur durch RCTs, sondern durch eine ganze Reihe unterschiedlicher Prä-Post-Designs zur Outcome-Analyse überprüft werden. Da es aber eine Reihe von bekannten oder unbekanntem Faktoren geben kann, die Einfluss beziehungsweise Wirkung entfalten, aber nur wenig mit der Maßnahme zu tun haben, besteht die besondere Leistung von RCTs drittens darin, sicherzustellen, dass sich keine anderen Einflüsse (sogenannte Störvariablen) finden, die den Zusammenhang von „A“ und „B“ statistisch irrelevant machen.

Die Stärke von RCTs besteht darin, dass der Einfluss von Störvariablen durch Verfahren der Randomisierung ausgeschlossen wird, das heißt, die TeilnehmerInnen werden zufällig auf eine Experimentalgruppe, die die zu untersuchende Maßnahme erhält, und eine Kontrollgruppe, die keine oder eine alternative Maßnahme erhält, verteilt. Bei einer ausreichend großen Stichprobengröße (namentlich sobald die Stichprobe groß genug ist, um

gemäß dem zentralen Grenzwertsatz⁴ davon auszugehen, dass die Wahrscheinlichkeiten der t-Verteilung kontinuierlicher Zufallsvariablen approximativ mit der Normalverteilung berechnet werden können, das heißt als Faustregel, ab $n > 30$) kann eine solche Zufallsaufteilung sicherstellen, dass sich die Störeinflüsse in der Experimental- und Kontrollgruppe gleich verteilen. Eine solche Gleichverteilung der Störvariablen sorgt dafür, dass auch deren systematischer Einfluss gleich verteilt und in diesem Sinne statistisch eliminiert wird. Diese statistische Eliminierung des Einflusses der Störvariablen gilt jedoch nur für die gesamte Gruppe und nicht für ihre jeweiligen Individuen. Die Wirkungsaussagen von RCTs treffen insofern nicht auf der Individualebene zu.

Was impliziert der Einsatz einer solchen experimentellen Wirkungsforschung nun für ein Praxisfeld wie die Frühen Hilfen im Kontext der Kinder- und Jugendhilfe? Aufschlussreich ist ein Blick in das Standardwerk zur Experimentalforschung in der Sozialwissenschaft: In „Experimental and Quasi-Experimental Designs for Generalized Causal Inference“ betonen William Shadish, Thomas Cook und Donald Campbell das Beschreiben der Folgen, die einer bewussten Behandlungsvariation zugeschrieben werden können, als einzigartige Stärke von RCTs [12]. RCTs liefern also valide Kausalbeschreibungen. Allerdings gelingt es ihnen weniger gut, über die Bedingungen und Mechanismen aufzuklären, die diesen Kausalbeziehungen zugrunde liegen. RCTs liefern also keine Kausalerklärungen. Eine Unterscheidung zwischen Kausalbeschreibungen und Kausalerklärungen ist gerade dann von sehr hoher praktischer Bedeutung, wenn es darum geht, Frühe Hilfen oder andere sozialpädagogische Leistungen „evidenzbasiert“ zu fundieren.

Die Relevanz von Kausalbeschreibungen ist enorm, wenn es darum geht festzustellen, ob ein bestimmtes Behandlungsprogramm eine bestimmte, erwartete Wirkung erreicht oder nicht erreicht

² Ob RCTs deswegen tatsächlich den „Goldstandard“ der Wirkungsforschung darstellen, ist gleichwohl strittig: Zumindest finden sich überzeugende methodische und methodologische Argumente dafür, dass sich neben RCTs eine Reihe weiterer Verfahren findet, die in einer validen Weise Kausalinferenzen beschreiben können [11].

³ Daher basieren RCTs auf mindestens zwei Messzeitpunkten: eine Messung vor der Maßnahme und eine nach der Maßnahme. Kann man mehr als zwei Messzeitpunkte heranziehen, ist es unter Umständen auch möglich, Wirkungsdynamiken und Verläufe sichtbar zu machen.

⁴ Der zentrale Grenzwertsatz lautet kurz gesagt, dass die Summen von stochastisch unabhängigen Zufallsvariablen annähernd normal verteilt sind.

hat.⁵ Wenn die Aufgabe einer Wirkungsevaluation in der bloßen Feststellung besteht, ob ein bestimmtes Programm – bemessen an einem klar definierten Endpunkt – wirksam oder unwirksam war und man darauf verzichten kann, die Zusammenhänge zu eruieren, die diesen Erfolg oder Misserfolg bedingen, fällt es schwer, sich eine zuverlässigere Methode als das randomisierte kontrollierte Experiment vorzustellen [14, 15, 16].

Es besteht also wenig Grund dafür, der Experimentalforschung per se skeptisch gegenüberzustehen. Allerdings ist der Nutzen von RCTs davon abhängig, wozu die Ergebnisse der Wirkungsforschung dienen sollen. Im Falle der Frühen Hilfen scheint das wesentliche Erkenntnisinteresse weniger darin zu bestehen festzustellen, in welchem Ausmaß eine bestimmte Maßnahme wirksam war, sondern vielmehr darin, wirksame Praktiken zu erkunden, um sie künftig durchzuführen.

Mein Argument lautet, dass die Kausalbeschreibungen von RCTs nur sehr bedingt dafür tauglich sind, die Praxis der Kinder- und Jugendhilfe anzuleiten. Gerade in einem Bereich, in dem ein hoher praktischer Bedarf besteht, nämlich in der Produktion validen Erklärungswissens, an dem sich die Frühen Hilfen strategisch orientieren könnten, liegt genau nicht die Stärke von RCTs.

⁵ Dies lässt sich gut am Beispiel des prominenten Elterntrainingsprogramms Triple P (Positive Parental Program) zeigen. Dieses Programm ist sehr erfolgreich vermarktet und verbreitet worden. Eine Reihe von Studien mit mäßiger Qualität hat dabei große Hoffnung über die Leistungsfähigkeit dieses Programms geschürt. Die Kausalbeschreibungen einer hochwertigen randomisiert kontrollierten Experimentalstudie zeichnen ein anderes Bild [13]: Das Triple-P-Programm, habe „keinerlei positive Wirkungen im Verhalten des Kindes ausgelöst, weder auf externalisierenden noch auf internalisierenden Verhaltensdimensionen“. Im Sozialverhalten der Kinder zeigte sich vielmehr „ein unerwünschter Effekt von Triple P. Die Lehrpersonen beobachteten bei Kindern von Eltern, die an einem Kurs teilgenommen hatten, tendenziell eine Zunahme von nicht-aggressivem Problemverhalten, während bei den Kindern der Kontrollgruppe ein Rückgang wahrgenommen wurde“ ([13], S. 5). Die Kausalbeschreibung dieser Effekte erscheint allemal ausreichend, um Triple P mit Skepsis zu begegnen.

Ohne auf die sozialetischen und theoretischen Kritiken einer auf experimenteller Wirkungsforschung fundierten Evidenzbasierung der Kinder- und Jugendhilfe im Einzelnen einzugehen [3], interessiert hier zunächst ein weitgehend unumstrittenes methodisches Argument: Die Kausalbeschreibungen von RCTs sind notorisch mit dem Problem „externer Validität“ konfrontiert⁶ [17].

Das Problem „externer Validität“ verweist weniger auf ein Studien-, sondern auf ein Situationskriterium, das sich auf die Verallgemeinerungsfähigkeit der Studienergebnisse bezieht. Gerade dass RCTs aus Gründen der Sicherstellung einer möglichst hohen internen Validität (im Sinne einer Validität der Ceteris-paribus-Bedingungen) methodisch darauf zielen, sowohl den Einfluss der Kontexte, in denen die Experimente stattfinden, als auch den der vermittelnden Faktoren zu minimieren, die zwischen Interventionen und Effekten intervenieren, bringt eine Reihe von Implikationen⁷ für die Wiederholung wirksamer Maßnahmen mit sich. Zwar klären RCTs über die Wahrscheinlichkeit auf, mit der definierte Maßnahmen, die gegenüber einer definierten Zielgruppe⁸ in einer ganz bestimmten Form und in einem ganz bestimmten Setting durchgeführt werden, im Vergleich zu anderen Maßnahmen oder einer Non-Intervention ein ganz bestimmtes Ziel erreichen, aber sie sind nicht in der Lage, die

⁶ Nancy Cartwright verweist auf den „familiar trade-off between internal and external validity. RCTs have high internal validity but the formal methodology puts severe constraints on the assumptions a target population must meet to justify exporting a conclusion from the test population to the target“ ([10], S. 11).

⁷ Diese Implikationen hängen nicht zuletzt mit dem sogenannten Black-Box-Problem zusammen [18], mit dem die Ergebnisse von Experimentalstudien konfrontiert sind: Sie geben Auskunft darüber, ob eine bestimmte Maßnahme wirkt, aber nicht darüber, warum diese wirkt.

⁸ Unter den artifiziellen Bedingungen von RCTs wird im Sinne der Therapietreue in der Regel auch die „Compliance“, das heißt das kooperative Verhalten des Patienten in Therapie, kontrolliert und sichergestellt. In der Praxis stellt die Sicherstellung von Compliance häufig ein größeres Problem dar, das zu Kompromissen (und das heißt auch zu „Verwässerungen“ des Interventionsprogramms) zwingt.

Bundesgesundheitsbl 2010 · 53:1061–1066
DOI 10.1007/s00103-010-1133-9
© Springer-Verlag 2010

H. Ziegler

Ist der experimentelle Goldstandard wirklich Gold wert für eine Evidenzbasierung der Praxis Früher Hilfen?

Zusammenfassung

Der folgende Beitrag fragt nach der Leistungsfähigkeit von experimentellen Wirkungsforschungen in der Praxis Früher Hilfen. Die Idee einer evidenzbasierten Neuausrichtung der Kinder- und Jugendhilfe auf Basis der Ergebnisse dieser Wirkungsforschung wird kritisiert. In einem abschließenden Abschnitt wird das Potenzial einer alternativen Form der Wirkungsforschung als Prozess-Mechanismus-Forschung mit Blick auf eine professionelle Praxis in der Kinder- und Jugendhilfe diskutiert.

Schlüsselwörter

Frühe Hilfen · Kinder- und Jugendhilfe · Wirkungsforschung · Evidenzbasierte Praxis · Professionalität

Is the experimental gold standard actually worth a mint for evidence-based early support systems?

Abstract

The article discusses the capacities of experimental impact research in the practice of early support systems or primary prevention. It criticizes the conception of evidence-based child and youth welfare services relying on results of such experimental research. In the final section, the potentials of an alternative form of impact research, such as process-mechanism research, are discussed with respect to professional practice in child and youth welfare.

Keywords

Early support systems · Child and youth welfare · Impact research · Evidence-based practice · Primary prevention

Bedingungen und generischen Mechanismen zu erfassen, die die Wirkung hervorbringen [19, 20]. Daher gelten die Kausalbeschreibungen von RCTs nur für die tatsächlich geprüften Kontexte und Programmgestaltungen. Eine Übertragung ihrer Ergebnisse auf andere Kontexte und Programmgestaltungen ist nicht möglich. In Struktur und Ausprägung veränderte Rahmenbedingungen machen die Aussagen der Kausalbeschreibung als „objektive“ Entscheidungsgrundlage wertlos.

Es ist deshalb folgerichtig, wenn der Wissenschaftliche Beirat „Psychotherapie“ in seinen im Juni 2009 verabschiedeten „Verfahrensregeln zur Beurteilung der wissenschaftlichen Anerkennung von Methoden und Verfahren der Psychotherapie“ standardisierte Diagnosen, ein „Therapiemanual, bei dem die Interventionen so beschrieben sind, dass das therapeutische Vorgehen vergleichbar und replizierbar ist“, eine „prospektive Festlegung und umfassende Beschreibung der Kontrollbedingung“ oder zumindest die „Ex-post-facto-Beschreibung der Kontrollbedingungen“ sowie ein Belegen der „Manualtreue/Treatmentintegrität durch externe Beobachter“ oder zumindest „durch Fragebögen“ verlangt ([21], S. 31).

Das Prinzip der Programmintegrität ist entscheidend, wenn man die Ergebnisse der experimentellen Wirkungsforschung zur Steuerung von Maßnahmen wie Frühen Hilfen verwenden möchte. Denn bei der Einführung evidenzbasierter Programme können bereits leichte Veränderungen (einschließlich „Ergänzungen“) der Programme oder ihrer Kontexte die Validität der Aussagen kausaldeskriptiver Wirkungsmodellberechnung als praktische Entscheidungsgrundlage erheblich reduzieren. Interventionen, so lautet daher die Forderung, „should be rigorously managed and delivered as designed“⁹ ([22], S. 27). Die Sicherstellung der Integrität und Replizierbarkeit der Programme erfolgt typischerweise über Manuale, die im Sinne schriftlich fixierter Prozessanweisungen den PraktikerInnen

vorschreiben, in einer definierten Situation A die Handlung B zu vollziehen.

Nimmt man die Prämissen und Gültigkeitskriterien der Experimentalforschung ernst, dann impliziert eine wirkungsorientierte Steuerung auf Basis von Kausalbeschreibungen der Wirkungsforschung nicht nur (transparente) Verfahrensstandards der Organisation, sondern eine Standardisierung der Kinder- und Jugendhilfepraxis im Sinne einer gradlinigen Realisierung manualisierter Vorgehensweisen. Eine solche technologische Applikation von Programmvorschriften steht im Widerspruch zur Betonung des fallinterpretativ-hermeneutischen Ermessens bei Handlungsentscheidungen, das bislang als Gütekriterium angemessener professioneller Praxis gilt.

Kritik der experimentalbasierten Wirkungsorientierung

Der Verweis auf eine solche Technologisierung der Praxis spricht nicht notwendigerweise gegen eine Neuausrichtung der Kinder- und Jugendhilfe auf Basis experimentell gewonnener Kausalbeschreibungen. Wenn der Nutzwert Früher Hilfen für ihre AdressatInnen auf diese Weise erhöht wird, haben möglicherweise schlicht die professionstheoretischen Prämissen ausgedient. Der Nutzen einer direkten Übertragung der Kausalbeschreibungen experimenteller Wirkungsforschung in die Praxis ist jedoch strittig. So sehen KritikerInnen die Gefahr, dass eine evidenzbasierte Praxis auf dieser Basis der Tatsache nicht gerecht wird, dass es die Kinder- und Jugendhilfe in der Regel nicht mit eindeutigen wissenschaftlich-empirisch auflösbaren, sondern mit widersprüchlichen, politisch und moralisch umkämpften Problemkonstellationen zu tun hat.¹⁰ Dies gilt gerade auch im Feld der Frühen Hilfen, in dem sich zum Beispiel das grundlegende Dilemma

von Unterstützung und Kontrolle in ein drücklicher Weise zeigt. Bei der Bearbeitung solcher Ambiguitäten erscheint neben dem Kriterium statistisch messbarer Wirksamkeit vor allem das nicht manualisierbare Kriterium der fall- und situationsbezogenen Angemessenheit wesentlich, das bei einer bloßen Anwendung formalisierter wirksamer Programme in den Hintergrund rückt.

Ein weiterer Einwand bezieht sich auf den Fokus der experimentellen Wirkungsforschung. Zwar ist es letztlich nicht der experimentellen Vorgehensweise per se geschuldet, aber dennoch kaum zu übersehen, dass im Mittelpunkt von RCTs in aller Regel die Wirkung einer spezifischen Methode beziehungsweise eines bestimmten Verfahrens steht. Dies gilt, obwohl spezifische manualisierbare Verfahren gerade im Bereich personenbezogener Humandienstleistungen nur einen bescheidenen Teil der Wirkungsvarianz erklären. Metaanalysen der Psychotherapieforschung sprechen für vergleichsweise geringe Effekte spezifischer Interventionen und eine hohe Effektstärke genereller Wirkfaktoren. Setzt man Bona-Fide-Therapien voraus, so klären spezifische Therapieverfahren in der Regel *ceteris paribus* zwischen 15% und 1% der Varianz der Ergebnisse auf [23, 24]. Aufgrund der vergleichsweise starken Einbettung in intersubjektive und lebensweltliche Bedingungskonstellationen und ihrer hohen Abhängigkeit von Fallverstehen und kontextadäquaten Vorgehensweisen ist kaum zu erwarten, dass der Effekt spezifischer Verfahren im Kontext der Frühen Hilfen durch die Kinder- und Jugendhilfe stärker sein wird als in der Psychotherapie.

Aus der bislang umfassendsten quasi-experimentellen Wirkungsstudie im Bereich der erzieherischen Hilfen im Kontext des Modellprogramms „Wirkungsorientierte Jugendhilfe“ [25] ist demgegenüber zu entnehmen, dass gut 30% der Wirkungsvarianz durch partizipative Kontexte, Beziehungsgestaltungen und Arbeitsbündnisse zu erklären ist. Diese „unspezifischen“ Wirkdimensionen haben eine Reihe („steuerbarer“) institutioneller Voraussetzungen und sind professionell gestaltbar. Allerdings ist die Gestaltung dieser Dimensionen fallspezifisch, von intersubjektiven Prozessen und Konstellati-

⁹ Selbst energische VertreterInnen einer „evidence-based practice“ konstatieren inzwischen, dass das Verhältnis „fidelity to the evidence base versus innovation“ eine der wesentlichen ungeklärten „tensions inherent in evidence-based programmes“ sei ([8], S. 5).

¹⁰ Dies hängt mit der erkenntnistheoretischen Einsicht zusammen, dass sich die Kinder- und Jugendhilfe auf menschliche Praktiken richtet, das heißt auf Gegenstände, die nicht nur – wie Naturereignisse – „Ursachen“ haben, sondern – als Handlungen – auf „Gründen“ basieren, die nur unter Berücksichtigung von Motiven und Gründen angemessen zu erfassen sind.

onen abhängig und insofern kaum sinnvoll manualisierbar.

Ein dritter Kritikstrang macht sich an der Tatsache fest, dass die experimentelle Wirkungsforschung nur standardisierbare Maßnahmen in kontrollierten Kontexten überprüfen kann. Wie ausgeführt basiert die Experimentalforschung auf der Sicherstellung, dass keine maßnahmeexternen Ereignisse und Kontexte die Wirkungsaussagen konfundieren. Die experimentelle Wirkungsmessung setzt hierzu methodische Verfahren ein, die letztlich dem Versuch gleichkommen, einen bestimmten Wirkzusammenhang von breiteren sozialen und institutionellen Kontexten zu isolieren. Das Bestreben, die Wirkungen von Kontextprozessen und anderen „Störvariablen“ zu eliminieren, kommt im Feld der Frühen Hilfen einem anti-sozialen Bias nahe [26]. Denn die Kontrolle dieser Einflüsse ist letztlich unmöglich, wenn die Interventionen überwiegend fall-spezifisch unterschiedlich gestaltete, soziale Prozesse sind, die in vielfältigen Kontexten stattfinden, die von den Planungs- und Administrationsstellen nicht vollständig zu kontrollieren sind [27].

Während zum Beispiel kognitiv-behaviourale Trainingsmaßnahmen auch bei einer Eliminierung von Kontexten Wirkung entfalten können, sind typische Vorgehensweisen der Kinder- und Jugendhilfe auf kaum standardisierbare soziale Prozesse, Kontexte und Konstellationen verwiesen. Werden diese als programmexterne Bedingungen durch Randomisierungen eliminiert und die Vorgehensweisen in manualisierter Form standardisiert, dann können solche Maßnahmen gar nicht mehr funktionieren. Der anti-soziale Bias ließe sich insofern auch im Sinne einer methodisch induzierten fehlerhaften Setzung der Nullhypothese, genauer, als „Type II error with regard to desirability of social interventions“ verstehen ([26], S. 275).

Die experimentelle Anstrengung „to cancel out difference“ und die Versuche die Wirkung von Maßnahmen von den weiteren Kontexten dieser Maßnahmen zu entkoppeln, wird von einigen sozialwissenschaftlichen ForscherInnen als absurde Anstrengung beschrieben „to write out what is essential to a program – social conditions favourable to it“ ([28], S. 52).

Auf die Relevanz sozialer Bedingungskonstellationen haben insbesondere jene Formen der „hard-to-do science“ einzugehen, zu denen im Anschluss an David C. Berliner auch die Forschungen im Bereich der Frühen Hilfen gerechnet werden können [29]. Die Schwierigkeit dieser Forschungen resultiert daraus, dass sich die Frühen Hilfen im hohen Maße auf partikulare, individual-biografische Problemlagen richten, die sich innerhalb lokaler lebensweltlicher Bedingungskonstellationen ereignen und in ihrer Sinnhaftigkeit erschließen. Dies limitiert ihre Generalisierung [30] und ihre Anschlussfähigkeit an Covering-law-Theorien¹¹, die den erkenntnistheoretischen Referenzrahmen jener experimentellen Erforschung von Wirksamkeit darstellen, die Kausalität im Anschluss an den Empiristen David Hume als „constant conjunction“ von Ereignissen modellieren.

Zum Potenzial einer alternativen Wirkungsforschung als „Prozess-Mechanismus“-Forschung

Diese Einwände schmälern in keiner Weise die Relevanz der Wirkungsforschung in der Kinder- und Jugendhilfe. Allerdings scheint hier eine Form der Wirkungsorientierung angemessen, die weniger darauf zielt, Praxis in der Anwendung von Methoden anzuleiten, die eine relative Erfolgswahrscheinlichkeit hinsichtlich spezifisch definierter Ziele versprechen, sondern vielmehr in der Lage ist, empirisch fundiertes Erklärungs- und Reflexionswissen auf Basis einer Analyse von „Kontext-Mechanismus-Ergebnis-Konfigurationen“ herauszuarbeiten [20, 28]. Diese Formen einer sogenannten „realistischen“

¹¹ Covering-law-Theorien, teilweise auch als Subsumtionsmodelle der Erklärung beschrieben, basieren auf der Annahme gleichermaßen universeller und empirisch gehaltvoller allgemeiner Gesetzmäßigkeiten, wie sie etwa in der Naturwissenschaft formuliert werden. Mit Blick auf die Wirkungsforschung besteht die Covering-law-Annahme darin, dass Ereignis A mit Ereignis B durch ein allgemeines Gesetz verbunden ist, das die entsprechenden Ereignisse „deckt“ („to cover“). Das deduktiv-nomologische Programm der Sozialwissenschaften in der Tradition von Hempel-Oppenheim geht explizit oder implizit von der Existenz solcher Covering laws aus.

Wirkungsforschung werden in der Sozialwissenschaft verstärkt diskutiert. Aufgabe der Wirkungsforschung wäre es, in diesem Fall Wirkmechanismen in Kenntnis zu setzen, um eine Grundlage für eine empirisch informierte, professionelle Prüfung der Angemessenheit einer Intervention bereitzustellen. Das Ergebnis solcher Wirkungsforschungen wären empirisch belastbare Theorien (mittlerer Reichweite) über Wirkungszusammenhänge, die es erlauben, professionelle Entscheidungen wirkungsorientiert zu fundieren

Diese Forschung müsste über einen seriellen Wirkungsbegriff einer „successionist causality“ hinausgehen [31]. Das Problem eines Begriffs von Wirkung, der das regelmäßige Zusammentreffen von unmittelbar beobachtbaren Ereignissen fokussiert, besteht darin, dass sich soziale Wirklichkeit nicht auf beobachtbare Eigenschaften von Objekten und Ereignissen reduzieren lässt. Forschungen auf Basis einer seriellen Modellierung von Wirkung können letztlich lediglich „Black-box-Erklärungen“ bereitstellen, „aber keine erklärungsdienlichen Mechanismen angeben, welche die beobachteten Zusammenhänge hervorbringen“ ([32], S. 103). Neuere sozialwissenschaftliche Konzeptionen von Wirkung gehen demgegenüber von einem „generativen“ Wirkungsbegriff aus, das heißt sie argumentieren, dass „X is a cause of Y if and only if there is a mechanism from X to Y“ ([33], S. 60), und unterstellen damit, dass Wirkungen eine „Folge des Wirkens der zugrunde liegenden Mechanismen¹² und nicht schlicht statistische Zusammenhänge [sind]“ ([34], S. 136). Da davon auszugehen ist, dass ein Wirkungsmechanismus unterschiedliche Effekte erzeugen kann und umgekehrt einem bestimmten Effekt verschiedene generative Ursachen zugrunde liegen können, erschließt sich der Zusammenhang zwischen Wirkungsmechanismen und Effekten darüber hinaus erst im Zusammenspiel mit spezifischen sozialen Kontextbedingungen. Daher reicht es „nicht aus Mechanismen

¹² Mechanismen sind „complexes of interacting individuals, usually classified into specific social categories, which generate causal relationships between aggregate-level variables“ ([33], S. 59).

[zu] benennen die einem empirischen Ereignis zugrunde liegen“ ([34], S. 136). Vielmehr gilt es, auch den Kontext zu spezifizieren, innerhalb dessen diese Mechanismen entstanden sind [34].

Für eine der Praxis Früher Hilfen angemessenen Form der Wirkungsforschung ist demnach die Frage entscheidend, wie bestimmte soziale Prozesse mit Ressourcenausstattungen, Gelegenheitsstrukturen, Handlungsbefähigungen und Handlungsentscheidungen der AkteurInnen sowie spezifischen sozialen Kontexten (zum Beispiel institutionalisierten Regeln, kulturellen Sinndeutungen, Machtungleichgewichten und so weiter) interagieren, vor deren Hintergrund sich die Effekte sozial wirksamer generativer Mechanismen realisieren. Eine solche Wirkungsforschung zielt darauf ab, Verursachungsprozesse und -mechanismen in ihrem Kontext zu analysieren und erklärende Deutungsangebote über Wirkungszusammenhänge zu machen. Ein empirisch fundiertes, nichtsdestoweniger theoretisches Wissen über Prozess-Mechanismus-Wirkungs-Konfigurationen taugt nicht zur manualisierten Anleitung der Praxis. Aber es kann einen Hintergrund für professionelle (Ermessens-)Entscheidungen darstellen, um evidenzbasiert zu begründen, welches Vorgehen mit Blick auf die je fallspezifischen Konstellationen und sozialen Einbettungen der AdressatInnen, angemessen ist und eine Korrespondenz in der Wirklichkeit jener emergenten Prozesse findet, die Professionelle zusammen mit den AdressatInnen bearbeiten und hervorbringen.

Ohne Zweifel können insbesondere quasi-experimentelle Modellierungen dabei eine wichtige Grundlage für Prüfungen von Kontext-Mechanismus-Ergebnis-Konfigurationen liefern. Sofern sich Wirkungsforschungen aber lediglich auf die Prüfung bestimmter standardisierbarer Verfahren reduzieren und versuchen, die Einflüsse von Kontexten und Mechanismen nicht systematisch zu erfassen, sondern methodisch zu eliminieren, besteht erstens die Gefahr, „das Falsche“ zu messen und eine Reihe gestaltbarer Prozesse systematisch auszuklammern, die für die Wirksamkeit der Praxis (mit)verantwortlich sind. Zweitens besteht die Gefahr eine managerielle Standardisierung und Technologisie-

rung der Praxis voranzutreiben, die hinter dem tatsächlichen Potenzial der Kinder- und Jugendhilfe zurückfällt, fallspezifisch angemessene und lebensweltlich nachhaltige Formen Früher Hilfen erbringen zu können.

Korrespondenzadresse

Prof. Dr. H. Ziegler

Fakultät für Erziehungswissenschaft, AG 8
Soziale Arbeit, Universität Bielefeld
Universitätsstr. 25, 33615 Bielefeld
holger.ziegler@uni-bielefeld.de

Interessenkonflikt. Der korrespondierende Autor gibt an, dass kein Interessenkonflikt besteht.

Literatur

1. Nationales Zentrum Frühe Hilfen (2009) Begriffsbestimmung „Frühe Hilfen“. <http://www.fruehehilfen.de/wissen/fruehe-hilfen-grundlagen/begriffsbestimmung>
2. Deutsches Institut für Urbanistik (2009) Bundesweite Bestandsaufnahme zu Kooperationsformen im Bereich Früher Hilfen. Ergebnisbericht der DIfU-Umfrage bei Jugend- und Gesundheitsämtern. Deutsches Institut für Urbanistik, Berlin
3. Otto H-U, Polutta A, Ziegler H (Hrsg) (2010) What Works – Welches Wissen braucht die Soziale Arbeit? Barbara Budrich, Opladen Farmington Hills I.E
4. Otto H-U, Polutta A, Ziegler H (2009) A second generation of evidence-based practice – reflexive professionalism and causal impact in social work. In: Otto H-U, Polutta A, Ziegler H (Hrsg) Evidence-based practice – modernising the knowledge base of social work? Barbara Budrich, Opladen Farmington Hills, S 245–252
5. Soydan H (2008) Towards the gold standard of impact research in social work – avoiding threats to validity. In: Otto H-U, Polutta A, Ziegler H (Hrsg) Evidence-based practice – modernising the knowledge base of social work? Barbara Budrich, Opladen Farmington Hills, S 111–137
6. Sackett D, Rosenberg W, Muir Gray J et al (1997) Was ist evidenzbasierte Medizin und was nicht? Munch Med Wochenschr 139:644–645
7. Farrington D (2003) Methodological quality standards for evaluation research. Ann Am Acad Pol Soc Sci 587:49–68
8. Nutley S, Homel P (2006) Delivering evidence-based policy and practice: lessons from the implementation of the UK Crime Reduction Programme. Evidence Policy 2:5–26
9. Tschuschke V (2005) Die Psychotherapie in Zeiten evidenzbasierter Medizin. Psychotherapeutenjournal 2:106–115
10. Sachverständigenrat für die Konzertierte Aktion im Gesundheitswesen (2003) Finanzierung, Nutzerorientierung und Qualität. Band I: Finanzierung und Nutzerorientierung. Band II: Qualität und Versorgungsstrukturen. Gutachten 2003. Kurzfassung, Bonn
11. Cartwright N (2007) Are RCTs the gold standard? Bio-Societies 2:11–20
12. Shadish W, Cook T, Campbell D (2002) Experimental and quasi-experimental designs for generalized causal inference. Houghton-Mifflin, Boston
13. Eisner M, Ribeaud D (2008) Das Zürcher Interventions- und Präventionsprogramm an Schulen „zipps“ – Eine Zusammenfassung der zentralen Ergebnisse. Bundesamt für Gesundheit, Bern

14. Boruch R (1997) Randomized experiments for planning and evaluation: a practical guide. Sage, Thousand Oaks
15. LaLonde R (1986) Evaluation the econometric evaluations of training programs with experimental data. Am Econ Rev 76:604–620
16. Pignotti M, Thyer B (2009) Why randomized clinical trials are important and necessary to social work practice. In: Otto H-U, Polutta A, Ziegler H (Hrsg) Evidence-based practice – modernising the knowledge base of social work? Barbara Budrich, Opladen Farmington Hills, S 99–109
17. Campbell D, Stanley J (1963) Experimental and quasi-experimental designs for research. Rand McNally, Chicago
18. Howe KR (2004) A critique of experimentalism. Qualitative Inquiry 10:42–61
19. Otto H-U, Ziegler H (2008) The notion of causal impact in evidence-based social work. Res Soc Work Pract 18:273–277
20. Pawson R, Tilley N (2009) Realist evaluation. In: Otto H-U, Polutta A, Ziegler H (Hrsg) Evidence-based practice – modernising the knowledge base of social work? Barbara Budrich, Opladen Farmington Hills, S 151–180
21. Wissenschaftliche Beirat Psychotherapie (2009) Methodenpapier des Wissenschaftlichen Beirats Psychotherapie nach § 11 PsychThG. Verfahrensregeln zur Beurteilung der wissenschaftlichen Anerkennung von Methoden und Verfahren der Psychotherapie. Version 2.7. <http://www.wbpsychotherapie.de/downloads/Methodenpapier2720090709.pdf>
22. Chapman T, Hough M (1998) Evidence based practice: a guide to effective practice. Home Office, London
23. Lambert M, Ogles B (2004) The efficacy and effectiveness of psychotherapy. In: Lambert M (Hrsg) Bergin and Garfield's handbook of psychotherapy and behavior change, 5. Aufl. John Wiley & Sons, New York, S 139–193
24. Wampold B (2001) The great psychotherapy debate. Models, methods and findings. Lawrence Erlbaum Associates, Mahwah
25. Albus St, Greschke H, Klingler B et al (2010) Wirkungsorientierte Jugendhilfe. Abschlussbericht der Evaluations des Bundesmodellprogramms „Qualifizierung der Hilfen zur Erziehung durch wirkungsorientierte Ausgestaltung der Leistungs-, Entgelt- und Qualitätsvereinbarungen nach §§ 78a ff SGB VIII. Waxmann, Münster
26. Hope T (2005) Pretend it doesn't work: the „anti-social“ bias in the Maryland Scientific Methods Scale. Eur J Criminal Policy Research 11:275–296
27. Hammersley M (2009) What is evidence for evidence-based practice? In: Otto H-U, Polutta A, Ziegler H (Hrsg) Evidence-based practice – modernising the knowledge base of social work? Barbara Budrich, Opladen Farmington Hills, S 139–150
28. Pawson R, Tilley N (1997) Realistic evaluation. Sage, London
29. Berliner D (2002) Comment: Educational research: the hardest science of all. Educational Researcher 31:18–20
30. Böttcher W, Dicke JN, Ziegler H (2009) Erziehungswissenschaft, Bildungspolitik und Bildungspraxis. Anmerkungen zu einem schwierigen Verhältnis. In: Böttcher W, Dicke JN, Ziegler H (Hrsg) Evidenzbasierte Bildung. Wirkungsevaluation in Bildungspolitik und pädagogischer Praxis. Waxmann, Münster, S 7–21
31. Harré R, Madden E (1975) Causal powers. Blackwell, Oxford
32. Schmid M (2006) Die Logik mechanistischer Erklärungen. VS Verlag für Sozialwissenschaften, Wiesbaden
33. Steel D (2004) Social mechanisms and causal inference. Philos Soc Sci 34:55–78
34. Mackert J (2006) Ohnmächtiger Staat? Über die sozialen Mechanismen staatlichen Handelns. VS Verlag für Sozialwissenschaften, Wiesbaden