

Evaluation und Komplexität

Wirkungskonstruktion in der Evaluation von Gesundheitsförderung und Prävention

Dissertation
zur Erlangung der Würde eines Doktors der Philosophie

vorgelegt der Philosophisch-Historischen Fakultät
der Universität Basel

von
Günter Ackermann

aus
Mümliswil-Ramiswil, Solothurn

Rheinfelden 2016

Genehmigt von der Philosophisch-Historischen Fakultät der Universität Basel, auf Antrag von Prof. Dr. Manfred Max Bergman und Prof. Dr. Sandro Cattacin.

Basel, den 28. Oktober 2015

Die Dekanin
Prof. Dr. Barbara Schellewald

Dank

Gemäss Niklas Luhmann (1992) ist wissenschaftliche Erkenntnis nicht im Subjekt verortet, sondern als kommunikativer sozialer Prozess zu verstehen. Entsprechend vielfältig und kaum nachvollziehbar sind die Einflüsse, die zur vorliegenden Arbeit geführt haben. Einige Beiträge waren aber sehr direkt und dafür möchte ich mich insbesondere bei folgenden Personen ganz herzlich bedanken:

Meinem Doktorvater Prof. Dr. Manfred Max Bergman danke ich dafür, dass er mein Forschungsvorhaben von Beginn weg wohlwollend unterstützt hat. Mit seiner Betreuung und seiner kritisch-differenzierten Sichtweise hat er meiner Arbeit einen hilfreichen Rahmen gegeben und mir gleichzeitig die Freiheit gelassen, den mich drängenden Fragen nachzugehen.

Dem Koreferenten Prof. Dr. Sandro Cattacin danke ich bestens für seine Gutachtertätigkeit.

Susanne Haab danke ich für die individuelle und sehr angenehm gestaltete forschungsmethodische Beratung.

Dr. Brigitte Ruckstuhl, Dr. Hanspeter Stamm und Martina Blaser danke ich für ihre Flexibilität und ihre kompetenten Kriseninterventionen, wenn ich mich selber in der Komplexität des Themas zu verlieren drohte.

Juliette Fong danke ich dafür, dass sie mich mit einfachen, entscheidenden Fragen immer wieder dazu herausgefordert hat, Komplexes zu reduzieren und die Dinge auf den Punkt zu bringen.

Prof. Dr. Thomas Abel danke ich für die gemeinsamen Arbeitstage und spätabendlichen Reflexionen mit Blick aufs Meer.

Dr. August Rohdewald danke ich herzlich für seine grosse Unterstützung bei der sprachlichen Bereinigung und beim Aufspüren von Helvetismen.

Für wertvolle Gespräche danke ich im Weiteren Prof. Dr. Peter Sommerfeld, Prof. Dr. Martin Hafen, Dr. Claudia Meier Magistretti, Petra Zeyen Bernasconi, Dr. Marlène Läubli-Loud, Dr. Hubert Studer sowie meinen Teamkolleginnen und -kollegen bei Gesundheitsförderung Schweiz: Lisa Guggenbühl, Dr. Kathrin Krause, Karin Clemann und Nicolas Sperisen.

Gesundheitsförderung Schweiz danke ich für die Möglichkeit, mein Promotionsstudium mit meiner Berufstätigkeit kombinieren zu können sowie für die Unterstützung im Rahmen eines Bildungsurlaubs.

Abstract

Projects and programs of health promotion and prevention are increasingly called upon to prove their effectiveness. Recent literature attests to a lively, ongoing debate about evidence and finding the “right” way of evaluating effectiveness in this field. Randomized controlled trials, considered the “gold standard” for assessing efficacy in evidence-based medicine, have been largely deemed insufficient for evaluating interventions of health promotion and prevention, since they fail to adequately account for the complexity of these interventions. The search for alternatives has prompted calls for approaches informed by sociology and systems theory.

This thesis comprehensively addresses the complexity of social systems and reflects on how to evaluate interventions’ effectiveness in light of such complexity, drawing on complexity- and sociological systems theory. In an empirical study rooted in Grounded Theory, nine evaluation reports of health promotion and prevention projects were subjected to in-depth content analysis, enabling identification of different strategies for handling complexity.

Proofs of effectiveness are revealed to be social constructions of evaluators, tasked with balancing the needs of mandating agencies, science, and practice. Concerning the construction of effectiveness, evaluators have significant room for maneuver. Ongoing decisions are made about how to frame, capture, and measure complexity as well as assess effectiveness. These decisions influence the construction of effectiveness and thus any resulting proven impacts. Impact assessments are continually caught between demands for simplicity, risking neglect of key points, and for complexity, which cannot be fully captured. The present thesis offers a detailed analysis of the spectrum of approaches and illustrates the strategies employed by evaluators to broadly capture complexity, on the one hand, and to strongly reduce it, on the other.

When conducting evaluations, evidence of effectiveness is still frequently derived from randomized controlled trials, but not exclusively. In addition to counterfactual comparisons, statements about effectiveness are tested by means of detailed reconstruction of possible causal relationships. Well-implemented (quasi-) experimental studies can generate valuable knowledge about the potential impact of interventions. However, they are of limited usefulness in decision-making and action on behalf of health promotion and prevention, since they seldom enable conclusions about the possible effects of target group-, context-, or intervention-specific variables, and fail to identify differentiated mechanisms of change or possible side effects. Complexity-sensitive evaluations can offer support in identifying distinct mechanisms of change, enabling sound statements about interventions’ effectiveness and differentiated management-related knowledge. Based on theoretical and empirical insights, the present paper recommends adopting a middle way between positions that maintain experimental studies as the “gold standard” and those that deem such studies unsuitable for assessing the effectiveness of health promotion and prevention. To enable robust and plausible impact assessments, counterfactual comparisons and detailed reconstructions of cause–effect relationships should be combined. Iterative processes of expanding and reducing complexity show promise in identifying impact-relevant structural and process-related features, capturing intermediate influences, estimating interventions’ contribution to observed changes, and ultimately generating plausible overarching conclusions about the effectiveness of a given program.

Abstract

Die Gesundheitsförderung und Prävention steht zunehmend unter Druck, die Wirksamkeit ihrer Projekte und Programme nachweisen zu müssen. In der einschlägigen Literatur wird derzeit eine engagierte Debatte über Evidenz und die ‚richtige‘ Evaluationsform für Wirkungsbeurteilungen in diesem Handlungsfeld geführt. Randomisierte kontrollierte Studien, welche sich in der evidenzbasierten Medizin als ‚Goldstandard‘ der Wirksamkeitsforschung etabliert haben, werden für die Evaluierung von Interventionen der Gesundheitsförderung und Prävention weitgehend als unangemessen zurückgewiesen, weil sie der Komplexität der Interventionen nicht gerecht würden. Auf der Suche nach Alternativen zum Forschungsverständnis der evidenzbasierten Medizin werden Rufe nach soziologischen und systemtheoretischen Ansätzen laut.

Diese Arbeit befasst sich eingehend mit der Komplexität sozialer Systeme und reflektiert auf der Grundlage der Komplexitäts- und der soziologischen Systemtheorie den Umgang von Wirkungsevaluation mit der Komplexität der von ihr untersuchten sozialen Systeme. In einer an die Grounded Theory angelehnten empirischen Untersuchung wurden neun Schlussberichte über Evaluationen von Gesundheitsförderungs- und Präventionsprojekten vertieft inhaltsanalytisch ausgewertet und unterschiedliche Strategien des Umgangs mit Komplexität herausgearbeitet.

Wirkungsnachweise zeigen sich dabei als soziale Konstruktionen der Evaluation im Spannungsfeld zwischen Auftrag Gebenden, der Wissenschaft und der Praxis. Bei der Wirkungskonstruktion hat die Evaluation einen grossen Handlungsspielraum. Sie fällt bei der Eingrenzung, Erfassung und Verarbeitung von Komplexität sowie bei der Wirkungsbeurteilung laufend Entscheide, welche die Wirkungskonstruktion und damit die nachgewiesenen Wirkungen beeinflussen. Wirkungsevaluation bewegt sich immer im Spannungsfeld zwischen Einfachheit, die Wichtiges vernachlässigt, und Komplexität, die nicht mehr zu bewältigen ist. In der Arbeit wird dieses Handlungsspektrum detailliert herausgearbeitet und es wird gezeigt, mit welchen Strategien Evaluationen Komplexität einerseits breit erfassen und andererseits stark reduzieren.

Wirkungsnachweise werden in der Evaluationspraxis auch, aber nicht alleine über randomisierte kontrollierte Studien erbracht. Neben kontrafaktischen Vergleichen werden Wirkungsaussagen vor allem mittels detaillierter Rekonstruktion von Wirkungszusammenhängen begründet. Gut gemachte (quasi-) experimentelle Studien können auf einer übergeordneten Ebene wertvolle Hinweise auf das Wirkungspotenzial von Interventionen geben. Da sie allerdings kaum Rückschlüsse auf die Wirkungsrelevanz zielgruppen-, kontext- oder interventionsspezifischer Besonderheiten ermöglichen und keine Erkenntnisse über differenzierte Wirkungsmechanismen und über nicht vorhersehbare Nebenwirkungen liefern, sind sie für fundiertes Entscheiden und Handeln in der Gesundheitsförderung und Prävention nur begrenzt tauglich. Hier bieten sich komplexitätssensible Evaluationen an, um über die Herausarbeitung von Wirkungsmechanismen fundierte Aussagen über die Wirksamkeit von Interventionen und differenziertes steuerungsrelevantes Wissen zu generieren. Auf der Basis der theoretischen und empirischen Erkenntnisse wird hier für einen Mittelweg plädiert zwischen Positionen, die experimentelle Studien grundsätzlich als ‚Goldstandard‘ für Wirkungsevaluationen proklamieren und Positionen, die diese Studien als unangemessen für die Gesundheitsförderung und Prävention zurückweisen. Um zu aussagekräftigen und möglichst plausiblen Wirkungsbeurteilungen zu gelangen, bietet es sich an, kontrafaktische Vergleiche und detaillierte Rekonstruktionen von Wirkungszusammenhängen zu kombinieren. Iterative Prozesse der Ausweitung und Reduktion von Komplexität erscheinen vielversprechend, um wirkungsrelevante Struktur- und Prozessmerkmale zu erkennen, intermediäre Wirkungsbeiträge zu erfassen, den Beitrag einer Intervention zu einer beobachteten Gesamtveränderung abzuschätzen und so zu einer plausiblen übergeordneten Wirkungsbeurteilung zu gelangen.

Inhaltsverzeichnis

1	Einleitung	11
2	Die Evidenzdebatte in der Gesundheitsförderung	15
2.1	Evidenz in der ‚Evidence Based Medicine‘	15
2.2	Randomisierte kontrollierte Studien im Fokus der Kritik.....	17
2.3	Das Evidenzdilemma in der Gesundheitsförderung	21
2.4	Erweiterung des Evidenzbegriffs	25
2.5	Konsequenzen für die Evaluation von Gesundheitsförderung und Prävention.....	28
3	Komplexität	31
3.1	Eine historische Annäherung an Komplexität.....	31
3.2	Die Komplexität sozialer Systeme	37
3.3	Merkmale komplexer sozialer Systeme.....	42
3.4	Die Komplexität der Gesundheitsförderung	51
3.5	Intervenieren in komplexen sozialen Systemen	55
4	Erkenntnis im Lichte von Komplexität	67
4.1	Systemtheoretischer Konstruktivismus	67
4.2	Komplexitätsreduktion aus systemtheoretisch-konstruktivistischer Sicht	73
4.3	Evaluation als beurteilendes soziales System	74
4.4	Wirkungen erkennen	75
4.5	Generalisierung von Wirkungen	82
4.6	Fazit	85
5	Methodik der empirischen Untersuchung	89
5.1	Wahl und Begründung der Analysemethodik zur Rekonstruktion der Evaluationspraxis	89
5.2	Der Analyseprozess	92
5.3	Potenzial und Grenzen der Untersuchung	102
6	Wirkungskonstruktion in Evaluationsberichten	107
6.1	Komplexität eingrenzen	108
6.2	Komplexität erfassen	119
6.3	Komplexität verarbeiten.....	132
6.4	Wirkungen beurteilen.....	147
6.5	Synthese.....	157

7	Umgang der Evaluationen mit Komplexität	161
7.1	Evaluationsspezifische Analyse	161
7.2	Interaktionsvielfalt	193
7.3	Komplexe Wirkungsmechanismen	198
7.4	Selbstorganisation	205
7.5	Unvorhersehbarkeit	208
7.6	Wirkungsbeurteilung	212
7.7	Synthese	219
8	Zusammenfassung, Schlussfolgerungen und Ausblick	225
8.1	Zusammenfassende Beantwortung der Forschungsfrage	225
8.2	Reflexionen zur Angemessenheit des Komplexitätsgrades in Evaluationen	227
8.3	Umriss einer komplexitätssensiblen Evaluation	232
8.4	Schlussfolgerungen für die Evaluation von Gesundheitsförderung und Prävention	236
8.5	Synthese	241
8.6	Ausblick	244
9	Anhang	247
9.1	Begriffe	247
9.2	Merkmale komplexer sozialer Systeme	250
9.3	Codesystem	251
9.4	Auszug Arbeitstabelle axiales Kodieren	255
9.5	Literaturverzeichnis	258

Abbildungsverzeichnis

Abb. 1	Die vier Bezugspunkte der Untersuchung.....	13
Abb. 2	Realisierte und potenziell mögliche Systemrelationen	39
Abb. 3	Systemrelationen mit Rückgriff auf das Medium Sinn.....	41
Abb. 4	Merkmale komplexer sozialer Systeme.....	43
Abb. 5	Hauptdeterminanten für Gesundheit	52
Abb. 6	Interventionssystem als sich herausbildendes neues soziales System	57
Abb. 7	Strukturelle Kopplungen	57
Abb. 8	Das Evaluationssystem als neu entstehendes Sozialsystem	75
Abb. 9	Der Prozess der Generalisierung	82
Abb. 10	Merkmale komplexer sozialer Systeme und Fragen an die Wirkungsevaluation.....	87
Abb. 11	Kodierschema in Anlehnung an Strauss und Corbin (1996)	98
Abb. 12	Stufen der Wirkungskonstruktion in komplexen Feldern.....	107
Abb. 13	Komplexität eingrenzen	108
Abb. 14	Komplexität und Design der untersuchten Evaluationen.....	124
Abb. 15	Einfluss von Modellen auf die Wirkungskonstruktion	126
Abb. 16	Strategien der Verarbeitung von Komplexität.....	132
Abb. 17	Umriss eines induktiv entwickelten, sehr differenzierten Wirkungsmodells	135
Abb. 18	Einfaches Wirkungsmodell einer Präventionskampagne	136
Abb. 19	Evaluationsfragen zum Wirkungsmodell der Präventionskampagne	136
Abb. 20	Unterschiedliche Vergleiche als Basis für Wirkungsbeurteilungen	137
Abb. 21	Betrachtung der Entwicklung über die Interventionszeit hinaus.....	139
Abb. 22	Strukturgleichungsmodell	142
Abb. 23	Strategien der Wirkungszuschreibung.....	148
Abb. 24	Berechnung der Projektwirkung	151
Abb. 25	Der Prozess der Wirkungskonstruktion	157
Abb. 26	Vorgehensweisen zur Wirkungsbeurteilung	159
Abb. 27	Einfaches Wirkungsmodell im Kontext einer Präventionskampagne	200
Abb. 28	Strukturelle Kopplung zwischen sozialen, psychischen und physischen Systemen	203
Abb. 29	Nutzung der drei Strategien der Wirkungszuschreibung für Plausibilisierungen	217
Abb. 30	Typen des Umgangs mit Komplexität.....	221
Abb. 31	Ein Blick in die Black Box	228
Abb. 32	Spielraum der Evaluation zwischen Ausweitung und Reduktion von Komplexität.....	231

Tabellenverzeichnis

Tab. 1	Simple, complicated and complex problems	58
Tab. 2	Übersicht über die Merkmale einfacher, komplizierter und komplexer Programme	59
Tab. 3	Überblick über die untersuchten Evaluationen	95
Tab. 4	Induktiv sich entwickelndes Kodierschema	97
Tab. 5	Deduktives Kodierschema	98
Tab. 6	Auszug aus dem Kategoriensystem	100
Tab. 7	Kategorien von Wirkungszielen	110
Tab. 8	Einschätzung der untersuchten Evaluationen	114
Tab. 9	Handlungsspielraum bei der Eingrenzung von Komplexität.....	118
Tab. 10	Unterschiedliche Datenquellen.....	119
Tab. 11	Kategorisierung der untersuchten Evaluationen	124
Tab. 12	Unterschiedliche Komplexität von Befragungsdaten	128
Tab. 13	Handlungsspielraum bei der Erfassung von Komplexität.....	131
Tab. 14	Dimensionen von Wirkungsmodellen	134
Tab. 15	Dimensionen von Beschreibungen.....	144
Tab. 16	Handlungsspielraum bei der Verarbeitung von Komplexität	146
Tab. 17	Forschungsdesign der untersuchten Evaluationen	147
Tab. 18	Handlungsspielraum bei der Beurteilung von Wirkungen	156
Tab. 19	Handlungsspielraum der Evaluation bei der Wirkungskonstruktion	158
Tab. 20	Zentrale Aspekte der Wirkungsevaluation im Umgang mit Komplexität	219
Tab. 21	Strategien der Reduktion und Ausweitung von Komplexität	220
Tab. 22	Vor- und Nachteile der vier Typen im Umgang mit Komplexität	223
Tab. 23	Umgang mit den Merkmalen komplexer sozialer Systeme	224
Tab. 24	Anhaltspunkte zur Bestimmung des Komplexitätsgrades einer Evaluation	229

1 Einleitung

Gesundheitliche Probleme gehören zu den bedeutendsten gesellschaftlichen Herausforderungen der heutigen Zeit. Die grossen Erfolge der Medizin und der Sozialhygiene zwischen der Mitte des 19. und der Mitte des 20. Jahrhunderts haben zwar viele Krankheiten weitgehend unter Kontrolle gebracht, denen die Menschheit bisher ausgeliefert war. Mit der Entdeckung von Antibiotika und Impfstoffen sowie der Verbesserung der hygienischen Verhältnisse konnte eine Reihe von gefürchteten Infektionskrankheiten massgeblich eingedämmt werden. Trotzdem bedrohen Infektionskrankheiten wie Aids, Hepatitis, Malaria oder Ebola weiterhin das Leben von Millionen von Menschen. Die Medizin ist hier gefordert, Impfstoffe und Behandlungsmethoden zu finden oder zu optimieren, um diese Krankheiten in den Griff zu bekommen und um weitere Ausbreitungen zu verhindern. Allerdings lassen sich diese Krankheiten nicht allein pharmakologisch bekämpfen. Eine erfolgreiche Prävention und Behandlung auf Bevölkerungsebene muss sich ebenfalls sozialen, politischen, ökonomischen, psychologischen und kulturellen Fragen widmen und diese Dimensionen in Lösungsstrategien einbinden. Es geht um Fragen der Sensibilisierung von Bevölkerungsgruppen, um die Bildung von Gesundheitskompetenzen, um Strukturanpassungen in Gesundheitssystemen, um Veränderung von Normen und Werten, um die Zugänglichkeit zu Medikamenten, um politische Mitwirkung etc. Dies gilt auch und in besonderem Masse für nicht-übertragbare Krankheiten wie Herz-Kreislaufkrankheiten, Krebs, Atemwegserkrankungen oder Diabetes. Diese Krankheiten, welchen zum grossen Teil nicht pharmakologisch vorgebeugt werden kann, sind mittlerweile zur Todesursache Nummer eins geworden. Über 36 Millionen Menschen sterben weltweit jährlich an nicht übertragbaren Krankheiten, was 63% aller Todesfälle entspricht. Etwa 40% davon sterben an diesen Krankheiten, bevor sie ihr 70. Altersjahr erreichen und erleiden dadurch gemäss Definition der Weltgesundheitsorganisation einen verfrühten Tod. Die meisten dieser verfrühten Todesfälle stehen in Zusammenhang mit vermeidbaren Risikofaktoren wie Tabak- oder Alkoholkonsum, ungesunder Ernährung oder unzureichender Bewegung (WHO, 2013). In der Schweiz entfielen im Jahr 2012 rund 75% der Todesfälle auf Herz-Kreislaufkrankheiten, Krebs, Atemwegserkrankungen oder Demenz, während infektiöse Krankheiten lediglich noch für ca. 1.2 % der Todesfälle verantwortlich waren¹. Die Prävention von (nicht-übertragbaren) Krankheiten und die Förderung der Gesundheit auf Bevölkerungsebene sind hierzulande nicht zuletzt auch unter dem Druck stetig steigender Gesundheitskosten zu einer wichtigen gesellschaftlichen Aufgabe geworden. Mittlerweile belaufen sich die Gesundheitskosten in der Schweiz auf über 10% des Bruttoinlandprodukts².

Aufgrund der hohen Komplexität der Entstehung von Krankheit und Gesundheit sind einfache Rezepte und Routinemassnahmen auf Bevölkerungsebene nur begrenzt tauglich, um den aktuellen gesundheitlichen Herausforderungen und Problemen erfolgreich begegnen zu können. Während die Gesundheitswissenschaften traditionell stärker auf die Identifizierung einzelner Krankheitsursachen wie beispielsweise Krankheitserreger fokussierten, setzen sich heute zunehmend umfassende Gesundheitsverständnisse durch, welche Gesundheit als das Ergebnis eines komplexen Zusammenspiels von individuellen Ressourcen und Verhaltensweisen, von sozialen, ökonomischen und kulturellen Einflüssen sowie von Umweltbedingungen materieller Art usw. verstehen³. Damit gelangen immer mehr soziologische und systemtheoretische Konzepte in den Fokus des Interesses (Rogers, 1968; Green, 2006). „Social programs are undeniably, unequivocally, unexceptionally social systems. They comprise, as with any social system, the interplays of individual and institution, of agency and structure, and of micro and macro social processes.“ (Pawson & Tilley, 1997, 63). Soziologische und insbeson-

¹ Bundesamt für Statistik, Todesursachenstatistik für das Jahr 2012 (<http://www.bfs.admin.ch/bfs/portal/de/index/themen/14/02/04/key/01.html>, 12.4.2015).

² Quelle: Bundesamt für Statistik: http://www.bfs.admin.ch/bfs/portal/de/index/themen/14/05/blank/key/internationaler_vergleich.html, 12.4.2015).

³ Vgl. dazu das Konzept der Hauptdeterminanten für Gesundheit (Dahlgren and Whitehead (1991)) aber auch den wegweisenden WHO-Bericht zu sozialen Einflussfaktoren auf die Gesundheit (CSDH (2008)).

dere systemtheoretische Perspektiven scheinen vielversprechende Alternativen zu einem medizinisch orientierten Interventions- und Forschungsverständnis, welches sich auf individuelle Behandlungen konzentriert, darzustellen. „Public health asks of sociology“ (Rogers, 1968) und „Public health asks of systems science“ (Green, 2006) sind programmatische Rufe, welche soziale Interaktionen und gesellschaftliche Zusammenhänge in den Fokus von Intervention und Evaluation der Gesundheitsförderung und Prävention stellen. Zur Lösung gesundheitlicher Probleme auf Bevölkerungsebene sind langfristig angelegte, multisektorale und interdisziplinäre Strategien gefordert, die gleichzeitig auf verschiedenen gesellschaftlichen Ebenen ansetzen und eine Vielzahl von Akteuren mit zum Teil unterschiedlichen Interessen und Kompetenzen koordinieren (WHO, 1998c; WHO, 2013). Die kantonalen Aktionsprogramme für ein Gesundes Körpergewicht, wie sie seit 2007 in der Schweiz durchgeführt werden, um Übergewicht und Adipositas bei Kindern und Jugendlichen vorzubeugen, sind gute Beispiele dafür. Um eine nachhaltige Wirkung zu erzeugen, sind die Aktionsprogramme langfristig angelegt und kombinieren Interventionsprojekte mit politischen Massnahmen, mit Sensibilisierungs- und Öffentlichkeitsarbeit sowie mit Vernetzung und Koordination der relevanten Akteure. Die Programme setzen auf unterschiedlichen Wirkungsebenen an und bauen auf der Zusammenarbeit mit lokalen Akteuren auf. Sie situieren sich in ganz unterschiedlichen Settings, umfassen unterschiedliche methodische Ansätze und sprechen eine grosse Vielfalt von Zielgruppen an. So werden zum Beispiel gleichzeitig kantonale Ernährungsrichtlinien erlassen, Mütter- und Väterberater/-innen weitergebildet, Schulen bewegungsfreundlicher und Schulwege sicherer gemacht, Schulkantinen für ausgewogene Ernährung sensibilisiert und zertifiziert etc. Während noch vor zehn Jahren begrenzte und oft isolierte Einzelprojekte die Landschaft der Gesundheitsförderung und Prävention bestimmt haben, hat sich das Feld heute stark in Richtung grösserer Programme entwickelt, welche die verschiedenen Projekte und lokalen Akteure unter einer gemeinsamen Strategie zusammenfassen und koordinieren (Kolip, Ackermann, Ruckstuhl, & Studer, 2012). Die breit abgestützten nationalen und kantonalen Strategien und Programme, welche in der Schweiz aktuell in den Themen Gesundes Körpergewicht, Alkoholprävention, Tabakprävention, HIV/STI-Prävention, psychische Gesundheit und Gesundheit im Alter durchgeführt werden oder in Vorbereitung sind, zeugen von dieser Entwicklung⁴.

Gleichzeitig mit dieser Ausweitung und Systematisierung der Gesundheitsförderung und Prävention steigen aber auch die Erwartungen an den Nachweis der Wirkungen entsprechender Programme, um die eingesetzten Ressourcen politisch legitimieren zu können (Tang, Ehsani, & Mc Queen, 2003; Mc Queen, 2001; Bödeker & Kreis, 2006). Die Evaluation von Gesundheitsförderung und Prävention, also die wissenschaftlich fundierte Beurteilung von entsprechenden Projekten und Programmen steht zunehmend unter Druck, Wirksamkeit von Projekten und Programmen nachweisen zu müssen, deren Komplexität eindeutige und generelle Wirkungsnachweise allerdings kaum zulässt. Multisektorale, interdisziplinäre, settingorientierte, partizipative Ansätze, die sich in unterschiedlichen Kontexten immer wieder anders entwickeln, sind das Wesen der Gesundheitsförderung und Prävention. Als Teilgebiet der Gesundheitswissenschaften steht Gesundheitsförderung und Prävention unter starkem Einfluss einer medizinischen Forschung (Bödeker, 2006), die sich auf die Überprüfung von Behandlungen am Individuum spezialisiert hat. Randomisierte kontrollierte Studien, welche auf einem systematischen Vergleich von zufällig gebildeten Gruppen von behandelten und nicht behandelten Personen gründet, haben sich in der medizinischen Forschung und vor allem Bereich der Wirksamkeit von Medikamenten, chirurgischen Eingriffen und weiteren medizinischen Behandlungen etabliert und bilden so die entscheidende Grundlage für ärztliche Entscheide. In der Gesundheitsförderung und Prävention stehen im Gegensatz dazu aber nicht einzelne Individuen im Fokus der Intervention und der Wirkungsüberprüfung, sondern grössere sozialräumliche Systeme wie Stadtteile, Gemeinden, Schulen, Betriebe etc. oder ganze Bevölkerungsgruppen. In der Literatur der Gesundheitsförderung und Prävention wird stark kritisiert, dass randomisierte kontrollierte Studien nicht oder nur sehr eingeschränkt geeignet seien, die Wirksamkeit komplexer sozialer Interventionen zu untersuchen (z.B. Elkeles, 2006; Nut-

⁴ Vgl. z.B. www.gesundheitsfoerderung.ch (Gesundes Körpergewicht, Gesundheit im Alter) oder www.bag.admin.ch (Alkohol- und Tabakprävention, HIV/STI-Prävention).

beam, 1998; Walach, 2006). Randomisierte kontrollierte Studien würden der Komplexität der Gesundheitsförderung und Prävention ganz generell nicht Rechnung tragen, sie seien zu eingeschränkt in ihrer Fragestellung, sie würden dem dynamischen Charakter von partizipativen, settingorientierten Ansätzen nicht gerecht werden und sie seien aus pragmatischen oder ethischen Gründen kaum realisierbar, so einige Argumente für diese Ablehnung oder Zurückhaltung. Die Gesundheitsförderung und Prävention sieht sich auf der einen Seite gefordert, die Wirksamkeit ihrer Interventionen nachweisen zu müssen. Auf der anderen Seite weist sie denjenigen Studientypus als für die eigenen komplexen Interventionen weitgehend unangemessen zurück, der sich als ‚Goldstandard‘ der Wirksamkeitsforschung etabliert hat.

Die Komplexität von Gesundheitsförderung und Prävention und die damit verbundenen skizzierten Herausforderungen lassen die Evaluation offenbar nicht davon abhalten, laufend Wirkungen zu untersuchen und nachzuweisen. In den letzten Jahren sind in der Schweiz eine Vielzahl von Evaluationen durchgeführt worden, welche sich u.a. auf die Beurteilung von Wirkungen von Projekten und Programmen der Gesundheitsförderung und Prävention eingelassen haben. Eine fundierte Auseinandersetzung und eine genaue Betrachtung scheinen lohnenswert, um Erkenntnisse darüber zu gewinnen, welche Möglichkeiten der Evaluation offenstehen, mit der Komplexität der von ihr untersuchten Systeme und mit den angesprochenen Spannungsfeldern umzugehen. Folgende Forschungsfrage lässt sich davon ableiten:

Wie geht Evaluation bei der Beurteilung von Wirkungen der Gesundheitsförderung und Prävention mit der Komplexität sozialer Interventionen um? Oder etwas provokativ formuliert: Wie geht Evaluation mit dem Spannungsfeld um, Wirkungen nachweisen zu müssen und es faktisch nicht tun zu können?

Im Fokus des Interesses stehen die folgenden vier miteinander in Verbindung stehenden Bezugspunkte: Komplexität, Gesundheitsförderung und Prävention, Wirkungen und Evaluation. Komplexität verweist auf die grundlegenden Eigenschaften sozialer Systeme, welche die Planung, Steuerung und Evaluation von Interventionen grundlegend erschweren. Gesundheitsförderung und Prävention grenzt den Gegenstandsbereich inhaltlich ein. Evaluation lenkt das Erkenntnisinteresse auf die Beurteilung (von Gesundheitsförderung und Prävention) und Wirkung bezeichnet dabei den Fokus bei der Beurteilung:

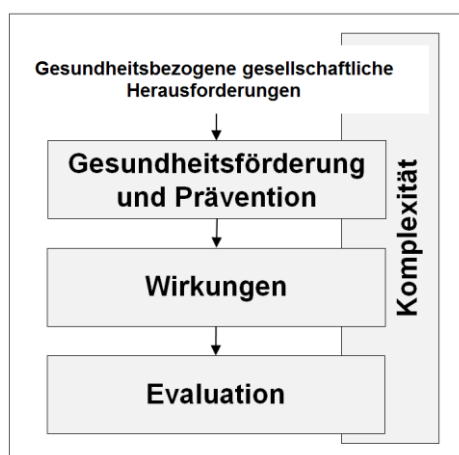


Abb. 1 Die vier Bezugspunkte der Untersuchung

Zur Beantwortung der Forschungsfrage wird in einem ersten theoretischen Kapitel (2) ein Überblick über die Evidenzdebatte in der Gesundheitsförderung und Prävention gegeben. Dabei wird die Verunsicherung deutlich, die bezüglich der Evaluation von Wirkungen zurzeit in diesem Feld herrscht. Da die Wurzeln dieser Verunsicherung in der Komplexität der Gesundheitsförderung und Prävention liegen, ist es wichtig, diese Komplexität in ihren Facetten verstehen und theoretisch fassen zu können. Was steckt hinter diesem mittlerweile fast inflationär verwendeten Begriff und was macht ein komple-

xes (soziales) System letzten Endes aus? Das zweite theoretische Kapitel (4) befasst sich mit der Komplexität sozialer Systeme. Die neuere soziologische Systemtheorie um Niklas Luhmann, welche u.a. Aspekte der interdisziplinär entwickelten Komplexitätstheorie („complexity science“) für das Verständnis sozialer Systeme fruchtbar gemacht hat, bietet einen angemessenen theoretischen Rahmen für das Verständnis komplexer Interventionsprogramme der Gesundheitsförderung und Prävention. Sie ermöglicht es, die vielfältigen sozialen Einflüsse auf Gesundheit und Krankheit und den umfassenden und dynamischen Charakter von Interventionen der Gesundheitsförderung analytisch zu fassen. In diesem Kapitel werden Dimensionen und Eigenschaften komplexer sozialer Systeme herausgearbeitet, die als Grundlage für die weitere Analyse dienen. Gleichzeitig wird der Frage nachgegangen, inwiefern Interventionen der Gesundheitsförderung und Prävention überhaupt komplex sind. Das Gewahr werden der Komplexität sozialer Systeme hat weitreichende erkenntnistheoretische Folgen, die in einem weiteren Kapitel (5) beleuchtet werden. Es wird der Frage nachgegangen, was unter Wirkung verstanden wird und inwiefern man Wirkung überhaupt erkennen und untersuchen kann.

Nach dieser theoretischen Auseinandersetzung folgt eine empirische Untersuchung, deren Methodik in Kapitel 6 detailliert dargelegt wird. Neun Schlussberichte von Evaluationen der Gesundheitsförderung und Prävention werden hinsichtlich ihres Umgangs mit Komplexität bei der Beurteilung von Wirkungen untersucht. Aus einem Verständnis heraus, dass grundsätzlich alle sozialen Systeme (mehr oder weniger) komplex sind⁵, wurden Evaluationen mit unterschiedlichem Komplexitätsgrad in die Untersuchung aufgenommen, auch solche, die andernorts als einfache oder komplizierte Interventionen gelten würden. Im Rahmen einer Inhaltsanalyse dieser Evaluationsberichte, die sich an die Grounded Theory anlehnt, wird in Kapitel 7 exemplarisch herausgearbeitet, wie Komplexität eingegrenzt, erfasst und verarbeitet wird und wie unterschiedlich Wirkungen in der Gesundheitsförderung und Prävention beurteilt werden.

In einem nächsten Schritt, in Kapitel 8, werden die Ergebnisse der empirischen Analyse vor dem Hintergrund der Theorie komplexer sozialer Systeme und der aktuellen Diskussionen in der Evaluationsliteratur reflektiert. In einem Schlusskapitel (9) werden die Forschungsfrage zusammenfassend beantwortet, Schlussfolgerungen für die Wirkungsevaluation von Gesundheitsförderung und Prävention gezogen, die Erkenntnisse zu Grundzügen einer komplexitätssensiblen Evaluation verdichtet und ein Ausblick auf lohnende Weiterentwicklungen und weiterführende Forschungsperspektiven gegeben.

⁵ Vgl. dazu Kapitel 4.4.

2 Die Evidenzdebatte in der Gesundheitsförderung

Unter dem Druck von Politik und Geldgebern, im Zuge steigender Professionalisierung sowie aus eigener fachlicher Motivation heraus sucht das Handlungsfeld der Gesundheitsförderung und Prävention⁶ nach wissenschaftlicher Legitimation (Tang et al., 2003). Der Evaluation, also der systematischen Beurteilung von Policies, Programmen, Projekten und Interventionen der Gesundheitsförderung auf Basis wissenschaftlicher Theorien und Forschungsmethoden wird dabei eine zentrale Rolle zugeschrieben. Evaluationen sollen in erster Linie Nachweise über die Wirksamkeit und in immer stärkerem Ausmass auch über die Wirtschaftlichkeit (Stevens, 2004) der Gesundheitsförderung erbringen. Evidenzbasierte Gesundheitsförderung und Prävention, also die Basierung der Gesundheitsförderung und Prävention auf wissenschaftlichen Erkenntnissen, steht für ein enges Zusammenspiel zwischen Wissenschaft und Politik. Der Wissenschaft kommt dabei eine unterstützende Doppelrolle zu: einerseits hat sie die Aufgabe, bestehendes wissenschaftliches Wissen zu sammeln, zu beurteilen und für Entscheidungsträger in Politik, Verwaltung und Organisationen aufzuarbeiten, als Grundlage für Entscheidungsfindung und Handeln. Andererseits hat sie aber auch die Rolle, Programme zu evaluieren, um so die Evidenzbasis zur Wirksamkeit von Interventionen, laufend zu erweitern (Broesskamp-Stone, 2009; Unicef, 2008). Die Gesundheitsförderung und Prävention sieht sich durch den Umstand, Wirkung nachweisen zu müssen und dies aufgrund der hohen Komplexität der untersuchten Projekte und Programme faktisch kaum zu können, mit der Gefahr konfrontiert, dass Interventionen bevorzugt finanziert werden, die von eher geringer Komplexität sind und entsprechend einfacher auf ihre Wirksamkeit hin zu evaluieren sind. Interventionen hingegen, welche komplexen sozialen Problemen eher Rechnung tragen, deren Wirksamkeit aber nur schwer belegbar ist, werden tendenziell vernachlässigt (vgl. Luber, 2004 und Rosenbrock, 2004). Die aktuell geführten Diskussionen rund um die Wirksamkeit von Gesundheitsförderung und Prävention orientieren sich – nicht zufällig – am Konzept der wissenschaftlichen Evidenz, welches in der evidenzbasierten Medizin tragend ist. Das Handlungsfeld der Gesundheitsförderung und Prävention hat sich als Teilgebiet der Gesundheitswissenschaften (Public Health) etabliert (Gutzwiler & Paccaud, 2007) und ist deshalb stark von einem medizinischen Forschungsverständnis geprägt. Der Rückbezug auf das aus der Medizin stammende Evidenzkonzept ist allerdings nicht unproblematisch und in der Gesundheitsförderung umstritten, wie nachfolgend gezeigt wird. Dass der Evidenzbegriff selbst in der evidenzbasierten Medizin nicht eindeutig definiert ist, erschwert eine Klärung für die Nutzung in der Gesundheitsförderung zusätzlich.

2.1 Evidenz in der ‚Evidence Based Medicine‘

Evidenz wird alltagssprachlich, philosophisch-erkenntnistheoretisch, juristisch oder medizinisch unterschiedlich definiert und verwendet. Während sich die im Deutschen gebräuchliche alltagssprachliche und philosophisch-erkenntnistheoretische Definition am lateinischen Begriff ‚evidentia‘ orientiert und offensichtliche, einleuchtende Sachverhalte meint, orientiert sich die Verwendung im medizinischen und juristischen Kontext am englischen Begriff ‚evidence‘, welcher Beleg oder Beweis bedeutet (Bödeker, 2006; Gerhardus, 2010; Mc Queen & Anderson, 2001). Unter Evidenz wird in der evidenzbasierten Medizin der wissenschaftliche Nachweis der Wirksamkeit von Interventionen verstanden:

⁶ Mit dem Begriffspaar „Gesundheitsförderung und Prävention“ sind Public Health-Interventionen gemeint, die das Ziel verfolgen, Gesundheit zu fördern und Krankheiten zu verhindern. Während sich Krankheitsprävention von ihrer Tradition her primär an zu verhindernden Risikofaktoren orientiert, nimmt die Gesundheitsförderung nicht die Krankheiten, sondern die Determinanten für Gesundheit und Wohlbefinden in den Blick (Altgeld and Kolip (2010)). In Programmen und Projekten der Gesundheitsförderung und (Primär-)Prävention, um deren Evaluation es in der vorliegenden Studie geht, kommen häufig beide Zugänge kombiniert, aber mit unterschiedlicher Ausprägung zur Anwendung. Da primärpräventive Ansätze sich heute weitgehend an den Grundlagen und Prinzipien der Gesundheitsförderung orientieren und es immer schwieriger wird, die Ansätze zu trennen, wird in der vorliegenden Arbeit weitgehend von „Gesundheitsförderung und Prävention“ gesprochen.

Im Kontext der *Evidenz* basierten Medizin hat der Begriff Evidenz allerdings eine besondere Bedeutung. Hier bezeichnet er – englischsprachig – nach dem naturwissenschaftlich gebräuchlichen Erkenntnisparadigma Zusammenhänge zwischen Intervention und Effekt, die durch belastbare, d.h. unverzerrte Studien „bewiesen“ wurden. (Bödeker, 2006, 3; Hervorhebungen im Original).

Evidenzbasierte Medizin gründet allerdings nicht, wie man vermuten könnte, allein auf wissenschaftlichen Wirksamkeitsnachweisen, sondern bezieht zusätzlich die klinische Erfahrung des behandelnden Arztes resp. der Ärztin mit ein, ein Umstand, welcher in der Debatte rund um Evidenzbasierung in der Gesundheitsförderung oft nicht beachtet wird:

Evidence based medicine is the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients. The practice of evidence based medicine means integrating individual clinical expertise with the best available external clinical evidence from systematic research. (Sackett, Rosenberg, Gray, Haynes, & Richardson, 1996, 71)

Nach Hart (2000) setzt sich evidenzbasierte Medizin zusammen „aus den Elementen wissenschaftliche Erkenntnis (systematisches Wissen), praktische Erfahrung (intuitives Wissen) und der Akzeptanz in der Profession.“ (Hart, 2000, 4). Evidenzbasierte Medizin erhält dadurch eine normsetzende Funktion für die Profession der Medizin (Hart, 2000).

(Wissenschaftliche) Evidenz fokussiert bezüglich Gegenstand auf Wirksamkeit, erkenntnistheoretisch auf empirische Nachweise. Obwohl durchaus unterschiedliche Verfahren Evidenz begründen können⁷, steht die Empirie bezüglich Nachweiskraft an oberster Stelle (Gay & Rossboth, 2007). Empirische Erkenntnisse werden wiederum in Bezug auf ihre Nachweiskraft priorisiert und zwar je nach Studientyp, mittels welchem sie gewonnen wurden.

Evidenzhierarchie

Die evidenzbasierte Medizin unterscheidet verschiedene Evidenzklassen, die sie in hierarchischer Weise ordnet. Obwohl sich die Klassifizierungen in den Details unterscheiden können, folgen sie einem ähnlichen Aufbau. Von höchster Aussagekraft sind gemäss einer weit verbreiteten Evidenzhierarchie systematische Reviews über kontrollierte randomisierte Studien, gefolgt von einzelnen randomisierten Studien (Evidenzklasse 1). In der Evidenzklasse 2 befinden sich systematische Reviews von Kohorten-Studien und entsprechende Einzelstudien, in der Evidenzklasse 3 systematische Reviews von Fall-Kontroll-Studien sowie entsprechende Einzelstudien. Serien von Einzelfallstudien sowie qualitativ weniger gute Kohorten- und Fall-Kontroll-Studien befinden sich in der Evidenzklasse 4 und sogenannte Expertenmeinungen bilden die unterste, fünfte Evidenzklasse (Oxford Centre for Evidence Based Medicine, 2009)⁸.

Diese Klassifizierungen und damit die in der klinischen Forschung weitgehend akzeptierte Vorrangstellung randomisierter kontrollierter Studien geht weitgehend zurück auf die Begründung in Cochranes Klassiker „Effectiveness and Efficiency“ (Cochrane, 1972). Cochrane legt dar, wie mittels Randomisierung und doppelter Blindung eine deutlich höhere Aussagekraft über die Wirksamkeit von Interventionen erreicht werden kann als mit Beobachtungs- oder quasi-experimentellen Studien. Der grosse Vorteil randomisierter kontrollierter Studien liegt darin, dass sich durch die zufällige Verteilung der Studienteilnehmenden die Interventions- und Kontrollgruppe lediglich durch das Faktum der Interven-

⁷ Gay und Rossboth unterscheiden empirische Evidenz, Evidenz durch Analogie und anekdotische Evidenz (Gay and Rossboth (2007)).

⁸ Es existieren mittlerweile unterschiedliche Varianten der Evidenzhierarchie, wobei sich die Grundlogik der Hierarchie und die Vormachtstellung randomisierter kontrollierter Studien gleich bleiben. Einen Vergleich über drei etablierte Stufenmodelle liefert Grossman (2008).

tion unterscheiden, während andere Charakteristika wie Geschlecht, Herkunft, Einstellungen etc. gleichmässig auf beide Gruppen verteilt sind. Dadurch können entsprechende methodische Biases weitgehend ausgeschlossen werden und in der Studie festgestellte Unterschiede zwischen den beiden Gruppen – im Rahmen der statistisch möglichen Aussagekraft – können auf die Intervention zurückgeführt werden. Andere Erklärungsfaktoren können damit weitgehend ausgeschlossen werden. Aus diesem Grund haben sich die randomisierten kontrollierten Studien in der medizinischen Forschung als ‚Goldstandard‘ etabliert (Sackett et al., 1996). Cochrane hat allerdings bereits auch auf statistische, ethische und messmethodische Grenzen randomisierter kontrollierter Studien hingewiesen sowie auf die Limitierung bezüglich Aufwand und Machbarkeit (Cochrane, 1972). Randomisierten kontrollierten Studien wird aufgrund ihres Designs eine sehr hohe interne Validität zugesprochen. Die externe Validität, also die Generalisierbarkeit über die Untersuchungspopulation hinaus, ist, obwohl die evidenzbasierte Medizin gerade darauf aufbaut (Sackett et al., 1996), in Studien allerdings oft unbefriedigend (Rothwell, 2005; Tones, 2000).

Obwohl sich die evidenzbasierte Medizin nicht auf randomisierte Kontrollstudien (RCT's) und entsprechende Metaanalysen beschränkt, wird in der Evidenzdebatte um die Übertragbarkeit des Evidenzbegriffs in das Feld der Gesundheitsförderung sehr schnell auf randomisierte kontrollierte Studien fokussiert (Bödeker, 2006; Grossman, 2008). Randomisierte kontrollierte Studien bzw. die gesamte Evidenzhierarchie sind auf die Frage der Wirksamkeit von Interventionen ausgelegt (Zeeb & Donath, 2010), andere Fragestellungen wie z.B. die Untersuchung von Krankheitsprävalenzen oder die Beurteilung der Qualität von Behandlungen werden durch diese Klassifizierungslogik nicht erfasst. Diese Verkürzung des Begriffs Evidenz auf Aspekte der Wirksamkeit liegt in der hohen Priorität von Wirksamkeitsaspekten in der medizinischen Forschung generell begründet (es geht dabei oft um Leben und Tod) und wird dadurch gefördert, dass sich für andere Erkenntnisaspekte keine vergleichbare Evidenzklassifizierung etabliert hat. Diese Schwerpunktsetzung innerhalb der evidenzbasierten Medizin, die sich nicht aus dem Begriff der Evidenz selber ableiten lässt, wird in der Evidenzdebatte der Gesundheitsförderung als einschränkend kritisiert, da Evaluation gemeinhin ein viel grösseres Spektrum aufweist als die Beurteilung von Wirksamkeit.

2.2 Randomisierte kontrollierte Studien im Fokus der Kritik

In der Gesundheitsförderung wird seit den 90-er Jahren eine engagierte Debatte rund um die Übertragbarkeit des Evidenzkonzeptes der evidenzbasierten Medizin auf das Feld der Gesundheitsförderung und um die Bedeutung randomisierter kontrollierter Studien für die Evaluation von Interventionen der Gesundheitsförderung und Prävention geführt (vgl. (WHO, 1998a) sowie die Sammelbände (WHO, 2001); (Bödeker & Kreis, 2006); (Luber & Geene, 2004); (Thorogood & Coombes, 2004); (Robert Koch Institut, 2012)). Vergleichbare Diskussionen werden auch in der Sozialen Arbeit geführt (Sommerfeld, 2005), ein Handlungsfeld, welches sich ebenfalls durch einen hohen Interventionscharakter auszeichnet und welches sich deshalb auch verstärkt mit Fragen der Effektivität und im Zuge der Verknappung staatlicher Ressourcen bei gleichzeitig steigendem Bedarf mit Fragen der Effizienz seiner Massnahmen konfrontiert sieht.

Die Evidenzdebatte in der Gesundheitsförderung dreht sich im Wesentlichen um die Frage der Angemessenheit randomisierter kontrollierter Studien für Interventionen der Gesundheitsförderung und Prävention (andere Studientypen der evidenzbasierten Medizin finden verhältnismässig wenig Beachtung). Der Grundtenor wird durch folgende offizielle Feststellung der WHO treffend zusammengefasst: "The use of randomized control trials to evaluate health promotion initiatives is, in most cases, inappropriate, misleading and unnecessarily expensive." (WHO, 1998a, 2). Allerdings fehlt bisher sowohl eine fundierte und breit akzeptierte theoretische Grundlage für die Gesundheitsförderung und Prävention, noch gibt es Konsens, was unter evidenzbasierter Gesundheitsförderung und Prävention im Detail zu verstehen ist (Mc Queen & Anderson, 2001; Bödeker, 2006; Elkeles, 2006).

Die meisten Exponentinnen und Exponenten in dieser Diskussion sprechen randomisierten kontrollierten Studien zwar nicht grundsätzlich Aussagekraft und Nützlichkeit ab (Elkeles, 2006), sie attestieren ihnen aber im Feld der Gesundheitsförderung lediglich marginale Bedeutung und zweifeln vehement deren prioritäre Stellung in der Evaluation von Gesundheitsförderungsmassnahmen an (Elkeles, 2006; Nutbeam, 1998; Tones, 2000; Saan, 2005; Walach, 2006; Grossman, 2008). Da sich die Hauptkritik in der Evidenzdebatte um die Angemessenheit randomisierter kontrollierter Studien dreht und auf Fragen der Messbarkeit komplexer Sachverhalte hinausläuft, soll im Folgenden zunächst vertieft auf diese Kritik eingegangen werden. Die zentralen Kritikpunkte in der Debatte rund um die Übertragbarkeit von randomisierten kontrollierten Studien auf das Feld der Gesundheitsförderung lassen sich wie folgt gruppieren:

Eingeschränkter Anwendungsbereich

Die eine Argumentationslinie zielt auf den eingeschränkten Anwendungsbereich randomisierter kontrollierter Studien und ihren Fokus auf die Wirksamkeit von Interventionen ab. Diese Argumentation ist gleichzeitig eine Kritik am eingeschränkten Evidenzbegriff, wie er sich in der evidenzbasierten Medizin etabliert hat. Evidenz, also wissenschaftlicher Nachweis, habe ein weit grösseres Spektrum abzudecken als die Beurteilung von Effektivität (Nutbeam, 1998). Nicht lediglich die Frage *ob* eine Intervention wirke sei von Interesse, sondern, mit Blick auf die Heterogenität der Zielgruppen, Kontexte und Interventionen, *wie, bei wem* und *unter welchen Bedingungen* (Saan, 2005; Oliver et al., 2005; Tones, 1997). Weitere Felder, in denen Evidenz gefragt sei, seien beispielsweise die Beurteilung von unerwünschten Nebenwirkungen von Interventionen, Kosten-Nutzen-Analysen oder die Verbesserung des Zugangs und der Implementierung von Programmen (Mc Queen, 2001; Trojan, 2006).

Eingeschränkt sei der Anwendungsbereich des engen Evidenzbegriffs und von randomisierten kontrollierten Studien aber auch noch in anderer Hinsicht. Mc Queen spricht von einem kulturellen und geographischen Bias und betont, dass diese Konzepte in westlichen Industrienationen entwickelt wurden und ausserhalb dieser kaum eine Bedeutung haben – ohne dass allerdings bereits etablierte alternative Konzepte vorliegen würden (Mc Queen, 2001).

Charakteristik des Forschungsgegenstandes

Die wichtigsten Argumente befassen sich mit der ungenügenden Passung zwischen der experimentellen Methodik und dem Forschungsgegenstand. In der Gesundheitsförderung gehe es primär um grössere komplexe soziale Systeme, welche stark vom Zusammenspiel unterschiedlicher Akteure und dadurch von unvorhersehbaren sozialen Dynamiken geprägt seien. Dies stehe im Widerspruch zur reduktionistischen Methodik der randomisierten kontrollierten Studien, welche in ihren Untersuchungen nicht soziale Systeme sondern einzelne Individuen fokussiere (Coombes, 2004; Bödeker & Kreis, 2006; Wright, 2006). Je stärker die Grundprinzipien der Gesundheitsförderung wie Partizipation, Empowerment, Salutogenese, der Settingansatz und die Kombination der für die Gesundheitsförderung relevanten fünf Handlungsfelder (Individuum, Gruppe, Organisation, Lebenswelt und Politik/Gesellschaft World Health Organization, 1986; WHO, 1998c) ernst genommen würden, desto weniger seien experimentelle Methoden geeignet, den Erfolg derartiger Interventionen zu beurteilen:

The dilemma emerging from this analysis is that the more powerful forms of health promotion action are those which appear to be longterm and least easily predicted, controlled and measured by conventional means. (Nutbeam, 1998, 38)

Eine besondere Kritik erwächst aus dem Konzept der Partizipation, welches für die Gesundheitsförderung konstitutiv ist. Der Einbezug der Zielgruppen in die Planung, Durchführung und Evaluation von Gesundheitsförderungsinterventionen mache es unmöglich, ein Programm zu isolieren. Insofern sei es nicht „das Programm“, welches wirke, sondern Wirkung entstehe durch die situative Aufnahme

von einzelnen Programmelementen durch die Zielgruppen. Die Bedeutung von kontextuellen und subjektiven Faktoren für die Wirksamkeit von Interventionen würden bei experimentellen Designs generell vernachlässigt (Walach, 2006; Wright, 2006). Mittels Randomisierung würden kontextuelle und subjektive Faktoren nivelliert, obwohl doch gerade diese für den Erfolg einer Intervention entscheidend seien (Wright, 2006). Dasselbe gilt auf Ebene der Interventionen: diese müssten für experimentelle Designs standardisiert werden, obwohl gerade die Flexibilität in der Umsetzung oft erfolgsbestimmend sei (Wright, 2006; WHO, 1998a).

Unklarheit bestehe darüber hinaus in Bezug auf die mit randomisierten kontrollierten Studien zu messenden ‚Outcomes‘. Während es in der Kuration und in der medizinischen Prävention in erster Linie um Morbiditäts- und Mortalitätsindikatoren gehe, seien die Wirkungen der Gesundheitsförderung auf unterschiedlichsten Ebenen angesiedelt und würden sich primär nicht an Morbiditäts- und Mortalitätsraten orientieren, sondern gemäss salutogenetischem Ansatz an gesundheitsförderlichen individuellen und sozialen Ressourcen. Da Gesundheitsförderung grundsätzlich als Prozess definiert sei⁹, stelle sich zudem die Frage, inwiefern sich Evaluationen überhaupt an ‚Outcomes‘ ausrichten sollen und nicht eher Prozesse in den Vordergrund der Beurteilung rücken sollten (Bödeker, 2006). Zu beurteilende Ergebnisse seien dieser Argumentation zufolge nicht gesundheitliche Auswirkungen sondern „ein Empowermentprozess, der zu veränderten Strukturen (Lebenswelten) führen kann.“ (Ruckstuhl, 2003). Summative Evaluationen seien weniger geeignet, derartige partizipative und emanzipatorische Prozesse zu beurteilen, als beispielsweise Prozessevaluationen oder Formen der Aktionsforschung (Tones, 1997).

Diese Argumentationslinie läuft darauf hinaus, dass randomisierte kontrollierte Studien grundsätzlich nicht geeignet sind, komplexe soziale Interventionen und entsprechende Interaktionsprozesse zu beurteilen. Wenn gesellschaftliche Faktoren in der Gesundheitsförderung so bedeutend seien, müssten diese auch mit passenden sozialwissenschaftlichen Methoden untersucht werden.

Grenzen der Machbarkeit

Viele der in die Diskussion eingebrachten Kritikpunkte betreffen Fragen der Machbarkeit. Randomisierte kontrollierte Studien seien oft zu aufwändig und dadurch zu teuer (Tones, 1997; Uhl, 2012), insbesondere wenn nicht Individuen, sondern ganze Settings (z.B. Schulen, Betriebe, Stadtteile) als Untersuchungseinheiten dienen (Clusterrandomisierung). Bei settingorientierten Interventionen scheitert die Durchführung oft daran, dass nicht ausreichend viele Interventions- und Kontrollsettings zur Verfügung stehen oder eine entsprechende Studie vom Aufwand her nicht realisierbar ist. Die Bildung von Kontrollgruppen sei gerade dann, wenn Settings im Fokus stehen, sehr schwierig, da für die Kontrollgruppensettings ein erheblicher Aufwand und kein direkter Nutzen resultiere¹⁰. Zum Teil sei die Bildung von Kontrollgruppen aufgrund der Interventionsart auch gar nicht möglich, da eine gezielte Adressierung einzelner Gruppen (in vergleichbarem Kontext), zum Beispiel bei massenmedialen Massnahmen, nicht machbar sei (Nutbeam, 1998) oder die notwendige Stichprobengrösse etwa bei der Evaluation nationaler ‚Policies‘ nicht erreichbar sei.

⁹ “Health promotion is the process of enabling people to increase control over, and to improve, their health. To reach a state of complete physical, mental and social well-being, an individual or group must be able to identify and to realize aspirations, to satisfy needs, and to change or cope with the environment. Health is, therefore, seen as a resource for everyday life, not the objective of living. Health is a positive concept emphasizing social and personal resources, as well as physical capacities. Therefore, health promotion is not just the responsibility of the health sector, but goes beyond healthy life-styles to well-being.” (World Health Organization (1986))

¹⁰ Letzterem kann man mit Wartegruppensettings allerdings teilweise entgegenkommen.

Methodische Aspekte

Ein zentraler Aspekt sei die hohe Dynamik in Feldern der Gesundheitsförderung, welche dazu führe, dass der Untersuchungsgegenstand nicht konstant bleibe und nur begrenzt standardisiert werden könne (Elkeles, 2006). Es stelle sich infolgedessen die grundsätzliche Frage, worüber Erkenntnis erworben werde, da es *das* Projekt oder *das* Programm so gar nicht gebe. In diesem Zusammenhang müsse zudem beachtet werden, dass die Zielgruppen einen vergleichsweise hohen Einfluss auf den Erfolg von Interventionen haben, da sie – zumindest in partizipativen Ansätzen – direkt in die Umsetzung der Interventionen einbezogen und nicht lediglich Empfänger seien (Nutbeam, 1998). In Bezug auf die Festlegung zu überprüfender Wirkungszusammenhänge würden Gesundheitsförderungsprojekte, wenn sie partizipativ angelegt sind, die Schwierigkeit mit sich bringen, dass abhängige und mögliche unabhängige Variablen im Detail nicht von vorneherein bekannt und setting- und kontextspezifisch höchst variant seien (Elkeles & Broesskamp-Stone, 2012).

Ein anderer Aspekt betrifft die Grenzen experimenteller Methodik in Bezug auf die langen Wirkhorizonte von Gesundheitsförderungsprojekten und -programmen. Je mehr externe Einflussfaktoren sich zu den Projekt- oder Programmelementen hinzugesellen, desto kleiner werde der relative Beitrag der Intervention und desto schwieriger sei er statistisch zu erfassen (Campbell et al., 2000). Mit Blick auf die Randomisierung lassen sich weitere Schwierigkeiten ausmachen. Unterschiede in den Interventionen, die sich erst nach der Randomisierung entwickeln, lassen sich zwar statistisch kontrollieren. Derartige Analysen können aber zu Fehlinterpretationen führen, weil im Nachhinein gebildete Vergleichsuntergruppen nicht mehr nach dem Zufallsprinzip verteilt sind und entsprechenden Verzerrungen unterliegen können.

Eine weitere Gruppe von Argumenten bezieht sich auf einzelne methodische Schwierigkeiten und Grenzen in Bezug auf umfassende bevölkerungsbezogene Interventionen. Hierzu gehören zum Beispiel die Gefahr der Kontamination der Kontrollgruppe (Tones, 1997), die höhere Fehleranfälligkeit dadurch, dass eine Blindung seitens der Zielgruppen in der Regel nicht möglich ist oder ‚Biases‘, die bei der Rekrutierung der Interventions- und Kontrollgruppen entstehen (Elkeles, 2006).

Ethische Aspekte

Zusätzlich werden ethische Argumente ins Feld geführt, vor allem, dass den Kontrollgruppen die Interventionen vorenthalten würden. Allerdings ist dieses Dilemma nicht für die Gesundheitsförderung spezifisch und dürfte sich in der Medizin aufgrund oft direkter existenzieller Bedrohungen insgesamt stärker akzentuieren (Pawson & Tilley, 1997). Von etwas anderer Art ist die Argumentation, dass Fragen der Chancengleichheit („equality“) oft zu kurz kommen, wenn nur auf Wirksamkeit fokussiert wird (Stevens, 2004).

Randomisierte kontrollierte Studien sind in der Gesundheitsförderung also starker Kritik ausgesetzt und eine Übertragung auf das Feld der Gesundheitsförderung erscheint vielen Autorinnen und Autoren als nicht angemessen. Diese Studien würden nicht die richtigen Fragen beantworten oder generell nicht zum Gegenstand der Gesundheitsförderung passen und sie seien aus praktischen oder methodischen Gründen nicht umsetzbar oder ethisch nicht zu verantworten. Eine differenzierte und umfassende Auseinandersetzung mit dem Potenzial und den Grenzen dieser Methoden für die Gesundheitsförderung sucht man auf Seite der Gesundheitsförderung jedoch vergebens, was gerade angesichts der Fortschritte in diesem Bereich und der Annäherung an komplexe Fragestellungen (Lee,

2006) bedauernd ist¹¹. Auch wird die evidenzbasierte Medizin teilweise verkürzt dargestellt, z.B. bei Tang 2003¹², um Fragen der Evidenz in der Gesundheitsförderung davon abheben zu können.

Da die Gesundheitsförderung randomisierte kontrollierte Studien mehrheitlich ablehnt, aber gleichzeitig ihre Wirkungen belegen muss, gerät sie in ein Dilemma, welches die Methodendiskussion rund um die Evaluation von Gesundheitsförderung und Prävention bestimmt.

2.3 Das Evidenzdilemma in der Gesundheitsförderung

Die Argumente gegen eine Übertragung des ‚Goldstandard‘ und damit der Evidenzklassen der evidenzbasierten Medizin auf die Gesundheitsförderung sind vielfältig, doch der Widerstand ist praktisch einhellig: Randomisierte kontrollierte Studien scheinen sich nur in Ausnahmefällen¹³ (in denen sie dann allerdings sehr willkommen sind) für die Evaluation von Interventionen im komplexen Handlungsfeld der Gesundheitsförderung zu eignen (Mc Queen, 2001). Die Komplexität sozialer Systeme und die unterschiedlichen Evaluationszwecke verlangen nach anderen als experimentellen Methoden:

Im Gegensatz zur evidenzbasierten Medizin, deren Datenbasis klinische und epidemiologische Studien mit eng umgrenzten Fragestellungen sind, die zu Aussagen über die Wirksamkeit bestimmter medizinischer Handlungen führen, muss in der Gesundheitsförderung die Datenbasis von (komplexen) Projekten und deren Evaluationen prinzipiell wesentlich breiter angelegt sein. (Elkeles & Broesskamp-Stone, 2012)

Daraus ergibt sich für die Gesundheitsförderung ein ‚Evidenzdilemma‘: obwohl die Gesundheitsförderung unter hohem Druck steht, ihre Wirksamkeit zu belegen und ihre Interventionen auf Evidenzen zu basieren, scheint gerade jener Studientyp dazu ungeeignet, welchem exklusiv zugestanden wird, Wirksamkeit nachweisen zu können¹⁴. Je einfacher, überschaubarer und je besser kontrollierbar die Interventionen und je kurzfristiger ihr Wirkhorizont – also je weniger eine Intervention den Grundprinzipien der Gesundheitsförderung entspricht, umso eher sind randomisierte kontrollierte Studien die Methode der Wahl (Mc Queen, 2001). Mit steigender Komplexität der Interventionen würden die methodischen Herausforderungen zunehmen und randomisierte kontrollierte Studien immer fragwürdiger werden (vgl. dazu auch Blättner & Niebuhr, 2015). Dabei werden folgende Komplexitätsmerkmale angeführt (Tones, 1997; Tones, 2000; Wright, 2006; Walach, 2006); (Bödeker, 2006; Trojan, 2006; Tang et al., 2003; Mc Queen, 2001):

- Hohe Anzahl und Mischung involvierter Akteure aus unterschiedlichen sozialen, wirtschaftlichen, kulturellen, ökologischen und politischen Kontexten
- Bündelung verschiedenster Strategien und Massnahmen (individueller, sozialer und umfeldbezogener Art)
- Ganzheitlicher Ansatz
- Interdisziplinarität und Multisektoralität

¹¹ Quantitative Wirkungsforschung bedient sich heute multivariater Analysen, Strukturgleichungsmodellen, Zeitreihen-, Cluster-, Netzwerk-, Mehrebenen- und vieler weiterer Analysen, um komplexe Wirkungszusammenhänge differenziert zu untersuchen (Lee (2006), 152).

¹² Mit Tangs Befund: „In practice, it is important to note that evidence alone cannot constitute effective practice, as is the case of evidence based medicine.“ wird u.a. vernachlässigt, dass in der evidenzbasierten Medizin die klinische Erfahrung der Ärzte eine wichtige Säule neben der wissenschaftlichen Evidenz darstellt (vgl. Sackett, Rosenberg, Gray, Haynes, and Richardson (1996)).

¹³ „When control of the setting and population under study can be achieved for the time of the trial, and when there is a focus on a single intervention with an expected dichotomous outcome of success or failure, the RCT is indeed a powerful methodology...“ (Mc Queen (2007), 37)

¹⁴ Bei Verzicht auf die Randomisierung, also bei quasi-experimentellen Studien, können festgestellte Unterschiede zwischen Interventionsgruppe und Kontrollgruppe nicht mehr mit Sicherheit (im Rahmen der statistischen Aussagekraft) auf die untersuchte Intervention zurückgeführt werden (Tones (1997)).

- Kombination verschiedener Handlungsfelder¹⁵
- Hoher Prozesscharakter der Interventionen mit grundsätzlicher Offenheit ihres Ausgangs
- Mitwirkung der Akteure (partizipative und emanzipatorische Prozesse)
- Grosser Einfluss vielfältiger kontextueller und subjektiver Faktoren
- Vielfältige Ergebnisparameter auf unterschiedlichen Ebenen und langfristige Wirkhorizonte

Die Gesundheitsförderung unterscheidet sich wesentlich durch diese Komplexität von individuellen medizinischen Behandlungen. Obwohl auch die evidenzbasierte Medizin mit Schwierigkeiten und Grenzen von experimentellen Studien konfrontiert ist (Cochrane, 1972), sind diese weniger grundsätzlicher Natur. Vor dem Hintergrund der Komplexität der Gesundheitsförderung lassen sich die zentralen Hindernisse für die Übertragbarkeit der evidenzbasierten Medizin auf die Gesundheitsförderung wie folgt benennen:

Die hohe Komplexität und daher geringe Standardisierbarkeit der Interventionen, die indirekte Wirkungsweise über Veränderung von Lebensstilen und Lebensweisen und die daher erschwerte Erfassung von Endpunkt-Effekten (Morbidity und Mortalität nur sehr bedingt geeignet), die langfristige Ausrichtung (Effekte erst nach geraumer Zeit zu erwarten), die Entwicklung/Entstehung von Gesundheit durch Interaktion Person-Umwelt. Randomisierung unter kontrollierten Bedingungen ist kontraproduktiv, da die Interaktionseffekte ausgeschaltet werden. (Noack, 2002 zit. in Trojan, 2006, 80)

Wright fasst zusammen, dass klinische Versuche als Massstab für die Wirksamkeit der lebensweltlichen Primärprävention (und damit auch der Gesundheitsförderung) nicht angemessen seien, weil:

Die Wirkungsweise dieser Arbeit wird nicht adäquat berücksichtigt, die Übertragbarkeit (externe Validität) jenes Interventionsversuches wird überschätzt, die zum Teil erheblichen Unterschiede innerhalb der Zielgruppe werden nicht beachtet, die Möglichkeit, Kausalität festzustellen wird überbewertet und der Aufwand, den solche Versuche verlangen, steht in keinem Verhältnis zum Ergebnis. (Wright, 2006, 62)

Mit dieser Kritik zeichnet sich eine Schwerpunktverschiebung des Begriffs Evidenz ab, weg von der Evidenzhierarchie und der Fokussierung auf Ergebnisse randomisierter kontrollierter Studien und hin zu einem weiteren Verständnis, welches auch andere, vielfach als weniger bedeutend eingestufte Quellen wissenschaftlichen Wissens (und teilweise anderen Wissens) mit einschliesst (Mc Queen, 2001). Die Evaluation der Gesundheitsförderung steht zwischen Evidenzansprüchen, die sich aus naturwissenschaftlicher Logik ergeben einerseits und sozialwissenschaftlichen Forschungsgegenständen und Fragestellungen andererseits:

When behaviour of individuals, organisations, or the political process is the focus, whether it relates to lifestyle practices, sales decisions or public policy formulation, the explanatory power of science will be limited. Above all, it may be problematic to fit these rules of the natural sciences to the social science disciplines. (Tang et al., 2003, 841)

Die forschungsmethodischen Diskussionen, mit denen sich die Gesundheitsförderung beschäftigt, führen damit gleichzeitig zu grundsätzlichen, erkenntnistheoretischen Fragestellungen, die in der Debatte bisher lediglich angeschnitten, aber nicht ausgiebig behandelt oder gar beantwortet werden (Bödeker, 2006). Bevor eine eigene forschungsmethodische und erkenntnistheoretische Position entwi-

¹⁵ In der Ottawa-Charta (World Health Organization (1986)) werden die fünf Handlungsfelder „Entwicklung einer gesundheitsfördernden Gesamtpolitik“, „Gesundheitsfördernde Lebenswelten schaffen“, „Gesundheitsbezogene Gemeinschaftsaktionen unterstützen“, „Persönliche Kompetenzen entwickeln“ und „Gesundheitsdienste neu orientieren“ unterschieden, eine Kombination derselben in Gesundheitsförderungsprogrammen hat sich als erfolgreich erwiesen und wird empfohlen (WHO (1998c)).

ckelt wird, um sich dann der Evaluation komplexer sozialer Interventionen zu nähern, soll zunächst ein Blick auf jene Ansätze geworfen werden, welche die Gesundheitsförderung aus dem Evidenzdilemma hinaus führen sollen – dass dabei ein weiterer Evidenzbegriff als jener der evidenzbasierten Medizin zur Anwendung kommen muss, dürfte aus der bisherigen Diskussion bereits ersichtlich geworden sein.

3 Erweiterung des Evidenzbegriffs

Einige Autorinnen und Autoren sprechen sich für eine Ausweitung des Evidenzbegriffs aus, wenn dieser in der Gesundheitsförderung Verwendung finden soll (Tang et al., 2003; Nutbeam, 1998; WHO, 1998a; Wright, 2006; Tones, 2000; Elkeles, 2006). Diese Erweiterung wird einerseits erkenntnistheoretisch, in Ablehnung des naturwissenschaftlich-positivistischen Erkenntnisparadigmas, begründet. Tang et al. (2003) kritisieren, dass die in den Naturwissenschaften gültigen Gütekriterien der Voraussagbarkeit, der Reproduzierbarkeit und Falsifizierbarkeit auf sozialwissenschaftliche Phänomene nur bedingt anwendbar seien, weil deren Einzigartigkeit und Komplexität kausalen Erklärungen allgemeiner Natur widersprechen würden. Wright führt an, dass die Sozialwissenschaften, im Gegensatz zu eher positivistischen Ansätzen der Medizin, durch mehrere erkenntnistheoretische und methodische Positionen gekennzeichnet seien wie beispielweise die verstehende, interpretative oder kritische Tradition (Wright, 2006). Subjektives und Situatives erlange in den Sozialwissenschaften grössere Bedeutung als beispielsweise Objektivität und Quantifizierung und dies lasse sich nicht angemessen mit experimentellen Verfahren erfassen (Wright, 2006).

Auf der anderen Seite bietet sich eine Ausweitung des Evidenzbegriffs auch mit Verweis auf die Vielfalt von Evaluationsgegenständen und Fragestellungen an. Die starke Fokussierung des Evidenzbegriffs auf den Nachweis der Wirksamkeit von Interventionen¹⁶ ist nicht nur für die Gesundheitsförderung eine unangemessene Einschränkung. Walach zeigt - am Beispiel der Prüfung einer pharmakologischen Substanz - dass das wissenschaftliche Erkenntnisinteresse auch in medizinischen Untersuchungen weit über die enge Frage der Wirksamkeit hinausreicht und z. B. Aspekte der Übertragbarkeit der experimentellen Ergebnisse auf den klinischen Alltag oder Fragen zu möglichen Nebenwirkungen umfasst. Damit wird schnell deutlich, dass die Beantwortung derartig unterschiedlicher Fragestellungen unterschiedlicher Methoden bedarf (Walach, 2006). Der komplexe Charakter von Interventionen der Gesundheitsförderung rückt in verstärktem Masse differenzierte Fragen zur Wirkung und Wirkungsweise in den Vordergrund. In Anlehnung an die „Realist Evaluation“ (Pawson & Tilley, 1997) gilt das Erkenntnisinteresse kontextspezifischen Wirkungsmechanismen und führt zu den oben bereits angesprochenen differenzierten Fragen, bei wem, unter welchen Bedingungen und auf welche Weise eine bestimmte Intervention Wirkungen erzeugt. Andere Studientypen als randomisierte kontrollierte Studien seien in vielen Fällen angemessener und würden wertvollere Erkenntnisse zur Wirksamkeit von Projekten liefern als randomisierte kontrollierte Studien (Grossman, 2008). Es sei wenig gewonnen zu erfahren, ob ein Projekt oder ein Programm wirkt – erstens, weil es *das* Projekt oder *das* Programm so oft nicht gibt und zweitens weil es erkenntnisreicher ist zu erfahren, unter welchen Umständen was, für wen, wie wirkt (Mc Queen, 2001). Realistische Evaluation hilft auch dann, höchst relevante förderliche und hinderliche Faktoren heraus zu arbeiten, wenn keine der angestrebten Hauptwirkungen nachgewiesen werden können (Hills & Carroll, 2004). Damit rückt mehr und mehr auch der Evaluationszweck ‚Lernen‘ anstelle von ‚Beweisen‘ in den Vordergrund des Evaluationsinteresses (Elkeles, 2006).

Die Verlagerung des Fokus vom Wirksamkeitsnachweis auf die Wirkungsweise von Interventionen führt zu Wirkungsketten und damit zu intermediären Ergebnisparametern resp. intermediären Zielen. Nicht epidemiologische Endparameter wie Morbiditäts- und Mortalitätsraten seien für Gesundheitsförderungsprojekte oder -programme primär relevant, sondern vielfältige intermediäre resp. indirekte Indikatoren wie beispielsweise Wissens- oder Kompetenzerwerb (Tones, 2000). Mit diesen intermedi-

¹⁶ Im englischen Sprachgebrauch werden in der Forschung zwei Begriffe für Wirksamkeit unterschieden: Der Begriff ‚Efficacy‘ ist eng mit experimenteller Forschung verknüpft, da er sich auf den Nachweis von Kausalitäten unter kontrollierten Bedingungen bezieht. Unter ‚Effectiveness‘ einer Intervention wird das Auftreten von Wirkungen unter natürlichen, also nicht kontrollierten, Alltagsbedingungen verstanden (Windeler, Antes, Behrens, Donner-Banzhoff, and Lelgemann (2008)).

ären Indikatoren gelangen zusätzliche Dimensionen ins Blickfeld, so z.B. Fragen der Inanspruchnahme und Akzeptanz von Interventionen oder auch politische und organisationale Rahmenbedingungen (Nutbeam, 2000; Spencer et al., 2008). Kontextspezifische Faktoren, also die zeitliche und örtliche Bedingtheit von Wirkungen, werden immer augenfälliger und führen fast zwangsläufig zur Erkenntnis, dass Evidenz letztlich immer nur kontextgebunden, also zeitlich und örtlich bedingt sein kann (Elkeles, 2006). Dieser Argumentationsstrang führt zu Ansätzen, die eine Ablösung des externen wissenschaftlichen, als Expertenwissen verstandenen Evidenzbegriffs fordern: Evidenz in der Gesundheitsförderung wird gemäss Wright (Wright, 2006) über partizipative Prozesse primär lokal erzeugt. Diese lokale oder partizipative Evidenz auf Basis lokaler Theorie¹⁷ stelle die Dominanz des extern erzeugten, all-gemeingültigen wissenschaftlichen Wissens über lokal erzeugtes Wissen grundsätzlich in Frage:

Der Begriff der « lokalen Evidenz » soll alle Nachweise des Erfolgs einer Intervention bezeichnen, die die Erfüllung der Ziele der Intervention angemessen belegen. Im Unterschied zur üblichen Anwendung des Begriffs « Evidenz » in der oben umrissenen Diskussion um die evidenzbasierte Medizin geht es hier nicht darum, « absolute » oder « allgemeine » Nachweise für ein bestimmtes Verfahren zu liefern, sondern um die Feststellung, ob die vor Ort konzipierten Interventionen dazu dienen, dem spezifischen lokalen Gesundheitsproblem entgegenzuwirken. So wie beim Konzept der lokalen Theorie liegt hier die Bestimmungsmacht über die Angemessenheit der erbrachten Nachweise im konsensuellen Verfahren auf lokaler Ebene. (Wright, 2006, 70; Hervorhebungen im Original)

Wright ist sich bewusst, dass eine so konzipierte lokale Theorie und lokale Evidenz dem Vorwurf ausgesetzt sein wird, unsystematisch und lokal interessengeleitet zu sein und dass wissenschaftliches und Alltagswissen in seiner Konzeption kaum mehr auseinander zu halten sind¹⁸. Er schlägt mit Rückgriff auf Tones vor, zur Absicherung und Systematisierung lokal erzeugten Wissens eine Art Peer-Review Verfahren zu etablieren, welches sich am Plausibilitätsprüfungsverfahren der Rechtssprechung orientiert. Tones (Tones, 2000) hat diesen Ansatz des ‚judicial review‘ aus der Rechtssprechung abgeleitet und auf die Gesundheitsförderung übertragen und er schlägt vor, Evidenzen, die aus unterschiedlichen Quellen stammen und u.a. durch verschiedene Formen der Triangulation gewonnen wurden, einer Plausibilitätsprüfung zu unterziehen. Allerdings wären hierzu die Gütekriterien – in Abgrenzung zu jenen der evidenzbasierten Medizin – noch zu entwickeln¹⁹. Ein Ansatz, welcher sich insbesondere auch der Frage des Verhältnisses zwischen lokalem Wissen und Wissen im „üblichen, wissenschaftlichen Sinne“ (Wright, Roche, von Unger, Block, & Gardner, 2010) annimmt, ist die sich an der Aktionsforschung orientierende partizipative Gesundheitsforschung, welche ihre Entsprechung in der ‚community-based participatory research‘ in den USA hat. Zentrale Aspekte dieses Ansatzes sind das voneinander Lernen, der wechselseitige Transfer von Expertise zwischen allen Forschungspartnern sowie geteilte Entscheidungsbefugnis und gemeinsames Eigentum an den Prozessen und Produkten des Forschungsunternehmens. Der Ansatz ist allerdings noch jung und erkenntnistheoretische und forschungsmethodische Grundlagen müssen noch entwickelt werden (Wright et al., 2010), der partizipative Ansatz sprengt den üblichen erkenntnistheoretischen und methodischen Rahmen der Qualitätssicherung und Evaluation (Wright, 2006, 59). Insgesamt sind partizipative Evaluationen

¹⁷ Eine lokale Theorie bezieht sich im Unterschied zur lokalen Evidenz auf die Beschreibung eines Gesundheitsproblems und dessen Ursachen: „Eine lokale Theorie ist eine durch einen partizipativen Prozess konsensuell formulierte Erklärung eines lokalen Gesundheitsproblems, die die Merkmale des Problems vor Ort konkret beschreibt, lokale Ursachen des Problems erläutert und die Grundzüge für eine Bekämpfungsstrategie umfasst, die einen unmittelbaren Bezug zum Problem und dessen Ursachen herstellt.“ (Wright (2006), 68)

¹⁸ Ein anderer Vertreter der Evidenzdebatte anerkennt den Umstand, dass Praxiserfahrungen bisher oft die beste empirische Evidenz darstellen, er vertritt hingegen klar die Haltung, dass ein Modell guter Praxis im strengen Sinne aber nur eines sein kann, „das auf der Basis einer möglichst rigorosen Fremdevaluation seine Wirksamkeit bewiesen hat.“ (Trojan (2006), 103)

¹⁹ Tones schlägt entsprechende Kriterien vor, so z.B. „(1) a strong, consistent, specific association; (2) a temporally correct „dose-response relationship“; (3) theoretical plausibility“ (Tones (2000), 232) und mit Rückgriff auf Janesick die beiden Validitätskriterien „transferability“ und „catalytic validity“ (Tones (2000), 232 ff).

für die Gesundheitsförderung von grosser Anziehungskraft, da sie dem Grundkonzept der Gesundheitsförderung entsprechen (Halkow, 2004) und auch die WHO fordert in ihrer Resolution zur Evaluation in der Gesundheitsförderung, diese entsprechend zu fördern (WHO, 1998a).

Mit der Erweiterung des Evidenzbegriffs in oben ausgeführten Dimensionen geht gleichzeitig ein Methodenpluralismus einher, der von den Exponentinnen und Exponenten der Evidenzdebatte propagiert wird (Tones, 2000; Wright, 2006; Nutbeam, 1998; Luber, 2004; WHO, 1998a; Thorogood & Coombes, 2004; Grossman, 2008). Tones fasst die Bedeutung dieses Pluralismus wie folgt zusammen: "... methodological pluralism is not only acceptable but essential" (Tones, 2000, 235). Nach Luber liegt die geforderte methodische Vielfalt, welche auch die Kombination verschiedener Methoden umfasst, in der Komplexität des Forschungsfeldes selbst begründet: "Gesundheitsförderung als komplexes Zusammenspiel von Interventionen braucht ein ebenso komplexes Zusammenspiel von Methoden der Evaluation" (Luber, 2004, 44). Ähnlich leitet Mc Queen (2000) aus dem ganzheitlichen Konzept der Gesundheitsförderung ab, dass auch die Evaluationsansätze ganzheitlich sein sollten.

Die Wahl der Methode sei keine vom Untersuchungsgegenstand und vom Erkenntnisinteresse unabhängig zu beantwortende Frage, sondern situativ immer wieder neu zu stellen, wie dies beispielsweise in der Evaluationsforschung üblich sei (Elkeles, 2006). Infolgedessen ist methodische Rigorosität auch keine Eigenschaft, welche experimentelle Studien für sich allein reklamieren können, sondern eine Frage ihrer Fundierung und der Gründlichkeit ihrer Anwendung. Um die Evidenzbasis in der Gesundheitsförderung steigern zu können, müsse die Diskussion auf diese Fragen gelenkt werden und weg von der wenig fruchtbaren Konkurrenz zwischen quantitativen und qualitativen Methoden (Coombes & Thorogood, 2004; Speller, Learmouth, & Harrison, 1997). Qualitativen sozialwissenschaftlichen Forschungsmethoden wird allerdings vielfach ein besonders grosses Potential zugesprochen, die differenzierten Aspekte der Gesundheitsförderung zu erfassen (Israel, Schulz, Parker, & Becker, 1998; Bodstein, 2007) und quantitative und qualitative Methoden müssten in Forschungsdesigns kombiniert werden (Nutbeam, 1998). Die verschiedenen Formen der Triangulation werden als geeignet erachtet, um Evidenzen aus verschiedenen Quellen miteinander zu verbinden und dadurch ein ganzheitliches und dichtes Bild über den Untersuchungsgegenstand zu erhalten (Nutbeam, 1998; Tones, 2000). In Forschungsdesigns können unterschiedliche Methoden zu unterschiedlichen Zeiten erforderlich werden (Oliver et al., 2005) oder aber die unterschiedlichen intermediären Ergebnisparameter und die verschiedenen Interventionen und Aspekte machen den Einsatz unterschiedlicher Methoden erforderlich (Ackermann & Pfyl, 2010; Speller et al., 1997). Dabei können durchaus auch experimentelle oder quasi-experimentelle Forschungsdesigns zur Beurteilung eingegrenzter Wirkungsfragen zum Einsatz kommen, wobei quasi-experimentelle Studien hinsichtlich Wirkungsnachweisen im Spannungsfeld zwischen Anspruch und (komplexer) Wirklichkeit eine interessante Mittlerposition einzunehmen scheinen (Caspari, 2012; Gertler, Martinez, Premand, Rawlings, & Vermeersch, 2011).

Der methodische Pluralismus, welcher für die Evaluation der Gesundheitsförderung propagiert wird, bricht die Evidenzhierarchie der evidenzbasierten Medizin auf und macht eine methodische Hierarchisierung fragwürdig:

Komplexe soziale Interventionen erfordern wegen ihrer mangelnden Standardisierbarkeit und ihrer starken Kontextabhängigkeit ein anderes Evidenzparadigma als das Standardmodell der hierarchischen Stufen der Cochrane Collaboration. (Trojan, 2006, 105)

Bisher hat sich keine neue Form der Hierarchisierung der Methoden für die Evaluation in der Gesundheitsförderung etabliert (Mc Queen, 2001). Es liegen aber erste alternative Klassifizierungen vor. Walach (2006) stellt der Evidenzhierarchie ein zirkuläres Modell entgegen, welches verschiedene Forschungsmethoden wie randomisierte kontrollierte Studien, Kohortenstudien, Pfadmodelle, Beobachtungsstudien oder nicht randomisierte Vergleiche als potenziell gleichberechtigt nebeneinander stellt. Einzelne Forschungsmethoden sind darin nicht hierarchisch geordnet, sondern „komplementär und zirkulär aufeinander bezogen“ (Walach, 2006, 49). Erkenntnisse, die mittels der einen Forschungsmethode gewonnen wurden, können dabei neue Fragen aufwerfen, welche mit anderen For-

schungsmethoden beantwortet werden müssen. Dieser Ansatz gründet auf den beiden Voraussetzungen, dass einerseits Studien mit hoher interner Validität mit solchen hoher externer Validität gekoppelt werden sollen²⁰ und dass zweitens Methoden nicht per se hierarchisch gegliedert sein können, sondern höchstens im Hinblick auf zu beantwortende Fragestellungen priorisiert werden können. Wenn eine Hierarchisierung der Evidenz gefordert sei, dann dürfe nicht die Methodenwahl das Kriterium sein, sondern die Güte, mit welcher die jeweilige Methode umgesetzt werde.

Eine weitere alternative Klassifizierung schlagen Tang et al. (2003) vor. Sie fordern, dass Evidenz über das enge naturwissenschaftliche Verständnis der Wirksamkeit hinausgehen müsse und schlagen eine Klassifizierung vor, welche auch Erkenntnisse über Interventionen enthält, deren Wirkung zwar gezeigt, aber deren Wirksamkeit - im Sinne nachgewiesener Kausalität - nicht belegt werden kann. Sie unterscheiden konkret zwischen Evidenz aus Interventionen mit A) nachgewiesener Wirksamkeit und allgemeiner Gültigkeit, B) nachgewiesener Wirksamkeit aber kontextuell begrenzter Gültigkeit, C) eingetretener Wirkung und allgemeiner Gültigkeit, ohne aber dass deren Wirksamkeit nachgewiesen ist und D) eingetretener Wirkung, aber kontextuell begrenzter Gültigkeit und ohne dass deren Wirksamkeit nachgewiesen ist. Auch dieser Klassifizierung ist eigen, dass nicht die Methodenwahl letztlich entscheidend ist für eine Bewertung der Evidenzgüte. Tang et al. weisen zudem darauf hin, dass wissenschaftliche Evidenz praktische Expertise zwar ergänzen, aber niemals aufheben könne, da letztere immer über die Anwendbarkeit im konkreten Falle zu entscheiden habe. Diese Ergänzung liegt allerdings auch bereits dem Evidenzbegriff der evidenzbasierten Medizin zugrunde, ein Umstand, der in der Evidenzdebatte in der Gesundheitsförderung aber oft übersehen wird (Tang et al., 2003).

Im Gegensatz zu einer gänzlichen Ablehnung experimenteller Forschung für das Gebiet der Gesundheitsförderung und Prävention bieten die Ausweitung des Evidenzbegriffs und eine Auflösung einer starren Evidenzhierarchie zugunsten einer differenzierteren kontextspezifischen Beurteilung der Güte und Angemessenheit wissenschaftlichen Wissens interessante Alternativen.

3.1 Konsequenzen für die Evaluation von Gesundheitsförderung und Prävention

Es lassen sich derzeit zwei Strömungen ausmachen, welche die Entwicklung der Evaluation von Gesundheitsförderung und Prävention beeinflussen. Auf der einen Seite zunehmender Druck von aussen hinsichtlich Evidenzbasierung und klarem Wirkungsnachweis auf Basis von (quasi-) experimentellen Studien. Auf der anderen Seite eine Abwendung von ebendiesen Studientypen in der Evidenzdebatte, die derzeit in der Gesundheitsförderung geführt wird. Die Diskussion innerhalb der eigenen Disziplin lässt folgende Entwicklungslinien erkennen:

- Der Evidenzbegriff der evidenzbasierten Medizin mit der zugrundeliegenden Hierarchisierung der Evidenzen und die damit einhergehende Vorrangstellung randomisierter kontrollierter Studien ist der Gesundheitsförderung nicht angemessen – neue Formen der Beurteilung und Klassifizierung von Evidenzen sind gefordert. Die Evaluation der Gesundheitsförderung – um grob zu pauschalisieren – verabschiedet sich damit von einem stark positivistischen Forschungsparadigma und favorisiert sozialkonstruktivistische Ansätze und pluralistische Evaluationen²¹.
- Das Konzept der Wirksamkeit, also der in experimentellen Versuchen bestätigten und generalisierbaren Wirkungen, spielt in Evaluationen der Gesundheitsförderung eine untergeordnete Rolle. Fragen der kontextspezifischen konkreten Wirkungen und Wirkungsweisen sowie damit einhergehend prozessorientierte Fragen erfolgreicher Implementierung stehen im Fokus der Evaluation,

²⁰ Beide Gütekriterien in einer Studie zu vereinen, sei nur in Ausnahmefällen möglich.

²¹ Der Begriff der pluralistischen Evaluation wurde von Means and Smith geprägt. Ausgehend davon, dass objektive Realität nicht existiert, müssen die Perspektiven verschiedener Akteure in eine Bewertung einbezogen werden (Means and Smith (1988)).

und systematische Plausibilisierungen treten an die Stelle der Beweise der Wirksamkeitsforschung (Elkeles, 2006). Stufflebeam (2003) umschreibt diese Verschiebung des Fokus mit "evaluation's most important purpose is not to prove, but to improve."

- Kontextuelle, Prozess- und subjektive Faktoren spielen bei der Beurteilung von Interventionen und deren Wirkungen eine entscheidende Rolle und dürfen nicht „wegrandomisiert“ werden. Kontextuellem und lokalem Wissen kommt entsprechend in Evaluationen der Gesundheitsförderung zentraler Stellenwert zu.
- Lokale Wissensträger/-innen werden ergänzend zu Vertreterinnen und Vertretern der Wissenschaft wichtig und sollen über partizipatorische Prozesse in die Herstellung von Evidenz eingebunden werden. Allerdings ist bisher nicht klar, auf welche Weise dies geschehen soll und mit welchen erkenntnistheoretischen und methodologischen Folgen. Die Entwicklungen im Rahmen der partizipativen Gesundheitsforschung scheinen das Potenzial zu haben, aus der Abwehrhaltung gegen die evidenzbasierte Medizin hinauszuführen und eine eigene Position und entsprechende Gütekriterien zu entwickeln (Wright et al., 2010).
- Die Evaluation in der Gesundheitsförderung ist aktuell von einer grossen Vielfalt von Ansätzen, Modellen, Designs und Methoden geprägt (WHO, 2001). Dies gründet einerseits in der Vielfalt der Forschungsgegenstände und Fragestellungen, mag andererseits aber auch eine gewisse Verunsicherung und das Fehlen von entsprechenden Leitkonzepten widerspiegeln.
- Ergänzend zu Ergebnisevaluationen und mit diesen in ganzheitlichen Konzepten zwingend verbunden sind Prozessevaluationen wichtig, die sich mit Fragen der Implementierung von Interventionen und der Rezeption durch die Zielgruppen befassen (Nutbeam, 1998; WHO, 1998a). Dies umso mehr, wenn Gesundheitsförderung als partizipativer Prozess und nicht als Wirkung definiert wird (Coombes & Thorogood, 2004)).

Die Gesundheitsförderung müsse eigene theoretische Grundlagen, Ansätze und Kriterien entwickeln und darauf aufbauend eine eigene Position erarbeiten, um nicht Gefahr zu laufen, mittels unangemessener Forschungsdesigns evaluiert und dann als nicht wirksam bewertet zu werden (Speller et al., 1997). Allerdings stehe sie diesbezüglich erst am Anfang (Elkeles, 2006). Bis auf wenige neuere Arbeiten (Robert Koch Institut, 2012; Jones, 2011b) fehlen systematische Auseinandersetzungen der Gesundheitsförderung und Prävention mit den Konzepten der Komplexitätstheorie. Auch in der Evaluationsliteratur sind diese Konzepte erst seit kurzem in der Diskussion (z.B. Patton, 2011; Forss, Marra, & Schwartz, 2011; Ackermann et al. 2009). Es bietet sich an, dass die Gesundheitsförderung und Prävention mit vergleichbaren Handlungsfeldern oder Disziplinen zusammenspannt, die mit gleich gelagerten Herausforderungen konfrontiert sind. Hier wäre insbesondere die Soziale Arbeit ein gewichtiger Partner, da sie der Gesundheitsförderung aufgrund ihrer Grundprinzipien²², aber auch aufgrund ihrer politischen Exponiertheit, zumindest teilweise ähnelt, gleichzeitig aber in methodologischen Diskussionen und in ihrer Professionalisierung ein gutes Stück weiter fortgeschritten ist.

Gleichzeitig solle auch bereits bestehende Evidenz – sowohl nach einem engeren als vor allem auch nach einem erweiterten Verständnis – besser genutzt werden (Nutbeam, 1998; Bödeker, 2006; Tones, 1997). Leider besteht hierbei noch immer eine Art ‚publication bias‘, dass also nur jene Studien Eingang in entsprechende Reviews finden, welche den Evidenzkriterien der relevanten Journals oder Forschergruppen (u.a. der Cochrane library) entsprechen – und diese orientieren sich noch allzu oft an der in der Gesundheitsförderung umstrittenen Evidenzhierarchie der evidenzbasierten Medizin.

Die Komplexitätsdiskussion kommt der Gesundheitsförderung aus genannten Gründen sehr gelegen. Die Komplexitätstheorie liefert die Grundlage dafür, das Unbehagen mit den naturwissenschaftlich-reduktionistischen Methoden und Modellen formulieren und wissenschaftlich begründen zu können und nach neuen methodischen Ansätzen zu suchen. Sozialwissenschaftliche Theorie und Forschung

²² Unter anderem: Chancengleichheit, Partizipation, Autonomie und Empowerment.

scheinen der Gesundheitsförderung besser zu entsprechen als die medizinische Forschung (Wright, 2006) und es werden Vorstellungen über Alternativen zu randomisierten kontrollierten Studien und zu klassischen epidemiologischen Studien skizziert²³:

Health promotion makes room for the stories which individuals and communities tell about their everyday experience of health, and legitimizes them as being as important to our understanding of health as statistics on morbidity and mortality rates. (Robertson, 1998, zit. in Tremblay & Richard, 2011, 6)

Der holistische Charakter der Gesundheitsförderung, Prinzipien wie Partizipation, Empowerment und Settingorientierung sowie die Vielschichtigkeit der Ansätze, Methoden und Wirkungszusammenhänge fordern dazu auf, traditionelle Vorstellungen zum Beispiel über Kausalität zu überdenken und im Rahmen der Komplexitätstheorie neu zu konzeptualisieren. Dabei wird Komplexität teilweise als neues Paradigma für die Forschung in der Gesundheitsförderung proklamiert (Tremblay & Richard, 2011; Rutter, 2012). Im Zentrum eines solchen Paradigmas würden gemäss Tremblay (2011) die subjektive Perspektive der Betroffenen und die diesbezügliche Reflexivität der Forschenden stehen:

In health promotion, the values of participation and empowerment demand the adoption of a relativistic or critical viewpoint that legitimates the experiential knowledge of individuals and recognizes reality as being subject to various contextual, historic and social contingencies. These positions contrast sharply with those of realism and dualism, which perceive the researcher as occupying an external position in relation to the subjects studied. The relativistic, critical and subjective approaches, on the other hand, allow access to the interpretation of reality as it is experienced by individuals themselves. (Tremblay & Richard, 2011, 5)

Ein für die Gesundheitsförderung spezifisches Forschungsparadigma lässt sich aus der Eigenheit des Handlungsfeldes nachvollziehbar herleiten, und der von Wright (2006) entwickelte Ansatz der lokalen oder partizipativen Evidenz gibt innovative Impulse in diese Richtung. Um das Potenzial eines derartigen Paradigmas allerdings beurteilen zu können, muss es zuerst an Konturen gewinnen. Bislang sind noch zu viele Fragen unbeantwortet, insbesondere was die Rolle und Berechtigung von Wissenschaft in derartigen Ansätzen betrifft²⁴.

Die Zurückweisung der Vorrangigkeit randomisierter kontrollierter Studien und die Forderung nach einem neuen Evaluationsverständnis wird vor allem mit der vergleichsweise hohen Komplexität sozialer Systeme begründet, die in der Gesundheitsförderung untersucht werden. Dabei wird zum Teil explizit auf die Komplexitätstheorie Bezug genommen:

The theory of complexity makes explicit that the „rules for evidence“ lie in the reductionistic approach to science and that approach is inadequate to understanding and explanation in health promotion. (Mc Queen, 2007, 36)

In einem nächsten Schritt soll deshalb untersucht werden, welche theoretischen Konzepte sich hinter dem Schlagwort Komplexität verbergen. Dies, um eine fundierte Diskussion zu Evaluation in komplexen Kontexten führen und die empirischen Befunde entsprechend einbetten zu können.

²³ Vgl. dazu auch die verschiedenen, allerdings erst zögerlichen, Ansätze in Bödeker and Kreis (2006).

²⁴ Vgl. in diesem Zusammenhang die wissenschaftskritische Haltung von Feyerabend (1980).

4 Komplexität

Der Begriff der Komplexität hat sich zu einem Modewort entwickelt. Er wird im alltagssprachlichen Gebrauch immer dann ins Feld geführt, wenn etwas bezeichnet werden soll, das der Mensch nicht gänzlich begreifen, vorhersehen, berechnen, steuern oder beherrschen kann. So werden beispielsweise der Klimawandel, die Entstehung von chronischen Krankheiten, die Ausbreitung des Internet oder die Entwicklungen an den Finanzmärkten als komplex bezeichnet. „Komplex“ wird sowohl für Phänomene in der Natur als auch für soziale Phänomene verwendet. Oft steht der Begriff in Zusammenhang mit Problemen oder Herausforderungen, die nicht oder kaum (mehr) zu bewältigen sind. Wenn auf Komplexität rekurriert wird, dann ist dies oft Ausdruck einer Ohnmacht Phänomenen gegenüber, die jenseits der menschlichen Beherrschbarkeit zu stehen scheinen und die Menschen in einer ‚Risikogesellschaft‘ leben lassen, in der der Ausnahmezustand zum Normalzustand zu werden droht (Beck, 1986). Komplexe Herausforderungen scheinen mit der Globalisierung stark zugenommen zu haben, die weltweite Vernetzung²⁵ begünstigt damit auch die Ausbreitung von negativen Phänomenen wie ansteckende oder chronische Krankheiten, Wirtschaftskrisen oder fundamentalistischer Terrorismus. Diese Komplexität stellt nicht nur die einzelnen Menschen in ihrer individuellen Lebensgestaltung vor grosse Herausforderungen, sondern auch die Institutionen, die für die Planung und Steuerung gesellschaftlicher Prozesse und für entsprechende politische Entscheidungen verantwortlich sind. Dörner (2008) zeigt in seinem Buch „Die Logik des Misslingens“ beispielhaft auf, wie traditionelle Problemlösestrategien in komplexen Zusammenhängen scheitern und dazu führen können, dass die anvisierten gesellschaftlichen Probleme trotz guter Absichten verschlimmert anstatt vermindert werden. Komplexität scheint gleichzeitig aber auch etwas Attraktives, Anziehendes zu haben und verheisst in einer immer stärker strukturierten Welt eine Offenheit und einen Möglichkeitsraum jenseits der Grenzen technischer Machbarkeit und Berechenbarkeit (Thrift, 1999, 34), Raum für Phantasie, Kreativität und Innovation. So sind es nicht nur negative Phänomene, die sich scheinbar urplötzlich entfalten und ausbreiten können, sondern auch positiv bewertete wie demokratische Bürgerbewegungen²⁶, der plötzliche Erfolg eines zuvor unbeachteten Buches oder der rasante Aufschwung sozialer Netzwerke im Internet. Von der Attraktivität des Komplexitätsthemas in der Bevölkerung zeugt der Erfolg populärwissenschaftlicher Literatur rund um dieses Thema (Watts, 2004; Westley, Zimmerman, & Patton, 2007; Waldrop, 1992; Taleb, 2007; Gladwell, 2002). Theorien rund um komplexe Systeme haben die Wissenschaften in vielerlei Hinsicht herausgefordert, wie hier später gezeigt wird. Ob und inwiefern sich die vielfältigen wissenschaftlichen und politischen Hoffnungen allerdings erfüllen, die in diese neue Lehre gesteckt werden, oder ob sie sich lediglich als eine rasch vergängliche Mode und nichts Weiteres als ‚business as usual‘ herausstellt (Thrift, 1999, 60), bleibt abzuwarten.

4.1 Eine historische Annäherung an Komplexität

Ein Blick in die Geschichte komplexitäts- und systemtheoretischer Ansätze soll zunächst zeigen, wie die zum Teil abstrakten Charakteristika komplexer Systeme grundgelegt sind. Die Komplexitätstheorie (im englischen weiter gefasst als ‚complexity science‘) bezeichnet allerdings keine einheitliche Theorie, sondern vereint Beiträge und Theorien (u.a. Chaostheorie, Katastrophentheorie,...) von Vertreterinnen und Vertretern unterschiedlichster Disziplinen (Walby, 2007), vor allem aus den Vereinigten Staaten, Westeuropa und Japan (Thrift, 1999), die sich mit komplexen Systemen befassen. Gleichsam zeichnet sich das Feld der Systemtheorie auch durch verschiedene disziplinübergreifende Strömungen aus, welche sich dadurch charakterisieren, die Wirklichkeit ‚vom System‘ her zu denken und zu

²⁵ Vgl. das ‚small-world‘ Phänomen, zurückgehend auf ein Experiment des Sozialpsychologen Stanley Milgram im Jahre 1967, nach welchem mittlerweile alle Menschen der Erde mit allen anderen über nur sehr kurze Ketten (‘six degrees’) verbunden sind, was die Verbreitung u.a. von Informationen stark begünstigt (s. Watts (2004)).

²⁶ Man erinnere sich an den Fall der Berliner Mauer, den Sturz des rumänischen Diktators Ceausescu oder die Massenbewegungen in Nordafrika im Jahr 2011.

erforschen. Im Gegensatz zur Komplexitätstheorie haben sich hier aber explizite wissenschaftliche Theorien etabliert wie die allgemeine Systemtheorie von Ludwig von Bertalanffy (1969) oder die soziologische Systemtheorie Niklas Luhmanns (1987). Komplexitäts- und systemtheoretische Konzepte treffen sich dort, wo es um die Analyse von komplexen Sachverhalten aus einer systemtheoretischen Perspektive, respektive um die Erforschung komplexer Systeme geht. Mit Blick auf die neuere soziologische Systemtheorie aber auch auf systemtheoretische Managementansätze, welche Konzepte wie Komplexität, Selbstorganisation, Emergenz etc. in ihre systemtheoretischen Ansätze integriert haben, lässt sich eine polarisierende Unterscheidung zwischen diesen beiden Strömungen, wie sie z.B. in Ramalingam und Jones (2008) aufgeführt ist²⁷, nicht aufrecht erhalten. Vielmehr kann von einer gegenseitigen Befruchtung der Systemtheorie und der Komplexitätstheorie (Capra, 2005) gesprochen werden.

Während sich die Soziologie seit je her mit komplexen Phänomenen beschäftigt hat und gerade dies ihre Stärke ausmacht (Willke, 2006), widmen sich die Naturwissenschaften erst etwa ab den 1960-er Jahren unter dem Begriff der ‚complexity science‘ explizit der Erforschung komplexer Phänomene. Eine massgebliche Grundlage dafür waren neue Lösungsansätze für nichtlineare Gleichungen mit mehreren Unbekannten, die dank der fortschreitenden Computertechnologie entwickelt werden konnten. Nichtlineare Gleichungen zeichnen sich durch iterative Prozesse aus, in welchen Zwischenergebnisse (output) laufend wieder als Anfangswerte (input) in die Gleichung einfließen. Bei mehreren unbekannt Variablen führen derartige Gleichungen schnell an Grenzen der Berechenbarkeit, obwohl sie mathematisch vollständig determiniert sind (Mainzer, n.d.). Die Ergebnisse dieser iterativen Prozesse können zu Werten führen, welche zufällig und chaotisch erscheinen, aber plötzlich auch zu regelmässigen Bewegungsmustern und Gleichgewichtszuständen um sogenannte Attraktoren (Anziehungspunkte) herum (Simon, 2008; Parker, Schaller, & Hansmann, 2003). Wenn bei derartigen ‚komplexen‘ Gleichungen die Anfangsparameter auch nur geringfügig verändert werden, können über die iterativen Prozesse grosse Abweichungen in den Ergebnissen resultieren (Simon, 2008). Je ‚komplexer‘ die Gleichungen und je zahlreicher die Iterationen sind, desto weniger sind die Ergebnisse voraussehbar. Die scheinbar plötzliche Etablierung von Ordnungsmustern kann aber auch bei einfacheren mathematischen Operationen nach einer Vielzahl von Iterationen beobachtet werden²⁸.

Während mathematische Gleichungen und Operationen determiniert und kontrollierbar sind (gleiche Anfangsparameter in derselben Gleichung oder Anweisung ergeben dieselben Ergebnisse), so gilt dies nur bedingt für physische Systeme. Die Zufuhr von Energie oder der Austausch mit anderen Systemen lässt sich zwar beeinflussen, aber nicht im Detail kontrollieren. So können kleinste Abweichungen über iterative Prozesse und Rückkoppelungen zu unvorhersehbaren Entwicklungen führen. Die als Schmetterlingseffekt berühmt gewordene Metapher von Edward Lorenz illustriert dies schön. Sie besagt, dass der Flügelschlag eines Schmetterlings irgendwo auf der Erde durch Verkettung und Kumulierung letztlich einen Tornado in einem anderen Teil der Erde verursachen kann (Parker et al., 2003)²⁹. Komplexe Systeme stehen jenseits der Berechenbarkeit mittels linearer Gleichungen, sie lassen sich aber durch nichtlineare Gleichungen annähernd beschreiben und verstehen (Capra, 2005), wenn auch nicht exakt berechnen und vorhersagen³⁰. Umfangreiche Forschungen auf dem Gebiet der Thermodynamik haben das nichtlineare Entwicklungsverhalten von komplexen (thermody-

²⁷ In dieser Unterscheidung wird ein überholtes Bild systemtheoretischer Ansätze in den Vergleich eingebracht. Systemtheoretisches Denken wird konnotiert mit Berechenbarkeit, Kontrolle, Voraussagbarkeit, klaren Regeln, Negativfeedback während komplexitätstheoretische Ansätze Selbstorganisation, Dynamik, Interdependenzen, Positivfeedback etc. betonen würden.

²⁸ Vgl. ‚Langtons Ameise‘, ein System, welches durch einfache Regeln über zahlreiche Iterationen sowohl zu chaotischen als auch zu sich wiederholenden Schlaufen und entsprechenden Ordnungsstrukturen führen kann.

²⁹ Zu diesem Phänomen siehe auch den amerikanischen Bestseller „Tipping point“ von Malcolm Gladwell (2002).

³⁰ Entsprechende Computersimulationen, mit denen das Verhalten komplexer Systeme nachgebildet wird (Vester (2007)), stossen in Bezug auf Berechenbarkeit und Voraussagekraft rasch an ihre Grenzen - mit jeder Iteration potenzieren sich kleinste Abweichungen in den Annahmen und können rasch zu völlig falschen Prognosen führen.

namischen) Systemen belegt und in Anlehnung an die Chaostheorie (Parker et al., 2003) gezeigt, wie aus (scheinbar) chaotischen Zuständen auch wieder Ordnungsmuster und Gleichgewichtszustände (dissipative Strukturen³¹) entstehen können. Ein Beispiel dafür ist, wenn sich beim Erhitzen von Flüssigkeiten die einzelnen Moleküle ab einer bestimmten Temperatur plötzlich zu Mustern zu ordnen beginnen (Prigogine & Stengers, 1984). Die Unvorhersehbarkeit von Entwicklungen und die Entstehung von Ordnung aus dem ‚Chaos‘ werden als zentrale Merkmale komplexer Systeme gesehen, wie in den drei Axiomen der Chaostheorie grundgelegt ist:

Das erste ist, dass chaotische Systeme in ihrem Verhalten zwar streng deterministisch sind (d.h. ihr Verhalten ist vollkommen durch Gesetze und Regeln bestimmt), dass aber – zweites Axiom – ihr Verhalten trotzdem praktisch unberechenbar ist, weil aufgrund der unüberschaubaren Menge und Schnelligkeit von Rückkoppelungsprozessen minimale Veränderungen der Ausgangsbedingungen maximale Auswirkungen haben können. Das dritte Axiom ist, dass in chaotischen Systemen nicht einfach Zufall produziert wird, sondern dass sich in derart dynamischen Systemen Muster und Ordnungen bilden können (die bereits erwähnten Attraktoren), wenn sie sich selbst überlassen, d.h. von der Umwelt isoliert, sind. (Simon, 2008, 29)

Komplexe Systeme verhalten sich in der Regel weder völlig chaotisch noch sind sie in komplettem Gleichgewicht. Sie bewegen sich ‚am Rande des Chaos‘, wo sie zwar Ordnungsmuster etabliert haben, aber ‚in kritischem Zustand‘ sensibel selbst auf kleinste Einflüsse in unvorhersehbarer Weise reagieren können (Simon, 2008). Das kybernetische Modell erster Ordnung, in welchem einem Thermostat ähnlich Gleichgewichtszustände mittels negativem Feedback wiederhergestellt werden, hat sich als unzureichend für die Erklärung von Mechanismen in diesem Zustand erwiesen. Vielmehr konnte auch positives Feedback nachgewiesen werden, welches Systeme in ihren von Gleichgewicht und Ordnung abweichenden Entwicklungen bestärkt (Urry, 2006, 114).

Die Eigenschaft von Systemen, Ordnungsmuster zu bilden, welche nicht in den einzelnen Elementen selber angelegt sind, sondern erst durch deren Interaktion entstehen, wird als Emergenz bezeichnet, eine Eigenschaft, welche in der gesamten Komplexitäts- und Systemtheorie von zentraler Bedeutung ist.

Die Übertragung dieser Erkenntnisse aus der Mathematik und aus der Erforschung nicht lebender komplexer Systeme auf lebende Systeme war ein weiterer wichtiger Schritt in der Entwicklung der Komplexitätstheorie. Auch Organismen bilden komplexe Systeme, deren Verhalten sich mit den Axiomen der Chaostheorie beschreiben lässt. Die einzelnen biologischen Mechanismen und Regeln lassen sich zwar isoliert beschreiben und verstehen, ihr komplexes Zusammenspiel und die vielfältigen Rückkoppelungsprozesse führen aber zu Verhaltensweisen, die nicht mehr aus den einzelnen Elementen oder Regeln berechnet oder vorhergesagt werden können. Auch entwickeln Organismen Ordnungsstrukturen und Gleichgewichtszustände, die nicht direkt aus den einzelnen Systemkomponenten ableit- und verstehbar sind, wie z.B. der rhythmische Herzschlag, der zwischen zwei Gleichgewichtszuständen oszilliert. Aber auch Organismen haben die Möglichkeit, in chaotisches Verhalten zu kippen, ein Herzkammerflimmern kann als derartiger Zustand beschrieben werden, um beim angeführten Beispiel zu bleiben (Mainzer, n.d.). Lebende Systeme sind einerseits offene Systeme, da sie in Austausch mit ihrer Umwelt stehen und u.a. Energie aufnehmen und auch wieder abgeben. Gleichzeitig sind sie aber operativ geschlossen, also über ihre Grenzen vollständig gegenüber ihrer Umwelt abgeschlossen und in ihrer Funktionsweise nur auf sich selber bezogen (Simon, 2008). Die operative Geschlossenheit lebender Systeme wurde von den Biologen Maturana und Varela zum Konzept der Autopoiesis ausgebaut. Autopoiesis meint die Eigenschaft eines lebenden Systems, die Elemente, aus denen es besteht, selber aus seinen eigenen Elementen zu produzieren. Das System reproduziert sich - seine Elemente aber auch die Systemgrenze - laufend selber und ist nicht Ergebnis eines externen Eingriffs (Maturana zit. in Baraldi, Corsi, & Esposito, 2008). Komplexe autopoietische Systeme

³¹ Strukturen, die in energieumsetzenden (dissipativen) Systemen entstehen.

lassen sich in ihrer Funktionsweise also nicht von aussen determinieren, sondern sie sind auf sich selbst bezogen und innengesteuert. Da sie gleichzeitig gegenüber ihrer Umwelt offen sind, lassen sie sich durch diese beeinflussen und irritieren (Simon, 2008). Das System entscheidet dabei selber, wie es mit dieser Irritation von aussen umgeht. Simon hat dies mit Rückgriff auf Gregory Bateson (1972) an einem etwas makabren Beispiel veranschaulicht, welches gleichzeitig auch den Unterschied zwischen nichtlebenden und lebenden Systemen verdeutlicht (Simon, 2008). Wenn man einen Stein tritt, so kann man dessen Flugbahn berechnen, indem man die Gesetze der Physik anwendet (,triviale Maschine' nach Foerster (2009)). Tritt man hingegen einen Hund, so kann man zwar auch die Gesetze der Physik anwenden, diese reichen aber nicht aus, um das Verhalten des Hundes im Detail vorauszusagen. Ob er wegläuft, bellt, beisst oder jault, hängt zu einem guten Teil von seinem Innenleben ab, welches von aussen nicht beobachtbar ist. Der Hund reagiert als operativ geschlossenes, autopoietisches System nicht trivial, in Abhängigkeit von seinen aktuellen internen Zuständen und Strukturen. Im Extremfall – und nur in diesem – ist es allerdings möglich, die Innensteuerung eines autopoietischen Systems auszuschalten und sein Verhalten von aussen zu determinieren – man muss es dazu aber zerstören. Solange ein autopoietisches System ein solches ist, ist eine Determinierung von aussen nicht möglich. Später wurde der Begriff der ,Black Box' eingeführt, welcher auf die Innensteuerung eines Systems Bezug nimmt, deren Mechanismen von aussen aber nicht beobachtbar sind (Kneer & Nassehi, 2000).

Die Erkenntnisse der Komplexitätstheorie fanden über die Naturwissenschaften hinaus grossen Anklang und wurden begeistert von Fachleuten aufgenommen, die sich explizit oder implizit mit komplexen Phänomenen beschäftigten und mit traditionellen Ansätzen und Methoden an Grenzen stiessen. Systemtheoretische Ansätze, welche sich in derselben Zeit formierten, bauten auf weiten Strecken auf denselben Grundlagen wie die Komplexitätstheorie auf³², behandelten analoge Phänomene (komplexe Systeme, Selbstorganisation, Emergenz, Nichtlinearität,...) und entwickelten sich in vergleichbare Richtungen (z.B. Übertragung von naturwissenschaftlichen Erkenntnissen auf soziale Phänomene), so dass eine Trennung mehr historischer und begrifflicher als inhaltlicher Art ist. Systemtheoretische Ansätze begannen sich primär in der Biologie zu entwickeln, wo sich in Abgrenzung zum Newtonschen Welt- und Wissenschaftsverständnis eine neue wissenschaftliche Betrachtungsweise für lebende Systeme etablierte, weg vom Studium isolierter Einzelphänomene hin zur Betrachtung von Systemen, also der Vernetzung von Einzelphänomenen (Kneer & Nassehi, 2000). Treibende Kraft dieser Entwicklung war Ludwig von Bertalanffy, welcher ausgehend von biologischen Systemen eine interdisziplinäre, allgemeine Systemtheorie proklamierte (von Bertalanffy, 1969). In dieser konzentrierte er sich ebenfalls auf komplexe Phänomene, die durch gegenseitige Wechselwirkungen hervorgerufen werden. Diese ,organisierte Komplexität' beobachtete er überall in lebenden und sozialen Systemen (Kneer & Nassehi, 2000). Auf von Bertalanffy geht zudem die Leitunterscheidung zwischen Systemen und ihrer Umwelt zurück, welche die traditionelle Unterscheidung zwischen Ganzem und Teil als zentrale Differenz ablöste und welche es ihm ermöglichte, zwischen offenen, also sich mit ihrer Umwelt im Austausch befindenden, und geschlossenen Systemen zu unterscheiden (Luhmann, 1987). Der grosse Fortschritt dieses Perspektivenwechsels lag nun darin, offene Systeme, und damit sind alle lebenden Systeme mit eingeschlossen, nicht nur als solche in ihrer Einheit wahrzunehmen, sondern sie in Relation zu ihrer Umwelt zu begreifen.

Die Übertragung der Erkenntnisse der Komplexitätstheorie auf soziale Systeme liess nicht lange auf sich warten, nicht zuletzt weil sich die soziale Welt durch Instabilität, Unvorhersehbarkeit und undurchschaubare Ursache-Wirkungszusammenhänge auszuzeichnen scheint (Mathews, White, & Long, 1999). In der Ökonomie und Soziologie haben sich einzelne Vertreter intensiv mit den Erkenntnissen der Komplexitätstheorie auseinandergesetzt, obwohl die Übertragung der Erkenntnisse aus den Naturwissenschaften auf soziale Systeme nicht unumstritten ist und zu einiger Kritik geführt hat.

³² Mathematische Erkenntnisse rund um nichtlineare Gleichungen gepaart mit den neuen Berechnungs- und Darstellungsmöglichkeiten der Computertechnologie (von Bertalanffy (1969)), aber auch Erkenntnisse aus der Physik (u.a. Thermodynamik) und der Biologie (u.a. offene, adaptive Systeme; Autopoiesis).

Kritiker haben vor allem bemängelt, dass es an Nachweisen für die Gültigkeit dieser Übertragung mangelt, eine Kritik, die aus heutiger Sicht und auf Basis zahlreicher Anwendungen und Forschungen in den Sozial-, Wirtschafts- und Politikwissenschaften allerdings mehrheitlich obsolet geworden ist (Ramalingam & Jones, 2008; Urry, 2006). Erkenntnisse aus der Komplexitäts- und Chaostheorie werden heute in verschiedenen Lebensbereichen angewendet, so zum Beispiel auch in der Managementlehre (Malik, 2008; Gomez & Probst, 1999; Antonacopoulou & Chiva, 2007), in der Sozialen Arbeit (Staub-Bernasconi, 2007; Lüssi, 1995) oder in der Therapie (Parker et al., 2003). Die Management- und Organisationslehre hat auf der Grundlage der Komplexitätstheorie traditionelle Steuerungsmodelle grundlegend hinterfragt (siehe eindrucksvoll Dörner, 2008) und neuartige Steuerungstheorien und -modelle entwickelt (z.B. Malik, 1993; Malik, 2008; Vester, 2007; Gomez & Probst, 1999) - es wird gar ein neues Paradigma des Managements proklamiert (Thrift, 1999). In diesen Ansätzen kommt den Kräften der Selbstorganisation grosse Bedeutung zu. Soziale Systeme, welche nach Störungen aus dem Gleichgewicht geraten, sind in der Lage, auf einer höheren Komplexitätsebene, selber wieder Ordnung und neue strukturelle Muster herzustellen. Die Entwicklung von Ordnung über chaotische Zustände zu erneuter Ordnung auf einem höheren Niveau kann zur Steuerung genutzt werden, allerdings ist damit auch eine Gefahr verbunden: Komplexe Systeme können sich, da prinzipiell nicht voraussehbar, auch in unerwünschte, negative Richtungen entwickeln und unerwünschte Ordnungsmuster etablieren (Parker et al., 2003). Man denke dabei zum Beispiel an Entwicklungen in Grossstädten, wo sich nach dem Verlust staatlicher Kontrolle in einzelnen Stadtteilen zuerst chaotische Zustände und danach neue, kriminelle Ordnungsstrukturen herausgebildet haben.

Komplexität als Gegenstand der Soziologie

Explizit und prominent wurde und wird Komplexität vor allem in systemtheoretischen Ansätzen thematisiert. Eine Wende im Verständnis sozialer Systeme führte Niklas Luhmanns Systemtheorie herbei, welche in „Soziale Systeme“ (Luhmann, 1987) dargelegt ist. Luhmann integrierte in seine Theorie Erkenntnisse aus den Komplexitätstheorien, so insbesondere das Konzept der Autopoiesis komplexer Systeme von Maturana und Varela, welches zu einem Grundpfeiler seiner Theorie wurde. Während sich die sogenannte Komplexitätstheorie fragmentarisch aus Erkenntnissen unterschiedlichster Disziplinen zusammensetzt (Thrift, 1999, 34), welche auch auf soziale Bereiche übertragen werden, legt Luhmann nun eine in sich konsistente Theorie komplexer sozialer Systeme vor. Luhmann versteht soziale Systeme analog zu lebenden Organismen als komplexe, autopoietische, gegenüber ihrer Umwelt offene aber gleichzeitig operativ geschlossene Systeme (Luhmann, 1987). Als autopoietische, sich selbst reproduzierende selbstreferentielle Systeme bestimmen sie selber über die Grenzziehung zu der sie umgebenden Umwelt und damit über ihre Identität und sie reproduzieren diese Grenzziehung laufend. Die Reproduktion des Systems geschieht selbstreferentiell, die einzelnen Elemente des Systems werden aus den systemeigenen Elementen hergestellt und nicht von aussen eingeführt. Neben der Einführung des Konzeptes der Autopoiesis sind weitere grundlegende Neuerungen für seine Theorie sozialer Systeme charakteristisch. Die Theorie baut darauf auf, nicht den Systembegriff selber, sondern die Differenz von System und Umwelt zum Ausgangspunkt der Betrachtungen zu machen und Luhmann begründet damit ein neues Paradigma im Verständnis von Systemen, das „Paradigma der Differenz von Identität und Differenz“ (Luhmann, 1987, 26, 243). Eine weitere entscheidende, wenn auch zunächst abstrakt anmutende Neuerung in der Systemtheorie Luhmanns ist, dass er nicht etwa Menschen als die konstitutiven Elemente sozialer Systeme versteht, sondern Kommunikationen³³. Durch diesen Schachzug war es ihm möglich, soziale Systeme deutlich von biologischen und psychischen Systemen abzugrenzen und so seine Theorie zu schärfen. Damit handelte er sich allerdings die Kritik von Seiten der kritischen Theorie ein, er würde den Menschen in die Umwelt sozialer

³³ Bereits in der Kybernetik Norbert Wieners werden Kommunikationen als ausschliessliche Grundbestandteile von Sozialsystemen definiert (Müller-Benedict (2011)).

Systeme und damit aus der Theorie verbannen³⁴. Es ist allerdings nicht so, dass Luhmann den Menschen wirklich aus seiner Theorie ausschloss. Im Gegenteil, durch seine konsequente System-Umwelt-Konzeption ist der Mensch zwingende Bedingung für die Entstehung und den Erhalt sozialer Systeme (Luhmann, 1987). Soziale Systeme reproduzieren sich, indem Kommunikationen an Kommunikationen anschliessen und sie bleiben solange erhalten, bis keine kommunikativen Anschlüsse mehr erfolgen (Luhmann, 1987). Oder wie Simon es ausdrückt: „So wie ein Spiel, das nicht weitergespielt wird, beendet ist, finden soziale Systeme ihr Ende, wenn die sie konstituierende Kommunikation aufhört“. (Simon, 2008, 89)

Jeder soziale Kontakt wird als soziales System begriffen und Luhmanns Theorie reklamiert auf dieser Grundlage einen Universalitätsanspruch für den gesamten Gegenstandsbereich der Soziologie (Luhmann, 1987, 33). Luhmanns Definition sozialer Systeme unterscheidet sich durch den Rückgriff auf Kommunikation als kleinste Systemkomponente von traditionellen Systembegriffen, welche Systeme primär als Elemente (in sozialen Systemen in der Regel Menschen) und sekundär deren Relationen untereinander verstehen. In Luhmanns auf Kommunikation basierender Konzeption liegt der Fokus von Beginn an auf der Relation und er ist eng mit der Notwendigkeit zur Selektion verbunden: Luhmann versteht Kommunikation als dreistelligen Selektionsprozess, der Information, Mitteilung und Verstehen miteinander kombiniert. Selektion ist also sowohl bei der Wahl der mitzuteilenden Information erforderlich als auch beim Medium der Mitteilung und letztlich kann die Information auch unterschiedlich verstanden werden (Kneer & Nassehi, 2000). Soziale Systeme haben die Funktion, über Selektionsprozesse Komplexität zu reduzieren, was zu organisierter Komplexität führt. Organisierte Komplexität meint dabei Komplexität mit selektiven Beziehungen zwischen den Elementen (Luhmann, 1987). Organisierte Komplexität unterscheidet sich dabei von organisierter Einfachheit (einfache Systeme) und nicht-organisierter Komplexität. Helmut Willke hat Luhmanns Systemtheorie explizit auf Komplexitätsaspekte hin ausgeführt, mit zusätzlichen Erkenntnissen der Komplexitätstheorie angereichert (Willke, 2006) und auf dieser Basis eine umfassende Interventions- und Steuerungstheorie entwickelt (Willke, 2005; Willke, 2001). Die Luhmannsche Theorie sozialer Systeme und die Weiterentwicklungen bei Willke bilden neben der Komplexitätstheorie die Grundlage für das Verständnis komplexer sozialer Systeme in dieser Arbeit und werden auch für die Analyse des Potenzials und der Möglichkeiten der Evaluation von Interventionen in sozialen Systemen leitend sein.

Paradigmenwechsel

Die Beschäftigung mit komplexen Systemen führte in den Wissenschaften - über die Grenzen einzelner Disziplinen hinweg - zu neuen Einsichten, Verständnissen und Ansätzen, welche herkömmlichen Auffassungen zum Teil widersprachen und vergleichsweise besser geeignet schienen, neu ins Blickfeld getretene komplexe Phänomene zu beschreiben und zu erklären. Sowohl von Vertreterinnen und Vertretern der Komplexitätstheorie (Prigogine, 1997; Sawyer, 2005) als auch von jenen der Systemtheorie (von Bertalanffy, 1969; Klir, 2001; Luhmann, 1987) wurden die neuen Erkenntnisse als wissenschaftliche Revolution resp. als Paradigmenwechsel im Kuhn'schen Sinne (Kuhn, 1976) proklamiert³⁵. Traditionelle Auffassungen in der Physik, Mathematik, Biologie, Ökonomie, Psychologie, Pädagogik, Soziologie usw. wurden korrigiert und Theorien und Modelle auf Basis der Erkenntnisse der komplexitäts- und systemtheoretischen Ansätze angepasst, reformuliert, erweitert oder gänzlich neu entwickelt. Das Ideal einer berechenbaren Welt und das vorherrschende Bild des Menschen als determinierbare ‚lebende Maschine‘ wurden fundamental in Frage gestellt (Simon, 2008; von Bertalanffy, 1969) und es galt, sich von Vorstellungen einfacher Ursache-Wirkungszusammenhänge, von

³⁴ Vgl. die Debatte Luhmann-Habermas, die sich an den Positivismus-Streit anschloss.

³⁵ Zur unterschiedlichen Verwendungsweise des Begriffs Paradigma siehe Bergman (2010). Er unterscheidet zwischen einem engen Paradigmenbegriff im Sinne Kuhns, welcher auf Inkommensurabilität gründet und einer abgeschwächten Form, welche Paradigmen eher als theoretische Rahmenkonzepte versteht. Letzteres Verständnis findet sich u.a. bei Denzin and Lincoln (2005, 7).

Determinismus und präziser Voraussagbarkeit zu verabschieden (Nowotny, 2005). Die Grenzen linearer Kausalität für das Verständnis komplexer Systeme wurden offensichtlich und an deren Stelle traten neu multidimensionale, zirkuläre Ursache-Wirkungs-Erklärungen, in welchen Ursachen und Wirkungen nicht mehr eindeutig getrennt werden konnten (Simon, 2008). Triviale Maschinen mit einfachen Input-Output-Relationen, welche geeignet sind, einfache technische Systeme zu berechnen, wurden um nichttriviale Maschinen (resp. Systeme) ergänzt, um anspruchsvollere Systeme zu beschreiben (von Foerster, 2009).

Die Komplexität sozialer Systeme

Komplexitäts- und systemtheoretische Erkenntnisse aus den Naturwissenschaften und der Mathematik haben die Sozialwissenschaften in ihrem Verständnis komplexer sozialer Systeme in vielerlei Hinsicht bereichert, aber auch herausgefordert. Hauptsächlich durch Analogieschlüsse wurden Konzepte wie Emergenz, Selbstorganisation, offene Systeme, Autopoiesis etc. von den Sozialwissenschaften entweder neu aufgenommen³⁶ oder aber sie haben zu Erweiterungen und Neudefinitionen geführt³⁷. Diese Erkenntnisse haben allerdings nicht zu einem Wissensdurchbruch im Sinne eines exakten Verständnisses komplexer Systeme geführt, wie man dies hätte erhoffen können, sondern sie haben zunächst einmal aufgezeigt, dass und weshalb Wissen über komplexe Systeme zwingend limitiert ist (Cilliers, 2002), mit Konsequenzen für die Vorhersehbarkeit und Steuerbarkeit derartiger Systeme (s. Kapitel 4.4). Die Konzepte der Komplexitäts- und Systemtheorie helfen in einem ersten Schritt nicht bei der Bestimmung, wie soziale Systeme zu steuern oder zu evaluieren sind, sondern sie zeigen im Gegenteil zunächst nur auf, weshalb herkömmliche Steuerungsmechanismen und Analysemethoden oftmals scheitern (vgl. insbesondere Dörner, 2008) und sie zwingen zu entsprechender Bescheidenheit.

In diesem Kapitel werden die Eigenschaften komplexer Systeme sowie deren Übertragung auf soziale Systeme, wie sie u.a. die neuere soziologische Systemtheorie vorschlägt, dargestellt und diskutiert.

Das Konzept der Komplexität hat neben der Komplexitätstheorie auch in der soziologischen Systemtheorie eine zentrale Bedeutung und wird dort gar als deren Angelpunkt bezeichnet (Willke, 2006, 16). Allerdings herrscht bezüglich der Definition von Komplexität alles andere als Einigkeit. Unterschiedliche disziplinäre Felder und unterschiedliche Problemstellungen haben zu vielfältigen Definitionen und Verständnissen von Komplexität geführt, die es, insbesondere mit Blick auf die Unterschiede mathematischer, naturwissenschaftlicher und sozialwissenschaftlicher Anwendungen, auseinanderzuhalten gilt.

Ein mathematisch-logisches Verständnis von Komplexität basiert auf grundsätzlicher Determiniertheit von komplexen Systemen, deren Entwicklungen lediglich über die Vielzahl von Iterationen als zunächst unberechenbar erscheinen. Mittels ausreichender Computerkapazität ist oder wäre es möglich, das Verhalten komplexer mathematischer Systeme vorzuberechnen und bei erneuter Durchführung derselben Iterationen mit denselben Anfangswerten die Ergebnisse exakt zu reproduzieren. Die Besonderheit, dass selbst kleinste Abweichungen in den Anfangsparametern zu grossen und überraschenden Abweichungen in den Ergebnissen führen können, darf nicht darüber hinwegtäuschen, dass Komplexität in mathematischen Anwendungen grundsätzlich deterministisch und berechenbar bleibt (Mainzer, n.d.). Auch die Bildung von Ordnungsmustern aus Zuständen scheinbaren Chaos über iterative Wechselwirkungsprozesse ist streng deterministisch und exakt reproduzierbar, wenn die Anfangsparameter und Operationen stabil gehalten werden. Dies im Unterschied zu physischen und sozialen Systemen. So ist es beispielsweise nicht denkbar, alle Zellen und deren Interaktionen und

³⁶ Z.B. das Prinzip der Autopoiesis, welches von Luhmann in seine Theorie sozialer Systeme integriert wurde (Luhmann (1987)).

³⁷ Das Konzept der Emergenz, welches bereits von Durkheim in die Soziologie eingeführt wurde, erhielt durch die Erkenntnisse der Komplexitätstheorie wissenschaftliche Bestärkung.

Wechselwirkungen in einem Körperorgan im Detail zu kennen, um so dessen Verhalten berechnen zu können. Die Vielzahl und die Verschiedenheit der Systemkomponenten einerseits und die Vielzahl möglicher Relationen andererseits scheinen entscheidende Dimensionen von Komplexität zu sein. In offenen Systemen (zu denen die lebenden Systeme zählen) werden durch den Kontakt mit Systemen der Umwelt zusätzliche Relationsmöglichkeiten geschaffen und damit die Komplexität potenziell erhöht. Wollte man das Verhalten dieser Systeme exakt voraussagen können, dann müsste es möglich sein, auch die relevante Umwelt bis ins Detail zu kennen. Allerdings, da es sich bei umgebenden Systemen in der Regel auch wieder um offene Systeme handelt, müssten auch deren relevante Umwelten identifiziert, abgegrenzt und im Detail bekannt sein usw. Es wird deutlich, dass spätestens dann vom Anspruch exakter Kenntnis, Berechenbarkeit und Voraussagbarkeit Abstand genommen werden muss, wenn es um komplexe *offene* Systeme geht. Nicht nur die Alltagserfahrung, sondern auch Forschungsergebnisse in den unterschiedlichsten Disziplinen bestätigen dies (von Bertalanffy, 1969; Prigogine, 1997).

Die Art und Anzahl der Systemelemente sowie deren Verknüpfung können als Voraussetzung für Komplexität angesehen werden, sie liefern aber keine hinreichende Definition für Komplexität. Rein technische Systeme, welche zwar eine grosse Vielzahl von Elementen und Verbindungen enthalten mögen (z.B. Computer oder Kraftwerke) sind nicht komplex, da ihr Verhalten durch die einzelnen Systemkomponenten und die definierten Verknüpfungen eindeutig determiniert ist. In Anlehnung an Cilliers wird in diesen Fällen, in welchen Systeme über die Kenntnis der einzelnen Systemkomponenten und Verknüpfungen exakt berechnet werden können, von komplizierten, nicht aber von komplexen Systemen gesprochen:

If a system - despite the fact that it may consist of a huge number of components - can be given a complete description in terms of its individual constituents, such a system is merely *complicated*. Things like jumbo jets or computers are complicated. In a *complex* system, on the other hand, the interaction among constituents of the system, and the interaction, and the interaction between the system and its environment, are of such a nature that the system as a whole can not be fully understood simply by analysing its components. Moreover, these relationships are not fixed, but shift and change, often as a result of self-organisation. This can result in novel features, usually referred to in terms of emergent properties. The brain, the language and social systems are complex. (Cilliers, 1998, iix)

In diesem Sinne sind auch ‚nicht triviale Maschinen‘, wie sie Foerster im Gegensatz zu ‚trivialen Maschinen‘ entworfen hat, zwar komplizierte, nicht aber komplexe Systeme. Sie gehören in die Kategorie mathematisch-logischer Komplexität. Foerster selber charakterisiert sie als synthetisch determiniert, analytisch unbestimmbar³⁸, vergangenheitsabhängig und unvoraussagbar (von Foerster, 2009, 66). Dass sie nicht voraussagbar sind, obwohl determiniert, sagt mehr über die Grenzen der Berechenbarkeit aus als über die Eigenschaften der Systeme selber³⁹. Daraus ergibt sich ein erster Ansatzpunkt für die Bestimmung von Komplexität, wie sie für den Bereich sozialer Systeme sinnvoll erscheint: Systeme, welche aus einer grossen Anzahl potenzieller Relationsmöglichkeiten ihrer Systemkomponenten untereinander als auch mit Systemen/Systemkomponenten in ihrer Umwelt nur einen sehr begrenzten Teil realisieren, diese Realisierung nicht durch bekannte Grössen und Gesetze determiniert und dadurch in ihrer Entwicklung über die Zeit nicht voraussagbar ist, können als komplex bezeichnet werden. Dabei erweisen sich die folgenden Aspekte als komplexitätsrelevant:

- die Art und Anzahl der Systemkomponenten
- die Art und Anzahl der potenziell möglichen Relationen

³⁸ Analytisch unbestimmbar, weil experimentell nicht erschliessbar (von Foerster (2009)).

³⁹ Nicht-triviale Maschinen nach Foerster sind ab einem bestimmten Grad ohne Computer nicht voraussagbar, allerdings bei entsprechender technischer Unterstützung exakt berechenbar. Was heute nicht berechenbar erscheint - obwohl determiniert - wird künftig mit grosser Wahrscheinlichkeit ebenfalls berechenbar sein.

- die Art und Anzahl der realisierten Relationen
- das Verhältnis zwischen den potenziell möglichen und den tatsächlich realisierten Relationen
- die Wandelbarkeit der Systemkomponenten und Relationen über die Zeit

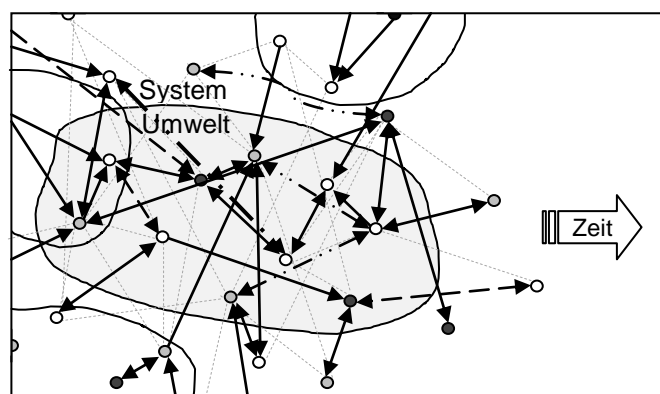


Abb. 2 Realisierte und potenziell mögliche Systemrelationen

Vor diesem Hintergrund ist Komplexität keine absolute, sondern eine relative Größe, sie muss im Sinne eines Konjunktivs verstanden werden: Es ist nicht die tatsächliche Realisierung von Relationen, welche Systeme komplex macht, sondern die Realisierung im Verhältnis zum Potenzial möglicher Relationen. Luhmann definiert zunächst:

Als komplex wollen wir eine zusammenhängende Menge von Elementen bezeichnen, wenn aufgrund immanenter Beschränkungen der Verknüpfungskapazität der Elemente nicht mehr jedes Element jederzeit mit jedem anderen verknüpft sein kann. (Luhmann, 1987, 46)

Und er ergänzt:

Als Komplexität soll hier, in erster Annäherung an den schwierigen Begriff, die Gesamtheit der möglichen Ereignisse verstanden werden, (...). Der Begriff der Komplexität bezeichnet stets eine Relation zwischen System und Welt, nie einen Seinszustand. (Luhmann zit. in Willke, 2006, 23)

Komplexität ist damit eng mit Selektion verbunden⁴⁰. Dem Umstand, dass jede Selektion auch anders ausfallen könnte, wird mit dem Begriff der Kontingenz Rechnung getragen. Kontingenz kann als Möglichkeitsraum verstanden werden, welcher Freiheiten zur Selektion, damit zusammenhängend aber auch Risiken birgt. Komplexität und Kontingenz sind eng miteinander verwandt. Während sich Kontingenz auf den Grad der Handlungs- und Entscheidungsmöglichkeiten bezieht, meint Komplexität den Grad gegebener Vielschichtigkeit im Sinne der „Gesamtheit der in einem System aneinander anschließbaren Differenzen“ (Willke, 2006, 32). Während in dieser Arbeit Kontingenz als eine wichtige Dimension von Komplexität verstanden wird, trennt Willke die beiden Konzepte wie folgt:

Es gibt Beziehungen von hoher Kontingenz und geringer Komplexität und umgekehrt: So kann man die anfänglichen Beziehungen zwischen Robinson und Freitag als extrem wenig komplex und hoch kontingent bezeichnen. Mangels anderer Bezugssysteme entfällt Vielfalt, Vernetzung und Folgelastigkeit der sozialen Beziehungen, während die möglichen Handlungsalternativen vom gegenseitigen Fressen bis zur Freundschaft reichen. (Willke, 2006, 32)

⁴⁰ Interessanterweise entsteht Komplexität überhaupt erst durch die Möglichkeit der Selektion und damit durch Reduktion von Komplexität (Luhmann (1987), 47).

Es gelingt Willke allerdings nicht, diese Trennung in seiner Konzeption konsequent einzuhalten, beispielsweise wenn er als zentrales Kriterium operativer Komplexität wachsende Handlungsoptionen anführt (Willke, 2006, 97). Willke definiert Komplexität als „den Grad der Vielschichtigkeit, Vernetzung und Folgelastigkeit eines Entscheidungsfeldes“ und verweist darauf, "dass es keine Komplexität an sich gibt, sondern nur in Bezug auf ein bestimmtes Problem, welches für ein bestimmtes System in einer bestimmten Situation Entscheidungen erfordert.“ (Willke, 2006, 23). Diese Definition, welche Komplexität in Bezug zu einem zu bewältigendem Problem oder einer Herausforderung setzt, spricht ebenfalls dafür, Kontingenz als eine Dimension von Komplexität zu begreifen.

Komplexität wird graduell verstanden und nicht im Sinne einer binären Logik „komplex-nicht komplex“. Ein soziales System - so lässt sich zusammenfassen - kann nicht per se komplex, wenig komplex oder gar nicht komplex sein, Komplexität ist nicht eine statische Eigenschaft eines Systems sondern eine (graduelle) Eigenschaft einer System-Umwelt-Relation im Hinblick auf eine Herausforderung, die das System zu bewältigen hat. Die Relativität von Systemkomplexität ist damit doppelt: erstens in Bezug auf das Verhältnis zwischen realisierten und potenziellen Relationen und zweitens im Hinblick auf eine zu bewältigende Herausforderung. Die Orientierung des Konzeptes Komplexität an einer zu bewältigenden Herausforderung erscheint aus drei Gründen eine sinnvolle und für diese Arbeit zweckdienliche Festlegung: Erstens grenzt sie den Begriff ein, schärft ihn und erhöht damit die Aussagekraft für den gewählten Bereich. Weil alle sozialen Systeme potenziell unendlich viele Relationsmöglichkeiten haben, wären grundsätzlich alle sozialen Systeme höchstkomplex. Ein System kann nun aber mit Bezug auf Herausforderungen in seiner Komplexität variieren. Es kann im Hinblick auf eine bestimmte Herausforderung hoch komplex sein, aber weniger komplex im Hinblick auf eine andere Herausforderung. So kann ein Familiensystem in Bezug auf die Bewältigung von Alltagsroutinen wenig komplex sein, hingegen in Bezug auf die Erziehung der jugendlichen Kinder höchst komplex. Zweitens wird damit der dynamische Charakter des Konzeptes der Komplexität betont⁴¹ und drittens trägt sie dem Umstand Rechnung, dass Komplexität ihre Relevanz im Bereich der Sozialwissenschaften vor allem in Zusammenhang mit zu lösenden Herausforderungen erhält, so auch in dieser Arbeit, in welcher es um Gesundheitsförderungsinterventionen zur Bewältigung komplexer gesundheitlicher Probleme geht. Nicht die realisierten Beziehungen und Handlungen von Akteuren in einem Gesundheitsförderungsprojekt machen allein dessen Komplexität aus, sondern die vielen potenziell möglichen Beziehungen und Handlungen, die zu gesundheitlichen Verbesserungen führen könnten.

Bei der Selektion der zu realisierenden Relationen kommt dem Medium Sinn eine zentrale Bedeutung zu. Luhmann führt Sinn als eine evolutionäre Errungenschaft psychischer und sozialer Systeme ein, die für beide Systemarten unerlässlich ist; Sinn wird dabei entweder als Bewusstsein (in psychischen Systemen) oder aber als Kommunikation (in sozialen Systemen) aktualisiert (Luhmann, 1987, 92 und 142):

Sinn ist das unerlässliche Medium für die Operationen der sozialen und der psychischen Systeme. Auf der einen Seite kann die Welt nur im Medium Sinn beobachtet werden. Auf der anderen Seite realisiert sich Sinn nur in sozialen und psychischen Systemen. Sinn und System setzen sich also gegenseitig voraus: sie sind nur zusammen möglich. (Baraldi et al., 2008, 171)

Sinn unterstützt ein System bei der Strukturierung und Bewältigung von Komplexität, denn „jeder bestimmte Sinn qualifiziert sich dadurch, dass er bestimmte Anschlussmöglichkeiten nahelegt und andere unwahrscheinlich oder schwierig oder weitläufig macht oder (vorläufig) ausschliesst. (Luhmann, 1987, 94)

⁴¹ So verweist Komplexität nicht nur auf das System selber sondern auch auf die Systemumwelt und die sich in der Relation System-Umwelt ergebenden und sich selber auch wandelnden Herausforderungen.

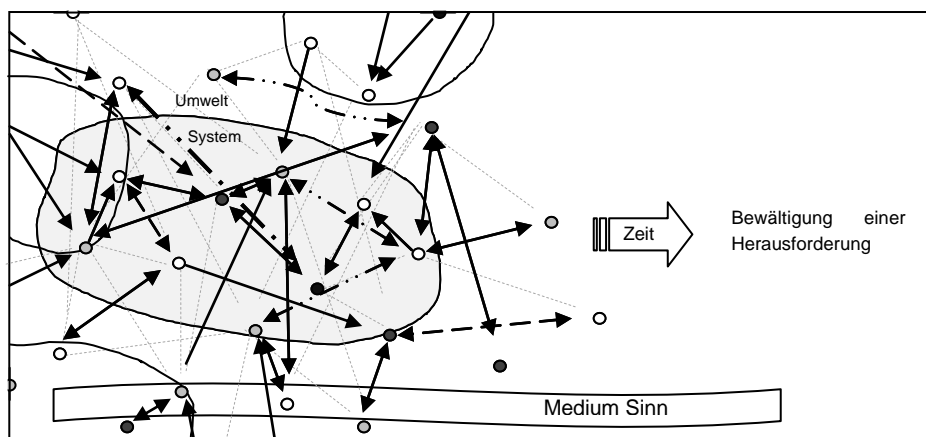


Abb. 3 Systemrelationen mit Rückgriff auf das Medium Sinn

Die Einführung des Mediums Sinn in Bezug auf die Bewältigung einer Herausforderung schränkt die potentiell möglichen auf die potentiell sinnvollen Relationen im Hinblick auf eine erfolgreiche Bewältigung ein:

Der Sinn von Grenzen liegt in der Begrenzung von Sinn. Nicht alles, was in der Welt passiert, nicht alle Ereignisse, Informationen und Zustände können von sozialen Systemen berücksichtigt und verarbeitet werden. Gegenüber einer komplexen Umwelt müssen Sozialsysteme ihre Aufmerksamkeit, ihre Zeit und Energie auf das systemspezifisch Sinnvolle begrenzen. So ist etwa für eine Partei nur das wichtig, was eine politische Frage ist oder werden kann; für ein Unternehmen nur das, was Auswirkungen auf seine Zahlungsfähigkeit hat; für eine Organisation nur diejenigen Ereignisse in seiner Umwelt, die ihre Ziele tangieren. (Willke, 2006, 55)

Dies führt zu folgender Komplexitätsreduktion: nicht mehr das Verhältnis der realisierten Relationen zu allen möglichen bestimmt den Grad der Komplexität (mit), sondern das Verhältnis der realisierten Relationen im Verhältnis zu den potenziell sinnvollen. Damit ist eine Eingrenzung geschaffen, welche für die weitere Analyse entscheidend ist. Die Entwicklung von komplexen Systemen und die Entstehung von Ordnung in solchen erscheinen nicht mehr im Lichte von undurchschaubarer Zufälligkeit, sondern sie sind sinnhaft konstituiert. Sinn wird damit zu einem Schlüssel bei der Analyse komplexer sozialer Systeme und soziale Systeme erhalten gerade erst über die Dimension Sinn ihren spezifischen Sinn. Soziale Systeme differenzieren sich nicht zuletzt deshalb heraus, um „eine die menschliche Aufnahmekapazität überwältigende Komplexität der Welt in spezifischer Weise zu erfassen und zu bearbeiten.“ (Willke, 2006, 6). Mit Rückgriff auf die bei Luhmann eingeführten drei Sinndimensionen⁴² und in Erweiterung um zwei weitere unterscheidet Willke sachliche, soziale, zeitliche, operative und kognitive Komplexität und liefert dadurch eine Differenzierung und ein Analyseschema, welches Komplexität als graduelles Konzept greifbar macht:

Sachliche Komplexität bezieht sich auf die Vielfalt von Einheiten, die aufeinander wirken und insofern nimmt sachliche Komplexität zu, „wenn die Zahl und die Dichte von Einheiten in einem bestimmten Raum-Zeit-Abschnitt steigt und wenn diese Einheiten aufeinander Wirkungen ausüben“ (Willke, 2006, 84). So steigert sich beispielsweise die Komplexität sozialer Systeme mit steigender Anzahl interagierender Personen und ihrer Interaktionen. Soziale Komplexität ist in Zusammenhang mit funktionaler Differenzierung zu verstehen und vergrößert sich bei zunehmender Differenzierung. Die unterschied-

⁴² Sachdimension „im Hinblick auf alle Gegenstände sinnhafter Intention (in psychischen Systemen) oder Themen sinnhafter Kommunikation (in sozialen Systemen)“ (Luhmann (1987), 114); Zeitdimension zur „Interpretation der Realität im Hinblick auf eine Differenz von Vergangenheit und Zukunft.“ (ebd., 116) und Sozialdimension zur Ermöglichung eines ständig mitlaufenden Vergleichs dessen „was andere erleben können bzw. erleben würden und wie andere ihr Handeln ansetzen könnten.“ (ebd., 121)

lichen sozialen Rollen, Interessen und Funktionen, welche Personen in eine Interaktion mitbringen, erhöhen Komplexität ebenfalls. Die etwas andere Bedeutung, welche Luhmann der sozialen Sinndimension beimisst⁴³, fängt Willke mit der operativen Komplexität auf. Operative Komplexität entsteht durch partielle Loslösung eines Systems von seiner Umwelt und bezieht sich auf die damit verbundene partielle Autonomie, welche es dem System ermöglicht, nicht nur reaktiv im Sinne eines stimulus-response-Schemas auf Umwelteinflüsse zu reagieren, sondern proaktiv „Zwecke und Ziele eigenständig zu setzen und zu variieren“ (Willke, 2006, 94)⁴⁴. Zeitliche Komplexität ist zu verstehen als potenzielle Beeinflussung der Gegenwart durch selektive Wahrnehmung von Vergangenheit und die Vorwegnahme von Zukunft. Menschen richten sich in ihrem Handeln nicht nur an der Gegenwart aus, sondern sie beziehen früher gemachte Erfahrungen und mögliche Konsequenzen in ihre Entscheidung mit ein. Kognitive Komplexität letztlich gründet auf der Reflexions- und damit auch Selbstreflexionsfähigkeit sozialer Systeme und konstituiert sich im Wissen, welches in sozialen Systemen kumuliert wird und auf welches zurückgegriffen werden kann. Beim Schreiben einer wissenschaftlichen Arbeit, wie der hier vorliegenden, ist man beispielsweise mit einer hohen kognitiven Komplexität konfrontiert, da wissenschaftliches Wissen mittlerweile stark ausdifferenziert, schwer überschaubar und auch nicht widerspruchsfrei ist.

Dass soziale Systeme von vergleichsweise hoher Komplexität sind, hängt maßgeblich mit den sozialen, zeitlichen und kognitiven Komplexitätsdimensionen zusammen, welche speziell diese Systeme betreffen und welche die Grenzen der Vergleichbarkeit sozialer mit technischen und biologischen Systemen deutlich machen. In einem nächsten Kapitel sollen die Merkmale komplexer sozialer Systeme identifiziert und im Detail beschreiben werden, um Komplexität auf diese Weise für die empirische Analyse und die darauf folgende Diskussion zu operationalisieren.

4.2 Merkmale komplexer sozialer Systeme

Aus der Komplexitäts- und systemtheoretischen Literatur lassen sich verschiedene Eigenschaften herauskristallisieren, welche komplexe soziale Systeme charakterisieren. Vor allem von Seiten der Komplexitätstheorie liegen Kategorisierungen vor (u.a. Ramalingam & Jones, 2008; Cilliers, 2005; Patton, 2011), auf welche hier zurückgegriffen wird. Ergänzend werden systemtheoretische Grundkonzepte (sozialer) Systeme mit aufgenommen, da sie für das Verständnis und die Analyse von sozialen Systemen bedeutend sind. Die hier vorgelegte Kategorisierung, welche die bisherigen Ausführungen mit zusammenfasst, unterscheidet zehn Eigenschaften, die für das Verständnis und die Analyse von komplexen sozialen Systemen als bedeutend erachtet werden:

⁴³ Luhmann würde die Sozialdimension, so wie sie Willke versteht, eher in seiner Sachdimension verorten. Unter der Sozialdimension versteht er die Berücksichtigung potentiell unterschiedlicher Perspektiven (Luhmann (1987), 119f).

⁴⁴ Willke bezeichnet soziale Systeme, welche (noch) nicht eigene Ziele definieren können, als „Quasi-Systeme“ (Willke (2006), 94).

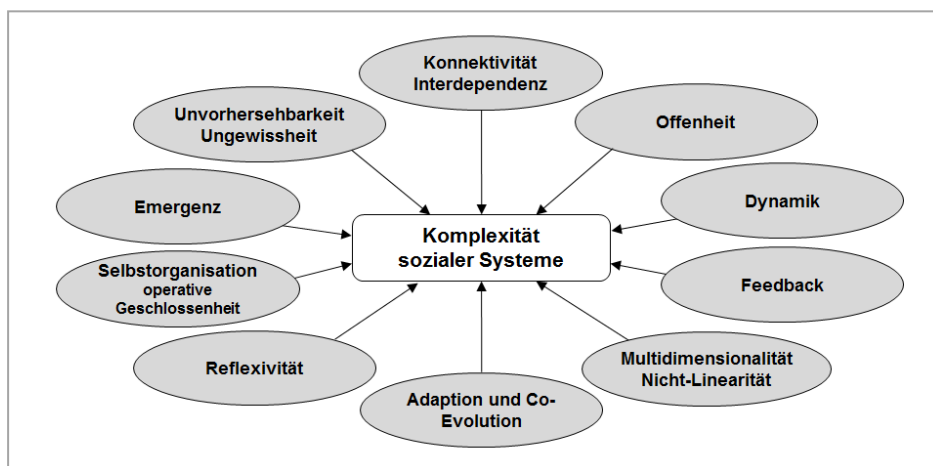


Abb. 4 Merkmale komplexer sozialer Systeme

Diese zehn Eigenschaften komplexer sozialer Systeme werden nachfolgend erläutert. Im Anhang 10.2 befindet sich eine gekürzte Zusammenstellung.

Konnektivität und Interdependenz

Die Konnektivität betont die Art und die Dichte der Verbindungen zwischen einzelnen Systemelementen und zur Systemumwelt, die Interdependenz die Abhängigkeit der einzelnen Systemkomponenten untereinander. Die beiden Elemente verweisen auf die sachliche Dimension von Komplexität. Von Interdependenz zwischen einzelnen Systemen oder Subsystemen wird dann gesprochen, wenn der Erfolg des einen Systems von anderen Systemen resp. Subsystemen abhängig ist (Page, 2008, 120). Ein bedeutendes komplexitätssteigerndes Element ist, dass Konnektivität und Interdependenz nicht eindimensional zu verstehen sind. Die einzelnen Systemelemente und Subsysteme können in unterschiedlichen Dimensionen miteinander verbunden sein und komplexe Strukturen auf der Makroebene des Systems herausbilden. Der Grad der Konnektivität und Interdependenz zwischen Systemelementen in unterschiedlichen Dimensionen hat grossen Einfluss auf die Art und Weise, wie sich ein System entwickelt. Während bei enger Konnektivität und Interdependenz („tightly coupled“) bereits kleine Veränderungen in einem einzelnen Teilbereich grosse Auswirkungen auf das Gesamtsystem haben können, sind Einflüsse innerhalb eines Systems mit schwacher Konnektivität und Interdependenz („loosely coupled“) diffuser und erstrecken sich über längere Zeiträume (Ramalingam & Jones, 2008, 10). In Sozialsystemen, welche von dichter Kommunikationsstruktur geprägt sind, sind Rückkoppelungen und Wechselwirkungen häufiger, was die Dynamik und die Komplexität erhöht. Ein sehr hoher Grad an strukturierenden Regeln (z.B. in hoch formalisierten Organisationen), welcher die Freiheitsgrade für Entscheidungen reduziert, vermindert allerdings die Kontingenz und damit auch die Komplexität der Entscheidungsfelder wieder⁴⁵.

Offenheit

Die Offenheit sozialer Systeme ermöglicht vielfältigste potenzielle System-Umwelt-Beziehungen (Komplexitätssteigerung) und zwingt das System zu entsprechenden Selektionen (Kontingenzsteigerung). Soziale Systeme können nur in ihrer System-Umwelt-Relation verstanden werden, da die Umwelt nicht bloss beeinflussender, sondern konstitutiver Faktor und damit Voraussetzung der Identität

⁴⁵ Kontingenz wird hier in Abgrenzung zu Willkes Verständnis (Willke (2006), 32) als eine wichtige komplexitätssteigernde Dimension verstanden.

der Systeme ist⁴⁶. Die Umwelt ist also für das System nicht weniger wichtig als das System selbst (Fetterman & Wandersman, 2005). Systeme erhalten ihren Sinn durch Abgrenzung und in Auseinandersetzung mit ihrer Umwelt (Willke, 2006, 7). Dadurch dass Systemgrenzen nur die Systemelemente abgrenzen, Relationen zwischen System und Umwelt aber bestehen können, entsteht Offenheit. Die System-Umweltbezüge führen in letzter Konsequenz dazu, dass soziale Systeme tendenziell unendlich komplex sind und in letzter Konsequenz nie vollständig erfasst werden können:

Furthermore, because complex systems are open systems, we need to understand the system's complete environment before we can understand the system, and, of course, the environment is complex in itself. There is no human way of doing this. (Cilliers, 2005, 258)

Hinzu kommt, dass einzelne Subsysteme in der Regel Teil unterschiedlichster übergeordneter Systeme sind, was dazu führt, dass für dasselbe System unterschiedlichste Umwelten und Umweltbezüge relevant werden können.

Dynamik

Unter Dynamik eines Systems wird die Änderung von Systemzuständen über die Zeit verstanden (Mainzer, n.d.). Hohe Dynamik ist dann gegeben, wenn viele Feedbackprozesse im System selber, aber auch in System-Umwelt-Relationen, zu ständigen Adaptionen führen. Es können sowohl einzelne Systemelemente variieren als auch die Art und Weise der Relationen zwischen diesen. Die hohe Sensitivität komplexer Systeme trägt zu einer hohen Systemdynamik bei, weil kleinste Veränderungen im System oder in seiner Umwelt grosse Bewegungen auslösen können. Die Dynamik eines Systems wird stark durch die Dynamik des Systemumfeldes beeinflusst. Laufen viele Interaktionen und Feedbackprozesse über die System-Umwelt-Grenze hinaus, erhöht sich dadurch die Dynamik des Systems (weil System und Umwelt nicht getrennt betrachtet werden können). Eine Organisation, die von einem stark bewegten Umfeld abhängig ist, kommt nicht umhin, sich selber laufend zu verändern oder zumindest Anpassungen laufend zu prüfen, wenn die an sie gestellten Herausforderungen auf die Dauer erfolgreich bewältigt werden sollen. Wenn sich Systeme laufend verändern, verändern sich auch die Ausgangslagen für neue Prozesse ständig, was zur Konsequenz hat, dass sich Entwicklungen nicht 1:1 replizieren lassen.

Systeme können sich in oder nahe von Gleichgewichtszuständen befinden, in chaotischem Zustand⁴⁷ oder aber in organisierter Komplexität, also zwischen Gleichgewicht und Chaos:

Complexity repudiates the dichotomies of determinism and chance, as well as nature and society, being and becoming, stasis and change. Systems are thus seen as being 'on the edge of chaos'. Order and chaos are in a kind of balance where the components are neither fully locked into place but yet do not fully dissolve into anarchy. (Urry, 2006, 113)

Wenn organisierte Komplexität als Komplexität mit selektiven Beziehungen zwischen den Elementen (Luhmann, 1987, 46) verstanden wird, dann ist es naheliegend, dass sich soziale Systeme wohl ausnahmslos in organisierter Komplexität - höheren oder geringeren Grades - befinden. Organisierte Komplexität zwingt laufend zu neuen Selektionen, was Systeme weg von stabilen in dynamische Ver-

⁴⁶ Die neuere soziologische Systemtheorie beginnt mit Rückgriff auf Spencer-Brown („draw a distinction“) mit der Leitunterscheidung System-Umwelt - und führt damit das neue Paradigma der Differenz von Identität und Differenz ein (Luhmann (1987) 242; King and Horrocks (2010)). Damit bilden nicht Einheiten (also Systeme) die Basis der Systemtheorie und des Verständnisses von Systemen, sondern die Differenz System-Umwelt (Luhmann (1987)).

⁴⁷ Chaos wird häufig mit Unordnung gleichgesetzt. Diesem weit verbreiteten Verständnis setzt Hayles ein differenzierteres entgegen, welches auf versteckte und potenzielle (im Sinne einer Vorstufe von Emergenz) Ordnungsstrukturen verweist (Byrne (1998), 16)

hältnisse bringt, in denen unvorhersehbare Entwicklungen nicht Ausnahme, sondern eher die Regel sind:

This complex systems world is a world of avalanches, of founder effects, self-restoring patterns, apparently stable regimes that suddenly collapse, punctuated equilibria, 'butterfly effects' and thresholds as systems tip from one state to another. (Urry, 2005, 237)

Feedback

Unter Feedback oder Rückkoppelung wird die Rückmeldung verstanden, die ein System auf seine Aktivitäten hin erhält. Es wird unterschieden zwischen negativem Feedback, welches Entwicklungen zu korrigieren trachtet, die ein System von einem Gleichgewichtszustand entfernen, und positivem Feedback, welches abweichendes Verhalten unterstützt und ein System dazu ermuntert, sich von einem Zustand des Gleichgewichts weg zu bewegen. Negatives Feedback wird in der Kybernetik vor allem mit einfachen Steuerungsmodellen in Verbindung gebracht (Beispiel Thermostat). Das Konzept des positiven Feedbacks hat in der Komplexitätstheorie grosse Bedeutung und gründet auf Beobachtungen, dass Feedback nicht nur stabilisierend sondern auch destabilisierend wirken und ein System über eine Vielzahl von iterativen Feedbackschlaufen von Gleichgewichtszuständen weg ‚an den Rand des Chaos‘ oder gar in chaotische Zustände drängen kann (Urry, 2005, 243):

Positive Rückkoppelungen in einem System sind Beziehungen, in denen eine Variable sich direkt oder indirekt selbst so beeinflusst, dass ihre Vergrösserung zu einer weiteren Vergrösserung führt und ihre Verkleinerung zu einer weiteren Verkleinerung. Tier- und Pflanzenpopulationen sind in einem gewissen Ausmass positiv rückgekoppelt: Je grösser eine Population ist, in desto grösserem Umfang vermehrt sie sich. Positive Rückkoppelungen sind gewöhnlich gefährlich für die Stabilität eines Systems. Ein System mit vielen positiven Rückkoppelungen gerät leicht «aus den Fugen». Eine Variable mit negativer Rückkoppelung hat die Tendenz, einen bestimmten Zustand aufrechtzuerhalten. Sie befindet sich in einem stabilen Gleichgewicht und tendiert dazu, ihren Gleichgewichtszustand wieder anzunehmen. In einem ökologischen System sind «Räuber-Beute-Beziehungen» zwischen zwei Tierpopulationen negativ rückgekoppelt. Das Anwachsen der Beutepopulation führt zu einem Anwachsen der Räuber-Population; dies führt wiederum zu einer Senkung der Beute-Population. Die Senkung der Beute-Population führt wieder zu einer Senkung der Räuber-Population. (Dörner, 2008, 110)

Während in wenig komplexen Systemen Feedback lineare Formen aufweisen kann, sind in komplexeren Systemen wechselseitige, zirkuläre und damit nicht-lineare Feedbackformen vorherrschend (Ramalingam & Jones, 2008, 15). Feedbackprozesse können rasch zu grossen und unvorhersehbaren Dynamiken führen, und oft reichen zunächst kleine positive Feedbacks, um ein System an einer entscheidenden Gabelung in die eine oder andere Richtung zu treiben (Urry, 2006, 114). In sozialen Systemen kann negatives Feedback dazu führen, eine mutige Entscheidung wieder rückgängig zu machen, positives Feedback kann eine solche entscheidend unterstützen und zunächst noch unsicheren ersten Schritten in eine neue Richtung grossen Schub verleihen. Feedback kann allerdings auf unterschiedliche Weise interpretiert werden, eine Rückmeldung kann je nach Interpretation sehr unterschiedlichen Charakter haben. Ramalingam und Jones (2008) geben ein Beispiel für diesen Sachverhalt:

Self-fulfilling prophecies are fundamental to understanding certain market movements. Specifically, when stocks are rising (a bull market), the belief that further rises are probable gives investors an incentive to buy, leading to further rises. At certain points, the increased price of the shares, and the knowledge that there must be a peak after which the market will all, ends up deterring buyers (negative feedback). Once the market begins to fall regularly (a bear market), some investors may expect further losing days and refrain from buying (positive feedback), but

others may buy because stocks become more and more of a bargain (negative feedback). (Ramalingam & Jones, 2008, 47)

Multidimensionalität, Nicht-Linearität und Zirkularität von Ursache-Wirkungsbeziehungen

Unter Linearität von Ursache-Wirkungsbeziehungen ist im Wesentlichen zu verstehen, dass Wirkungen gleichmässig auf Veränderungen von ursächlichen Parametern reagieren. Grafisch lässt sich das Verhältnis als Linie in einem Koordinatensystem zwischen Ursache (x-Achse) und Wirkung (y-Achse) veranschaulichen. Der doppelte Umfang an Ursache führt zu doppelter Wirkung, die Entwicklung erfolgt gleichmässig. Unabhängige und abhängige Variablen sind nach diesem Schema eindeutig identifizierbar und trennbar. In komplexen Systemen ist diese Linearität nicht gegeben. Die Entwicklung des Ursache-Wirkungs-Verhältnisses ist nicht linear, nicht vorhersehbar und nicht replizierbar. So kann beispielsweise eine therapeutische Intervention zu einer Beruhigung des betreffenden Familiensystems führen, eine Verstärkung dieser Intervention kann aber zu plötzlichen Widerständen führen und bei einer Wiederholung der Intervention zu einem späteren Zeitpunkt ist es hoch wahrscheinlich, dass unterschiedliche Effekte ausgelöst werden. Rückkoppelungen der Wirkungen auf die Ursachen sprengen die einfache Dichotomie zwischen unabhängigen und abhängigen Variablen und führen zu wechselseitigen Beeinflussungen und zirkulären Abhängigkeiten⁴⁸: „Dabei kommt es zu charakteristischen Rückkopplungen von Systemelementen, bei denen Wirkungen von Ursachen selber wieder zu Ursachen werden, um ihre Ursachen zu beeinflussen.“ (Mainzer, n.d., 4)

Unter systemischer Perspektive verdienen Rückkoppelungen zwischen der Makro- und der Mikroebene besondere Beachtung. Das Verhalten einzelner Systemelemente beeinflusst das Systemverhalten, gleichzeitig wirkt das Systemverhalten aber wieder auf die einzelnen Elemente zurück. Das Wahlverhalten einzelner Bürger führt zu bestimmten Wahlrends und diese können wiederum das Wahlverhalten der Bürger beeinflussen (Mainzer, n.d., 4).

Komplexe Systeme sind grundsätzlich durch Multidimensionalität ihrer Ursache-Wirkungs-Beziehungen geprägt. Eine Ursache kann Wirkungen auf unterschiedlichen Ebenen auslösen und eine Veränderung in einem System kann auf einer Vielfalt und einem Zusammenspiel unterschiedlichster Faktoren beruhen. Ein unüberschaubares Netz von grossen und kleinen Ursachen ist auf mannigfaltige Weise mit grossen und kleinen Wirkungen verknüpft, wobei Ursachen selber auch zu Wirkungen werden können und umgekehrt. In grösseren sozialen Systemen ist es unmöglich, alle potenziellen Ursachen und Wirkungen in Bezug auf die Bewältigung einer Herausforderung zu identifizieren, geschweige denn die Richtungen, Stärken und die Dynamik ihrer Relationen. Dieses Netz an wechselseitig abhängigen Ursache-Wirkungsbeziehungen führt zu einer potenziell hohen Sensitivität der Wirkungsparameter:

The behaviors of complex systems are sensitive to their initial conditions. Simply, this means that two complex systems that are initially very close together in terms of their various elements and dimensions can end up in distinctly different places. This comes from nonlinearity of relationships – where changes are not proportional, small changes in any one of the elements can result in large changes regarding the phenomenon of interest. (Ramalingam & Jones, 2008, 27)

⁴⁸ „Die Zirkelkausalität dieser vielfältigen Wechselwirkungen wird mathematisch durch gekoppelte nichtlineare Gleichungen der einzelnen Systemelemente ausgedrückt.“ (Mainzer (n.d.), 4)

Adaption und Co-Evolution

Adaption setzt die Offenheit von Systemen voraus und meint die Fähigkeit von Systemen, auf Feedback aus der Umwelt, aber auch auf systeminternes Feedback zu reagieren und Verhalten und Strukturen entsprechend zu adaptieren:

Interacting elements and agents respond and adapt to each other, and to their environment, so that what emerges is a function of ongoing adaptation both among interacting elements and the responsive relationships interacting agents have with their environment. (Patton, 2011, 131)

Systeme und Subsysteme können dabei auf sehr unterschiedliche Weise auf Feedback reagieren. In einem sozialen System können diese Reaktionen beispielsweise erwartungsgemäss ausfallen, oder aber, was häufig beobachtet werden kann, diametral entgegengesetzt. Die Adaptionleistung ist eine wichtige Voraussetzung für Veränderung in Systemen und gleichzeitig eine Voraussetzung für Komplexität schlechthin: wenn Adaptionmöglichkeiten nicht gegeben wären, würde Feedback ins Leere laufen und Systeme und Subsysteme könnten nicht von aussen beeinflusst werden, was sie in ihrer Existenz bedrohen würde. Adaption setzt Lernfähigkeit von Systemen voraus und baut darauf auf. Es ist ein dynamisches Konzept, weil Verhalten und Strukturen sich ständig wandeln und weiter entwickeln. Soziale Systeme beeinflussen sich in ihrer Entwicklung dabei gegenseitig, was zu zirkulären Rückkoppelungsprozessen zwischen Systemen und ihren Umwelten führt. Derartige Co-Evolution ist sowohl bei hierarchisch geordneten Systemen gegeben, wenn sie sich gegenseitig in ihrer Entwicklung beeinflussen (Rückkoppelungen zwischen der Mikro- und Makroebene (Mainzer, n.d.) als auch bei nicht hierarchisch verbundenen Systemen.

Reflexivität

Soziale (und psychische) Systeme unterscheiden sich von anderen Systemen unter anderem durch ihre Fähigkeit zur Reflexion, also durch die Fähigkeit, sich selbst zu thematisieren und sich als Umwelt anderer Systeme zu begreifen. Dies gibt die Möglichkeit, potenzielle Entwicklungen zu antizipieren und sich entsprechend darauf einzustellen:

Das System gewinnt damit die Möglichkeit, seinen Reichtum an intern produzierten Handlungsalternativen - seine Kontingenz oder operative Komplexität - in bewusst gesteuerter Weise auf die Möglichkeiten und Restriktionen sowohl seiner internen Bedingungen als auch seiner Umwelt einzustellen und optimale Handlungsstrategien zu konzipieren. (Willke, 2006, 104)

Ein System kann mittels Reflexion Umwelterwartungen und -entwicklungen erahnen und sich antizipierend daran ausrichten. Reflexion führt darüber einerseits zu Selbstbeschränkungen, andererseits eröffnet kognitive Komplexität, welche durch Reflexionsprozesse gefördert wird, neue Anschlüsse und Möglichkeiten und steigert dadurch potenziell die Komplexität.

Selbstorganisation

Selbstorganisation gilt als Schlüsseleigenschaft komplexer Systeme mit einem grossen anwendungsorientierten Potenzial⁴⁹. Systeme haben die Funktion, Umweltkomplexität für ihre Mitglieder zu reduzieren resp. Voraussetzungen dafür zu schaffen, dass der Umgang mit Komplexität gelingt. Das Prinzip der Selbstorganisation schliesst an das Prinzip der operativen Geschlossenheit an. Komplexe Systeme sind nicht von aussen determinierbar oder steuerbar, sondern sie entscheiden selbst, inwiefern sie sich von aussen beeinflussen lassen (Willke, 2005). Das Prinzip der Selbstorganisation geht

⁴⁹ Insbesondere in den Managementwissenschaften haben sich Steuerungsmodelle etabliert, die Selbstorganisation als Leitkonzept haben (u.a. Malik (1993)).

davon aus, dass es weder eine externe Steuerung noch eine interne zentralisierte Kontrollinstanz⁵⁰ gibt, welche die Entwicklung eines Systems steuern. Vielmehr entstehen und verändern sich Systemstrukturen über komplexe Interaktionen zwischen dem System und seiner Umwelt: „The capacity for self-organisation is a property of complex systems which enables them to develop or change internal structure spontaneously and adaptively in order to cope with, or manipulate, their environment.“ (Cilliers, 1998)

Diese Adaptionen an Umweltbedingungen führen auf der Makroebene zu Strukturen, die den einzelnen Elementen auf der Mikroebene nicht bewusst sind und die auch nicht direkt auf diese zurückführbar sind. So kann sich ein Fischschwarm in Form und Grösse an seine Umweltbedingungen anpassen, ohne dass das einzelne Tier sich dieser Anpassung bewusst ist, sondern sich nur nach seinen eigenen Interessen richtet (Cilliers, 1998) oder aber ‚der Markt‘ kann auf der Makroebene Muster etablieren, die den einzelnen Käufern und Verkäufern weder bewusst sind noch von Zentralinstanzen gesteuert sind⁵¹.

Selbstorganisation ist gemäss Cilliers (1998) immer an eine Form der Erinnerung des Systems gebunden. Das System nimmt seine Selektionen auf der Grundlage seiner eigenen Geschichte vor, welche zudem auch selektiv gespeichert ist. Nur so kann sich ein System weiter entwickeln. Für soziale Systeme, die aus sich aneinander anschliessenden Kommunikationen bestehen, ist dies einleuchtend. Ohne Erinnerung keine Anschlussmöglichkeiten und ohne Anschlüsse kein soziales System. Die Regeln, welche Selbstorganisationsprozesse leiten und entsprechende Muster etablieren lassen, gehen aus Selektionsprozessen (auf der Mikroebene) hervor. Selektionen, die sich in der Vergangenheit bewährt haben, werden wiederholt und damit verstärkt (Vester, 2007). Das Zusammenspiel verschiedener solcher Selektionsmuster führt auf der Makroebene zu Ordnungsstrukturen, die auf der Mikroebene nicht bewusst oder intendiert sind.

Komplexe Systeme, deren Aufgabe es ist, erfolgreich mit Umweltkomplexität umzugehen, bilden tendenziell weder allzu rigide und unflexible Ordnungsmuster aus, noch tendieren sie zu chaotischem Verhalten. In beiden Fällen wäre ihr Überleben gefährdet. Sie versuchen sich vielmehr in einem Zwischenbereich organisierter Komplexität zu optimieren⁵². Dieser Zwischenbereich ist aber risikobehaftet: „Selbstorganisation ist notwendig, um die zunehmende Komplexität zu bewältigen, sie kann aber auch zu unkontrollierter Eigendynamik und zu Chaos führen“ (Mainzer, n.d., 33). Die paradoxe Frage, die sich daraus für die Managementwissenschaften, aber dergleichen auch im Hinblick auf eine Interventionstheorie ergibt, lautet: „Wie muss ein System organisiert werden, damit es sich selbst organisieren kann?“

Operative Geschlossenheit

Soziale Systeme sind trotz Offenheit gegenüber ihrer Umwelt operativ geschlossen bzw. selbstreferentiell:

Ein System wie etwa eine Familie, eine Klasse, eine Legislative, eine wissenschaftliche Disziplin oder eine Kirche definiert für sich selbst diejenige Grenze, die es ermöglicht, die eigene Identität nach intern produzierten und prozessierten Regeln zu erzeugen und gegenüber einer externen Realität durchzuhalten. (Willke, 2006, 61)

⁵⁰ Das Internet kann beispielsweise als ein komplexes sich selbst organisierendes Informationssystem verstanden werden, in dem keine zentrale Leitungsvermittlung und keine externe Steuerung stattfindet (Mainzer (n.d.), 30).

⁵¹ Karl Marx ist einer der ersten, die diese Selbstorganisation des Marktes analysiert hat, allerdings vor allem deren negative Konsequenzen, die er durch eine Zentralinstanz zu korrigieren suchte (zit. in Mainzer (n.d.), 21).

⁵² Weaver unterscheidet zwischen Problemen einfacher Zusammenhänge (z.B. technische Apparate), Problemen unorganisierter Komplexität (bei einer extrem hohen Anzahl sich erratisch verhaltender Elemente) und Problemen organisierter Komplexität (hier sind selektive Beziehungen entscheidend) (zit. in Willke (2005)).

Die operative Geschlossenheit ist eng mit dem Konzept der Autopoiesis von Varela und Maturana verbunden. Dieses versteht autopoietische Systeme als „operativ geschlossene Systeme, die sich in einer ‹basalen Zirkularität› selbst reproduzieren, indem sie in einer bestimmten räumlichen Einheit die Elemente, aus denen sie bestehen, in einem Produktionsnetzwerk wiederum mit Hilfe der Elemente herstellen, aus denen sie bestehen“ (Maturana zit.in Willke, 2006, 62).

Autopoietische Systeme bauen in ihren Operationen immer und ausschliesslich auf ihren aktuellen internen Strukturen und Prozessen auf, sie sind also selbstbezogen und innengesteuert (Simon, 2008), in ihrer Tiefenstruktur geschlossen, damit gänzlich unabhängig und unbeeinflussbar von ihrer Umwelt und dadurch nur um den Preis ihrer eigenen Zerstörung von aussen steuerbar (Willke, 2006; Simon, 2008).

Die operative Autonomie, welche autopoietischen und damit allen sozialen Systemen zukommt⁵³, ist insofern komplexitätsrelevant, als dadurch ein System eigene Ziele setzen und verfolgen kann. Dies führt einerseits zu zusätzlichen Handlungs- und Entscheidungsoptionen und trägt damit zu erhöhter Kontingenz bei, andererseits gleichzeitig aber dazu, dass „Systeme mit der Ausbildung einer gewissen Eigenkomplexität immer stärker auf sich selbst reagieren und mit ihren eigenen Prozessen beschäftigt sind - und sich deshalb nur noch sehr selektiv mit ihrer Umwelt auseinandersetzen“ (Willke, 2005, 70).

Emergenz

In komplexen Systemen interagieren die einzelnen Systemkomponenten miteinander und mit ihrer Umwelt in vielfältiger Weise, was zu immer neuen Verknüpfungen, Wechselwirkungen und Rückkopplungen und dadurch zu Systemeigenschaften und Strukturen führt, die aus den Eigenschaften ihrer Einzelelemente nicht mehr erklärbar sind (Willke, 2006):

Adding two elementary actions to one another can induce dramatic new effects reflecting the onset of cooperativity between the constituent elements. This can give rise to unexpected structures and events whose properties can be quite different from those of the underlying elementary laws. (Nicolis zit. in Urry, 2005, 238)

Durch das Aufeinandertreffen von Systemen und Systemkomponenten respektive durch entsprechende vielfältige Feedbacks (Ramalingam & Jones, 2008) werden Prozesse in Gang gesetzt, die nicht wie bei mathematischen komplexen Systemen im Vorfeld bereits als Operationen festgelegt sind, sondern sich in der Interaktion erst entwickeln. Da die potenziellen Anschlüsse unzählig sind und sich über die Zeit potenzieren, entstehen Prozesse, die so vielfältig sind, dass sie im Detail nie vorausgesagt werden können. Die Charakteristika der einzelnen Systemkomponenten können also nicht einfach aufaddiert werden, um zu Charakteristika des Gesamtsystems zu gelangen - das Ganze ist - je nach Interpretation - mehr als die Summe seiner Teile (Thrift, 1999), respektive zumindest etwas anderes (Urry, 2006), oder wenn wir die komplexitätsreduzierende Funktion von Systemen berücksichtigen, gar weniger (Willke, 2006). Bei diesen emergenten, kollektiven Eigenschaften oder unerwarteten Mustern handelt es sich bei physischen Systemen häufig um für uns selbstverständliche, ‚einfache‘ Eigenschaften:

The flavour of sugar is not present in the carbon, hydrogen and oxygen atoms that comprise it. Such large-scale patterns or characteristics emerge from, but are not reducible to, the micro-dynamics of the phenomenon in question. (Urry, 2006, 113)

⁵³ Die Übertragung des Konzeptes der Autopoiesis auf soziale Systeme ist umstritten und wird bis heute von Maturana selber abgelehnt, da kein soziales System in der Lage sei, den materiellen Prozess der autopoietischen Reproduktion seiner Komponenten zu leisten (Mayring (2008), 287). Luhmann entgeht diesem Einwand, indem er soziale Systeme ausschliesslich auf Kommunikation aufbaut und die materielle Dimension in die Systemumwelt verlagert.

In sozialen Systemen kann das Konzept der Emergenz gut mit Eigengesetzlichkeit oder Eigendynamik (Willke, 2005) begriffen werden. Soziale Systeme wie etwa Familien oder Unternehmen entwickeln Eigengesetzlichkeiten und überindividuelle Eigendynamiken, die sich mit alleinigem Blick auf die handelnden Individuen nicht begreifen lassen (Willke, 2005).

Die Kenntnis der einzelnen Systemkomponenten ist nicht nur keine ausreichende Bedingung für das Verstehen eines Systems, im Gegenteil, es scheint auch nicht erforderlich, alle einzelnen Systemkomponenten im Detail zu kennen, um Aussagen über ein komplexes System machen zu können. So ist es beispielsweise nicht notwendig, die Eigenschaften und das Verhalten jedes einzelnen Automobilisten zu kennen, um Verkehrsströme zu analysieren, oder die Intentionen jedes einzelnen Demonstranten zu kennen, um Dynamiken einer Massendemonstration im Grossen und Ganzen verstehen zu können. Diese emergenten Eigenschaften von Systemen sind von den einzelnen Systemkomponenten insofern partiell unabhängig, als diese Ordnungsmuster weitgehend bestehen bleiben, auch wenn einzelne Systemkomponenten ausgewechselt oder erneuert werden (Cilliers, 2005). Diese emergenten Strukturen auf der Makroebene können von geringerer Komplexität sein als die einzelnen Systemkomponenten, aus welchen das System besteht. Das System reduziert mittels Selbstorganisation Komplexität. So kann eine Organisation durch das Ausbilden von Strukturen und Definieren von Prozessen Kontingenz und dadurch Komplexität in hohem Mass reduzieren. Die Struktur auf der Makroebene beeinflusst das Verhalten der Systemelemente auf der Mikroebene (Byrne, 1998), indem klare Verhaltenserwartungen definiert werden. Ein bekanntes Beispiel ist die sogenannte Schwarmintelligenz. Die einzelnen Vögel bauen die in Schwärmen beobachtbaren Muster nicht bewusst auf, es gibt keinen ‚Bauplan‘ und keine zentrale Führung. Die Interaktionen der einzelnen Vögel führen zu den Mustern und das Schwarmverhalten wirkt gleichzeitig wieder auf das Verhalten der einzelnen Vögel zurück (vgl. Mainzer, n.d., 18). Allerdings darf man sich diese Mechanismen nicht deterministisch vorstellen, weil die einzelnen Elemente bzw. Subsysteme immer auch die Möglichkeit haben, anders als vorgegeben zu reagieren - die operative Geschlossenheit gilt auch für soziale Subsysteme.

Unvorhersehbarkeit / Ungewissheit

Das multidimensionale Netz von Ursache-Wirkungszusammenhängen in komplexen Systemen, die vielfältigen Rückkoppelungen und Feedbackprozesse sowie die hohe Sensitivität und Dynamik führen dazu, dass Entwicklungen in komplexen sozialen Systemen nicht im Detail berechenbar oder vorhersehbar sind. Dazu müssten alle einzelnen Systemkomponenten und alle potenziellen und bereits realisierten Beziehungen zwischen allen Elementen in allen Dimensionen bekannt sein, alle bereits herausgebildeten Strukturen auf der Makroebene sowie alle potenziellen und realisierten Feedbackprozesse zwischen der Makro- und der Mikroebene (sachliche Komplexität), alle Subsysteme mit ihren spezifischen Eigenschaften und Umweltbezügen, die Eigenkomplexitäten des Systems und seiner Subsysteme (operative Komplexität), die Systemumwelt in ihrer Komplexität und mit ihren Anschlussmöglichkeiten, der Einfluss der Zeit auf die Systementwicklung (zeitliche Komplexität), die unterschiedliche Bedeutung, welche verschiedene Akteure einzelnen Aspekten beimessen (soziale Komplexität), und dies auf der Basis des vorhandenen Wissens in all seinen Dimensionen (kognitive Komplexität). Es leuchtet ein, dass dies ein unmögliches Unterfangen wäre. Ob komplexe soziale Systeme in letzter Konsequenz aber trotzdem determiniert sind und ihre Entwicklung vorausgesagt werden könnte, wenn alle beschriebenen Elemente in allen Dimensionen bekannt wären, bleibt eine philosophische Frage von untergeordneter Bedeutung für die Analyse komplexer Systeme (Mc Queen, 2007). Denn unabhängig von der Frage der Determiniertheit bleiben Aussagen über die Entwicklung von komplexen sozialen Systemen, und insbesondere auch Prognosen, aufgrund eines begrenzten Kenntnisstandes immer probabilistisch (Mc Queen, 2007). Durch die Analyse von Ordnungsstrukturen, Austauschbeziehungen und Regeln auf der Makroebene von Systemen können Veränderungen in Systemen zwar nicht mit Gewissheit vorausgesagt werden, approximative Aussagen über die Entwicklung des Gesamtsystems sind - ohne Gewähr - aber möglich. Selbst wenn einzelne Systemkomponenten und Relationen auf der Mikroebene nicht im Detail bekannt sind, kann die Analyse von Struktu-

ren und Prozessen auf der Makroebene zu einem - zwar limitierten - Verständnis von Systemverhalten führen. So sind wirtschaftliche Prognosen selbst über die begrenzte Kenntnis von einzelnen Aspekten und Zusammenhängen möglich, wenn auch, wie uns die Geschichte immer wieder lehrt, Überraschungen ebenso selbstverständlich sind wie prognostiziertes Verhalten und man deshalb gut daran tut, eine bescheidene Position einzunehmen:

When dealing with complexity, modest positions are inescapable. This does not imply that they should be relative, vague or self-contradictory, nor does it imply a reason to cringe in false modesty. We can make clear, testable assertions about complex systems. We can increase the knowledge we have of a certain system, but this knowledge is limited and we have to acknowledge these limits. (Cilliers, 2005, 263)

Grundlage für die Komplexität sozialer Systeme sind die unzähligen gelebten und potenziellen Interaktionen und Feedbacks zwischen den verschiedenen Systemen und Systemkomponenten. Diese sind derart unüberschaubar und dynamisch, dass Entwicklungen in sozialen Systemen nicht mit Bestimmtheit voraussehbar und voraussagbar sind. Eine bescheidene Intervention in einem sozialen System kann unverhältnismässige und zum Teil unerwartete Effekte auf unterschiedlichsten Systemebenen auslösen. Ursache-Wirkungsbeziehungen sind nicht-linear und häufig zirkulär. Durch die vielfältigen Interaktionen zwischen den Komponenten eines komplexen Systems können Eigendynamiken und Unvorhersehbares entstehen, das in den einzelnen Komponenten selber nicht angelegt ist. Diese Innovation, welche in der Komplexitäts- und Systemtheorie mit dem Begriff der Emergenz bezeichnet wird, kann deshalb auch bei genauer Kenntnis der einzelnen Komponenten nicht aus diesen hergeleitet bzw. vorausgesagt werden, sondern wächst gewissermassen über das Bestehende hinaus. Dabei besitzen komplexe Systeme die Fähigkeit, sich selbst zu organisieren und insbesondere nach Irritationen aus der Umwelt eigenständig neue Ordnungsmuster zu etablieren. Das Prinzip der Selbstorganisation bedeutet auch, dass sich soziale Systeme nicht von aussen determinieren lassen, was die Grenzen sozialer Intervention deutlich macht.

Gesundheitsförderung nimmt für sich in Anspruch, von besonderer Komplexität zu sein und begründet damit ihren Widerstand gegen randomisierte kontrollierte Studien für die Wirkungsevaluation ihrer Interventionen. Das nächste Kapitel ist deshalb der Komplexität der Gesundheitsförderung gewidmet.

4.3 Die Komplexität der Gesundheitsförderung

Das Verständnis von Gesundheit ist stark kulturell geprägt und gesellschaftlichem Wandel unterworfen. Obwohl in den modernen Industriegesellschaften noch immer ein bio-medizinisches und reduktionistisches Gesundheitsverständnis vorherrscht, haben sich im Public Health Sektor ganzheitlichere Gesundheitsverständnisse entwickelt. Die in den Nachkriegsjahren verbreitete Gesundheitsdefinition der Weltgesundheitsorganisation, welche Gesundheit als „Zustand vollständigen körperlichen, geistigen, und sozialen Wohlbefindens und nicht nur das Freisein von Krankheit und Gebrechen“ versteht (WHO, 2006), zeigt eine erste Ausweitung und die Berücksichtigung der sozialen Dimension. Im WHO-Gesundheitsförderungsglossar von 1998 (WHO, 1998b) wird der umfassende Charakter des Gesundheitsbegriffs betont:

A comprehensive understanding of health implies that all systems and structures which govern social and economic conditions and the physical environment should take account of the implications of their activities in relation to their impact on individual and collective health and well-being. (WHO, 1998b, 1)

Dieses umfassende Verständnis ist im Konzept der Gesundheitsdeterminanten konkretisiert, welches Gesundheit als das Zusammenwirken vielfältigster Faktoren auf individueller und gesellschaftlicher Ebene versteht. Der Begriff ‚Determinanten‘ ist vor dem Hintergrund der Komplexitätsdiskussion allerdings ungünstig gewählt, da komplexe Systeme eben nicht determiniert, sondern höchstens beein-

flusst werden kann. Da sich der Begriff in Theorie und Praxis aber breit etabliert hat, wird er hier übernommen. Als Gesundheitsdeterminanten werden jene Faktoren bezeichnet, deren Einfluss auf die Gesundheit wissenschaftlich nachgewiesen ist. Abb. 5 gibt einen Überblick über die vielfältigen Einflussfaktoren auf die Gesundheit. Bei diesen Hauptdeterminanten der Gesundheit handelt es sich um allgemeine Kategorien von gesundheitsbeeinflussenden Faktoren. Sie zeigen die Bedeutung der unterschiedlichen gesundheitsrelevanten materiellen, sozialen und individuellen Bereiche und spannen damit ein Feld von potenziell gesundheitsbeeinflussenden Faktoren auf.

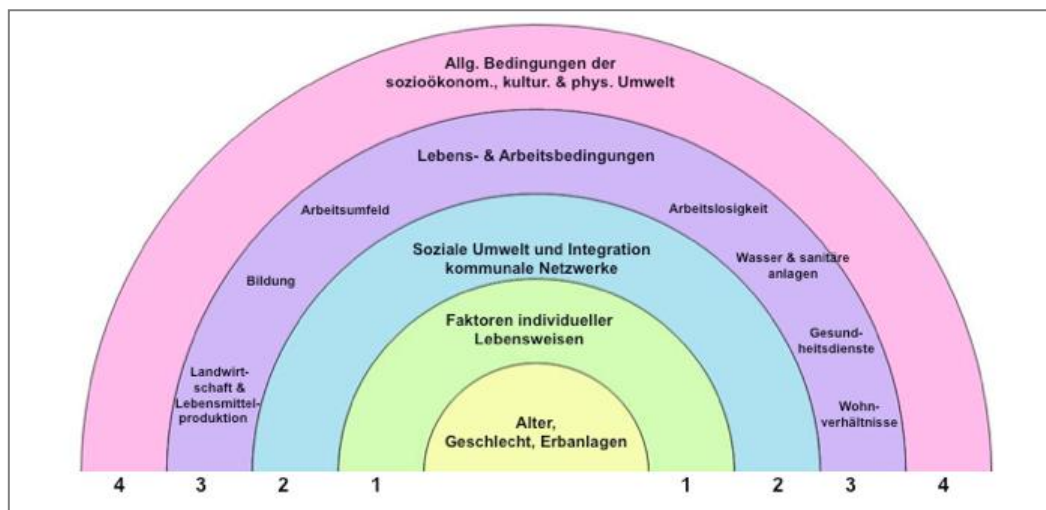


Abb. 5 Hauptdeterminanten für Gesundheit

Englische Quelle: Dahlgren & Whitehead 1991 *Policies and strategies to promote social equity in health*. Stockholm: Institute of Future Studies

Im Zusammenhang mit Fragen der Ungleichverteilung von Gesundheit sind in den letzten Jahren vor allem die sozialen Determinanten der Gesundheit in den Vordergrund der Analyse und der Diskussion gerückt (WHO, 2003; CSDH, 2008). Das schärfte den Blick für die Komplexität der Herausforderung Gesundheit und liess Gesundheit nicht mehr nur als individuelle oder rein medizinische, sondern als gesamtgesellschaftliche Aufgabe erscheinen, welche die Mitwirkung aller Politikbereiche erfordert. Parallel zu diesen Entwicklungen hat sich das Handlungsfeld⁵⁴ der Gesundheitsförderung herausgebildet, welches sich sehr breit dem Phänomen Gesundheit widmet. Obwohl nicht jede einzelne Intervention von gleich grosser Komplexität ist, kann Gesundheitsförderung als grundsätzlich komplexes Handlungsfeld begriffen werden, wobei Multisektoralität, Interdisziplinarität, Partizipation, Empowerment, Setting-Ansatz usw. zu dieser Komplexität beitragen: Das Grundlagendokument dieses neuen Handlungsfeldes, die Ottawa Charta von 1986 (World Health Organization, 1986), betont die gesellschaftlich breit abgestützte Verantwortung für Gesundheit. Sie fordert die Berücksichtigung politischer, ökonomischer, sozialer, kultureller und biologischer Faktoren, die Koordination von Verantwortlichen in Regierungen, im Gesundheits-, Sozial- und Wirtschaftssektor, in nichtstaatlichen und selbstorganisierten Verbänden, in lokalen Institutionen, in der Industrie und den Medien sowie die Beteiligung und Befähigung von Menschen in allen Lebensbereichen als Einzelne, als Familien und Gemeinschaften.

Auch die nachfolgenden offiziellen Deklarationen der Weltgesundheitsorganisation (WHO) zur Gesundheitsförderung bestärken diese systemische Sichtweise und die Verantwortung aller gesellschaftlichen und politischen Handlungsebenen für Gesundheit. Sie rufen zu einem koordinierten Vorgehen dieser Politikbereiche auf lokaler, regionaler, nationaler und internationaler Ebene auf (WHO, 2008). Mit diesem multisektoralen Ansatz und dem ganzheitlichen Verständnis von Gesundheit wird die Ge-

⁵⁴ Die Gesundheitsförderung kann als Handlungsfeld bezeichnet werden, sie konnte sich aber bisher noch nicht als Disziplin etablieren (Mc Queen (2000); Tremblay and Richard (2011)).

sundheitsförderung zu einem interdisziplinären Unternehmen, welches psychologische, soziologische, sozialmedizinische, politik- und wirtschaftswissenschaftliche, ethnologische, rechtliche usw. Fragen aufwirft und einen disziplinübergreifenden Diskurs erfordert.

Der statische, Gesundheit als ‚Zustand‘ verstehende Gesundheitsbegriff, welcher die WHO-Definition von 1976 prägt (s.o.), wurde mehr und mehr zu einem dynamischen und systemischen Konzept weiterentwickelt. Dieses geht von einem Gleichgewicht aus, welches durch das Individuum im Austausch mit seiner Umwelt laufend hergestellt werden muss:

Gesundheit wird von Menschen in ihrer alltäglichen Umwelt geschaffen und gelebt: dort, wo sie spielen, lernen, arbeiten und lieben. Gesundheit entsteht dadurch, dass man sich um sich selbst und für andere sorgt, dass man in die Lage versetzt ist, selber Entscheidungen zu fällen und eine Kontrolle über die eigenen Lebensumstände auszuüben sowie dadurch, dass die Gesellschaft, in der man lebt, Bedingungen herstellt, die all ihren Bürgern Gesundheit ermöglichen. (WHO, 1986, 5)

Huber et al. (2011) kritisieren an der WHO-Definition vor allem, dass die Erreichung ‚vollständigen Wohlbefindens‘ vor dem Hintergrund einer immer älter werdenden Bevölkerung und dem Vorherrschen vieler chronischer Belastungen ein heute nicht mehr angemessenes Konzept darstelle und sie schlagen alternativ dazu ein systemtheoretisches Gesundheitskonzept vor:

... moving from the present static formulation towards a more dynamic one based on the resilience or capacity to cope and maintain and restore one's integrity, equilibrium, and sense of wellbeing. The preferred view on health was „the ability to adapt and to self manage“. (Huber et al., 2011, 2; Hervorhebung im Original)

Ohne bereits eine klare Definition zu liefern, geben sie mit ihrem Konzept eine Richtung vor, welche Selbstorganisation in den Mittelpunkt stellt, und zwar in Anlehnung an die WHO-Definition in physischer, psychischer und sozialer Hinsicht. Dieses Konzept ist mit Blick auf komplexitäts- und systemtheoretische Erkenntnisse sehr aktuell. Gesundheitsförderungskonzepte bauen seit langem auf einem vergleichbaren Verständnis auf, wenn bisher auch versäumt wurde, eine entsprechende Gesundheitsdefinition vorzulegen. Gesundheitsförderung wird mit der Ottawa Charta als Initiierung eines Prozesses definiert, welcher Individuen dazu befähigen soll, Verantwortung für Ihre Gesundheit zu übernehmen, sowohl auf individueller als auch auf gesellschaftlicher Ebene:

Gesundheitsförderung zielt auf einen Prozess, allen Menschen ein höheres Maß an Selbstbestimmung über ihre Gesundheit zu ermöglichen und sie damit zur Stärkung ihrer Gesundheit zu befähigen. (WHO, 1986, 1)

Diese Befähigung, welche mit dem Prinzip des Empowerment konzeptionell gefasst ist (Ninacs, 2008), ist nicht bloss auf einen möglichst gesundheitsfördernden Umgang mit einer gegebenen Umwelt durch Entwicklung entsprechender personaler Kompetenzen ausgerichtet, sondern geht weit darüber hinaus: „Menschen können ihr Gesundheitspotential nur dann weitestgehend entfalten, wenn sie auf die Faktoren, die ihre Gesundheit beeinflussen, auch Einfluss nehmen können.“ (WHO, 1986, 2).

Zentrales Element dieses Ansatzes ist eine systemische Perspektive, welche die wechselseitige und vielfältige Beeinflussung von individuellen, sozialen und materiellen Faktoren und die Wahrnehmung von Verantwortung für Gesundheit auf allen Ebenen proklamiert (WHO, 2008). Diese Perspektive widerspiegelt sich auch in verschiedenen Wirkungsmodellen der Gesundheitsförderung (s. Übersicht in: Spencer et al., 2008) und hat u.a. das für die Gesundheitsförderung charakteristische Begriffspaar ‚Verhalten und Verhältnisse‘ geprägt. Verhaltensorientierte Ansätze fokussieren auf individuelles Verhalten und die Weiterentwicklung entsprechender personaler Kompetenzen und Verhaltensmuster, während verhältnisorientierte Ansätze Umweltbedingungen ins Zentrum der Interventionslogik rücken. Eine handlungsorientierte Konkretisierung des Verhältnisansatzes findet sich im Setting-Ansatz der Gesundheitsförderung wieder, welcher die Bedeutung von sozialräumlichen Systemen wie Familie,

Stadt, Schule, Betrieb etc. für die Gesundheit betont und auf entsprechende Organisationsveränderungen hinarbeitet, um gesundheitsfördernde Bedingungen zu fördern (Barić & Conrad, 1999). Die Beeinflussung von Umweltfaktoren wird über partizipative Prozesse angestrebt, um die Betroffenen zu Beteiligten zu machen und dadurch das Potenzial für Wirksamkeit und Nachhaltigkeit zu steigern. So sind die Zielgruppen der Gesundheitsförderung nicht lediglich zu ihren Bedürfnissen zu befragen⁵⁵, sondern an der Planung, Durchführung und Evaluation von Interventionen mitwirken und mitentscheiden zu lassen (Wright, 2010). Potvin versteht Partizipation in Interventionen der Gesundheitsförderung einerseits als Quelle von Komplexität und Dynamik von Gesundheitsförderung, gleichzeitig aber auch als wohlüberlegte und erfolgversprechende Strategie⁵⁶ zum Umgang mit der Komplexität und der Ungewissheit gesellschaftlicher Entwicklungen (Potvin, 2007). Dies ist ein Statement dafür, dass die Steuerung komplexer Systeme nur durch Systeme mit mindestens ebenso grosser Komplexität gelingen kann. Die Gesundheitsförderung baut in ihren Interventionen auf der hohen Reflexivität moderner gesellschaftlicher Systeme auf und nimmt Abstand von einer von aussen auf Zielsysteme aufgedrückten Programmlogik. Gesundheitsförderungsinterventionen müssen an ihre Zielsysteme anschlussfähig sein und Wirkungsentfaltung geschieht entsprechend kontextspezifisch lokal:

A conception of health promotion programs that addresses the challenges of reflexive modernity should rest on the idea that in a reflexive world, a predictable chain of events leading to local transformations cannot be reliably triggered by implementing a program based upon universal knowledge. Meaning, programs cannot be reduced to technical solutions that can be imported in local contexts once they have proved effective. (Potvin, 2007, 123)

Es gilt mittlerweile als evidenzbasiert, dass ganzheitliche Interventionsansätze, welche die in der Ottawa-Charta definierten fünf Handlungsebenen⁵⁷ kombinieren, am wirksamsten sind (WHO, 1998c) und Byrne konstatiert, dass die grössten medizinischen Erfolge durch Public Health Ansätze mit komplexem Gesundheitsverständnis erreicht worden sind (Byrne, 1998, 3).

Vor dem Hintergrund der Komplexität der Gesundheit und der Gesundheitsförderung, in welchen den sozialen Determinanten eine entscheidende Bedeutung zukommt, erstaunt es nicht, dass der Ruf nach soziologischer („Public health asks of sociology“ Rogers, 1968) und Jahrzehnte später systemtheoretischer Unterstützung („Public health asks of systems science“ Green, 2006) laut wurde. Die systemtheoretische Analyse scheint einen geeigneten Rahmen zu bieten, um die gesellschaftlichen Konstrukte Gesundheit und Krankheit zu verorten (Hafen, 2007 b), das Herausdifferenzieren des Krankheitssystems (Luhmann zit. in Pelikan, 2007a) und verwandter pathogenetisch und salutogenetisch orientierter gesellschaftlicher Teilsysteme nachvollziehen zu können (Pelikan, 2007b), die vielfältigen Interaktionen und den dynamischen Charakter der Gesundheitsförderung analytisch zu fassen (Naaldenberg et al., 2009) sowie der Komplexität des Gegenstandes angemessene Analyse- und Praxismodelle zu entwickeln (Dür, Pelikan, & Waldherr, 2010).

Die Vielfalt der für die Gesundheit und die Gesundheitsförderung relevanten Bereiche und Akteure sowie deren zahlreiche Verknüpfungen führen zusammen mit dem Anspruch einer koordinierten, ganzheitlichen Herangehensweise zu einer potenziell überaus hohen Konnektivität und Interdependenz der einzelnen Systemelemente und dadurch zu einem hohen Grad an sachlicher und sozialer Komplexität. Der multisektorale und interdisziplinäre Charakter der Gesundheitsförderung verweist auf ein kaum überschaubares Feld an Einflussfaktoren und potenziell relevanten Interaktionsmöglichkei-

⁵⁵ Diese noch immer in vielen Bereichen vorherrschende Form des Einbezugs wird als unechte Partizipation (Braun and Clarke (2006)) oder als Vorstufe für Partizipation (Wright (2010)) eingestuft.

⁵⁶ Nicht alle potenziell Betroffenen sind in alle Aspekte eines Programms einzubeziehen, sondern legitimierte Vertreterinnen und Vertreter, welche relevante Bevölkerungsgruppen und Organisationen repräsentieren (Hsieh (2005)).

⁵⁷ 1) Eine gesundheitsfördernde Gesamtpolitik entwickeln, 2) gesundheitsförderliche Lebenswelten schaffen, 3) gesundheitsbezogene Gemeinschaftsaktionen unterstützen, 4) persönliche Kompetenzen entwickeln und 5) die Gesundheitsdienste neu orientieren (Gertler, Martinez, Premand, Rawlings, and Vermeersch (2011)).

ten. Die Gesundheitsförderung kann dabei als ein in hohem Mass offenes System verstanden werden, welches in vielfältigem Austausch mit wesentlichen gesellschaftlichen Funktionssystemen (Wirtschaft, Politik, Wissenschaft, Recht etc.⁵⁸) steht, laufend Feedback aus diesen Umweltbezügen zu verarbeiten und sich daran zu orientieren hat, wenngleich auch die letztliche Entscheidungsmacht, wie mit Umwelteinflüssen umgegangen werden soll, beim sozialen System Gesundheitsförderung selber liegt⁵⁹. Die Mitwirkung von entscheidungsrelevanten Schlüsselpersonen, Fachleuten aus der Praxis, Zielgruppen etc. lässt eine Dynamik entstehen, welche nicht berechenbar ist und nicht mit Sicherheit vorausgesehen werden kann. Partizipation und Empowerment auf individueller und sozialer Ebene⁶⁰ führen dazu, dass sich Strukturen und Prozesse der Gesundheitsförderung in der Umsetzung (emergent) laufend neu konstituieren (Naaldenberg et al., 2009). Unvorhersehbarkeit und Ungewissheit sind infolgedessen keine Ausnahmeerscheinungen, sondern im Wesen der Gesundheitsförderung selber angelegt. Die Möglichkeiten der Planung, Steuerung und Kontrolle sind vor diesem Hintergrund und angesichts der operativen Geschlossenheit sozialer (zu intervenierender) Systeme grundsätzlich begrenzt und erfordern eine hohe Flexibilität in der Umsetzung. Adaption und Co-Evolution werden damit zu Schlüsselkriterien der Gesundheitsförderung. Ursache-Wirkungszusammenhänge sind hoch komplex und nicht-linear⁶¹. Die Zirkularität der Beziehungen, z.B. zwischen Verhalten und Verhältnissen oder zwischen körperlichen, psychischen und sozialen Faktoren lässt eine eindeutige Festlegung von Ursachen und Wirkungen, also unabhängigen und abhängigen Variablen nicht zu, vielmehr sind einzelne Faktoren in einem unübersichtlichen Netz in unterschiedlichen Dimensionen sowohl Ursachen als auch Wirkungen. Dies und der tendenziell holistische Ansatz der Gesundheitsförderung widersprechen einem reduktionistischen Wissenschaftsverständnis, welches über die Analyse einzelner Elemente komplexe Phänomene wie Gesundheit oder Gesundheitsförderung verstehen oder erklären will:

... it may also be argued that the „whole is the whole“, that is simply that it is a whole entity and can only be understood in its wholeness. The implications for this notion are many, but most importantly imply that deconstruction is not a pathway to understanding of the whole. (Mc Queen, 2007, 34)

Gesundheitsförderung vereint, wenn wir das Handlungsfeld als Ganzes betrachten, in hohem Mass Eigenschaften komplexer sozialer Systeme, und komplexitäts- und systemtheoretische Ansätze bieten eine adäquate theoretische Grundlage für die Analyse dieses ganzheitlichen Gesundheitsverständnisses und Interventionsansatzes. Um die Grundlage für die Untersuchung von Evaluation sozialer Interventionen zu schaffen, soll zunächst auf das Verhältnis zwischen sozialen Systemen und Interventionen eingegangen werden und danach gefragt werden, welche Konsequenzen sich aus der Komplexitätstheorie für soziale Interventionen ergeben.

4.4 Intervenieren in komplexen sozialen Systemen

Soziale Interventionen, also zum Beispiel Projekte oder Programme der Gesundheitsförderung, sind soziale Systeme, die sich in ihrem Wesen nicht von Organisationen, Familien oder anderen sozialen Systemen unterscheiden⁶². Aus systemtheoretischer Perspektive kann davon ausgegangen werden,

⁵⁸ Vgl. Luhmann (1987).

⁵⁹ Dies liegt in der operativen Geschlossenheit komplexer sozialer Systeme begründet und der Umstand, dass sich Gesundheitsförderung bislang weder als Profession noch als Disziplin etabliert hat, keinen Abbruch.

⁶⁰ Vgl. community-Ansätze und die Strategie der ‚capacity-building‘ (Egger, Spark, and Donovan (2005).

⁶¹ Ein lineares Ursache-Wirkungsverhältnis würde zum Beispiel bedeuten, dass eine zweimalige Durchführung eines Projektes auch doppelte Wirkung erzeugt. Man kann sich gut vorstellen, dass die Repetition nicht nur zu einem Mehr an Wirkungen, sondern auch zu Widerständen führen kann.

⁶² Tremblay and Richard beschreiben mit Blick auf Partizipation Gesundheitsförderungsinterventionen als soziale Systeme: „As well, health promotion interventions are defined as dynamic systems that evolve over time, espe-

dass sich im Hinblick auf Interventionen zunächst „zwei oder (mehrere) verschiedene, nichttriviale, operativ geschlossene und selbstreferentielle Systeme gegenüber stehen“ (Willke, 2005, 88), das intervenierende System und das System, das interveniert werden soll (Zielsystem). Ein intervenierendes System ist beispielsweise eine Organisation, welche den Auftrag hat, ein Gesundheitsförderungsprojekt an einer Schule durchzuführen. Das Zielsystem wäre entsprechend das Schulsystem. Intervenierende Systeme und Zielsysteme sind sich gegenseitig Umwelt. Als autopoietische Systeme können sie nur selbstreferenziell, nicht aber im anderen System, operieren. So kann die Gesundheitsförderungsfachstelle zwar versuchen, durch gekonnte Impulse Prozesse in der Schule in Gang zu bringen, ihr Einfluss ist aber begrenzt und die Schule entscheidet letztlich selber, inwiefern sie auf die Angebote der Gesundheitsförderung einsteigt. Stimmen sich Zielsysteme und intervenierende Systeme aufeinander ab, so spricht man von struktureller Kopplung. Obwohl Luhmann von struktureller Kopplung nur in Zusammenhang mit dem Verhältnis von psychischen, biologischen und sozialen Systemen sowie in Bezug auf das Verhältnis der gesellschaftlichen Funktionssysteme spricht (Luhmann, 1987; Luhmann, 2008) und Willke in seiner Interventionstheorie ganz auf dieses Konzept verzichtet (Willke, 2005), scheint die Verwendung des Konzepts der strukturellen Kopplung im vorliegenden Zusammenhang und in Anlehnung an Blank (2009) sehr wertvoll. Ziel- und intervenierende Systeme haben grundsätzlich eine unterschiedliche Logik, die sie aber strukturell aufeinander abstimmen können, um gegenseitiges Lernen zu ermöglichen. So können in einem Zielsystem beispielsweise Erwartungsstrukturen aufgebaut werden, welche an die Logik des intervenierenden Systems anschließen. Dadurch werden die Chancen auf Beeinflussung durch das intervenierende System gesteigert bzw. Lernen erleichtert. Wenn die systemtheoretische Prämisse ernstgenommen wird, dass nicht Menschen, sondern Kommunikationen soziale Systeme bilden, muss man sich vom Bild lösen, dass Menschen gleichzeitig mehreren Sozialsystemen angehören. Luhmann operiert in seiner Theorie mit den Begriffen Person und Rolle (Luhmann, 1987; Hafen, 2007 a). Eine Person bezeichnet nicht etwa den Menschen als physisches System, sondern der Begriff der Person wird als soziale Struktur eingeführt, um Verhaltenserwartungen adressieren zu können. Je nach sozialem Zusammenhang können an jemanden unterschiedliche Verhaltenserwartungen gerichtet werden. Jemand kann entsprechend z.B. als Privatperson, als Fachperson oder als Kirchgänger/in auftreten. Um Erwartungsbündel zu bezeichnen, die überindividuell sind und sich an unterschiedliche Personen in ähnlichen sozialen Zusammenhängen richten, dient der Begriff der Rolle. So werden an Lehrpersonen beispielsweise ganz allgemein Verhaltenserwartungen adressiert, unabhängig von individuellen Eigenschaften.

Ein Zielsystem (z.B. eine Schule) und ein intervenierendes System (z.B. eine Fachstelle der Gesundheitsförderung) sind grundsätzlich zwei getrennte, autopoietisch operierende Systeme, die sich gegenseitig Umwelt sind. Durch Beobachtung kann das eine System Impulse vom anderen System aufnehmen und diese Impulse bei Bedarf dazu nutzen, das eigene Systemverhalten in Selbstorganisation zu modifizieren. Wenn eine Schule und eine Fachstelle beginnen, zu kommunizieren und eine Zusammenarbeit aufzubauen, entsteht ein neues operativ geschlossenes System mit eigenen Zielen, Prozessen und Regeln, das Interventionssystem.

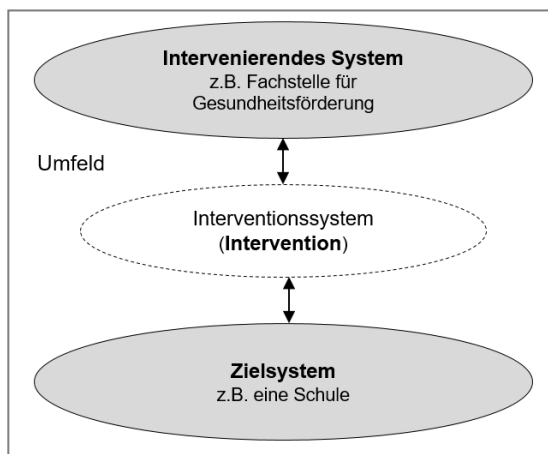


Abb. 6 Interventionssystem als sich herausbildendes neues soziales System

Eine Lehrperson ist über an sie gerichtete Verhaltenserwartungen primär dem Sozialsystem Schule und damit dem Bildungsauftrag verpflichtet, sie kann gleichzeitig aber auch dem Interventionssystem verpflichtet sein, z.B. als Multiplikatorin für die Botschaften der Gesundheitsförderung.

Interventionen der Gesundheitsförderung und Prävention beabsichtigen einerseits strukturelle Veränderungen in den sozialen Zielsystemen. Ihre letztlichen Zielsysteme aber, bei denen sich Wirkung entfalten soll, sind psychische und physische Systeme (Menschen). So sollen beispielsweise über die Veränderungen der Schulstrukturen günstige Bedingungen dafür geschaffen werden, dass Kinder eine positive Einstellung zu Bewegung und ausgewogener Ernährung entwickeln (physische Systeme) und dadurch ihr Verhalten gesundheitsförderlicher ausrichten, so dass letztlich ein individuell feststellbarer Gesundheitsnutzen wie bessere Grobmotorik (physisches System) oder ein besseres gesundheitliches Wohlbefinden (psychische Systeme) resultiert. Die Gesundheitsförderung und Prävention wirkt oftmals auch über die Veränderung der materiellen Umwelt auf die Gesundheit einzelner Menschen ein:

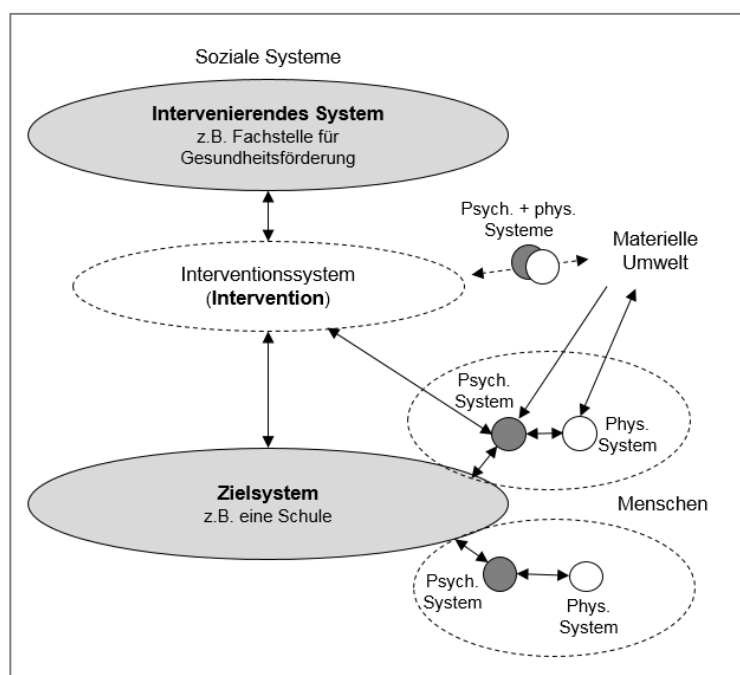


Abb. 7 Strukturelle Kopplungen

Einfach, kompliziert oder komplex?

Die Komplexität von Zielsystemen, intervenierenden Systemen und Interventionssystemen ist keine absolute Grösse, sondern muss relativ zu den Herausforderungen, die bewältigt werden sollen, definiert werden (s. Kapitel 0). Die Komplexität des Interventionssystems definiert sich infolgedessen im Hinblick auf die zu realisierende Intervention, deren Ziele, Strategien, Ressourcen, Vernetzung usw. Die Komplexität der Zielsysteme bemisst sich aus Sicht der intervenierenden Systeme ebenfalls an der Intervention, andere Dimensionen sind für das intervenierende System nicht relevant oder werden erst dann bedeutend, wenn sie einen Bezug zur Intervention erhalten, diese fördern oder hemmen könnten. So ist beispielsweise das Schulsystem für ein Gesundheitsförderungsprojekt nicht primär in seiner Funktion der Vermittlung mathematischen Wissens relevant, wohl aber darin, wie Lehrpersonen ihre Zusammenarbeit untereinander und mit den Eltern organisieren.

Es gibt unterschiedliche Ansätze, komplexe soziale Systeme von anderen Formen abzugrenzen. Wenn wir Luhmanns Begriffsbestimmung folgen⁶³, so sind kaum soziale Systeme vorstellbar, welche nicht komplex sind. Für Willke sind durchaus „einfache“ Interventionen (also Interventionssysteme⁶⁴) vorstellbar (Willke, 2005, 52), wobei offen bleibt, ob sein in Anführungszeichen verwendetes „einfach“ wenig komplex meint und nicht etwa nicht komplex. Einen interessanten Ansatz zur Differenzierung liefern Glouberman und Zimmerman. Sie unterscheiden zwischen einfachen, komplizierten und komplexen Herausforderungen. Im Unterschied zum Befolgen eines Rezeptes (einfache Herausforderung) und einer Mondlandung (komplizierte Herausforderung) sind komplexe Herausforderungen wie beispielsweise die Erziehung von Kindern auch bei wiederholter Durchführung von grosser Unsicherheit geprägt. Kinder reagieren auf Erziehungsmassnahmen individuell und situationsabhängig potenziell unterschiedlich, und das stereotype Befolgen rezeptartiger Handlungsanweisungen führt oft nicht zur gewünschten Wirkung. Frühere einschlägige Erfahrungen können zwar helfen, komplexe Herausforderungen erfolgreich zu meistern, allerdings nicht durch strikte Repetition gelungener Interventionen, sondern durch einen flexiblen und situationsangepassten Rückgriff auf unterschiedlichste Erfahrungsaspekte.

Following a recipe	Sending a rocket to the moon	Raising a child
The recipe is essential	Formulae are critical and necessary	Formulae have a limited application
Recipes are tested to assure easy replication	Sending one rocket increases assurance that the next will be OK	Raising one child provides experience but no assurance of success with the next
No particular expertise is required. But cooking expertise increases success rate	High levels of expertise in a variety of fields are necessary for success	Expertise can contribute but is neither necessary nor sufficient to assure success
Recipes produce standardized products	Rockets are similar in critical ways	Every child is unique and must be understood as an individual
The best recipes give good results every time	There is a high degree of certainty of outcome	Uncertainty of outcome remains
Optimistic approach to problem possible	Optimistic approach to problem possible	Optimistic approach to problem possible

Tab. 1 Simple, complicated and complex problems
(Glouberman & Zimmerman, 2002)

⁶³ „Als komplex wollen wir eine zusammenhängende Menge von Elementen bezeichnen, wenn aufgrund immenser Beschränkungen der Verknüpfungskapazität der Elemente nicht mehr jedes Element jederzeit mit jedem anderen verknüpft sein kann.“ (Luhmann (1987), 46)

⁶⁴ Jede Intervention wird zum System, weil sie auf sozialem Kontakt gründet und jeder soziale Kontakt als System begriffen wird (Luhmann (1987), 33).

Diese Differenzierung wurde von verschiedenen Autorinnen und Autoren aufgenommen und vor allem für die Evaluation von Interventionsprojekten fruchtbar gemacht (Rogers, 2008; Funnell & Rogers, 2011; Patton, 2011). Die Entscheidungskriterien für die Abgrenzung der drei Kategorien wurden von Rogers und Funnell (Funnell & Rogers, 2011) weiterentwickelt und differenziert. Sie eignen sich dazu, einzelne Aspekte von Interventionen⁶⁵ in Bezug auf Einfachheit, Kompliziertheit und Komplexität zu unterscheiden. Mit Einfachheit werden beispielsweise standardisierte Massnahmen in Verbindung gebracht, welche bei allen Menschen und überall bei gleicher Anwendung dieselben Wirkungen erzeugen, was idealtypisch zu verstehen ist. Kompliziertheit bezieht sich vor allem darauf, dass unterschiedliche Komponenten, Akteure, Sichtweisen,... koordiniert werden müssen. Von Komplexität wird in Zusammenhang von Emergenz, Adaption, Co-Evolution und Unvorhersehbarkeit gesprochen:

	Simple	Complicated	Complex
What it looks like			
Focus	Single set of objectives	Different objectives valued by different stakeholders. Multiple, competing imperatives. Objectives at multiple levels of a system	Emergent objectives
Governance	Single organization	Spezific organizations with formalized requirements	Emergent Organizations working together in flexible ways
Consistency	Standardized	Adapted	Adaptive
How it works			
Necessariness	Only way to achieve the intended impacts	One of several ways to achieve the intended impacts	
Sufficiency	Sufficient to produce the intended Impacts	Works only in conjunction with other interventions (previously, concurrently, or subsequently)	
Change trajectory	Simple relationship, that is already understood	Complicated relationship; needs expertise to understand and predict	Complex relationship (including tipping points) ; cannot be predicted ; can be understood only in retrospect
Unintended outcomes	Readily anticipated and adressed	Likely only in particular situations; need exertise to predict and address	Cannot be anticipated

Tab. 2 Übersicht über die Merkmale einfacher, komplizierter und komplexer Programme
(Funnell & Rogers, 2011, 456)

Die hier vorgeschlagene kategoriale Abgrenzung ist allerdings wenig trennscharf und lässt viel Raum für Interpretation. Zudem lassen sich Interventionen nicht eindeutig klassifizieren, da diese in der Regel sowohl Aspekte einfacher, komplizierter als auch komplexer Art enthalten können (Funnell & Rogers, 2011). Eine differenzierte Analyse von Interventionen im Hinblick auf komplexe Aspekte

⁶⁵ Es wird vorgeschlagen die Aspekte *Focus*, *Governance*, *Consistency*, *Necessariness*, *Sufficiency*, *Change trajectory* und *unintended outcomes* einzeln zu bewerten (Funnell and Rogers (2011), 74-91).

erscheint hingegen sinnvoll. Allerdings dürfte es vor dem Hintergrund der grundsätzlichen Multidimensionalität, Kontingenz und Dynamik sozialer Systeme wenig hilfreich sein, diese anders als komplex zu definieren (vgl. dazu auch Mowles, 2014). Die vorgeschlagenen Unterscheidungskriterien können zwar durchaus für die Bestimmung des Grades von Komplexität verwendet werden (z.B. geringe, mittlere und hohe Komplexität, relativ zu den zu bewältigenden Herausforderungen). Für die Unterscheidung zwischen komplizierten und komplexen Systemen soll aber an der in Kapitel 0 eingeführten Differenz zwischen technischen und lebenden Systemen festgehalten werden. Technische Systeme sollen, auch wenn vielfältige Zusammenhänge zwischen einzelnen Komponenten bestehen, höchstens kompliziert aber nicht komplex genannt werden, weil ihr Verhalten durch die einzelnen Systemkomponenten und die definierten Verknüpfungen eindeutig determiniert ist und über die Kenntnis der Systemkomponenten und Verknüpfungen exakt berechnet werden kann. Komplizierte Systeme lassen sich über die Kenntnis der einzelnen Elemente und Operationen verstehen, das ‚Ganze‘ lässt sich über das Aufaddieren der einzelnen ‚Teile‘ erschliessen. Lebende und insbesondere soziale Systeme hingegen zeichnen sich dadurch aus, dass ihr Verhalten nie exakt berechnet oder vorausgesagt werden kann, weil die realisierten und potenziellen Relationen im Innern und mit ihrer Umwelt in ihrer Vielzahl und ihren Dimensionen nicht fassbar sind. Für soziale Systeme erscheint es also konsequenter, von grundsätzlicher, aber je nach Situation variierender Komplexität auszugehen. Je grösser und unüberschaubarer das Netz an Elementen und Beziehungen, je grösser die im Hinblick auf eine Herausforderung potenziellen oder realisierten Umweltbezüge, je grösser die Dynamik des Systems und seiner relevanten Umweltsysteme, desto komplexer ist ein System und desto wahrscheinlicher sind unvorhersehbare Entwicklungen. Dies trifft in besonderem Masse auf Interventionssysteme zu, weil es sich bei diesen per se um zu bewältigende Herausforderungen handelt. Ganz grundsätzlich zwischen sozialen Systemen und (sozialen) Interventionen zu unterscheiden, wie dies Shiell et al. tun (2008), erscheint vor dem Hintergrund der Systemtheorie wenig sinnvoll, weil soziale Interventionen nichts anderes sind als soziale Systeme, deren Herausforderung darin besteht, in bestimmten anderen sozialen Systemen Veränderungen zu bewirken.

Im Anschluss an die Differenzierung zwischen einfachen, komplizierten und komplexen Interventionen (Glouberman & Zimmerman, 2002) setzt sich in der Evaluationsliteratur mehr und mehr die Tendenz durch, Intervention und zu intervenierendes Zielsystem in Bezug auf ihre jeweilige Komplexität zu trennen. Kuhn et al. unterscheiden beispielsweise zwischen "der Komplexität der Interventionen selbst, der Komplexität des Systems, in das interveniert wird, und der Komplexität der Beziehungen zwischen Intervention und System" (Kuhn, Lampert, & Ziese, 2012). Auf dieser Grundlage werden dann einfache soziale Interventionen in komplexen Systemen präsentiert, wie beispielsweise die Durchführung von Rückenschulungen in einem Betrieb. Dabei gilt allerdings zu beachten, dass eine geplante Intervention zwar anfänglich einfach anmuten kann (wenige involvierte Akteure, wenige Massnahmen, begrenzte Interaktionen), in ihrer Konstituierung in der sozialen Realität ist sie jedoch immer der Komplexität sozialer Systeme unterworfen und sie lässt sich davon nicht trennen. Selbst eine als einfache Intervention bezeichnete Rückenschulung in einem Betrieb ist folglich komplex, sobald sie in die Umsetzung geht und sie dürfte strukturell eng an das System „Betrieb“ gekoppelt sein. Craig et al. (2008) gehen ebenfalls davon aus, dass es kaum soziale Interventionen gibt, welche nur einfach sind, die Komplexität dürfte aber stark variieren. Als Gradmesser für Komplexität geben sie - mit ihrem Blick auf experimentelle Studien gerichtet - die folgenden Kriterien an:

Number of interacting components within the experimental and control interventions, number and difficulty of behaviours required by those delivering or receiving the intervention, number of groups or organisational levels targeted by the intervention, number and variability of outcomes, degree of flexibility or tailoring of the intervention permitted. (Craig, Dieppe, Michie, Nazareth, & Petticrew, 2008, 979)

Mit Verweis auf diese Merkmale kommt Trojan bei seiner Reflexion der Komplexität von Gesundheitsförderungsprogrammen zu folgendem Schluss:

Die Idee, die Bausteine der komplexen Intervention als « einfache Interventionen » zu bezeichnen, haben wir fallengelassen. Es stellte sich nämlich bald heraus, dass auch die kleinen, vermeintlich einfachen Interventionen oft die Kriterien der Komplexität erfüllen, die von massgeblicher Seite in England zusammengestellt wurden. (Trojan, 2012, 35)

Dies lässt sich einfach an einem Beispiel erläutern. Eine Schulungseinheit im Bereich HIV- oder Tabakprävention, die sich an einzelne Schulklassen richtet, kann im Sinne obiger Dreiteilung als einfache Intervention eingestuft werden. Die Schulungseinheit ist standardisiert, die Ziele klar festgelegt, es bedarf einer geringen Koordination etc. Wenn man sich die einzelnen Durchführungen aber im Detail ansehen würde, so dürfte die ganze Komplexität sozialer Interventionen schnell sichtbar werden. Obwohl die Einheit zu weiten Teilen dem Konzept gemäss durchgeführt werden kann, kann im Detail kaum von einer standardisierten Intervention gesprochen werden. In jeder Schulklasse dürfte eine solche Schulungseinheit ganz unterschiedliche Dynamiken entwickeln und Fragstellungen aufwerfen, die unterschiedliches Reagieren erfordern. In der einen Klasse mag eine konstruktive Auseinandersetzung mit der Thematik gelingen, in einer anderen mag sich eine unvohersehbare Dynamik der Ablehnung entwickeln, die möglicherweise durch ein Ereignis ausgelöst wird, das in einem anderen Kontext völlig unbedeutend wäre. Und diese Dynamik dürfte sich als höchst wirkungsrelevant erweisen.

Auch Bödeker (2012) folgert ähnlich und hält die Unterscheidung zwischen einfachen und komplexen Interventionen für kaum möglich (Bödeker, 2012). Andere wiederum fokussieren auf die Flexibilität und die geringe Standardisierung komplexer Interventionen (Hawe, Shiell, & Riley, 2004), die dazu führen, dass sich grundsätzlich jede Intervention unterscheidet und eine Kategorisierung kaum sinnvoll ist.

Folgerungen der Komplexitätsdiskussion für Interventionen

Aus der Komplexitätsdiskussion ist deutlich geworden, dass jedes soziale System und damit auch jede Intervention grundsätzlich einzigartig sind. Sowohl die systeminternen Strukturen als auch die System-Umwelt-Bezüge können sich nie genau entsprechen. Aber auch die zu intervenierenden Systeme unterscheiden sich stets. Dies führt weg von einem stereotypen, rezeptartigen Interventionsverständnis hin zu einem kontextsensiblen, adaptiven Verständnis. Interventionen, die in einem System erfolgreich waren, sind es nicht zwingend in einem anderen und es muss vor einer Übergeneralisierung lokaler Erfahrungen gewarnt werden (Dörner, 2008). In klinischen Untersuchungen konnte mittlerweile der Einfluss des Kontextes auf die Intervention selbst bei Schmerzmitteln eindeutig nachgewiesen werden (Walach, 2006). Für die Planung und die Durchführung von Interventionen bedeutet dies, dass der Kontext, also das zu verändernde System und seine relevanten Umwelten möglichst gut bekannt sein müssen, um die Intervention darauf abstimmen zu können. Da sich Zielsystem und seine Umwelt als dynamische Systeme laufend weiter entwickeln, muss sich das Interventionssystem flexibel diesen Veränderungen anpassen können und sich mit dem Zielsystem entwickeln (vgl. Adaptation und Co-Evolution). Erfahrungen aus komplexen Interventionsprojekten und Computersimulationen haben gezeigt, dass wenige wohltdosierte Steuerungsmassnahmen dabei in der Regel effektiver sind als rasches und hartes Eingreifen (Dörner, 2008, 232), da die Gefahr besteht, Systeme durch zu starkes Eingreifen zu übersteuern. Durch die zeitliche Verzögerung des Eintretens der Wirkung bedingt läuft man Gefahr, so lange und so heftig in die eine Richtung zu steuern, dass der Steuerungsimpuls bereits viel zu gross ist, wenn das System zu reagieren beginnt. Bei der Korrektur in die Gegenrichtung passiert dann oft wieder dasselbe in umgekehrter Form. Erfolgversprechender ist es, kontinuierlich nur kleine Änderungen vorzunehmen (Dörner, 2008). Die Multidimensionalität und Zirkularität von Ursache-Wirkungs-Zusammenhängen erfordern ein kontinuierliches Beobachten des Zielsystems, um die unterschiedlichen relevanten Parameter im Auge zu behalten und das Auftauchen unerwarteter Effekte zu erkennen, damit frühzeitig reagiert werden kann:

Erst die Vernetztheit, also die zwischen den Variablen des Systems existierenden Verknüpfungen, macht die *gleichzeitige* Beachtung sehr vieler Merkmale notwendig und bringt es mit sich, dass man in solchen Realitätsausschnitten fast nie nur *eine* Sache machen kann. (Dörner, 2008, 61; Hervorhebungen im Original)

Die sorgfältige Beobachtung verschiedenster Variablen ist deshalb wichtig, weil nicht mit Sicherheit vorausgesagt werden kann, welche Elemente von Interventionen letztlich die gewünschten oder auch unerwünschte Wirkungen erzeugen:

Komplexe Systeme reagieren auf die Veränderung vieler Systemparameter bemerkenswert gering: hier sind Interventionen weitgehend bedeutungslos. Diese Parameter können noch so genau gemessen oder gesteuert werden - sie beeinflussen das Verhalten des Systems kaum. Aus der Sicht des Beobachters scheint es, als sei das System träge, als reagiere es auf Interventionen nicht. Allerdings reagieren Systeme organisierter Komplexität stark auf einige wenige Parameter oder Strukturveränderungen. Jedes System hat Stellen oder Druckpunkte («that pushes my buttons»), auf die es sehr sensibel reagiert. Um diese Stellen zu finden, bedarf es einer genauen Untersuchung der Systemdynamik. (Willke, 2005, 69 ff, Hervorhebungen im Original)

Allerdings führt Willke an, dass der menschliche Verstand aufgrund seiner genetischen Ausstattung und Funktionsweise nicht geeignet sei, die Dynamik komplexer Systeme zu verstehen und oft Interventionen nahelege, die einen zu behebenden Missstand gar vergrössern würden (Willke, 2005, 70)⁶⁶. Computersimulationsprogramme können helfen, Systemdynamiken verstehen zu lernen und besonders sensible Parameter erkennen zu können (Vester, 2007). Solche Simulationen sind im Hinblick auf prognostische Aussagen allerdings sehr beschränkt, da die Bestimmung der Inputvariablen und Operationen immer sehr selektiv ist und selbst kleinste - z.B. nicht berücksichtigte - System- oder Umweltvariablen grosse Dynamiken auslösen und das System entscheidend verändern können.

Vester (2007) hat auf Basis von Dörners Untersuchungen (Dörner, 2008) typische Fehler identifiziert, welche beobachtet werden können, wenn Menschen komplexe Herausforderungen zu bewältigen haben: Anstatt das System als Ganzes im Auge zu behalten, wird versucht, einzelne Probleme isoliert zu lösen, indem man sich auf wenige identifizierte Aspekte konzentriert. Da diese Problemlösungen zu Beginn oft Erfolge zeigen, wird das gewählte Vorgehen bestärkt und andere Aspekte wie beispielsweise Nebenwirkungen oder zusätzliche Probleme werden ausgeblendet. Es werden zwar viele Daten über ein System gesammelt, es gelingt aber nicht, die Dynamik des Systems, seine Funktionsweise und die Organisationsprinzipien zu erfassen. Die oben bereits angesprochene Übersteuerung wird als weiterer typischer Fehler identifiziert - wenn kleine Steuerungsversuche nicht sofort Effekte zeigen, ist man versucht, die Interventionsstärke massiv zu erhöhen, ohne dem System Zeit zu geben, sich in die gewünschte Richtung zu entwickeln. Als zusätzlicher Faktor wird die eigene Überschätzung angeführt, die dazu treibt, persönliche Sichtweisen in den Vordergrund zu stellen. Die Gründe für diese Fehler hängen gemäss Vester damit zusammen, dass wir es einerseits gewohnt sind, Dinge zu zerlegen, um sie verstehen und erklären zu können, damit aber übergeordnete Mechanismen resp. emergente Strukturen übersehen. Wenn solche aber erkannt werden, würden sie verdrängt und es würde lieber auf bekannte, herkömmliche Managementformen zurückgegriffen. Zudem liege es nicht in der Natur des Menschen, in langfristigen Zielhorizonten zu denken.

Obwohl es bisher keine anerkannte Theorie der Intervention oder des Managements komplexer Systeme gibt (Willke, 2005, 67), lassen sich wesentliche Grundlinien für eine solche ausmachen. So sind in Konsequenz der operativen Geschlossenheit Veränderungen in selbstreferentiellen Systemen nur denkbar als Eigenleistungen des zu verändernden Systems:

⁶⁶ Anmerkung: Es ist zwar gewiss so, dass der menschliche Verstand diesbezüglich limitiert ist, allerdings ist es bemerkenswert, wie er dennoch hoch komplexe Herausforderungen - wie beispielsweise die Erziehung eines Kindes oder die Koordination eines Grossprojektes - erfolgreich und mit Voraussicht, wenn auch ohne Gewähr, meistern kann.

Ganz im Gegensatz zu herkömmlichen Vorstellungen von Intervention bedeutet dies, dass jede externe Intervention darauf angewiesen ist, sich als Ereignis im Bereich der möglichen Perzeptionen des intervenierten Systems darzustellen und als Information in die operativen Kreisläufe dieses Systems eingeschleust zu werden. Wie diese Informationen sich auswirken, hängt demnach in erster Linie nicht von den Absichten der Intervention, sondern von der Operationsweise und den Regeln der Selbststeuerung des Systems ab, in das interveniert werden soll. Damit ist jede direkte Verhaltenslenkung nach dem Muster der Lenkung trivialer Maschinen ausgeschlossen. Jede Beeinflussung eines psychischen oder sozialen Systems setzt deshalb voraus, dass Umweltereignisse als perzipierte Differenzen auf den Bahnen des Regelwerks der Selbststeuerung dieses Systems prozessiert werden. Um bleibende Veränderungen zu bewirken, müssen die aus diesen Differenzen intern gewonnenen Informationen Veränderungen dieses Regelwerks induzieren, ohne dabei den Zusammenhang des Regelwerks als Netzwerk der Autopoiese des Systems zu zerstören.(...) Veränderung ist - legt man die Theorie selbstreferentieller Systeme zugrunde - nur als Eigenleistung des zu verändernden Systems möglich. Sie ist andererseits ohne externen Anstoss unwahrscheinlich. (...) Gelingende Intervention heisst demnach, dass externe Anstösse interne Entwicklungsmöglichkeiten des zu intervenierenden Systems beobachtbar machen und für das System in den Verfügungsraum möglicher Optionen bringen. Im Kern laufen damit alle Interventionen in komplexe Systeme auf *Selbständerungen* dieser Systeme heraus. Allerdings sind entsprechend angelegte Interventionen eine *conditio sine qua non* für die Möglichkeit der Selbständerung. (Willke, 2005, 88 ff; Hervorhebungen im Original)

Harry Jones hat auf der Grundlage der Komplexitätstheorien⁶⁷ Prinzipien für die Steuerung von komplexen Interventionsprojekten im Rahmen der Entwicklungszusammenarbeit entwickelt (Jones, 2011b). Sein Ansatz orientiert sich konsequent an der Selbstorganisation komplexer sozialer Systeme und läuft darauf hinaus, die Voraussetzungen zur Selbstorganisation und autonomen Problemlösung zu fördern. Diese Prinzipien bieten sich als Grundlage für eine Interventionstheorie komplexer sozialer Systeme an und können als solche auch als Evaluationskriterien für die Beurteilung von Interventionen in komplexen Zusammenhängen dienen. Sie geben auch erste Antworten auf die Frage, wie ein System organisiert werden muss, damit es sich selber organisieren kann.

Der Schlüssel für eine erfolgreiche Programmsteuerung liegt gemäss Jones in einem angemessenen Umgang mit Wissen und es gilt, die Art und Weise der Vermittlung zwischen Wissen und Entscheidung so zu definieren, dass das System ‚empowert‘ wird und Prozesse der Selbstorganisation gefördert werden. Die Ansätze zur Bewältigung von Komplexität können in den Eigenschaften komplexer Systeme selbst gefunden werden - wenn also beispielsweise Dynamik, Empowerment oder Selbstorganisation massgeblich zur Komplexität beitragen, sind auch Dynamik, Empowerment und Selbstorganisation Schlüssel zum Umgang mit dieser Komplexität. An dieser Stelle sollen zunächst jene Prinzipien angeführt werden, welche als strukturelle Rahmenbedingungen für Selbstorganisation angesehen werden können, weitere Prinzipien, welche den expliziten Umgang mit Wissen fokussieren und Monitoring und Evaluation betreffen, werden im Kapitel 8 aufgegriffen:

Implementing agencies need to collaborate and facilitate decentralised action and self-organisation. This can be done in the following ways:

- *Decentralisation and autonomy*: one priority is to decentralise policy-making and implementation, distributing decision-making power and involvement and allowing increased autonomy for units lower in a hierarchy.
- *Engaging local institutions*: implementing agencies may need to work with and through local organisations to address an issue at different scales and anchor interventions in local realities. For complex problems this may work best through co-management and power sharing.

⁶⁷ s. das vorausgegangene Grundlagendokument Ramalingam and Jones (2008)

- *Convening and boundary management*: implementing agencies may be able to play a role in facilitating processes that build trust and collaboration between stakeholders. They must act as trustworthy stewards of these processes, including providing transparent mechanisms for conflict resolution.
- *Building adaptive capacity*: capacity-building is central to enable actors to capitalise on any autonomy for addressing problems. Support for adaptive capacity and networks is crucial to stimulate emergent responses.
- *Removing the barriers to self-organisation*: many types of barriers and systemic issues prevent actors from adapting to emerging problems, such as the barriers to enabling national legislation or political systems, or issues of power, discourse and social capital.
- *Supporting networked governance*: agencies need a networked approach to governance, appropriate structures are needed to hold units accountable, and relationship management and participatory processes are crucial.
- *Leadership and facilitation*: even where the capacity to act is distributed, leadership is a critical variable. In the face of complex problems, leadership must be facilitative and enabling, working through attraction rather than coercion, and communicating a vision of change around which responses can emerge.
- *Incremental intervention*: where a central agency needs to intervene, this should be handled incrementally, starting from existing networks and taking an evolutionary approach that 'seeds' decentralised action and supports emerging responses, rather than implementing idealistic blueprints.

(Jones, 2011a, 3; für eine ausführlichere Beschreibung und Begründung der Prinzipien s. Jones, 2011b).

Wie das Prinzip „Leadership and facilitation“ deutlich macht, ist Selbstorganisation in Zusammenhang mit Intervention nicht gleichbedeutend mit Verzicht auf Führung, denn Selbstorganisation soll in eine bestimmte Richtung gelenkt werden. Vielmehr geht es darum, Selbstorganisationsprozesse angemessen zu nutzen, um bestimmte Ziele in komplexem Umfeld zu erreichen. Oder wie es Jones ausdrückt:

Shaping policy will always be a matter of degrees, and a negotiation between bottom-up and top-down structures, between planned and emergent responses and between technical and participatory guidance. (Jones, 2011b, 21)

Ein Interventionssystem, welches im Zielsystem gerichtete Selbstorganisationsprozesse anregen will, muss sich konsequenterweise dem Zielsystem öffnen, das heisst, es muss Personen, die sich an der Kommunikation des Zielsystems beteiligen, auch an der Kommunikation im Interventionssystem beteiligen. Diesen an beiden Sozialsystemen beteiligten Personen kommt eine grosse Bedeutung in Interventionen zu - sie erhöhen die Chance, dass Kommunikationen des Interventionssystems erfolgreich in das Zielsystem übersetzt und dort anschlussfähig werden. Sie ermöglichen, neue Sichtweisen in das zu intervenierende Sozialsystem hineinzutragen, das System mit Impulsen von aussen zu irritieren und Anregungen für Veränderungen anbieten zu können resp. Lernen zu ermöglichen.

Interventionen der Gesundheitsförderung und Prävention können unterschiedliche Formen und Komplexitätsgrade haben. Erfüllen sie die Prinzipien der Gesundheitsförderung, so kann auch bei lokal begrenzten Interventionen grundsätzlich von einer hohen Komplexität ausgegangen werden (s. vorangegangenes Kapitel). Es gilt, vielfältige Akteure mit unterschiedlichen Hintergründen, Interessen und Rollen zusammenzubringen und sie an Interventionen zu beteiligen, welche nicht einzelne Individuen sondern ganze Bevölkerungsgruppen fokussieren und auf vielfältige (strukturelle) Veränderungen in sozialen Systemen abzielen. Auch wenn aufgrund von einschlägiger Erfahrung Vorhersagen über das Verhalten zu intervenierender Systeme gemacht werden können, bleiben diese doch immer probabilistisch und sind von Unsicherheit geprägt. Da jedes Zielsystem und jedes Interventionssystem einzigartig ist, sind die Möglichkeiten der Standardisierung von Programmen und Projekten begrenzt,

diese müssen den Zielsystemen und dem weiteren Kontext angepasst, und deren Entwicklungen müssen laufend beobachtet werden.

Die prinzipielle Unterschiedlichkeit und Komplexität sozialer Systeme muss aber nicht zu einer resignativen Haltung führen. Selbst wenn nicht alle Elemente eines Sozialsystems im Detail erfasst werden können, sind dennoch (approximative⁶⁸) Aussagen über das System und sein Verhalten möglich. Nicht die Kenntnis jedes einzelnen Individuums, seiner Geschichte, seiner Erwartungen und seiner Handlungsoptionen sind erforderlich, um das System verstehen zu können, sondern eine Analyse der Strukturen und Regeln auf der Makroebene des Systems, also beispielsweise der kollektiven Geschichte, der Normen und Werte, der Systemdynamik etc. Mainzer erläutert dies am Beispiel der Steuerung des Autoverkehrs:

Wir müssen z. B. nicht das tatsächliche Mikroverhalten jedes einzelnen Autofahrers kennen, um für bestimmte Verkehrsdichten ein Makroverhalten wie Stop-and-Go-Wellen oder Verkehrsinfarkt voraussagen zu können. Intelligente Verkehrsleitsysteme müssen lernen, solche Trends rechtzeitig aus Dichtemustern zu erkennen und sich dem Verkehrsfluss anzupassen. Ebenso muss intelligentes Management lernen, mit Instabilitäten sensibel umzugehen und geeignete Rahmenbedingungen zu setzen, damit sich eine gewünschte Geschäftsdynamik selbst organisiert. (Mainzer, n.d., 24)

Wenn auch Ordnungsstrukturen auf der Makroebene nicht aus den einzelnen Elementen eines Systems abgeleitet werden können, so ist es doch so, dass die Ordnungsstrukturen auf der Makroebene nicht anders als durch Interaktionen auf der Mikroebene (systemintern und in Interaktion mit der Umwelt) zustande kommen - nur sind die Mechanismen aufgrund der Komplexität des Systems nicht einsichtig. Gleichzeitig ist es aber auch so, dass die Ordnungsstrukturen der Makroebene wieder auf die Mikroebene zurückwirken. Simon nennt dieses Prinzip das Versklavungsprinzip, nach welchem das Verhalten der einzelnen Systemelemente durch wenige Ordnungsparameter bestimmt wird, was die Freiheitsgrade der Einzelemente drastisch begrenzen kann (Simon, 2008, 24). Um beim Beispiel des Verkehrs zu bleiben: wenn man in Länder fährt, in denen andere Verkehrsregeln und „ungeschriebene“ Gesetze gelten, tut man gut daran, sich diesen anzupassen, um überhaupt vorwärts zu kommen.

Es wird mehr und mehr deutlich, dass soziale Systeme und damit auch soziale Interventionen grundsätzlich von hoher Komplexität sind. Es stellt sich aber gleichzeitig die Frage, ob diese grundsätzliche Komplexität in jeder Hinsicht relevant ist. Petticrew (2011) verneint dies. Er geht davon aus, dass es der Blick der Forschenden ist, der Interventionen als komplex oder einfach erscheinen lässt. So schlägt er in Abgrenzung zum gängigen Weg vor, Einfachheit oder Komplexität nicht an der Intervention festzumachen, sondern an der Forschungsfrage und entsprechend auch „Einfachheit“ zuzulassen:

However, there is another possibility: that in fact there are no ‘simple’ or ‘complex’ interventions, and that simplicity and complexity are instead pragmatic perspectives adopted by researchers to help describe and understand the interventions in question. After all, on close examination most interventions, including medical interventions, reveal themselves to be complex. Despite this, it is often helpful to treat them as simple, in order to answer appropriately simple questions. (Petticrew, 2011, 397)

Dieser konstruktivistische Ansatz verträgt sich tendenziell gut mit der systemtheoretischen Perspektive Willkes, wonach soziale Systeme nicht an sich komplex sind, sondern sich die Komplexität sozialer Systeme immer nur im Hinblick auf eine bestimmte Herausforderung bestimmen lässt. Inwiefern sich dieser Ansatz allerdings für die Evaluation und Wirksamkeitsbeurteilung als fruchtbar erweisen wird

⁶⁸ Aussagen auf der Makroebene eines komplexen Systems sind prinzipiell „nur“ approximativ (Cilliers (2005), 258).

oder als zu reduktionistisch zurückgewiesen werden muss, wird auf der Basis des empirischen Teils gründlicher zu reflektieren sein.

Aus der Perspektive der Komplexitätstheorie, resp. entlang der Merkmale komplexer sozialer Systeme ergeben sich für die Evaluation von Interventionen folgende Fragestellungen, die zu vier Bereichen zusammengefasst werden können:

Interaktionsvielfalt

Konnektivität und Interdependenz: Wie berücksichtigt die Evaluation systeminterne und systemübergreifende Zusammenhänge?

Offenheit: Auf welche Art und Weise wird der Kontextbedingtheit untersuchter sozialer Systeme Rechnung getragen?

Komplexe Wirkungsmechanismen

Feedback: Inwiefern werden positive und negative Feedbacks zwischen Systemen in der Evaluation berücksichtigt?

Multidimensionalität/Nicht-Linearität/Zirkularität: Welches Verständnis von Ursache-Wirkungsbeziehungen liegt der Evaluation zugrunde und wie geht die Evaluation mit der multidimensionalen, nicht-linearen und zirkulären Wirkungsmechanismen um?

Dynamik: Inwiefern wird die hohe Dynamik sozialer Systeme in der Evaluation eingefangen und berücksichtigt?

Selbstorganisation

Adaption, Co-Evolution: Inwiefern werden systemübergreifende Entwicklungsprozesse in der Evaluation berücksichtigt?

Reflexivität: Inwiefern berücksichtigt die Evaluation systemeigene Perspektiven und die Eigenreflexion der untersuchten Ziel- und Interventionssysteme?

Selbstorganisation/operative Geschlossenheit: Inwiefern wird Selbstorganisation von Ziel- und Interventionssystemen untersucht und in die Wirkungsbeurteilung aufgenommen?

Emergenz

Emergenz: Inwiefern werden in der Evaluation emergente Systemeigenschaften der untersuchten Settings berücksichtigt? Werden Eigenschaften und Dynamiken auf Ebene von Gesamtsystemen (z.B. in Settings) untersucht oder wird lediglich auf isolierte Subsysteme fokussiert?

Unvorhersehbarkeit/Ungewissheit: Wie wird mit Unvorhersehbarkeit und Ungewissheit umgegangen? Ist die Evaluation so angelegt, dass Unvorhersehbares, Ungeplantes und Nebenwirkungen sichtbar werden?

5 Erkenntnis im Lichte von Komplexität

5.1 Systemtheoretischer Konstruktivismus

Die Entdeckung der Komplexität und der damit einhergehende systemtheoretische Blick auf die Welt vertrugen sich schlecht mit einem Wissenschaftsverständnis Newtonscher Prägung, welches von einer deterministischen Realität, von linearen Ursache-Wirkungszusammenhängen und einer omnipotenten Berechenbarkeit der Natur (Mainzer, n.d., 2) ausging. Der auf dem Newtonschen Weltbild aufbauende Positivismus schien in den zentralen Fragestellungen diametral den Erkenntnissen der Komplexitäts- und der neueren Systemtheorie entgegenzustehen. Er ging davon aus, dass:

- die physische und die soziale Wirklichkeit durch universelle Gesetze bestimmt resp. determiniert werden,
- die Veränderungen in der Natur und im Sozialen exakt vorausgesagt werden können, wenn diese Gesetze bekannt sind,
- diese Gesetze empirisch über Induktionsschlüsse gefunden werden können,
- komplex erscheinende Phänomene erklärt werden können, indem sie in ihre einfacheren Einzelelemente zerlegt und diese analysiert werden,
- dass beobachterunabhängiges, objektives Erkennen möglich ist.

(Lincoln & Guba, 1985, 19 f)

Die damit verbundene reduktionistische Forschungspraxis, welche Phänomene zu erklären suchte, indem sie diese in Einzelelemente zerlegt analysierte, um die Ergebnisse dann aufzuaddieren (Ramalingam & Jones, 2008), widersprach dem ganzheitlichen Verständnis systemtheoretischer Ansätze und der Einsicht, dass durch Interaktionen Systemeigenschaften entstehen, die nicht auf die einzelnen Elemente zurückführbar sind. Reduktionistische Ansätze und lineare Ursache-Wirkungsverständnisse konnten keine überzeugenden Antworten liefern auf die überall sichtbar werdenden Phänomene unvorhersehbarer Dynamiken und emergenter Ordnungen. Insbesondere in den Sozialwissenschaften wuchs die Erkenntnis, dass Realität nicht direkt zugänglich und objektiv erfahbar ist, sondern dass unterschiedliche Sozialsysteme unterschiedliche Realitäten entwickeln. Die kognitive und die damit verbundene zeitliche Komplexität, welche sich speziell auf psychische und soziale Systeme bezieht, unterstützen derartige relativistische Konzepte. In erkenntnistheoretischer Gegenposition zum Positivismus entwickelten sich naturalistische und konstruktivistische Konzepte. Der erkenntnistheoretische Naturalismus beansprucht für sich, Wirklichkeit „naturnaher“, angemessener beschreiben zu können, wobei unter Wirklichkeit keine objektiv gegebene Wirklichkeit verstanden wird, sondern multiple, sozial konstruierte Realitäten⁶⁹. Lincoln Guba (1985) setzen die Axiome naturalistischer Erkenntnisgewinnung einer positivistischen gegenüber:

„Axiom 1: The nature of reality (ontology)

Positivist version: There is a single tangible reality „out there“ fragmentable into independent variables and processes, any of which can be studied independently of the others; inquiry can converge onto that reality until, finally, it can be predicted and controlled.

Naturalist version: There are multiple constructed realities that can be studied only holistically; inquiry into these multiple realities will inevitably diverge (...) so that prediction and control are unlikely outcomes although some level of understanding (verstehen) can be achieved.

⁶⁹ Eine andere Konnotation des „naturalistisch“ besteht darin, dass naturalistische Forschung ihre Gegenstände in ihrer natürlichen Umgebung untersuchen will, anstatt unter kontrollierten Bedingungen. Damit soll auf Kosten interner Validität die externe Validität gesteigert werden.

Axiom 2: The relationship of knower to known (epistemology)

Positivist version: The inquirer and the object of inquiry are independent; the knower and the known constitute a discrete dualism.

Naturalistic version: The inquirer and the “object“ of inquiry interact to influence one another; knower and known are inseparable.

Axiom 3: The possibility of generalization

Positivist version: The aim of inquiry is to develop a nomothetic body of knowledge in the form of generalizations that are truth statements free from both time and context (they will hold anywhere and anytime).

Naturalistic version: The aim of inquiry is to develop an idiographic body of knowledge in the form of “working hypotheses“ that describe the individual case.

Axiom 4: The possibility of causal linkages

Positivist version: Every action can be explained as the result (effect) of a real cause that precedes the effect temporally (or is at least simultaneous with it).

Naturalistic version: All entities are in a state of mutual simultaneous shaping so that it is impossible to distinguish causes from effects.

Axiom 5: The role of values in inquiry (axiology)

Positivist version: Inquiry is value-free and can be guaranteed to be so by virtue of the objective methodology employed.

Naturalistic version: Inquiry is value-bound in at least five ways (...) ⁷⁰.“
(Lincoln & Guba, 1985, 37ff)

Mit diesen Axiomen nahmen die Autorin und der Autor Abschied von der Vorstellung einer objektiv erfassbaren, berechenbaren und kontrollierbaren Welt. Ausgehend von diesen Axiomen und wiederum in Abgrenzung zu einem positivistischem Verständnis haben Lincoln und Guba (1985) in ihrem Evaluationsklassiker „Fourth generation evaluation“ 20 Theoreme konstruktivistischer Forschung formuliert⁷¹ und für die Evaluation konkretisiert⁷². Aus erkenntnistheoretischer Sicht bedeutend sind u.a., dass wissenschaftliche Wahrheit relativ konzipiert wird (2. Theorem), wissenschaftliche Tatsachen immer theorie- und werteabhängig und entsprechend nie neutral sind (Theoreme 5 und 6) und dass Phänomene jeglicher Art immer soziale und kontextabhängige Konstruktionen sind (Theorem 10). Für die Evaluation bedeutet dies konsequenterweise, dass Erkenntnisse immer werteabhängig sind (Theorem 16), Rechenschaft und Verantwortung nicht eindeutig zugewiesen werden können (Theorem 17) und Evaluationsergebnisse nicht von einer Aussenposition objektiv „gefunden“ und dann kommuniziert werden können, sondern in einem Aushandlungsprozess gemeinsam mit den verschiedenen Stakeholdern entstehen (Theoreme 18 und 19). Laut dem konstruktivistischen Forschungsparadigma kommt Evaluationsergebnissen keine grundsätzlich übergeordnete Bedeutung im Sinne wahren Wissens zu, sondern sie sind lediglich eine gleichwertige Konstruktion unter anderen (Theorem 20). Die Aufgabe der Forschung und damit auch der Evaluation innerhalb eines konstruktivistischen Forschungsparadigmas ist es, in einem Handlungsfeld existierende Konstruktionen zu rekonstruieren, es handelt sich dabei um Konstruktionen 2. Ordnung⁷³ (Simon, 2008). Mit Blick auf Wirkungszusammen-

⁷⁰ Influenced by inquirer-values, by the choice of the paradigm, by the choice of the substantive theory, by the values that inhere in the context and inquiry is either value-resonant or value-dissonant.

⁷¹ Die Autoren weisen darauf hin, dass die Theoreme (Lehrsätze) nicht im Euklidischen Sinne streng aus den Axiomen hergeleitet wurden und wissenschaftlich überprüfbar sind.

⁷² Auch die Geschichte der Evaluation beginnt mit positivistischen Positionen: „Evaluation activity at that time was essentially without theory, following a simple, measurement-focused paradigm, which assumed there was one reality that could be known within some probability, that objectivity could be value-free, and that measurement could be quantified.“ (Lee (2006), 140 f.; „At that time“ weist auf das beginnende 20. Jahrhundert hin).

⁷³ Vgl. systemtheoretisch: Beobachtungen 2. Ordnung.

hänge geht es darum, im Untersuchungsfeld vorherrschende Annahmen über Ursache-Wirkungszusammenhänge kritisch zu hinterfragen und an alternativen Erklärungsmodellen zu überprüfen:

Die Welt mit den Methoden der Wissenschaft zu verstehen, bedeutet demnach nicht, »reine Beobachtungen« korrekt zu interpretieren, sondern die ohne bewusstes Zutun entstandenen Zusammenhangsvorstellungen, unter Bezugnahme auf vorhandenes Wissen, zunächst zu dekonstruieren und dann – im Sinne von Popper (1934), darauf aufbauend – kritisch nach alternativen Erklärungen zu suchen sowie sich anschließend systematisch auf die Suche nach Pro- und Kontraargumenten für die jeweiligen Erklärungsansätze zu machen. (Uhl, 2012, 66ff)

Die durch die Erkenntnisse der Komplexitätstheorie und die neuen erkenntnistheoretischen Ansätze hervorgerufene Verunsicherung war in den Naturwissenschaften, die sich stärker an positivistischen und kritisch-rationalistischen Verständnissen orientierten, vergleichsweise grösser als in den Sozialwissenschaften, welche sich seit je her durch eine grössere Vielfalt an erkenntnistheoretischen und forschungsmethodischen Positionen auszeichnete. Sozialkonstruktivistische (Berger & Luckmann, 2009) und interpretative, qualitativ-rekonstruktive Forschungsansätze (Bohnsack, 2003) haben neben objektivistischen Strömungen seit langem einen festen Platz in den Sozialwissenschaften, spätestens seit Max Weber ist in der Soziologie und in von ihr beeinflussten Disziplinen *Verstehen* ein gleichberechtigtes Anliegen neben *Erklären* und die Auseinandersetzung mit komplexen Phänomenen gehört seit je her zum Kerngeschäft der Soziologie (s. Kapitel 4.1). Die Komplexitätstheorie hat aber auch die Soziologie herausgefordert und zu neuen Sichtweisen und Ansätzen geführt. Mit der neueren Systemtheorie liegt ein theoretischer Ansatz vor, der auf den Schlüsselkonzepten der Komplexitätstheorie aufbaut (Komplexität, Selbstorganisation, Nichtlinearität, Emergenz etc.).

Niklas Luhmann hat nicht nur eine Theorie sozialer Systeme und damit eine Gesellschaftstheorie mit universalistischem Anspruch entwickelt, sondern er hat auf dieser Grundlage auch eine systemkonstruktivistische Erkenntnistheorie entworfen, welche Wissenschaft selber als komplexes soziales System beschreibt. (Wissenschaftliche) Erkenntnis wird nicht im Subjekt bzw. im individuellen Bewusstsein verortet, sondern als kommunikativer Prozess sozialer Systeme konzipiert (Luhmann, 1992). Sie entsteht in einem dynamischen Prozess rekursiver Beobachtungen und bildet sich als strukturelles Resultat der Autopoiesis von Beobachtungssystemen in der Struktur sozialer Systeme ab. Erkenntnis ist somit Konstruktion: „Und deshalb ist alles, was für einen Beobachter Realität ist, Realität dank der Unterscheidung, die er verwendet, also Konstruktion!“ (Luhmann, 1992, 493). Luhmann geht davon aus, dass die Wirklichkeit zwar real existiert, dem Menschen aber nicht kognitiv zugänglich ist⁷⁴. Wissenschaftliche Erkenntnis und Wahrheit werden nicht als Entsprechung zur realen Welt gesehen, sondern können immer nur selbstreferentiell begründet werden. Wissenschaftliche Wahrheit kann nicht mit direktem Bezug auf eine Fremdreferenz hergeleitet werden, sondern entsteht durch Anschlüsse innerhalb des beobachtenden Systems Wissenschaft⁷⁵:

Der Konstruktivismus reflektiert Erkenntnis als geschlossenes System ohne Zugang zur Aussenwelt. Er reduziert den Aussenkontakt auf eine nur für Beobachter sichtbare strukturelle Kopplung, die im System Irritationen auslösen kann, die sich in dessen Strukturen zeigen und zu Neuspezifikationen dieser Strukturen mit Mitteln der systemeigenen Operationen führen können. (Luhmann, 1992, 530)

Mit seinem systemtheoretischen Konstruktivismus grenzt sich Luhmann von objektivistischen Erkenntnistheorien ab, welche aus seiner Sicht die Frage des Verhältnisses zwischen Beobachter und Gegenstand nicht befriedigend beantworten können, und von subjektivistischen Erkenntnistheorien,

⁷⁴ „Die Menschen leben in einer wirklichen Welt, diese ist ihnen aber kognitiv unzugänglich.“ (Luhmann (1992), 307)

⁷⁵ Über die grundlegenden Mechanismen der Kondensierung und Konfirmation (Luhmann (1992)).

die das Problem ungelöst haben, Erkenntnisse zu begründen, die über das einzelne Subjekt hinausgehen (Luhmann, 1992). Der systemtheoretische Konstruktivismus nimmt erkenntnistheoretisch eine Mittelposition ein zwischen einem radikalen Objektivismus und einem relativistischen Subjektivismus. Indem Luhmann Erkenntnis in sozialen Systemen verortet und wissenschaftliche Wahrheit als Konstruktion des Wissenschaftssystems begreift, entzieht er sich gekonnt der Pflicht, den Wahrheitsgehalt von Erkenntnis direkt an der Wirklichkeit messen zu müssen⁷⁶, und liefert gleichzeitig eine fundierte Begründung sozialer, also über das einzelne Subjekt hinausreichender Erkenntnis.

Damit verträgt sich seine Erkenntnistheorie gut mit den Grundzügen der Komplexitätstheorie: Wissenschaftliche Erkenntnis kann als emergente Eigenschaft des dynamischen sozialen Systems Wissenschaft resp. entsprechender Subsysteme verstanden werden. Wissenschaft muss sich selbstorganisierend immer wieder neu konstituieren, und zunächst vermeintlich bescheidene Entdeckungen oder Forschungsergebnisse können über daran anschließende Arbeiten und über vielfältige Wechselwirkungen zu überraschenden Erkenntnisfortschritten oder sogar zu wissenschaftlichen Revolutionen führen⁷⁷.

Ein anderes Argument unterstützt Luhmanns konstruktivistischen Ansatz: Die Komplexität sozialer Systeme und deren prinzipielle Nicht-Fassbarkeit führen zur Frage, wie Erkenntnis über komplexe Systeme dennoch möglich ist, und welche Bedeutung ihr zukommt. Auf der Suche nach einer Antwort fühlt man sich in erkenntnistheoretisch konstruktivistische Sichtweisen gedrängt: da Komplexität nie vollständig erfasst werden kann, ist man zur Selektion gezwungen. Die Selektion der Aspekte, die beschrieben werden, ist aufgrund der hohen Komplexität von Beobachter zu Beobachter verschieden, und die Beschreibungen variieren dabei potenziell auch in ihrer Komplexität:

Since different descriptions of a complex system decompose the system in different ways, the knowledge gained by any description is always relative to the perspective from which the description was made. This does not imply that any description is as good as any other. It is merely the result of the fact that only a limited number of characteristics of the system can be taken into account by any specific description. Although there is no a priori procedure for deciding which description is correct, some descriptions will deliver more interesting results than others. (Cilliers, 2005, 258)

Eine Beschreibung eines komplexen sozialen Systems geht also immer einher mit Komplexitätsreduktion, eine Reduktion, die geprägt ist vom Vorwissen, den Erfahrungen, den Normen und Werten etc. des beobachtenden Systems. Komplexität selber kann deshalb keine objektive Grösse sein. Dörner (Dörner, 2008, 61) versteht Komplexität als etwas Subjektives. Dies widerspricht allerdings dem systemtheoretischen Konstruktivismus, dem ich in dieser Arbeit folgen werde. Komplexität ist demnach weder objektiv noch subjektiv, sondern vielmehr soziale Konstruktion. Da die in Folge dieser Reduktion weggelassenen Elemente in Wirklichkeit⁷⁸ mit dem System und mit dessen Umwelt in vielfältiger Weise interagieren, kann nicht mit Sicherheit abgeschätzt werden, welches der Einfluss der jeweils vorgenommenen und unvermeidlichen Reduktion auf die Erkenntnis ist (Cilliers, 2005). Es wird rasch deutlich, dass in Bezug auf soziale Systeme einerseits keine beobachterunabhängige Erkenntnis möglich⁷⁹ ist und dass es andererseits nur probabilistische Aussagen, resp. bescheidene Positionen geben kann.

⁷⁶ Auch wenn wissenschaftliche Erkenntnis Wirklichkeit nicht abbilden kann, so muss sie sich doch an ihr bewähren: „Der Systembegriff bezeichnet also etwas, was wirklich ein System ist, und lässt sich damit auf eine Verantwortung für Bewährung seiner Aussagen an der Wirklichkeit ein“ (Luhmann (1992), 30).

⁷⁷ Eindrucksvolle Beispiele liefern u.a. Kuhn (1976) und Fleck (1980).

⁷⁸ Dies ist keine ontische Aussage, sondern ich gehe vielmehr davon aus, dass Komplexität ein Konstrukt im Prozess des Erkennens ist, wie es auch Cilliers beschreibt: „[Complex systems, like human systems] are not complex “as such” but only complex as described and defined by a given level of inquiry.“ (Cilliers (2002))

When dealing with complexity, modest positions are inescapable. This does not imply that they should be relative, vague or self-contradictory, nor does it imply a reason to cringe in false modesty. We can make clear, testable assertions about complex systems. We can increase the knowledge we have of a certain system, but this knowledge is limited and we have to acknowledge these limits. (Cilliers, 2005, 263)

Diese Limitierung darf gemäss Cilliers allerdings nicht als Schwäche gewertet werden, sondern das Eingestehen der aus der Komplexität erwachsenden (unumgänglichen) Grenzen und Schwierigkeiten in Bezug auf Erkenntnisgewinn zeuge von einer verantwortungsvollen erkenntnistheoretischen Position, die nicht mit Relativismus im Sinne eines ‚anything goes‘ und Vagheit gleichgesetzt werden dürfe (Cilliers, 2005, 262). Probabilistische Aussagen treten an die Stelle wissenschaftlicher Gewissheit und Beweise: „Chance or probability is no longer a convenient way of accepting ignorance, but rather part of a new, extended rationality.“ (Prigogine, 1997, 155)

Ein positivistisches Wissenschaftsverständnis, welches von objektiver und beobachterunabhängiger Erkenntnis ausgeht, kann mit seinem reduktionistischen Ansatz keine adäquate Antwort liefern auf Fragestellungen komplexer Natur. Die Frage, ob ein Projekt wirkt, kann so nicht beantwortet werden, weil es *das* Projekt selbst bei hoher Standardisierung nicht gibt, sondern es entfaltet sich in neuem Kontext immer wieder neu und anders. Dies lässt auf den ersten Blick bescheidenere aber gleichzeitig vielversprechendere Fragestellungen aufkommen, wie sie u.a. der Ansatz der ‚Realist Evaluation‘ ins Zentrum rückt: wie wirkt ein Projekt für wen, unter welchen Umständen, in welchem Umfang? Welches sind die konkreten Mechanismen, welche zwischen System und Kontext resp. Umwelt spielen? (Pawson & Tilley, 1997). Der kritisch-rationalistische Ansatz Poppers, welcher induktive Verifizierung wissenschaftlicher Theorie ablehnt und einzig den Weg der Falsifizierbarkeit als gültig anerkennt, verträgt sich schlecht mit derart komplexen Phänomenen. Da entsprechende Erkenntnisse immer stark kontext- und zeitgebunden sowie vergangenheitsabhängig sind und komplexe Systeme sich in dauerndem Wandel befinden, sind entsprechende Theorien immer an diese Bedingungen gebunden. Infolgedessen ist es streng genommen nicht möglich, eine Theorie wirklich zu falsifizieren, da die Grundbedingungen, unter denen sie zustande gekommen ist, nicht exakt reproduziert werden können. Falsifikation ist ein ungeeignetes Mittel, um Theorien über komplexe Zusammenhänge und mit probabilistischem Charakter zu prüfen. Es wird sich immer ein Einzelfall, eine Ausnahme finden, welche der Theorie widerspricht, weil die Theorie nie die gesamte Komplexität abdecken kann. Die Bildung resp. Bestätigung einer Theorie durch Induktion unterliegt derselben Schwierigkeit. Empirisch hergeleitete Theorien lassen sich nicht von den Kontextbedingungen trennen, unter denen sie zustande gekommen sind, diese Kontextbedingungen sind nie gleich und deshalb sind Generalisierungen, wie sie Theorien liefern, immer Abstraktionen und Komplexitätsreduktionen. Die Dichotomie zwischen (vorläufig) wahren und falschen - sprich falsifizierten - Theorien und jene zwischen Induktion und Deduktion wird diesen Umständen nicht gerecht. Vielmehr ist ein Theoriebegriff gefordert, der mit Stimmigkeit operiert:

Der systemtheoretische Ausgangspunkt, dass ein System nicht isoliert, sondern nur in Zusammenhang und in Auseinandersetzung mit seiner Umwelt analysiert werden kann, bietet einen Ansatz, um aus dem logischen Dilemma von Induktion und Deduktion herauszukommen. Denn der Grundsatz der Umweltbedingtheit eines Systems führt zu zwei ebenso einfachen wie umfassenden Fragen: Wie ist das, was ist überhaupt möglich? Genauer: was sind die konstitutiven Bedingungen der Möglichkeit eines bestimmten Systems in seiner spezifischen Umwelt? Und: Wozu ist das, was ist, überhaupt notwendig? Genauer: was sind die spezifischen Funktionen beobachtbarer Strukturen oder Prozesse für ein bestimmtes System in seiner gegebenen Umwelt? Die Beantwortung beider Fragen verlangt ein Hin-und-Her-Pendeln zwischen Deduktion und Induktion im Sinne eines iterativen Suchprozesses. Der Forscher muss sich zunächst hypo-

thetische Vorstellungen über den Gesamtgegenstand, sein hypothetisches System, bilden, diese Vorstellungen am Verhalten der Teile prüfen, daraus verbesserte Vorstellungen über das Ganze entwickeln, diese wieder in Bezug zum Verhalten der Teile setzen etc. bis schliesslich Widersprüche soweit abgebaut sind, dass sich ein gewisses Mass an Stimmigkeit ergibt. (Willke, 2006, 126)

Die systemtheoretisch-konstruktivistische Konzeption versteht wissenschaftliche Erkenntnis als Ergebnis sozialer Interaktion und verortet es in sozialen Systemen. Sie versteht Wissenschaft als soziales System unter anderen. Durch die Aufhebung der Trennung zwischen Beobachter und beobachtetem Objekt handelt sie sich allerdings gleichzeitig das Problem ihres eigenen blinden Flecks⁸⁰ ein. Auch abgesehen von dieser erkenntnistheoretischen Grundproblematik, welche Luhmann in „Die Wissenschaft von der Gesellschaft“ ausführlich behandelt (Luhmann, 1992), ist eine Beobachtung sozialer Systeme, welche sich durch eine komplexe Eigendynamik und ein partiell autonomes Innenleben auszeichnen, von aussen nur sehr bedingt möglich. „Geht der Beobachter aber nach «innen», lässt er sich auf diese internen Prozesse ein, dann verschwindet die Differenz von Beobachter und System: der Beobachter ist selber involviert und das System ist selber Beobachter.“ (Willke, 2005, 69). Der Mensch ist selber immer Teil der Welt, an welcher er teilhat, Beobachter und Gegenstand lassen sich nicht trennen und situations- und kontextfreies Beobachten ist nicht möglich (Montuori, 2003, 243). Beobachtung und damit Erkennen sind immer von der Komplexität der untersuchten Systeme mit beeinflusst, diese Komplexität wird für das beobachtende System als Umweltkomplexität relevant.

Die Analyse Ludwik Flecks zur „Entstehung und Entwicklung einer wissenschaftlichen Tatsache“ (Fleck, 1980) unterstützt die systemisch-konstruktivistische Konzeption von Wissenschaft. Er zeigt in seinen Studien, wie wissenschaftliches Wissen in einem kollektiven Prozess zustande kommt und sich in kollektiven Denkstilen manifestiert⁸¹. Wissenschaftliche Tatsachen sind immer denkstilgebunden und historisch, zeitlich und kontextuell bedingt. Ein Denkstil ist laut Fleck immer dann gegeben, wenn zwei oder mehrere Menschen Gedanken austauschen. Dies entspricht im Wesentlichen der sozialen Konstruktion von Erkenntnis in Luhmanns Theorie, wobei die auf Kommunikation beruhenden Denkstile eher den sozialen Systemen Luhmanns entsprechen als die aus Individuen bestehenden Denkkollektive. Denkstile zeigen, wie Fleck gut nachvollziehbar macht, wesentliche Eigenschaften komplexer Systeme: wissenschaftliches Wissen ist als Kollektivarbeit weit mehr als die Summe individueller Arbeiten (Fleck, 1980, 129) und kann gut mit dem Begriff der Emergenz umschrieben werden. Vielseitige Wechselwirkungen und Rückkoppelungen zwischen den unterschiedlichen Akteuren im Wissenschaftssystem tragen dazu bei, dass sich Erkenntnis, aufbauend auf Vorerfahrungen und Vorwissen eben dieses Kollektivs, weiterentwickelt. Veränderungen in den Denkstilen geschehen laut Fleck in der Regel zwar nicht im Sinne Kuhns als abrupte Revolutionen, sondern stetig, den Beteiligten oft unbewusst und unvorhersehbar.

Die systemtheoretisch-konstruktivistische Konzeption erscheint als angemessene erkenntnistheoretische Position, weil sie im Einklang steht mit den Erkenntnissen der Komplexitäts- und Systemtheorie, diese also auf sie selber anwendbar sind. Diese erkenntnistheoretische Position bleibt allerdings nicht folgenlos für das Verständnis von in der Evaluation und in der Wirksamkeitsforschung zentralen Begriffen wie Kausalität, Wirksamkeit und Generalisierung.

⁸⁰ Indem Wissenschaft Gesellschaft beobachtet, beobachtet sie sich auch selbst und Aussagen über die Gesellschaft können nie von der eigenen Position unabhängig formuliert werden (Luhmann (1992)).

⁸¹ „Definieren wir «Denkkollektiv» als Gemeinschaft der Menschen, die im Gedankenaustausch oder in gedanklicher Wechselwirkung stehen, so besitzen wir in ihm den Träger geschichtlicher Entwicklung eines Denkgebietes, eines bestimmten Wissensbestandes und Kulturstandes, also eines besonderen Denkstils.“ (Fleck (1980), 54ff)

5.2 Komplexitätsreduktion aus systemtheoretisch-konstruktivistischer Sicht

Soziale Systeme sind in ihrer Komplexität nicht vollständig fassbar. Sowohl im Alltag wie auch in der Wissenschaft muss mit dieser Begrenzung ein angemessener Umgang gefunden werden, und es gilt eine Balance zu finden im Spannungsfeld zwischen Zulassen von Komplexität und Reduktion von Komplexität.

To fully understand a complex system, we need to understand it in all its complexity. Furthermore, because complex systems are open systems, we need to understand the system's complete environment before we can understand the system, and, of course, the environment is complex in itself. There is no human way of doing this. The knowledge we have of complex systems is based on the models we make of these systems, but in order to function as models – and not merely as a repetition of the system – they have to reduce the complexity of the system. This means that some aspects of the system are always left out of consideration. The problem is compounded by the fact that that which is left out, interacts with the rest of the system in a non-linear way and we can therefore not predict what the effects of our reduction of the complexity will be, especially not as the system and its environment develops and transforms in time. (Cilliers, 2005, 258)

Komplexe Systeme bewegen sich in organisierter Komplexität zwischen Gleichgewicht (Stillstand) und Chaos. Organisierte Komplexität, welche durch selektive Beziehungen zwischen den Systemelementen zustande kommt (Luhmann, 1987, 46), ist Grundlage für den Fortbestand und die Weiterentwicklung eines Systems. Diese Organisation kann allerdings sehr unterschiedlich gestaltet sein. Sowohl im Alltag wie auch in der Wissenschaft nutzen wir ständig verschiedene Formen der Komplexitätsbewältigung, wobei im Unterschied zum Alltag in der Wissenschaft der Umgang mit Komplexität und insbesondere Komplexitätsreduktion weitgehend methodisch kontrolliert vorgenommen werden. Komplexitätsreduktion soll regelgeleitet und für andere nachvollziehbar erfolgen. Die Art des Umgangs mit Komplexität ist in den verschiedenen Forschungsmethoden sehr unterschiedlich. Bei stark standardisierten Erhebungsmethoden wird Komplexität bereits vor der Erhebung stark reduziert, während bei qualitativen Verfahren zunächst mehr Offenheit zugelassen wird und Komplexitätsreduktion stärker in die Auswertung verlagert wird.

Die systemtheoretisch-konstruktivistische Perspektive zwingt hier zu einem Zwischenstopp. Was eigentlich genau kann mit Komplexitätsreduktion gemeint sein, wenn Komplexität doch selber konstruiert und nicht als Eigenschaft von Systemen gegeben ist? Was genau kann dann noch reduziert werden? Luhmann bringt hier Klarheit:

Von Reduktion von Komplexität sollte man dagegen in einem engeren Sinne immer dann sprechen, wenn das Relationsgefüge eines komplexen Zusammenhanges durch einen zweiten Zusammenhang mit weniger Relationen rekonstruiert wird. (Luhmann, 1987, 49)

Es geht also um Rekonstruktion von Konstruktion. Vom System wahrgenommene respektive konstruierte Komplexität wird quasi in einer zweiten Lesart reduktiv rekonstruiert. Aus einer Reihe wahrgenommener Bezüge werden einige für die Weiterbearbeitung als relevant ausgewählt. Auf die Evaluation übertragen bedeutet dies, dass sie für die Aufgabe der Beurteilung z.B. von Wirksamkeit einige Elemente, Beziehungen, Dimensionen als besonders relevant benannt und in den Fokus der Beurteilung stellt. Diese arbiträren Selektionen bestimmen den gesamten Prozess und letztlich auch die Ergebnisse. Evaluationsergebnisse sind immer Konstruktionen, sie sind denkstilgebunden und entsprechend relativistisch. Relativistisch in einem zweifachen Sinne. Erstens wird von sozialen Systemen immer nur selektiv wahrgenommen und zweitens wird dieses Wahrgenommene selektiv reduziert. Um beurteilen zu können, inwiefern vorgenommene Reduktionen einem Evaluationsgegenstand angemessen sind, muss erstens der Gegenstand, so wie er wahrgenommen resp. beobachtet wird, be-

geschrieben werden, und zweitens müssen die darauf verwendeten Selektionen transparent, also nachvollziehbar gemacht werden.

5.3 Evaluation als beurteilendes soziales System

Evaluation wird genauso wie auch soziale Intervention als autonomes, selbstorganisierendes soziales System verstanden, welches aber aus der Beobachterperspektive 2. Ordnung soziale Intervention beobachtet, beschreibt und beurteilt: „Evaluation is a complex process, working within a context, that is even more complex, both of which must be understood for evaluation to achieve its purpose.“ (Lee, 2006, 146).

Und der Zweck der Evaluation kann mit Mertens breit gefasster Definition von Evaluation gesehen werden als die Reduktion von Ungewissheit (und damit Reduktion von Komplexität) in der Entscheidungsfindung: „Evaluation is the systematic investigation of the merit or worth of an object (program) for the purpose of reducing uncertainty in decision making.“ (Mertens zit. in Lee, 2006, 139)

Das Evaluationssystem gehört für das Interventionssystem zu seiner Umwelt, Ereignisse im Evaluationssystem können lediglich als Irritation wahrgenommen und intern verarbeitet werden, sie können ein autopoietisch operierendes System aber nicht kausal determinieren (Willke, 2005, 31). Als autonomes System hat das Interventionssystem die Möglichkeit, Kommunikationsangebote des Evaluationssystems anzunehmen oder aber abzulehnen (Blank & Löwenbein, 2009). Aus komplexitäts- und systemtheoretischer Sicht kommt einem Interventionssystem die Aufgabe zu, Zielsysteme in ihrer Selbstorganisation zu unterstützen, insbesondere indem deren Reflexivität gefördert wird. Analog dazu ist es Aufgabe der Evaluation, Interventionssysteme in ihrer Reflexivität und Selbstorganisation zu unterstützen. Lernen gewinnt in der Evaluation neben Legitimationszwecken zunehmend an Bedeutung, kann aber nur dann erfolgreich sein, wenn das Interventionssystem offen ist für Neues (Blank & Löwenbein, 2009). Eine Intervention wird als soziales System verstanden, welches sich dann konstituiert, wenn das intervenierende System mit dem Zielsystem zu kommunizieren beginnt, mit der Absicht, bei Letzterem eine geplante Veränderung herbeizuführen. Analog dazu konstituiert sich Evaluation als soziales System dann, wenn es zur Kommunikation zwischen dem evaluierenden System und dem Interventionssystem kommt. Genauso wie bei erfolgreicher Intervention Ziel- und Interventionssysteme strukturell gekoppelt sind, ist bei erfolgreicher Evaluation das Evaluationssystem mit dem Interventionssystem und gegebenenfalls direkt mit dem intervenierenden System oder dem Zielsystem strukturell gekoppelt und ermöglicht so Lernprozesse in den entsprechenden Systemen.

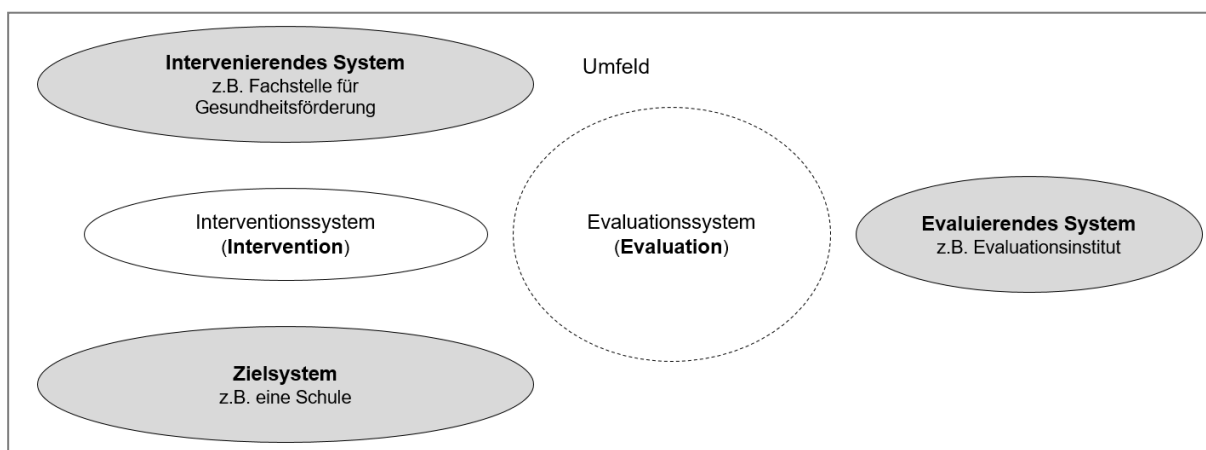


Abb. 8 Das Evaluationssystem als neu entstehendes Sozialsystem

Selbst wenn Evaluation angefordert wird, die Wirksamkeit von Projekten zu beurteilen, kann dies nicht aus einer neutralen Aussenposition heraus geschehen. Soll die Evaluation dem Evaluationsgegenstand angemessen sein, so bedarf es der Nähe zum Gegenstand. Was Wirksamkeit ist und wie sie am besten beurteilt werden kann, ist aus konstruktivistischer Sicht nicht a priori gegeben, sondern bedarf der Aushandlung zwischen Stakeholdern und Evaluierenden. Guba und Lincoln (1989) entwerfen Evaluation vor diesem Hintergrund als kontinuierlichen gemeinsamen und emergenten Prozess mit unsicherem Ausgang, in welchem nicht Wahrheit *gefunden*, sondern gemeinsam immer wieder *kreiert* wird.

5.4 Wirkungen erkennen

Das bisher Gesagte führt darauf hinaus, dass Wirkungen ebenso soziale Konstruktionen sind wie wissenschaftliche Erkenntnis ganz allgemein. Ursache-Wirkungsbeziehungen bzw. Kausalitäten können nicht direkt an der Wirklichkeit festgemacht werden, sondern sie entstehen durch Zuschreibungen von Beobachtern. Diese Position geht auf David Hume zurück. Nach seiner Auffassung hat noch nie jemand eine Ursache gesehen, sondern diese lediglich aus beobachteten Wirkungen abgeleitet. Der Umstand, dass ein Ereignis bisher immer (N mal) auf ein anderes Ereignis gefolgt ist, muss nicht zwingend bedeuten, dass dies auch im Fall N+1 so ist (Lincoln & Guba, 1985). Mc Queen (2013) folgert richtig, dass Wirkungszuschreibungen mehr von der verwendeten Theorie abhängen als letztlich von den beobachteten Phänomenen. Kausalität hat in der wissenschaftlichen Literatur seit je her einen prominenten Platz. Ohne einen systematischen Überblick über die verschiedenen theoretischen Ansätze hier geben zu können⁸², sollen einige für die vorliegende Thematik relevante Aspekte aufgezeigt werden.

Auf der einen Seite wird Kausalität über die Formulierung von Grundbedingungen fassbar zu machen versucht. Folgende Elemente lassen sich aus der Diskussion herauskristallisieren (Lincoln & Guba, 1985):

- Zeitlichkeit: Die Ursache muss vor der Wirkung auftreten
- Geografische Nähe: Die Ursache muss in geografischer Nähe der Wirkung auftreten
- Abhängigkeit: Wenn die Ursache reduziert wird, muss sich auch die Wirkung reduzieren und wenn die Ursache wegfällt, muss auch die Wirkung wegfallen
- Ähnliche Beziehungen: Wenn ähnliche Ereignisse wie die Ursache auftreten, dann müssen auch ähnliche Wirkungen die Folge sein

⁸² Eine differenzierte Darstellung und Diskussion verschiedener Ansätze liefern Lincoln and Guba (1985).

Diese Kriterien sind umstritten. In Bezug auf Zeitlichkeit wird vor allem angeführt, dass eine Wirkung auch zeitgleich mit einer Ursache auftreten kann, zum Beispiel beim Ziehen eines Wagens. Bei einer zeitlichen Distanz kann nicht ausgeschlossen werden, dass in dieser Zwischenzeit nicht andere Faktoren die Wirkung verursachen. Zudem sei im Lichte der Relativitätstheorie gar ein Wirkungseintreten vor einer Ursache möglich (Lincoln & Guba, 1985). In sozialen Systemen führt die kognitive Komplexität zudem dazu, dass potenzielle Wirkungen antizipiert werden können und diese künftigen Wirkungen auf heutige Entscheidungen (Ursachen) Einfluss nehmen können (vgl. auch das Phänomen der ‚self-fulfilling prophecy‘). Die Bedingung der geografischen Nähe kann angezweifelt werden, da Kommunikation über Distanz Ursache und Wirkung zusammenbringen kann. Eine Reduktion der Ursache muss gerade in komplexen Zusammenhängen nicht unbedingt eine Reduktion der Wirkung nach sich ziehen, da es sich nicht um lineare Beziehungen handelt, und das Kriterium der ähnlichen Beziehungen ist vor dem Hintergrund von Komplexität ebenfalls nicht haltbar, da kleine Änderungen in der Ausgangssituation grosse Änderungen in den Ergebnissen nach sich ziehen können.

Die Identifikation von Ursachen und Massnahmen scheint auf den ersten Blick oft einfach, ist bei genauerem Hinsehen aber nicht trivial, wie folgendes Beispiel zeigt:

We say, that malaria is caused by the bite of the anopheles mosquito. But the bite does not in fact always produce the disease, so the cause must not be the bite, but the actual release of plasmodia into the bloodstream. But again, malaria does not inevitably ensue, so the cause must be the attack by the plasmodia on the victim's red blood corpuscles. Malaria is still not inevitable, however, so the cause must be the loss of hemoglobin. But even, this does not always produce malaria, so the cause must be that the tissues are deprived of oxygen. But this last "cause" of malaria is just what malaria is ! Thus, what we do in sorting out causes and effects is to impose an intelligible structure of discrete events upon the continuous stream of occurrence ; we do it in the way that is most useful for our purposes. (Blanshard zit. in Lincoln & Guba, 1985, 137)

An diesem Beispiel lässt sich deutlich sehen, wie Kausalität sozial konstruiert ist, auch anders konstruiert werden könnte, und nicht natürlich gegeben ist. Je nach historisch-kontextuellem Hintergrund werden ganz unterschiedliche Ursachen für ein Phänomen ursächlich verantwortlich gemacht. So waren bei den Griechen, Römern und Germanen Zeus, Jupiter und Thor Verursacher von Blitz und Donner, manchmal wurde der Blitz als Ursache für den Donner angesehen und manchmal wiederum umgekehrt der Donner für den Blitz. Heute gilt das Phänomen Gewitter weitgehend physikalisch erklärt, wenn auch niemand beweisen kann, dass die physikalische Erklärung letztlich wahrer ist als beispielsweise eine göttliche (s. dazu insbesondere Feyereabend, 1980). Die physikalische Erklärung ist aber innerhalb des Wissenschaftssystems gut anschlussfähig, sie bewährt sich in der Interaktion mit der Wirklichkeit und gilt deshalb heute als weitgehend akzeptiert.

Ursachen und Wirkungen lassen sich auch deshalb nur schwer identifizieren, weil meistens mehrere Faktoren gleichzeitig Wirkungen beeinflussen (multiple causation⁸³) oder zusammenspielen müssen, damit eine Wirkung eintritt (conjunctural causation) resp. eine Ursache nur dann Wirkung entfalten kann, wenn andere Bedingungen mit erfüllt sind:

Causality can be described by means of necessary, unnecessary, sufficient and insufficient conditions. We can consider that a fallen lit cigarette has caused the house to burn down. The cigarette in itself is an insufficient condition because favorable circumstances (context) for the ignition of the fire are needed: for instance readily inflammable material and lack of a fire extinguishing system. But the cigarette is a necessary part of the causal chain or cluster (combination of mechanisms and contexts) that in the end produced the fire. (Högnabba et al., 2005, 115)

⁸³ Zum Unterschied zwischen „multiple causation“ und „conjunctural causation“ s. Kolberg (n.d.).

Dieser Umstand stellt eine grosse Herausforderung für die interne Validität von Untersuchungen dar, da Ursachen nicht eindeutig isoliert und nicht alle möglichen alternativen Einflussfaktoren kontrolliert werden können. Auf komplexe Interventionen bezogen, weisen Shiell et al. (2008) auf die grosse Schwierigkeit hin, Ursachen für beobachtete Veränderungen zu identifizieren, vor allem auch deshalb, weil das Zusammenspiel verschiedener Faktoren sehr wahrscheinlich ist:

This is a complex intervention because the content of each consultation is tailored to the individual needs of the patients in the group and is adapted as the programme of consultations unfolds and each client responds in their different ways to treatment. We do not know whether the treatment effect is the result of the content of the consultation, the personal style of the therapist, the dynamic of the group, a combination of all three, or some other consideration. (Shiell, Hawe, & Gold, 2008, 1282)

Die konsequente Berücksichtigung dieser kontextuellen Faktoren steht im Zentrum des Ansatzes der ‚Realist Evaluation‘⁸⁴. Pawson und Tilley verwenden dort ein generatives Konzept von Kausalität, welches in Abgrenzung zu successionistischen Konzepten Wirkungen nicht als beobachtete Aussenwirkungen auffasst, sondern als Zusammenspiel von realen internen und externen Zuständen (Pawson & Tilley, 1997).

Wirkungsfragen resp. Kausalitäten werden nicht nur an Einzelfällen diskutiert. In den Gesundheitswissenschaften werden Kausalitäten vor allem statistisch, in Beobachtungs- oder Interventionsstudien hergeleitet und für bestimmte Bevölkerungsgruppen oder Phänomene generalisiert. Die als gesichert geltenden Zusammenhänge zwischen Rauchen und Lungenkrebs oder zwischen Bewegung und Übergewicht beispielsweise sind derartige generalisierte Kausalitäten. Sie können nicht beobachtet werden (das Rauchen und die Krankheit werden zunächst als isolierte Tatsachen beobachtet), sondern werden aufgrund von empirischen Beobachtungen und logischen Schlüssen konstruiert. Auch randomisierte kontrollierte Studien, welche darauf ausgelegt sind, Kausalitäten zu prüfen, können diese nicht ‚beweisen‘, sondern Kausalitäten immer nur mit einer gewissen Wahrscheinlichkeit postulieren. Selbst bei Vollerhebungen und statistisch erdrückendem Zusammenhang bleibt Unsicherheit: die vermuteten Ursache-Wirkungszusammenhänge werden erstens praktisch nie bei allen Einzelfällen gefunden und zweitens können konfundierende Faktoren nie ausgeschlossen werden⁸⁵. Statistisch hergeleitete Ursache-Wirkungsbeziehungen, auch wenn sie als gesichert gelten, lassen sich leicht hinterfragen, so beispielsweise, wenn langjährige Raucher gefunden werden, die vollkommen gesund sind.

Eine weitere Erschwernis erwächst der Konstruktion von Wirkungen durch Rückkoppelungen. Die Vorstellung von linearen Ursache-Wirkungsbeziehungen trägt im Zusammenhang mit komplexen Systemen nicht. Die beiden traditionellen Kausalitätsprinzipien, nämlich, dass gleiche Ursachen gleiche Wirkungen verursachen und dass kleine Abweichungen in den Anfangsbedingungen zu kleinen Abweichungen in den Ergebnissen führen und umgekehrt, stimmen mit den Erkenntnissen der Komplexitätstheorie nicht überein (Simon, 2008). Ursache-Wirkungsbeziehungen in komplexen Systemen sind vielmehr nicht-linear und zirkulär zu verstehen. Wenn beispielsweise eine einzelne Schneeflocke letztlich eine Lawine auslöst und so das Fass zum Überlaufen bringt, kann dann diese einzelne Schneeflocke als Ursache für die Lawine angesehen werden? Die Zirkularität von Kausalität lässt sich, obwohl auch dies eine Konstruktion⁸⁶, am Beispiel Bewegung und Übergewicht verdeutlichen: Mangeln-

⁸⁴ Realistische Evaluation fragt beispielsweise nicht ‚did x cause y‘, sondern ‚what happened and why‘ (Ramalingam and Jones (2008), 14).

⁸⁵ Potenziell konfundierende Faktoren können zwar statistisch kontrolliert werden, aber auch nur dann, wenn sie bekannt sind. In komplexen Systemen ist es nicht möglich, alle potenziellen Einflussfaktoren zu kennen.

⁸⁶ Zirkulare Kausalität ist rein logisch nicht möglich, weil eine Wirkung nicht wieder als Ursache für ein bereits abgeschlossenes ursächliches Ereignis vorkommen kann (Tang (2007)). Zirkuläre Kausalität, wie sie hier verwendet wird, ist zu verstehen als iterative Abfolge von sich beeinflussenden Ereignissen, wobei in dieser Abfolge

de Bewegung kann zu Übergewicht ursächlich beitragen, Übergewicht kann umgekehrt das Bewegungsverhalten ursächlich beeinflussen und so weiter (Galea, Hall, & Kaplan, 2009). In komplexen zirkulären Kausalitäten ist es unmöglich, Ursachen und Wirkungen strikte getrennt zu halten, was vor allem auch statistische Berechnungen vor grosse Herausforderungen stellt. Das Phänomen der Emergenz führt traditionelle Kausalitätskonzepte vor weitere Schwierigkeiten: emergente Ordnungsmuster zeichnen sich gerade dadurch aus, dass sie nicht auf die einzelnen Systemelemente zurückführbar sind und deshalb als Wirkungen ohne Ursachen erscheinen oder aber als Wirkungen von nicht-definierbaren Ursachen (vgl. das emergente Aroma des Zuckers, welches in den einzelnen Molekülen selber nicht angelegt ist). Da in komplexen sozialen Systemen emergente Eigenschaften vorherrschend sind, scheinen traditionelle Kausalitätskonzepte schlecht für deren Analyse geeignet.

Das traditionelle Konzept der Kausalität hat, obwohl weit verbreitet und etabliert, viele grundsätzliche Schwächen, welche im Zusammenhang mit komplexen sozialen Systemen deutlich in Erscheinung treten und es schwierig erscheinen lassen, an einfachen Ursache-Wirkungsvorstellungen fest zu halten. Willke fasst diese Schwächen wie folgt zusammen:

Nun wird unverkennbar, dass es in den Beziehungen zwischen den interdependenten Elementen eines organisierten Ganzen, das worauf die neuzeitliche naturwissenschaftlich orientierte Wissenschaft aufbaut, so gut wie gar nicht gibt: klare Ursache-Wirkungsbeziehungen. Prägend für solche Systeme sind vielmehr negative und positive Rückkoppelungen, enge und lose Verknüpfungen, Reaktivitäten und Kontextbrüche durch unterschiedliche Systemebenen, Nonlinearitäten und konterintuitive Kombinationswirkungen, reversible und irreversible Prozessverläufe, Fluktuationen und emergente Gesamteigenschaften. (Willke, 2005, 67)

Lincoln und Guba kommen zum Schluss, dass dieses Konzept nicht weiter tragbar ist und sie ersetzen es durch ein Konzept, welches sie „mutual simultaneous shaping“ nennen (Lincoln & Guba, 1985). Dieses Konzept geht davon aus, dass alle Elemente eines Systems sich kontinuierlich gegenseitig beeinflussen und das Verhalten eines Elementes durch die spezifische Konfiguration aller anderen Elemente beeinflusst wird. In der Analyse dieser Systeme werden nun durch den Beobachter die für den eigenen Zweck am meisten plausiblen Elemente herausgegriffen und als formende (shaping) Faktoren deklariert, und zwar wird dabei unterschieden zwischen förderlichen (enabling) und hinderlichen Faktoren (blocking or masking). Das Konzept geht nicht von klar deklarierbaren Ursachen und Wirkungen aus, und die „formenden“ Faktoren werden nicht deterministisch verstanden, sondern im Sinne von probabilistischen Zuschreibungen im Hinblick auf eine beobachtete Situation. Die Abschwächung, welche die Autorin und der Autor mit diesem Konzept vornehmen, ist angesichts der Komplexitätstheorie zu unterstützen, allerdings kann dies auch ohne Umbenennung geschehen, zumal der Begriff der Kausalität weit verbreitet ist und auch von anderen Autoren in ähnlich abgeschwächter Form genutzt wird. Zudem scheint es schwierig, auf Kausalität unterstellende Begriffe ganz zu verzichten. Unsere Sprache ist geprägt von derartigen Zuschreibungen (bewirken, bedingen, zur Folge haben,...) und auch die von Lincoln und Guba verwendeten Begriffe „shaping“, „enabling“, „management“ etc. implizieren letztlich auch Annahmen über Ursache-Wirkungsbeziehungen.

Einen vergleichbaren Zugang zum Problem der Kausalität liefert die funktionalistische Analyse (Luhmann, 1970)⁸⁷. Die Unendlichkeitsproblematik der Kausalität, welche besagt, dass jede Wirkung unendlich viele Ursachen und jede Ursache unendlich viele Wirkungen hat, verunmöglicht es, Kausalität als invariante Beziehung zwischen einer Ursache und einer Wirkung festzulegen. Der Ausschluss aller anderen Kausalfaktoren sei nicht gerechtfertigt und eine Isolierung einer einzelnen Ursache-Wirkungsbeziehung empirisch nicht zu leisten. Luhmann schlägt vor, sich von der Vorstellung invarianter Ursache-Wirkungsbeziehungen zu lösen und nicht die 1:1 Kausalbeziehung zu suchen, sondern

ähnliche resp. als gleich eingestufte Ereignisse mehr als einmal vorkommen, also sowohl andere beeinflussen als auch von diesen beeinflusst werden.

⁸⁷ Vgl. insbesondere das Kapitel „Funktion und Kausalität“, S. 9-30.

von einer interessierenden Ursache oder Wirkung als funktionalem Bezugsgesichtspunkt auszugehen und nach äquivalenten Kausalbeziehungen zu fragen (1:n-Beziehungen):

Die funktionalistische Aussage betrifft nicht eine Beziehung von Ursache und Wirkung, sondern ein Verhältnis mehrerer Ursachen zueinander bzw. mehrerer Wirkungen zueinander, also die Feststellung funktionaler Äquivalenzen. (Luhmann, 1970, 18)

Und weiter:

Die Analyse selbst konzentriert sich entweder auf die Erforschung möglicher Ursachen unter dem Leitgesichtspunkt einer Wirkung oder auf die Erforschung von Wirkungen unter dem Leitgesichtspunkt einer Ursache. Beides zugleich durchzuführen, ist unmöglich, weil jede funktionalistische Analyse einen gewählten Bezugsgesichtspunkt voraussetzt, der nicht geändert werden kann, ohne dass die Ergebnisse sich verschieben. Zwischen Ursachen und Wirkungen besteht in diesem Sinne eine „Unbestimmtheitsrelation“. Der Sinn der Kausalität schliesst prinzipiell eine eindeutige Feststellung einer Ursache und einer Wirkung zugleich aus. Was die ontologische Auslegung der Kausalität anstrebte, ist nicht erreichbar. Diese Einsicht ergibt den Ausgangspunkt für die funktionalistische Kausaltheorie. Für sie sind exklusive Kausalgesetze allenfalls ein analytischer Grenzfall: Dass weder im Bereich der Ursachen noch im Bereich der Wirkungen andere Möglichkeiten bestehen, ist denkbar als Grenzfall absolut reduzierter Äquivalenz. Der Sinn des kausalen Beziehens liegt aber nicht darin, diesen Grenzfall zu erreichen und andere Möglichkeiten auszuschliessen, sondern darin, sie zu erfassen und zu ordnen. (Luhmann, 1970, 17)

Erkenntnis über kausale Zusammenhänge bedeutet in der funktionalistischen Analyse nicht Bestätigung von einzelnen deduktiv abgeleiteten Wirkungshypothesen, sondern Verstehen multipler Ursache-Wirkungsbeziehungen. Die empirische Verifikation orientiert sich nicht an der Frage, ob A immer (bzw. mit angegebener Wahrscheinlichkeit) B bewirkt, sondern ob A, C, D,... in ihrer Eigenschaft, B zu bewirken, funktional äquivalent sind⁸⁸. Diese bescheidene aber anspruchsvolle Herangehensweise an kausale Zusammenhänge geht Komplexität differenziert an und verzichtet auf Kausalversprechungen, die sie nicht halten kann. Sie bildet einen interessanten Ansatz für die Überprüfung von Wirksamkeit sozialer Interventionen, indem sie von einer festgestellten Wirkung ausgehend nach äquivalenten Ursachen fragt (im Sinne der formenden Faktoren bei Lincoln und Guba) oder umgekehrt von einer Intervention nach äquivalenten Wirkungen. Die detaillierte und vergleichende funktionale Analyse liefert laut Luhmann gleichzeitig eine erkenntnistheoretische Brücke zwischen einer Wirklichkeit, die nicht direkt, also unvermittelt zugänglich ist und sozial konstruierter und damit potenziell beliebiger Realität:

Nach einer alten, einsichtigen Regel treten Wahrheiten in Zusammenhängen auf, Irrtümer dagegen isoliert. Wenn es der funktionalen Analyse gelingt, trotz grosser Heterogenität und Verschiedenartigkeit der Erscheinungen Zusammenhänge aufzuzeigen, kann dies als Indikator für Wahrheit gelten, auch wenn die Zusammenhänge nur für den Beobachter einsichtig sind. Jedenfalls wird es bei dieser Technik des Einsichtgewinns schwerer und schwerer, die Überzeugung festzuhalten, die Ergebnisse könnten auf eine fehlerhafte Methode, auf Irrtum, auf reine Imagination zurückzuführen sein. Damit ist keineswegs gesagt, dass die semantische Form, in der sie präsentiert werden, der Realität „entspricht“; wohl aber, dass sie Realität „greift“, das heisst, sich als Ordnungsform im Verhältnis zu einer ebenfalls geordneten Realität bewährt. (Luhmann, 1987, 91)

⁸⁸ Luhmann unterscheidet dabei zwischen disjunktiven und konjunktiven Äquivalenzen, je nachdem ob es sich um alternative oder um zusammenwirkende Ursachen für bestimmte Wirkungen handelt. Für die empirische Überprüfbarkeit dieser beiden Formen zeigt er Ansätze aber auch damit verbundene Schwierigkeiten und noch unge löste Probleme auf (Luhmann (1970), 23 f).

Es ist naheliegend, den Begriff der Kausalität und damit Ursache-Wirkungsbeziehungen im Lichte der Komplexitätstheorie neu zu definieren. Ansatzpunkte dazu bieten die folgenden Elemente:

- Als Wirkungen deklarierte Ereignisse werden in der Regel von einer Vielzahl von untereinander abhängigen Faktoren beeinflusst, systemintern und aus der Umwelt.
- Wirkungszusammenhänge können nicht verstanden werden ohne ein Verständnis der System-Umwelt-Beziehungen (Elder-Vass, 2007).
- Da Wirkungen durch das Zusammenspiel von verschiedenen Faktoren entstehen, können sie nicht reduktionistisch durch Einzelfaktoren gänzlich erklärt werden.
- Ursachen und Wirkungen sind in komplexen Systemen nicht eindeutig identifizier- und trennbar. Rückkoppelungen führen dazu, dass Wirkungen wiederum zu Ursachen der Ursachen werden.
- Jede Ursache-Wirkungsdeklaration ist eine stark reduktionistische Beschreibung durch einen Beobachter, wobei dieser „Beobachter“ im Sinne der neueren Systemtheorie ein soziales System ist.
- Die durch den Beobachter vorgenommenen Reduktionen führen dazu, dass Ursache-Wirkungsbeziehungen nie vollumfänglich erfasst werden können und stets mit einer gewissen - zuweilen recht grossen - Unsicherheit belastet sind⁸⁹.
- Ursache-Wirkungsbeziehungen in komplexen Systemen sind nicht eindimensional und linear, sondern multidimensional, nicht-linear und zirkulär.
- Kausalität ist wie jede Erkenntnis kontext- und zeitgebunden. In diesem Sinne kann dasselbe Ereignis unter anderen Bedingungen oder zu einer anderen Zeit unterschiedliche Folgen haben.⁹⁰

Trotz dieser Limitierungen kann das Konzept der Kausalität zur Erklärung von Phänomenen herangezogen werden. Allerdings nur mit den Schwächen (und Stärken), die ein sozial konstruiertes und probabilistisches Konzept mit sich bringt. Der Begriff des Potentials kann zu einem Schlüsselbegriff in einem angepassten Kausalitätskonzept werden. Gewisse Ereignisse haben das Potenzial, unter gewissen Bedingungen andere Ereignisse hervorzurufen. So hat Rauchen das Potenzial, unter gewissen Bedingungen (also in Zusammenspiel mit anderen Faktoren) Lungenkrebs herbeizuführen, und ausreichende Bewegung hat das Potenzial, sich unter entsprechenden Bedingungen (z.B. bei gleichzeitiger gesunder Ernährung, bei Abwesenheit von gewichtsbelastenden Krankheiten) positiv auf das Körpergewicht und darüber hinaus, wieder unter entsprechenden anderen Bedingungen, auf die Gesundheit auszuwirken. Ursache-Wirkungszusammenhänge dürfen in komplexen Systemen nicht als isolierte und lineare Kausalketten aufgefasst werden, sondern müssen im Sinne von vielschichtig verwobenen Netzen von sich gegenseitig bedingenden Faktoren begriffen werden, als Systeme, deren Elemente untereinander und mit ihrer Umwelt⁹¹ in vielfältigen (u.a. kausalen) Beziehungen stehen. Faktoren, die gewissermassen von aussen auf komplexe Systeme einwirken, tun dies im Sinne von Irritationen und Perturbationen und stossen so Selbstorganisationsprozesse an. Eine deterministische Ursache-Wirkungsbeziehung über die Systemgrenzen hinaus, verbietet das Prinzip der operativen Geschlossenheit.

Methodische Folge daraus ist, dass komplexe Systeme nicht künstlich in einzelne Kausalketten und unabhängige und abhängige Variablen⁹² aufgeteilt werden sollten, sondern dass es darum gehen muss, den Funktionszusammenhang eines Systems zu erfassen (Willke, 2005). Es wird gezeigt werden, dass es hier keine einfachen Lösungen und kein Schwarz-weiss gibt, sondern dass es eine gros-

⁸⁹ Vgl. Mc Queen et al. (2007), p. 34: „This does not imply, that causality cannot be understood, but that the understanding lies more with the interpreter of causality than with the events being described“.

⁹⁰ „With non-linearity there is no consistent relationship between causes and effects. The same ‘cause’ can in specific circumstances produce quite different kinds of effect. Minor changes in the past can produce potentially large effects in the present since small events are not ‘forgotten’“ (Gleick, 1988, zit in Urry (2005), 238).

⁹¹ Die Theorie sozialer Systeme sieht vor, dass Kausalitäten Systemgrenzen passieren können (Luhmann (1987), 52) und zwar im Sinne von Irritationen, welche Selbstorganisationsprozesse auslösen können.

⁹² Die Beziehung zwischen unabhängigen und abhängigen Variablen ist nur dann in diese Richtung kausal, wenn ein Vertauschen der unabhängigen (Wenn-Teil, Ursache) und der abhängigen Variable (Dann-Teil, Wirkung) sprachlich und inhaltlich nicht sinnvoll ist (Bortz and Dörig (2003), 16).

se methodische Herausforderung darstellt, im Bereich zwischen der stark reduktionistischen Analyse einzelner Kausalzusammenhänge und dem Verstehen eines nie gänzlich erfassbaren Gesamtsystems, nach nützlichen Informationen zu suchen, um entsprechende Entscheidungen fällen zu können. Und dies gründet immer auf begrenzter Rationalität: „Entscheidungsverhalten findet unter den Bedingungen von Komplexität statt. Ihre nichtlineare Kausalität erlaubt nur beschränkte Rationalität.“ (Mainzer, n.d., 33)

Die Frage des Nachweises von Wirksamkeit steht und fällt mit der Zuverlässigkeit von Kausalitätsaussagen. Von einer wirksamen Intervention wird dann gesprochen, wenn die Intervention die gewünschten Wirkungen herbeiführt⁹³. Interventionen, respektive einzelne Massnahmen, deren Wirksamkeit überprüft werden sollen, gehen zunächst als Ursachen in entsprechende Modelle ein. Die Komplexität sozialer Systeme führt allerdings, wie wir gesehen haben dazu, dass nicht von einzelnen Ursache-Wirkungsketten gesprochen werden kann sondern vielmehr von einem Netz an untereinander bedingenden Faktoren. Da es sich bei Kausalitätsbeziehungen in komplexen sozialen Systemen also stets um von Beobachtern konstruierte Modelle mit starker Komplexitätsreduktion handelt, ist bei einer Wirksamkeitsbeurteilung folgendes zu beachten:

- Systemische Sichtweise: Wirkungen werden durch ein Zusammenspiel vielfältiger Faktoren verursacht. Wirkungsevaluation muss versuchen, dieses Zusammenspiel zu rekonstruieren anstatt sich auf einzelne Ursache-Wirkungszusammenhänge und isolierte Wirkungsketten zu versteifen. Diese Sichtweise versteht Ursache-Wirkungsbeziehungen als nicht-linear, multidimensional und zirkulär. Die funktionale Analyse kann dazu genutzt werden, die Ursachen einzelner Wirkungen und die Wirkungen einzelner Ursachen differenziert zu untersuchen.
- Mechanismen: Interaktionen in sozialen Systemen können zwar unendlich vielfältig sein, sie entstehen aber nicht zufällig, sondern sind gewissen Gesetzen unterworfen und bilden entsprechende emergente Ordnungsstrukturen aus. Wirkungsevaluation muss versuchen, die ein System und seine Umweltbeziehungen bestimmenden Regeln und Mechanismen auf der Makroebene aufzudecken.
- Nebenwirkungen: Es ist davon auszugehen, dass Interventionen immer auch nicht-beabsichtigte und in den Modellen nicht-berücksichtigte Wirkungen entfalten. Wirkungsevaluation darf sich nicht auf die Untersuchung von Zielerreichung und auf absehbare Wirkungen beschränken, sondern muss explorativ nach unvorhergesehenen (Neben-) Wirkungen fragen.
- Wirksamkeit: Die Frage, ob ein Projekt oder ein Programm wirksam ist oder nicht, ist eine verkürzte, wenn nicht falsche Frage für Wirkungsevaluationen in komplexen Systemen. Vielmehr geht es darum zu fragen, welche Wirkungen unter welchen Bedingungen wie zustande kommen und welche Bedingungen mit hohem Wirkungspotential sind. Nicht *das* Projekt oder *das* Programm wirkt, sondern Wirkung entsteht erst in Interaktion zwischen den verschiedenen beteiligten Systemen.

Einen interessanten aber nicht unproblematischen Ansatz zur Wirksamkeitsbeurteilung von Interventionen im Lichte divergierender Kontexte schlagen Hawe et al. (2004) vor. Anstatt die Wirksamkeit einzelner standardisierter Projekt- oder Programmelemente zu untersuchen, soll die Wirksamkeit von Funktionen, welche unterschiedliche Formen annehmen können, geprüft werden. Die Funktion kann sein, Patientinnen oder Patienten situationsgerecht über Krankheiten zu informieren. Die Form kann je nach Situation variieren, wird also kontextspezifisch massgeschneidert. Auf diese Art geht dann beispielsweise die Massnahme „massgeschneiderte Information“ als unabhängige Variable in eine randomisierte kontrollierte Studie ein. Diese Lösung besticht auf den ersten Blick. Problematisch ist sie hingegen in Bezug auf die Aussagekraft der Ergebnisse. Nicht die Wirkung einer bestimmten Mass-

⁹³Gesonderter Reflexion bedarf die Frage, wer konkret die „erwünschten Wirkungen“ definiert. Wird dies aus Sicht der Politik, der Trägerschaft oder der Leitung von Projekten und Programmen definiert und inwiefern fliessen die Bedürfnisse der Betroffenen in diese Definition mit ein? In Gesundheitsförderungsprogrammen, die mit den Ansprüchen der Gesundheitsförderung ernst machen, hat die Perspektive der Betroffenen höchste Priorität.

nahme wird bestätigt, sondern die Wirkung unterschiedlichster Massnahmen in unterschiedlichsten Kontexten. Bei der Übertragung der Ergebnisse auf neue Kontexte stellt sich dann die Frage, was konkret übertragen werden kann. Hier muss auf den kontextspezifischen Einzelfall zurückgegriffen werden, über dessen Wirksamkeit allerdings keine verallgemeinerte Aussage vorliegt. Wir wissen bestenfalls, dass massgeschneiderte Information wirkungsvoll ist, wir wissen welche Information in welchem Kontext angewendet wurde, wir wissen auch, ob eine massgeschneiderte Information im Einzelfall wirksam war, wir haben aber keine statistisch aussagekräftige Aussage über die Wirksamkeit einer bestimmten massgeschneiderten Information. Der Ansatz von Hawe et al. liefert auf unterschiedlichen Ebenen wertvolle Informationen, die für eine Wirksamkeitsbeurteilung im Hinblick auf einen Transfer von Massnahmen von Interesse sind. Er zeigt aber auch das Dilemma zwischen Generalisierung und Einzelfall auf, in welchem sich die Sozialwissenschaften bewegen und das mit den Diskussionen rund um Komplexität verstärkt ins Bewusstsein gerufen wird.

5.5 Generalisierung von Wirkungen

Neben der Frage der Wirkungen ist die Frage der Generalisierbarkeit von Aussagen, respektive die Frage der externen Validität von Ergebnissen für eine übergeordnete Beurteilung der Wirksamkeit von Interventionen von zentraler Bedeutung. Der Prozess der Generalisierung lässt sich wie folgt abbilden, wobei die induktive Herleitung von Generalisierungen sowohl für theoretische als auch für statistische Generalisierung gilt⁹⁴:

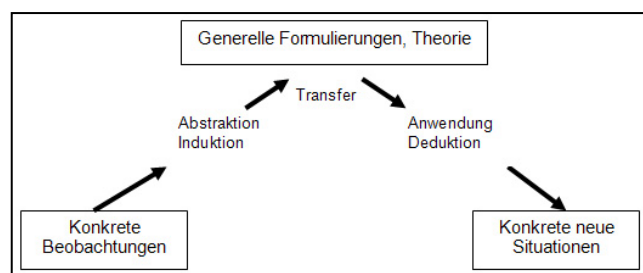


Abb. 9 Der Prozess der Generalisierung
(Mayring, 2007)

Generalisierung wird immer dann zur Notwendigkeit, wenn Aussagen gemacht werden sollen, die über die einzelnen Beobachtungen (Einzelfälle, statistische Stichproben) hinausreichen. Und dies ist bei Wirksamkeitsbeurteilungen in der Regel der Anspruch. Von Wirksamkeitsevaluationen wird erwartet, dass sie die grundsätzliche Wirksamkeit einer Intervention im Hinblick auf eine Übertragbarkeit und Multiplikation der Intervention belegen. Bezüglich Generalisierung führt die Komplexität sozialer Systeme allerdings in schwieriges Gelände: da jedes soziale System grundsätzlich einzigartig ist und sich in unterschiedlichen Kontexten unterschiedlich entfaltet, sind verallgemeinerte Aussagen prinzipiell problematisch, und es kann nicht gesagt werden, ob eine Intervention oder ein Projekt der Gesundheitsförderung generell wirksam ist - es kommt darauf an. Mit Blick auf die Evidenzdiskussion, die sich an der evidenzbasierten Medizin und ihrer experimentellen Methodik orientiert, muss aber relativiert werden. Wirksamkeitspostulate, die auf experimentellen Designs gründen, sind auch immer probabilistische Aussagen. Es sind Aussagen über den Durchschnitt der beobachteten Fälle (welche auf eine

⁹⁴ Theoretische bzw. analytische Generalisierung meint Verallgemeinerungen auf Basis von Erkenntnissen aus Einzelfällen, statistische Generalisierung meint die Verallgemeinerung von Aussagen einer statistischen Stichprobe auf eine Grundgesamtheit.

Grundgesamtheit hochgerechnet werden), die zwar Prognosen über Einzelfälle erlauben, nicht aber die Wirksamkeit für einen Einzelfall mit Sicherheit prognostizieren können. So kann nach einer experimentellen Wirksamkeitsstudie die Wirksamkeit eines Medikamentes zwar bejaht werden, das Medikament wirkt aber nicht bei allen Patientinnen und Patienten gleich und bei manchen in der Regel auch gar nicht. Ob das Medikament im Einzelfall ganz konkret Wirkung zeigt, und welche, kann nur mit einer gewissen Wahrscheinlichkeit vorausgesagt werden. Dasselbe gilt für Nebenwirkungen.

Die Idee der Generalisierung steht in engem Zusammenhang mit der Entdeckung von universellen Gesetzen in den Naturwissenschaften. Da die Existenz universeller Gesetze in den Sozialwissenschaften stark angezweifelt wird (Mayring, 2007), und derartige Gesetze auch in den Naturwissenschaften stark relativiert werden mussten, gleichzeitig aber nicht auf Aussagen mit verallgemeinerndem Charakter verzichtet werden soll, bietet sich eine differenziertere Betrachtungsweise an: Fragen der Generalisierung lassen sich in einem Kontinuum zwischen nomothetischen Aussagen (allgemeingültige Gesetze) und idiographischen Aussagen (Aussagen über Einzelfälle)⁹⁵ abbilden und es geht darum, sowohl nach Verallgemeinerungen zu suchen als auch dem Einzelfall gerecht zu werden. Mayring (2007) beschreibt insgesamt acht Klassen von Generalisierungen auf diesem Kontinuum, die je nach wissenschaftstheoretischer Position, nach Forschungsfeld und nach Forschungsfrage relevant werden können. Von besonderem Interesse sind hier vor allem kontextspezifische Generalisierungen, welche Wissen systemtheoretisch, also vom Kontext abhängig, verstehen. In diese Kategorie fällt beispielsweise Mertons Konzept der Theorien mittlerer Reichweite (Merton, 1968). Gemäss Lincoln und Guba (1985) haben es die Rechtswissenschaften vorbildlich geschafft, mit dem Spannungsfeld zwischen allgemein gültigen Gesetzen und dem Einzelfall umzugehen⁹⁶. Aufgrund der grundsätzlichen Verschiedenheit der Einzelfälle plädieren Lincoln und Guba allerdings dafür, das Konzept der Generalisierung in den Sozialwissenschaften gänzlich fallen zu lassen. Generalisierte Aussagen seien aufgrund der Einzigartigkeit der Phänomene nicht möglich, Aussagen mit Verallgemeinerungscharakter könnten lediglich den Charakter von Arbeitshypothesen haben, nicht aber den Charakter von Schlüssen:

Local conditions, in short, make it impossible to generalize. If there is a „true“ generalization, it is that there can be no generalization. And note that the „working hypothesis“ are tentative both for the situation in which they are first uncovered and for other situations; there are always differences in context from situation to situation, and even the single situation differs over time. (Lincoln & Guba, 1985, 124)

Vor dem Hintergrund der Komplexität sozialer Systeme ist diese Argumentation nachvollziehbar. Generalisierungen im Sinne allgemeingültiger Gesetze haben in den Sozialwissenschaften keinen Platz. Das Konzept der Generalisierung deshalb gänzlich zu verwerfen und es durch den Begriff der Arbeitshypothese zu ersetzen, erscheint aber keine notwendige Konsequenz, auch deshalb nicht, weil es auch unter dem Begriff der Arbeitshypothese letztlich um die Formulierung von verallgemeinerten Aussagen - einfach mit provisorischem Charakter - geht⁹⁷. Vielmehr scheint es wichtig, Generalisierungen in den Sozialwissenschaften prinzipiell als bedingte Aussagen zu verstehen, und zwar sowohl bei statistischen als auch bei theoretischen Generalisierungen. Damit ist auch Poppers Einwand gegen Generalisierung durch induktive Theorienbildung⁹⁸ begegnet. Die Herausforderung der Generalisierung von wissenschaftlichen Aussagen ist Gegenstand breiter Diskussionen, (Lincoln & Guba,

⁹⁵ Die Kategorisierung geht zurück auf Wilhelm Windelband und wird von Lincoln und Guba aufgegriffen (Lincoln and Guba (1985), 116).

⁹⁶ „Perhaps the professional group that has dealt best with this problem is the law, built largely on precedent cases (case law) that are powerful precisely because they take particulars into account.“ (Lincoln and Guba (1985), 117)

⁹⁷ Generalisierung scheint ein Kernbereich wissenschaftlicher Forschungsarbeit zu sein (Mayring (2007)).

⁹⁸ Zum Induktionsproblem aus Sicht Karl Poppers siehe Caspari (2012).

1985; Mayring, 2007; Kolberg, n.d.; Becker, 1990). In diesem Zusammenhang sollen einige Aspekte reflektiert werden, die sich spezifisch und ergänzend aus der Komplexität sozialer Systeme ergeben.

Die Nichtlinearität von Ursache-Wirkungsbeziehungen in komplexen Systemen zeigt die Anfälligkeit von Generalisierungen. Wenn selbst kleine Unterschiede in den Anfangsparametern oder im Verlaufe einer Intervention zu grossen Veränderungen über die Zeit führen können, wird die Aussagekraft von verallgemeinernden Aussagen stark relativiert. Bei einem Transfer einer Intervention in einen anderen Kontext ist davon auszugehen, dass sich die Kontexte in derart kleinen (oder grösseren) aber bedeutenden Punkten unterscheiden können. Dies kann nicht Anlass sein, gänzlich auf entsprechende Verallgemeinerungen oder Prognosen zu verzichten. Es bedeutet aber, dass selbst dann, wenn eine vergleichende Kontextanalyse vorgenommen wird, welche die Passung („fittingness“) der Kontexte untersucht (Lincoln & Guba, 1985), stets mit Unvorhersehbarem, Überraschendem gerechnet werden muss. Auch wenn sich die Kontexte nicht wesentlich unterscheiden, ist davon auszugehen, dass in der Interaktion zwischen Kontext und Projekt entscheidende Unterschiede entstehen. So kann beispielsweise ein Tabakpräventionsprojekt in einer ländlichen Schule in Bezug auf Wirkungen sehr erfolgreich sein, in einer vergleichbaren ländlichen Schule kann aufgrund einer früheren gescheiterten Intervention, die möglicherweise nur zwei einzelnen Lehrpersonen noch bewusst ist, eine Dynamik des Widerstandes ausgelöst werden, die das Projekt scheitern lässt. Bei der Generalisierung von Erkenntnissen über Verhaltensweisen komplexer Systeme ist also grundsätzlich Vorsicht und Bescheidenheit angesagt. Die Wahrscheinlichkeit ist gross, dass es sich bei einem, in einem Einzelfall beobachteten, Phänomen tatsächlich um ein Einzelphänomen handelt, das so nicht in anderen, auch nicht in ähnlichen Systemen vorkommen muss. Die Beobachtung von mehreren Einzelfällen z.B. durch die Strategie des ‚theoretical sampling‘ kann hier Verallgemeinerungen besser abstützen, idealerweise bis hin zu einer Sättigung der Evidenz (Mayring, 2007). Dabei können zwei Strategien zur Anwendung kommen: die Berücksichtigung von mehreren möglichst gleichen Fällen oder aber die Berücksichtigung von möglichst verschiedenen Fällen, in denen ein Phänomen beobachtet wird. Gerade letztere Strategie wird als besonders vielversprechend zur Testung von Hypothesen eingeschätzt: „If the assumed association between variables still hold under conditions different from the original ones (i.e. in a most different system), our belief in the association is strengthened.“ (Kolberg, n.d., 15)

Die Zirkularität von Ursache-Wirkungsbeziehungen und die adaptiven Fähigkeiten von sozialen Systemen führen überdies dazu, dass sich Projekte und Programme während ihrer Umsetzung laufend verändern, so dass auch bei standardisierten Anfangsbedingungen bald nicht mehr gesagt werden kann, auf was sich die Verallgemeinerung bezieht. Die grosse Wandelbarkeit von untersuchten Systemen (z.B. Interventionen) und ihren Kontexten ist insbesondere auch für experimentelle Studien ein Fallstrick: es können immer nur Anfangsbedingungen randomisiert werden. Wenn sich die Kontexte während der Untersuchung weiterentwickeln, tun sie dies nicht notwendigerweise gleichermassen in der Interventions- und der Kontrollgruppe, die zufällige Zuteilung von Kontextunterschieden geht verloren, und die Veränderungen der untersuchten Interventionen gefährden die interne Validität der Ergebnisse. Die Kontextbedingtheit von kausalen Zusammenhängen ist das Haupthindernis für Generalisierungen schlechthin, da einerseits verschiedene Kontextfaktoren das Systemverhalten mit beeinflussen und andererseits Veränderungen in komplexen Systemen erst durch das Zusammenspiel verschiedenster Faktoren emergent zustande kommen. Werden trotzdem Generalisierungen vorgenommen, dann sind diese immer mit Einschränkungen verbunden: erstens handelt es sich immer um Komplexitätsreduktionen, welche gewisse Kontextbedingungen in den Vordergrund rücken, die Verallgemeinerungen sind also idealtypischer Art. Zweitens handelt es sich bei Generalisierungen in den Sozialwissenschaften – und nicht nur dort - immer um vorläufige Aussagen, welche den aktuellen Stand des Wissens repräsentieren, also nie um abschliessende, definitive Erkenntnisse.

5.6 Fazit

Die Entdeckung der Komplexität hat die Wissenschaften herausgefordert, mit tiefgreifenden erkenntnistheoretischen Folgen. Die neuen Erkenntnisse widersprachen dem seit Newton vorherrschenden Bild einer vollständig determinierten Welt, welche über eine objektive Analyse ihrer Teilelemente voraussagbar und letztlich auch steuerbar sei. Komplexität schien einem solchen Verständnis diametral entgegenzulaufen und soziale Systeme verhielten sich oft so, wie es eigentlich gar nicht hätte sein dürfen. Die Erkenntnisse der Komplexitätstheorie haben zu einer Abkehr deterministischen, mechanistischen und reduktionistischen Denkens geführt und Komplexität wird vielerorts als neues Paradigma ausgerufen. Nicht nur in den Sozialwissenschaften, sondern auch in der Physik haben komplexitäts- und systemtheoretische Erkenntnisse gezeigt, dass beobachterunabhängige Erkenntnis eine Illusion ist. Kausalitäten, Wirksamkeiten und Generalisierungen sind Konstruktionen, die gemäss systemtheoretisch-konstruktivistischer Sicht sozialen Systemen zugerechnet werden können. Erkenntnis ist kontext-, zeit-, oder wie Fleck (1980) es ausdrückt, denkstilgebunden. Es gibt dabei immer mehrere Möglichkeiten, wie soziale Systeme oder Kausalitäten beschrieben werden und es stellt sich aus konstruktivistischer Sicht nicht die Frage, welche dieser Konstruktionen der Wahrheit entspricht, sondern welche dieser Konstruktionen sich in der Bewältigung der komplexen Herausforderungen am besten bewährt. Für die Evaluation von sozialen Interventionen bedeutet dies, dass Wirksamkeit nicht objektiv feststellbar und nachweisbar ist. Evaluation kann Modelle bzw. Konstruktionen über Wirkungszusammenhänge entwerfen, diese müssen dem Gegenstand allerdings angemessen sein und müssen sich in Auseinandersetzung mit der Realität bewähren. In komplexen Zusammenhängen ist naheliegend, dass diese Modelle entsprechend komplex sein müssen. Sollen sie die Beurteilung von dynamischen Prozessen ermöglichen, so müssen sie Dynamik erfassen können und sollen sie unvorhersehbare Entwicklungen berücksichtigen, so müssen sie eine entsprechende Offenheit vorsehen. Die Komplexität sozialer Systeme ist allerdings derart hoch, dass sie nie vollständig empirisch erfasst werden kann. Bei empirischen Beschreibungen von komplexen Phänomenen handelt es sich immer um Komplexitätsreduktionen und die Kunst besteht darin, einen angemessenen Umgang mit Komplexität und Komplexitätsreduktion zu finden. Zu wenig Reduktion birgt die Gefahr, den Überblick zu verlieren und nicht handlungsfähig zu sein, und zu starke Reduktion läuft Gefahr, Wesentliches zu übersehen und unangemessene Schlüsse zu ziehen. Evaluationen, welche komplexe Interventionen beurteilen sollen, müssen ihren Umgang mit dieser Komplexität reflektieren und begründen. Einfache Wirkungsmodelle scheinen auf den ersten Blick nicht geeignet, um komplexe Ursache-Wirkungsbeziehungen abzubilden und um handlungsleitend zu sein, hoch komplexe Wirkungsmodelle allerdings, in denen alle Faktoren mit allen anderen in unterschiedlichsten Dimensionen auf vielfältige Weise verbunden sind, sind dies ebenso wenig und führen eher zu Ohnmacht als zu proaktivem Handeln. Erkenntnistheoretisch gilt es, sich der Konstruktion, Begrenztheit und Vorläufigkeit von Erkenntnis rund um komplexe Systeme stets bewusst zu sein und der Reflexion dieser Erkenntnisbedingungen einen hohen Stellenwert einzuräumen.

Auf methodologischer Ebene geht es nicht darum, die eine richtige Methode zu finden, um Wirksamkeit von sozialen Interventionen beurteilen zu können. Im Gegenteil, methodischer Pluralismus ist die Grundlage, komplexen Sachverhalten angemessen begegnen zu können. Während standardisierte Methoden Komplexität sehr stark reduzieren, ein relativ oberflächliches Bild mit meist groben Kategorien zeichnen, dadurch aber wichtige statistische Aussagen ermöglichen, haben qualitative Methoden das Potenzial, in die Tiefe zu gehen, diese groben Kategorien zu hinterfragen und statistische Aussagen in ihrem Kontext zu vertiefen und zu interpretieren. Während sich die Gütekriterien quantitativer Forschung für die Prüfung quantitativer Tests durchaus bewähren, sind sie als Richtschnur für die Erfassung komplexer Sachverhalte ungeeignet. Quantitative Methoden haben einen entsprechend eingegrenzten Anwendungsbereich und sollten zur Beurteilung komplexer Interventionen nie ausschliesslich verwendet werden. Sie eignen sich für die Beurteilung von Einzelaspekten mit begrenzter, stark reduzierter Komplexität. Wirksamkeitsevaluationen bewegen sich zwischen (quasi-) experimentellen Studien, welche die Wirksamkeit einzelner isolierter Massnahmen nachweisen wollen und quali-

tativen Einzelfallstudien, welche Phänomene in ihrer Komplexität und Tiefe ergründen wollen. Es gibt unterschiedliche Modelle, sich in diesem Kontinuum zu bewegen und entsprechende Vorschläge werden im nächsten Kapitel ausführlich behandelt. Grundsätzlich wird das gesamte Methodenarsenal sozialwissenschaftlicher Forschung dazu genutzt, zu generalisierten Aussagen über Wirkungen zu gelangen. Vor dem Hintergrund der Komplexität sozialer Systeme muss eine Priorisierung gewisser Methoden als Goldstandard abgelehnt werden. Verschiedene Ansätze und Designs haben unterschiedliche Stärken und Aussagekraft in Bezug auf Kausalität und Generalisierungen. Werden standardisierte, quantitativ orientierte Ansätze mit qualitativen Ansätzen kombiniert, stellt sich allerdings die Frage nach der Kompatibilität der sie leitenden Paradigmen.

Bevor im empirischen Teil einzelne Evaluationen dahingehend untersucht werden, wie sie zu Wirkungsbeurteilungen kommen und wie sie dabei mit der Komplexität sozialer Systeme umgehen, soll zusammenfassend dargestellt werden, wie die einzelnen Merkmale komplexer sozialer Systeme im Kontext der Wirkungsentstehung zusammenhängen und was sie für die Wirkungsevaluation bedeuten können:

Merkmale komplexer sozialer Systeme

Interaktionsvielfalt

- Konnektivität und Interdependenz
- Offenheit



Komplexe Wirkungsmechanismen

- Feedback
- Multidimensionalität und Nicht-Linearität
- Dynamik



Selbstorganisation

- Adaption und Co-Evolution
- Reflexivität
- Selbstorganisation/ Operative Geschlossenheit



Emergenz

- Emergenz
- Unvorhersehbarkeit/ Ungewissheit

Interventionswirkungen

In Interventionssystemen spielen hinsichtlich einer zu bewältigenden Herausforderung vielfältigste endogene und exogene, bekannte und auch nicht bekannte Wirkfaktoren zusammen und bilden damit die Basis für komplexe Wirkungsmechanismen

Interventionssysteme zeichnen sich aus durch dynamische, komplexe Wirkungsmechanismen, welche Veränderungen in den Zielsystemen anstossen bzw. mit verursachen

Interventionswirkungen sind das Ergebnis von Selbstorganisationsprozessen, also Ergebnis einer aktiven Auseinandersetzung der Zielsysteme mit den Veränderungsabsichten der intervenierenden Systeme.

Interventionswirkungen sind zwar planbar, aber nicht mit Bestimmtheit vorhersehbar. Sie sind das Ergebnis komplexer Wirkungsmechanismen und Selbstorganisationsprozesse.

Wirkungsevaluation im Lichte komplexer sozialer

Wie berücksichtigt die Evaluation systeminterne und systemübergreifende Zusammenhänge? Inwiefern wird der Kontextbedingtheit der untersuchten sozialen Systeme Rechnung getragen?

Welches Verständnis von Ursache-Wirkungsbeziehungen liegt der Evaluation zugrunde? In welcher Form und Tiefe werden die komplexen Wirkungsmechanismen analysiert und inwiefern werden dabei unterschiedliche Wirkungsdimensionen und Dynamiken berücksichtigt?

Inwiefern berücksichtigt die Evaluation systemeigene Perspektiven und die Eigenreflexion der untersuchten Ziel- und Interventionssysteme? Inwiefern werden Selbstorganisationsprozesse von Ziel- und Interventionssystemen untersucht und in die Wirkungsbeurteilung aufgenommen?

Inwiefern werden in der Evaluation emergente Systemeigenschaften der untersuchten Settings berücksichtigt? Ist die Evaluation so angelegt, dass Unvorhersehbares, Ungeplantes und Nebenwirkungen sichtbar werden?

Abb. 10 Merkmale komplexer sozialer Systeme und Fragen an die Wirkungsevaluation

6 Methodik der empirischen Untersuchung

Nachdem in einer theoretischen Auseinandersetzung die Evidenzdebatte in der Gesundheitsförderung erläutert, die Wesensmerkmale komplexer sozialer Systeme herausgearbeitet und erkenntnistheoretische und forschungsmethodische Positionen reflektiert wurden, sollen in einem nächsten Schritt Evaluationen der Gesundheitsförderung und Prävention dahingehend untersucht werden, wie sie konkret mit der Komplexität der untersuchten Interventionen umgehen. Es soll die Frage beantwortet werden, wie Evaluationen mit dem Spannungsfeld umgehen, es einerseits mit sehr komplexen sozialen Systemen zu tun zu haben und andererseits diese Komplexität nie wirklich fassen zu können. In diesem Kapitel wird die Methodik der empirischen Untersuchung dargelegt und begründet, es werden der Analyseprozess beschrieben und das Potenzial und die Grenzen der Untersuchung reflektiert.

6.1 Wahl und Begründung der Analysemethodik zur Rekonstruktion der Evaluationspraxis

Die empirische Analyse untersucht Evaluationspraxis hinsichtlich ihres Umgangs mit Komplexität. Es interessiert, inwiefern die Evaluation die Komplexität der von ihr untersuchten Interventionen wahrnimmt und wie sie darauf reagiert. Wie kommt die Evaluation zu Wirkungsaussagen und wie geht sie dabei mit der Komplexität sozialer Systeme um? Da in der Studie der Umgang der Evaluation mit Komplexität zwar analysiert, aber nicht bewertet wird, handelt es sich nicht um eine Meta-Evaluation im eigentlichen Sinne (Widmer, 2006). Als Datenquelle für die Untersuchung dienen Schlussberichte von Evaluationen, welche die Wirkung von Interventionen der Gesundheitsförderung und Prävention untersuchen. Die Analyse wurde mit MAXQDA (Version 11), einer Software für qualitative Datenanalyse durchgeführt⁹⁹.

Evaluationsberichte erscheinen auf den ersten Blick eine geradezu ideale Datenquelle, um den Umgang von Evaluation mit Komplexität zu untersuchen. In Evaluationsberichten muss explizit dargelegt werden, welche Aspekte in den Fokus der Evaluation gelangen und wie diese bearbeitet werden. Es muss für die Leserschaft nachvollziehbar sein, wie die Evaluation zu einer Wirkungsbeurteilung kommt, welche Aspekte mit welcher Begründung in eine derartige Beurteilung eingeflossen sind und welche nicht. Sowohl die Evaluationsstandards nationaler Evaluationsgesellschaften (z.B. (American Evaluation Association AEA, 2004); (DeGEval, 2008); (Widmer, Landert, & Bachmann, 2000)) als auch die Gütekriterien wissenschaftlicher Forschung verlangen eine nachvollziehbare Darlegung der gewählten Untersuchungsmethodik und eine transparente Herleitung der Ergebnisse. Zudem wird in der Regel dargelegt, welchem Zweck die Evaluation dient und welche Interessen und Erwartungen damit verbunden sind. Dies sind wichtige Kontextinformationen, welche helfen, die Texte richtig einordnen und besser verstehen zu können (Charmaz, 2006).

Evaluationsberichte sind meist entlang einer etablierten Struktur aufgebaut, um die Nachvollziehbarkeit der Ergebnisse zu gewährleisten. Dazu gehört die Beschreibung der Ausgangslage und des Auftrags, der Ziele der Evaluation und des Evaluationsgegenstandes, der Methodik sowie der Ergebnisse und Empfehlungen. In der Regel werden die Ergebnisse in einer Zusammenfassung verdichtet. Evaluationsberichte werden meist mit einem Anhang abgeschlossen, welcher z.B. das Literaturverzeichnis, Datentabellen, Detailanalysen, Erhebungsinstrumente, aber auch Personenprofile des Evaluationsteams enthalten.

Evaluationsberichte sind in der Regel in einer Fachsprache geschrieben, die für ein nicht akademisch gebildetes Fachpublikum verständlich und nachvollziehbar sein soll. Gleichzeitig muss ein Evaluati-

⁹⁹ Vgl. www.maxqda.com.

onsbericht wissenschaftlichen Gütekriterien genügen. Zusammenfassungen sind oft in einer auch für Laien verständlichen Sprache abgefasst, um auch von Entscheidungsträgern verstanden werden zu können, die keine entsprechende Fachexpertise mitbringen.

Im Gegensatz zu Interviews, die mit Evaluationsfachpersonen geführt werden könnten und ebenso zusammenfassend sein würden, haben Evaluationsberichte den Vorteil, dass die Methodik sehr bewusst und systematisch beschrieben ist. Dies bringt auch Gefahren mit sich, die bei einer Interpretation der Ergebnisse mit berücksichtigt werden müssen: Bei Evaluationsberichten handelt es sich um bewusst abgefasste, systematische Darstellungen, die zu einem gewissen Teil strategisch gelenkt sein dürften. Evaluationsberichte dienen auch dazu, dem Auftraggeber Professionalität zu vermitteln, sich als attraktiver Auftragnehmer zu positionieren, die eigene Reputation zu stärken und sich Türen für mögliche Folgeaufträge offen zu halten. In diesem Zusammenhang sind Evaluationsberichte der Gefahr der Beschönigung von Evaluationsergebnissen ausgesetzt, insbesondere dann, wenn der Auftraggeber der Evaluation selber Träger der untersuchten Intervention ist (Eisner & Ribeaud, 2007). Insgesamt stehen Evaluationsberichte im Spannungsfeld zwischen wissenschaftlichen Gütekriterien, denen sie genügen müssen, und Erwartungen von Auftraggebern.

Für die vorliegende Arbeit wird in Anlehnung an die Grounded Theory Strauss'scher Prägung ein iteratives, qualitativ orientiertes Analyseverfahren gewählt, welches auf den Dreiklang von Induktion, Deduktion und Verifikation setzt, wobei Verifikation im Sinne einer Überprüfung der Plausibilität der an der Empirie entwickelten Erkenntnisse zu verstehen ist (Strübing, 2008). Im Gegensatz zu Glasers objektivistisch konzipierter Version der Grounded Theory, welche davon ausgeht, dass die Theorie bereits in den empirischen Daten liegt und mit richtig angewandter Methodik lediglich entdeckt resp. rekonstruiert zu werden braucht, konzipiert Strauss Grounded Theory als soziale Konstruktion auf Basis der Empirie, aber mit grossem Einfluss der Forschenden und ihrem Vorwissen (Strübing, 2008). Neben verfahrenstechnischen Unterschieden zum Beispiel bezüglich der Anwendung von Kodierfamilien (Glaser) im Gegensatz zu einem inhaltlich offeneren Kodierparadigma (Strauss) ist dies der zentrale Grund für die Anlehnung an Strauss.

Zunächst werden die Evaluationsberichte offen und noch nicht durch vorgegebene Kategorien eingeschränkt analysiert. Es wird angenommen, dass sich dadurch Formen und Details herausarbeiten lassen, die so in der gängigen Evaluationsliteratur nicht diskutiert werden. Komplexität wird in der Evaluationsliteratur erst seit wenigen Jahren verstärkt zur Kenntnis genommen und diskutiert und es gibt erst wenige Arbeiten, welche die Erkenntnisse der Komplexitätstheorie explizit auf das Feld der Evaluation übertragen (z.B. Robert Koch Institut, 2012; Patton, 2011; Forss et al., 2011). Wo es noch viele offene Fragen gibt und der Stand der Diskussion noch nicht allzu differenziert ist, da bieten sich offene, explorative methodische Zugänge an. Mit fortschreitender Analyse, je mehr sich also die Konzepte herauskristallisieren und wieder an die Texte herangelegt werden, desto stärker erhält die Analyse auch deduktiven Charakter. Die Wesenszüge der gewählten qualitativen Inhaltsanalyse lassen sich in Anlehnung an Corbin und Strauss (1990) wie folgt beschreiben:

Erkenntnisgewinn durch einen iterativen Prozess der komparativen Analyse

Um Unterschiede wie auch Ähnlichkeiten im Umgang mit Komplexität in den einzelnen Evaluationen erkennen und auf dieser Grundlage Besonderheiten und Typologien herausarbeiten zu können (Strübing, 2008), werden die Textstellen im Laufe der Kodierung ständigen Vergleichen ausgesetzt. Diese Vergleiche sind eine Voraussetzung für gehaltvolle konzeptuelle Kategorien, indem neue Textstellen mit den sich entwickelnden konzeptuellen Kategorien ständig abgeglichen werden, um das Kategoriensystem laufend weiter zu entwickeln und zu verfeinern (Glaser & Strauss, 2010). Ein zentrales Element in diesem Analyseprozess ist das kontinuierliche Erstellen von Memos. Mittels dieser Memos, welche den analytischen Prozess dokumentieren, werden die konzeptuellen Kategorien Schritt für Schritt entwickelt und verfeinert (Charmaz, 2006). Dimensionen und spezifische Merkmale einzelner Kategorien werden hier ebenso festgehalten wie sich abzeichnende Beziehungen zwischen einzelnen

Kategorien und Subkategorien. Die Memos nehmen die aus den Daten gewonnen Erkenntnisse auf und sind gleichzeitig Referenz für den weiteren Abgleich mit den Daten. Sie sind auch der Ort, an welchem kreatives Denken seinen Niederschlag findet. Hier werden Ideen festgehalten, offene Fragen formuliert und diskutiert und ad hoc- Hypothesen entwickelt, die deduktiv wieder auf die Daten bezogen werden. Memos geben der Analyse und damit der entstehenden Theorie den notwendigen Gehalt und entsprechende Fundierung. Im Laufe der empirischen Analyse hat sich die Form des Memo-Schreibens verändert. Während zu Beginn (vorläufige) Erkenntnisse primär zu einzelnen Kategorien und Subkategorien festgehalten wurden, traten in einer späteren Phase Bezüge zwischen einzelnen Kategorien stärker in den Vordergrund – die Memos werden auf Ebene übergeordneter Kategorien angesiedelt, oftmals in Tabellenform verfasst und entlang des Kodierparadigmas (Strauss & Corbin, 1996) aufgebaut.

Entwicklung von Erkenntnissen mit theoretischer Relevanz

Im Gegensatz zu anderen Verfahren der qualitativen Inhaltsanalyse verfolgt die Grounded Theory grundsätzlich den Anspruch der Theoriegenerierung (Glaser & Strauss, 2010). Durch akribische comparative Analyse und die kausale Verknüpfung von Konzepten und Kategorien sollen nach und nach Erkenntnisse mit theoretischer Relevanz generiert werden. Erst eine Analyse im Sinne von Ursachen und Wirkungen hat das Potenzial für Theoriebildung, indem kausale Aussagen möglich werden in Form von: „Unter diesen Bedingungen passiert das und das, während unter anderen Bedingungen das und das passiert.“ (Strauss & Corbin, 1996, 107). In der vorliegenden Studie stehen die von der Evaluation gewählten Strategien im Umgang mit Komplexität im Mittelpunkt der kausalen Verknüpfungen. In Anlehnung an das Kodierparadigma von Strauss und Corbin (Strauss & Corbin, 1996) werden die entwickelten Kategorien nach und nach zu einem Netz von Phänomenen, Bedingungen, Strategien und Konsequenzen geflochten, wobei dieselben Kategorien je nach Perspektive und Phase des Evaluationsprozesses sowohl Strategien, Bedingungen für andere Strategien, als auch Konsequenzen sein können. So entstehende kausale Bezüge, zunächst als Arbeitshypothesen zu verstehen, wurden durch ständiges Vergleichen mit zusätzlichen Textstellen laufend validiert und gefestigt. Die entwickelten Zusammenhänge wurden mittels verschiedener Diagramme sichtbar gemacht. Da es in der Analyse um die Rekonstruktion des Umgangs der Evaluation mit Komplexität ging, lag der Fokus von Beginn an auf Strategien und die Textstellen in den Evaluationsberichten wurden entsprechend fokussiert kodiert. Dies war notwendig, weil nicht der Inhalt der Evaluationen (z.B. Alkoholprävention oder betriebliche Gesundheitsförderung), sondern die Methodik, das Evaluationshandeln, im Vordergrund der Analyse stand.

Theoretisches Sampling

Das zentrale Kriterium des theoretischen Samplings, wie es u.a. in der Grounded Theory verwendet wird, ist die Selektion resp. die Entwicklung der Stichprobe aufgrund theoretischer Überlegungen bzw. "auf der Basis von Konzepten, die eine bestätigte theoretische Relevanz für die sich entwickelnde Theorie besitzen." (Strauss & Corbin, 1996). Die Stichprobe und die Selektionskriterien für die Erweiterung der Stichprobe entwickeln sich also mit dem Forschungsprozess resp. mit den sich entwickelnden theoretischen Konzepten mit. Entsprechend wird die Stichprobengröße nicht zu Beginn des Forschungsprozesses festgelegt, sondern laufend um neue Evaluationsberichte erweitert, welche neue Erkenntnisse zu den sich entwickelnden Konzepten versprechen. Um die Konzepte nicht einseitig zu halten, sondern möglichst breit abzustützen, wurden gezielt sehr unterschiedliche Fälle in die Stichprobe aufgenommen. Zu Beginn waren die sich entwickelnden Konzepte sehr provisorisch, vage und oberflächlich. Über kontinuierliche Vergleiche und die entsprechenden Reflexionen und Vertiefungen wurden sie immer gehaltvoller, klarer und fundierter. Die einzelnen Konzepte und Kategorien wurden so mehr und mehr gestärkt oder aber sie mussten modifiziert werden. Die Stichprobe wurde so lange um neue Fälle erweitert, bis sich ein Gefühl der „theoretischen Sättigung“ einstellte, bis also neue

Fälle keine wesentlichen neuen Erkenntnisse zu den sich entwickelnden Konzepten und zum sich abzeichnenden Prozess der Wirkungskonstruktion mehr ergaben (Glaser & Strauss, 2010). Eine Übersicht über den Prozess des theoretischen Samplings und die sich aufbauende Stichprobe liefert Kapitel 6.2.

Kodierprozess

Die Grounded theory basiert auf einer möglichst grossen theoretischen Offenheit. Theoretische Konzepte sollen weitgehend aus den Daten selbst herausgearbeitet und nicht bereits an die Texte herangetragen werden. Entsprechend wurde zunächst induktiv an die Texte herangegangen. Die einzelnen Textstellen wurden hinsichtlich der erkenntnisleitenden Frage primär offen kodiert, respektive konzeptualisiert, wobei mit fortschreitender Entwicklung des Kategoriensystems deduktive Elemente zunahmen, indem entwickelte Kategorien als Option für weitere Kodierungen genutzt wurden. Jedem entwickelten Konzept und jeder Kategorie liegt Vorwissen in Form von mehr oder weniger bewussten mentalen Modellen zugrunde. Obwohl teilweise falsch verstanden, geht die Grounded Theory nicht davon aus, jegliches Vorwissen beiseite zu lassen. Die Art und Weise, wie an die Texte herangegangen wird, die Auswahl der kodierten Textstellen und insbesondere die geforderte theoretische Sensibilität (Strauss & Corbin, 1996) bringen kontinuierlich Hintergrundwissen, das u.a. aus der Fachliteratur gewonnen wird, in die Analyse mit ein.

Im Prozess des offenen Kodierens wurden die Evaluationsberichte gründlich untersucht, die Texte aufgebrochen und die einzelnen Textstellen auf Ähnlichkeiten und Unterschiede hin verglichen (Strauss & Corbin, 1996). Für die Textstellen wurden Konzepte gesucht, welche das Gefundene nicht lediglich zusammenfassend beschreiben, sondern auf einer abstrakteren Ebene theoretisch zu fassen versuchen. Einzelne Konzepte wurden im Laufe der fortschreitenden Analyse zu übergreifenden Kategorien gebündelt. In einer nächsten Phase wurde das offene Kodieren um axiales Kodieren erweitert, also um die kontinuierliche Suche nach Verknüpfungen zwischen den entwickelten Kategorien. In dieser Phase diente das Kodierparadigma von Strauss und Corbin (1996) als Rahmenmodell. Ausgehend von der wahrgenommenen Komplexität der Interventionen der Gesundheitsförderung und Prävention (Phänomene) wurde nach Strategien im Umgang mit dieser Komplexität gefragt sowie nach Bedingungen und Konsequenzen der Anwendung dieser Strategien. In einer letzten Phase, der Synthese, wurden die verschiedenen Elemente in Bezug zu einem sich herauskristallisierenden übergeordneten Thema, analog einer Kernkategorie, analysiert (vgl. selektives Kodieren). Hier ging es darum, die verschiedenen Erkenntnisse, die in der Analyse erarbeitet wurden, zu einem dichten Ganzen zusammenzufügen.

6.2 Der Analyseprozess

Die Entwicklung der Stichprobe

Die Evaluationsberichte, die in die Analyse aufgenommen wurden, wurden fortlaufend mit Blick auf den Erkenntnisfortschritt ausgewählt (theoretisches Sampling). Allerdings wurden einige Kriterien festgelegt, welche dem Samplingprozess einen Rahmen geben sollten: mit Blick auf das Erkenntnisinteresse wurden nur Evaluationsschlussberichte von Projekten oder Programmen der Gesundheitsförderung und Prävention aufgenommen, in denen die Wirkungsbeurteilung eine zentrale Rolle spielt. Zudem musste in den Berichten die Methodik, auf welcher die Ergebnisse beruhen, transparent dargestellt sein. Aus praktischen Überlegungen wurden nur Berichte in deutscher Sprache berücksichtigt, um nicht die Analyse durch Übersetzungs- und Interpretationsschwierigkeiten zusätzlich zu erschweren. Bei der Suche nach geeigneten Evaluationsberichten wurde zunächst auf die Schweiz fokussiert. Dies schien angemessen, weil die Schweiz eine vergleichsweise gut entwickelte Evaluationskultur besitzt (Mader, 2009; Widmer & Beywl, 2009) und problemlos eine Vielfalt an Evaluationen aus dem Bereich

der Gesundheitsförderung und Prävention gefunden werden konnte, die sich auf unterschiedliche Art und Weise der Wirkungsbeurteilung annehmen. Die Eingrenzung auf deutschsprachige Evaluationsberichte, die grösstenteils aus der Schweiz stammen, bringt einen Selektionsbias mit sich, der aber insofern vertretbar ist, als es sich bei dieser Studie um ‚case studies‘ handelt und nicht eine statistische Repräsentativität angestrebt wird. Zudem wird keine Verallgemeinerung auf die Evaluationskultur ‚in der Welt‘ angestrebt, sondern die Studie ist im europäischen, deutschsprachigen Kontext zu sehen.

In einer ersten Phase (initiales Sampling (Charmaz, 2006)), welche primär durch offenes Kodieren bestimmt ist, wurden drei sehr unterschiedliche Evaluationsberichte vollständig offen durchkodiert. Nach und nach wurden Evaluationsberichte hinzugenommen, die durch unterschiedliche Evaluationsgegenstände und methodische Zugänge neue Einblicke in die Thematik versprachen (offenes Sampling). Zu Beginn wurden jeweils die gesamten Evaluationsberichte offen kodiert. Das anfängliche Ansinnen, lediglich die Methodenkapitel sowie die Ergebniskapitel zu analysieren, die sich explizit der Evaluation von Wirkungen widmen, wurde wieder verworfen, weil schnell klar wurde, dass sich die Wirkungsbeurteilung nicht von z.B. prozessevaluativen Elementen trennen lässt. Mit fortschreitender Analyse hatte sich dieser Eindruck bestätigt und die Abhängigkeit der Wirkungsbeurteilung von Elementen der Prozessevaluation wurde immer deutlicher sichtbar. Mit Fortschreiten des Analyseprozesses und mit zunehmender Herausarbeitung von theoretischen Konzepten wurde im Sampling immer mehr darauf geachtet, entstehende Konzepte zu validieren (Sampling von Beziehungen und Variationen) (Strauss & Corbin, 1996). Dazu wurden nicht nur neue Elemente aufgenommen, sondern auch nochmals Evaluationen in die Stichprobe integriert, die ähnliche Elemente wie bereits vorhandene enthielten, idealerweise kombiniert mit neuen Aspekten. Es wurden allerdings nicht mehr immer ganze Evaluationsberichte kodiert und ausgewertet, sondern es wurden Kapitel ausgelassen, wenn sie aufgrund ihres Fokus weder Validierungsmöglichkeiten noch neue Erkenntnisse versprachen. Als besonders vielversprechende Themen, die mit fortschreitender Analyse vertieft und weiter fundiert werden, kristallisierten sich die folgenden heraus:

- Spannungsfelder
- Selektion und Grenzziehung (später induktiv versus deduktiv)
- Komplexitätsreduktion
- Berechnen versus Beschreiben
- Wirkungskonstruktion
- Plausibilisierung

Evaluationsberichte, welche diesbezüglich neue Erkenntnisse oder aber Validierungen der bestehenden Kategorien und Beziehungen versprachen, wurden prioritär behandelt.

Die folgende Tabelle gibt einen Überblick über die Stichprobe und den im Sampling gegangenen Weg:

ID	Referenz	Titel	Inhaltliche und methodische Spezifika und Gründe für die Aufnahme in die Stichprobe
E1	(Gmel, Labhart, Venzin, & Marmet, 2010)	Kurzinterventionen bei risikoreich Alkohol konsumierenden Berufs- und Mittelschülerinnen und -schülern im Kanton Zürich. Abschlussbericht.	Im Fokus der Evaluation steht die Wirkung von Gruppenkurzinterventionen der verhaltensorientierten Alkoholprävention, welche mittels einer randomisierten kontrollierten Studie und ergänzenden Gruppen- und Einzelinterviews beurteilt wird.
E2	(Jenny et al., 2011)	Projekt SWiNG – Schlussbericht der Evaluation.	Die untersuchten Interventionen der verhältnisorientierten Stressprävention zielen primär auf strukturelle Veränderungen in den untersuchten Betrieben ab. In dieser umfangreichen Fallstudie in neun Betrieben, welche Komplexität explizit zum Thema macht, werden vielfältige Untersuchungsmethoden kombiniert, u.a. wird eine quantitative Wirksamkeitsanalyse mit einer qualitativ orientierten Beurteilung des Umsetzungsprozesses kombiniert.
E3	(Furrer & Balthasar, 2011)	Evaluation der bfu- Präventionskampagne im Schneesport.	Ergänzend zu den beiden lokal umgesetzten, aber sehr unterschiedlichen Projekten und Evaluationen E1 und E2 wurde als dritter Fall des initialen Samplings die Evaluation einer landesweit umgesetzten Schneesport-Präventionskampagne ausgewählt. Es handelte sich dabei um eine quantitativ angelegte Ein-Gruppen Studie (ohne Kontrollgruppe) mit ebenfalls ergänzenden Interviews.
E4	(Gloor & Meier, 2004)	Evaluationsbericht zum Projekt „MuM – mutige Mädchen – Integration von gewaltbereiten Mädchen im unteren Kleinbasel“.	Die vierte Evaluation wurde der Stichprobe hinzugefügt, da sie ihre Wirkungsbeurteilung (verhaltensorientierte Massnahmen der Gewaltprävention) nicht auf quantitativen Analysen aufbaut, sondern ausschliesslich auf qualitativen Interviews. Zudem stützt sie sich in der Wirkungsbeurteilung voll und ganz auf die Befragten ab. Sie verzichtet auf Vorher-Nachher-Befragungen und Gruppenvergleiche. Dies versprach vollkommen neue Konzepte und Kategorien – von der (induktiven) Grenzziehung bis hin zur Wirkungsbeurteilung selber.
E5	(Gmel & Wicki, 2010)	Effekt der Einschränkung der Erhältlichkeit von Alkohol auf Alkohol-Intoxikationen im Kanton Genf. Abschlussbericht.	Einen ganz anderen Weg geht diese Evaluation. Sie versprach eine Bereicherung, da es sich um reine Verhältnisprävention handelt. Die Wirkungsbeurteilung von Einschränkungen der Verfügbarkeit von Alkohol verzichtet im Gegensatz zu den ersten vier Evaluationen auf eigene Erhebungen und stützt sich stattdessen auf vorhandene Monitoringdaten ab. Es handelt sich um eine Black Box Evaluation: Basis der Wirkungsbeurteilung sind Veränderungen in wenigen Ergebnisindikatoren, Wirkungswege werden nicht analysiert.
E6	(Peters & Wapf, 2006)	Evaluation der Wirkungen lokaler Alkoholpolitiken – eine vergleichende Fallstudienanalyse. Schlussbericht.	Die fortschreitende Analyse hat gezeigt, dass Wirkungsbeurteilungen häufig entweder auf quantitativen Vergleichen oder aber auf qualitativen Beschreibungen beruhen. Diese Evaluation baut die Wirkungsbeurteilung auf qualitativen Vergleichen auf, weshalb sie für die Analyse interessant erschien. Zudem handelt es sich nochmals um einen verhältnispräventiven Ansatz, der aber ganz anders evaluiert wird als E2 oder E5. Es geht um die Wirksamkeit von Alkoholpolitiken in Gemeinden.
E7	(Sommerfeld et al., 2006)	Evaluation „Midnight Basketball“ Projekte.	In der bisherigen Analyse haben sich Wirkungsmodelle als eine wichtige Form des Umgangs mit Komplexität erwiesen. Diese Evaluation versprach einen Erkenntnisgewinn, da sie ein sehr komplexes Wirkungsmodell erarbeitet hat, welches das Interventionsgeschehen und entsprechende Wirkungszusammenhänge sehr detailliert rekonstruiert. Bei der Intervention „Midnight Basketball“ handelt es sich um ein Bündel gesundheitsförderlicher Massnahmen rund um die Öffnung von Turnhallen für Kinder und Jugendliche an Wochenendabenden. Die Sekundäranalyse der Monitoringdaten wurde nur insofern mit berücksichtigt, als sie Hinweise zur Wirkungsbeurteilung lieferte.
E8	(Wieser et al., 2010)	Synthesebericht – Ökonomische Evaluation von Präventions-	Wiederum in eine entgegengesetzte Richtung geht diese Evaluation. Wie die E5 verzichtet auch diese Evaluation auf die Analyse von Wirkungszusammenhängen (also Black Box) und untersucht

		massnahmen in der Schweiz.	die Gesamtwirkung von Präventionsmassnahmen auf einer gesamtgesellschaftlichen Ebene. Es handelt sich um eine retrospektive quantitative Fallstudie mit Vergleichen zu hypothetischen Kontrollsituationen. Thematisch geht es um Verkehrsunfall-, Tabak-, und Alkoholprävention. Die Kosten-Nutzen-Analysen, die auf den Wirkungsanalysen aufbauen, sind für die vorliegende Arbeit nicht von Interesse, entsprechend wurden sie nicht in die Analyse aufgenommen.
E9	(De Bock & Fischer, 2011)	Gesundheitsförderung im Kindergarten. Evaluation des Programms „Komm mit in das gesunde Boot“ der Baden-Württemberg Stiftung in Kindergärten in Baden-Württemberg.	Mit dieser Evaluation sollte nochmals eine randomisierte kontrollierte Studie in die Stichprobe aufgenommen werden, um zu prüfen, ob sich die bisher entwickelten Konzepte und Kategorien bewähren. Es wurde explizit nach einer grösseren clusterrandomisierten Studie gesucht, welche Gesundheitsförderungsinterventionen in Settings untersucht. Da in der Schweiz keine entsprechende Studie gefunden werden konnte, wurde diese deutsche Studie in die Stichprobe aufgenommen.

Tab. 3 Überblick über die untersuchten Evaluationen

Beziehungen des Autors zu den untersuchten Evaluationen

Das Feld der Gesundheitsförderung und Prävention in der Schweiz ist gut überschaubar. Neben dem Bundesamt für Gesundheit ist Gesundheitsförderung Schweiz als nationale Stiftung die zweite grösste Auftraggeberin von Evaluationen der Gesundheitsförderung und Prävention in der Schweiz. Der Evaluationsauftrag ist der Organisation per Gesetz zugeschrieben (KVG Art. 19). Da ich selber seit 2003 für Gesundheitsförderung Schweiz im Bereich Evaluation tätig bin und bis 2012 für den Bereich verantwortlich war, habe ich zu vielen Evaluationen in diesem Feld persönliche Beziehungen. Da ich auf einige interessante Evaluationsansätze in Zusammenhang mit Komplexität nicht verzichten wollte, war es unumgänglich, auch einige Evaluationen in die Untersuchung aufzunehmen, zu denen ein persönlicher Bezug besteht: Bei den Evaluationen von Jenny et al. (2011, E2) und Sommerfeld et al. (2006, E7) habe ich die Auftraggeberin vertreten, in der Evaluation Wieser et al. (2010, E8) war ich Mitglied der wissenschaftlichen Begleitgruppe. Die Nähe zu diesen drei Evaluationen gibt mir auf der einen Seite ein Kontextwissen, welches hilft, die Untersuchungsergebnisse verstehen und einordnen zu können und die Gefahr von Missverständnissen bei nicht eindeutiger Dokumentation zu minimieren. Andererseits besteht bei derartigen Verflechtungen aber auch immer das Risiko des Verlustes kritischer Distanz. Dieses Risiko ist insofern reduziert, als diese Zusammenarbeiten seit längerer Zeit abgeschlossen sind und keine Kontakte mehr zu den betreffenden Evaluierenden bestehen. Zudem wirken der Umstand, dass es sich um sehr unterschiedliche Evaluationen handelt, und die akribisch dokumentierte Analysearbeit, einer unkritischen Behandlung der bekannten Evaluationen entgegen.

Zu den anderen untersuchten Evaluationen bestanden zu keinem Zeitpunkt persönliche Verbindungen.

Die Entwicklung erster Konzepte und Kategorien

In der initialen Kodierphase wurden drei Evaluationsberichte zunächst offen codiert. Leitend war die Frage, wie Evaluation ganz generell mit der Vielschichtigkeit und Komplexität des Evaluationsgegenstandes umgeht, ohne bereits auf Wirkungsfragen zu fokussieren. Diese erste Analysephase förderte einen grossen Reichtum an Konzepten zutage. Insgesamt wurden gegen 2'000 Codes in 180 Seiten Text vergeben. In einem iterativen Strukturierungsprozess bildeten sich aus der induktiv begonnenen Codierung schliesslich neun Oberkategorien heraus, welche zu vier Bereichen zusammengefasst wurden:

Häufig verwendete Konzepte \Rightarrow	Kategorien \Rightarrow	Oberkategorien
Orientierung an Literatur, an theoretischen Modellen, an Interventionszielen, an wissenschaftlichen Standards, am Praxisnutzen, an Annahmen, an Problemen,...	Orientierung im Feld	Komplexität eingrenzen
Mehrere Wirkungsebenen berücksichtigen, viele Einflussfaktoren berücksichtigen, verschiedene Perspektiven einholen, eine hohe Anzahl von Personen einbeziehen, umfangreiche Daten sammeln,...	Auswahl aus der Vielfalt	
Messen, subjektiv einschätzen lassen, direkt nach Kausalitäten fragen, operationalisieren, multimethodisch herangehen, zusammenfassend erfragen, Kategorien im Voraus festlegen, pro- und retrospektiv fragen,...	Komplexität erfassen	Komplexität erfassen
Gruppieren, priorisieren, filtern, gewichten, fokussieren, zusammenfassen, systematisieren, betonen,...	Ordnung schaffen	Komplexität verarbeiten
Verschiedene Gruppen vergleichen, Interventions- mit Kontrollgruppe vergleichen, Vorher-Nachher Vergleiche, grössere Zeiträume beachten, Veränderungen über die Zeit untersuchen, Faktoren einander gegenüberstellen, Korrelationen untersuchen, Entwicklungen darstellen,...	Komplexität verarbeiten	
Grenzen der Erfassung anerkennen, Hinweise auf Uneindeutigkeit/Limiten, Nebenwirkungen beachten, plausibilisieren,...	Umgang mit Uneindeutigkeit	Wirkungen beurteilen
Potenziale untersuchen, Hinweise auf Ursache-Wirkungszusammenhänge, Wirkungsmechanismen untersuchen, Prozesse analysieren, Kontextsensibel, Wechselwirkungen,...	Wirkungsmechanismen-Wirkungspotenziale	
Vermutungen anstellen, Vermutungen überprüfen, Annahmen treffen, Ergebnisse verallgemeinern, Wirkungen in die Zukunft projizieren, Empfehlungen abgeben, Verbesserungen vorschlagen,...	Wirkungen diskutieren	
Spannungsfeld randomisierte kontrollierte Studien - andere Ansätze	Explizite Wirksamkeitsdiskussionen	

Tab. 4 Induktiv sich entwickelndes Kodierschema

Im Prozess der offenen Kodierung wurden laufend Stichworte zu einzelnen Kategorien in Memos festgehalten, so zum Beispiel zu sich abzeichnenden Dimensionen und Ausprägungen aber auch zu sich abzeichnenden Verbindungen zu anderen Kategorien. Es ist nicht weiter erstaunlich, dass sich in diesem Rekonstruktionsprozess evaluativen Handelns viele Aspekte zeigten, die zum Methodenrepertoire der Evaluation oder der empirischen Sozialforschung ganz allgemein gehören. Evaluatives Handeln ist konzeptionell geleitetes Handeln, das sich auf ebendiese Grundlagen bezieht. Die Analyse sollte natürlich nicht hier stehenbleiben, sondern Feinheiten herausarbeiten und die Analyse in den Kontext von Wirkungsbeurteilung in komplexem Umfeld stellen.

Ergänzend zum beschriebenen, an die Grounded Theory angelehnten Kodierprozess wurde ein zweiter Kodierprozess durchgeführt. Die im Theorieteil herausgearbeiteten Merkmale komplexer sozialer Systeme (vgl. Kapitel 4.2) dienten dabei als Kategorien für eine zusätzliche deduktive Kategorisierung. Hinweise in den Texten auf z.B. Selbstorganisation, Dynamik, Unvorhersehbarkeit oder Emergenz der untersuchten sozialen Systeme wurden den entsprechenden Kategorien zugeordnet. Durch dieses parallele Vorgehen war es einerseits möglich, evaluatives Handeln in komplexem Kontext von Grund auf und mit grosser Offenheit zu beleuchten und andererseits durch den theoretisch geschärften Blick zentrale komplexitätsrelevante Aspekte nicht aus den Augen zu verlieren. Durch die Verschränkung dieser beiden Analysestrategien, also durch Kreuzung der beiden Kategoriensysteme, kann der Blick gezielt darauf gelenkt werden, wie Evaluation mit spezifischen Aspekten komplexer sozialer Systeme umgeht, wie sie beispielsweise die Dynamik der untersuchten Systeme methodisch fasst oder inwiefern sie offen ist, um Unvorhersehbares resp. nicht intendierte Wirkungen zu erkennen. Insgesamt wurden in den ersten Berichten rund 250 Textstellen den vier Oberkategorien und elf Kategorien zugeordnet, die den Merkmalen komplexer sozialer Systeme entsprechen, wobei diese in aufbauender

Reihenfolge verstanden werden können, von der Grundbedingung "Interaktionsvielfalt" bis zur Ergebniskategorie "Unvorhersehbarkeit":

Kategorien	Konzepte (Merkmale komplexer sozialer Systeme)
Interaktionsvielfalt (60)	Konnektivität und Interdependenz Offenheit (Kontextfaktoren, System-Umwelt-Interaktionen) ¹⁰⁰
Komplexe Wirkungsmechanismen (120)	Feedback Nicht-Linearität/Zirkularität Adaption/Co-Evolution Dynamik
Selbstorganisation (60)	Reflexivität Selbstorganisation/operative Geschlossenheit
Unvorhersehbarkeit (20)	Emergenz Unvorhersehbarkeit

Tab. 5 Deduktives Kodierschema

Analytische Tiefe durch Verknüpfungen

Nach der offenen Kodierung der ersten drei Evaluationsberichte, also mit Hinzunahme der weiteren Evaluationsberichte, wurde damit begonnen, nach und nach Beziehungen zwischen den einzelnen Kategorien herzustellen. Dabei war die folgende übergeordnete Fragestellung, welche an das Kodierschema bei Strauss und Corbin 1996 anlehnt, leitend: Unter welchen Bedingungen und mit welchen Konsequenzen werden welche Strategien angewendet, um angesichts der Komplexität der untersuchten Systeme (Phänomen) zu Wirkungsbeurteilungen zu gelangen? Die einzelnen Konzepte und Kategorien wurden schrittweise zueinander in Beziehung gebracht, indem in den Textstellen zum Beispiel nach Bedingungen für die gewählten Formen des Umgangs mit Komplexität sowie nach Konsequenzen, die sich daraus ergeben, gesucht wurde. Die Bedingungen und Konsequenzen wurden jeweils um Vermutungen ergänzt, die als vorläufige Hypothesen aufgenommen wurden. Das folgende Schema zeigt, wie die Oberkategorien der eigenen Analyse mit theoretischen Dimensionen in Beziehung stehen, wie sie bei Strauss und Corbin (1996) verwendet werden:

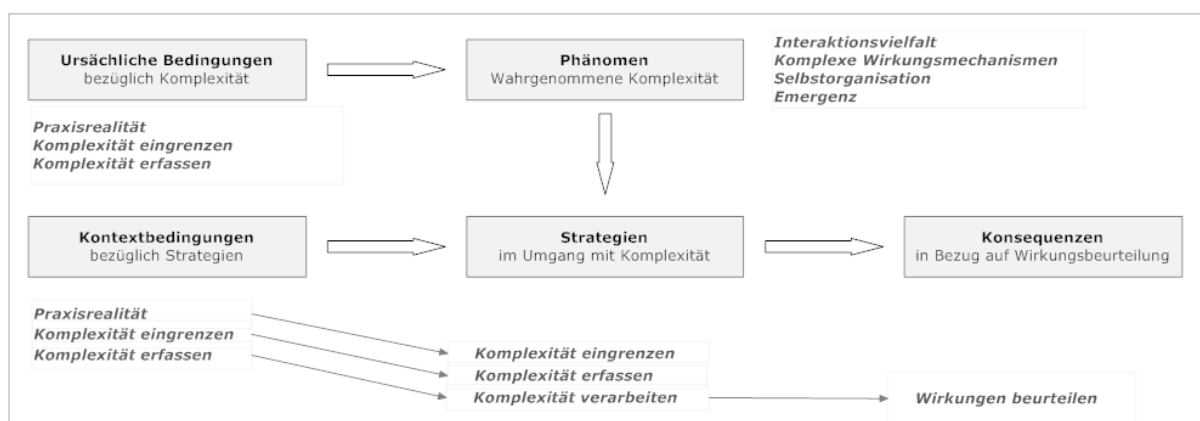


Abb. 11 Kodierschema in Anlehnung an Strauss und Corbin (1996)

Als Phänomen gilt die Komplexität, wie sie von der Evaluation wahrgenommen wird und wie sie sich in den entsprechend kodierten Textstellen der Evaluationsberichte zeigt. Unter ursächlichen Bedingun-

¹⁰⁰ Die beiden Kategorien sind im Verlauf der Codierung zusammengelegt worden, da die entsprechenden Textstellen aufgrund der Relativität der Grenzziehung System-Umwelt praktisch alle doppelt hätten kodiert werden müssen.

gen sind nicht die Bedingungen gemeint, die zur Komplexität selber führen, sondern Bedingungen für die Wahrnehmung von mehr oder weniger Komplexität. Dies sind beispielsweise Faktoren der Praxisrealität, welche den Blick einschränken können oder aber alle weiteren Faktoren, die bei der Eingrenzung oder Erfassung von Komplexität eine Rolle spielen: die theoretischen Modelle, die die Evaluation verwendet, die Art der Grenzziehung oder aber die Wahl der verwendeten Erhebungsmethoden. Fast alle kodierten Textstellen liefern Hinweise auf Strategien im Umgang mit Komplexität, da sie aufzeigen, wie Evaluation Komplexität eingrenzt, erfasst und verarbeitet. Diese Elemente kommen allerdings auch in anderen Dimensionen des Schemas vor, denn so kann die Art der Eingrenzung von Komplexität eine Kontextbedingung dafür sein, wie Komplexität erfasst wird oder die Art der Erfassung von Komplexität kann eine Kontextbedingung dafür sein, wie Komplexität verarbeitet wird. Die Textstellen, die unter „Wirkungen beurteilen“ kodiert wurden, können letzten Endes als Konsequenzen der gewählten Strategien verstanden werden. Ein Beispiel: Eine Strategie im Umgang mit Komplexität ist es beispielsweise, sich bei der Erfassung relevanter Elemente auf Andere abzustützen. Da der Untersuchungsgegenstand in seiner Komplexität der Evaluation nicht unmittelbar zugänglich ist, werden Dritte um ihre Erfahrungen, Einschätzungen, Beurteilungen, Begründungen etc. gebeten. Die zur Verfügung stehenden Ressourcen, die Interessen der Auftraggeber oder auch die verwendeten wissenschaftlichen Modelle können Kontextbedingungen sein für die konkrete Ausgestaltung der Strategie „Auf Andere abstützen“. Es können zum Beispiel mehr oder weniger viele Perspektiven berücksichtigt werden und diese nur oberflächlich oder aber sehr tiefgehend. Letzten Endes haben alle diese Entscheidungen Konsequenzen für die Wirkungsbeurteilung, so dass zum Beispiel eine Wirkungsbeurteilung, die sich auf detaillierte Informationen aus verschiedenen Perspektiven abstützt, an Plausibilität gewinnen kann. Aber auch die Art des Untersuchungsgegenstandes, also ob es sich zum Beispiel um eine Verhaltens- oder um eine verhältnispräventive Massnahmen handelt, fließt als Kontextbedingung für die Evaluation mit ein.

Die beobachteten und vermuteten Bedingungen, Strategien und Konsequenzen wurden tabellarisch dokumentiert und laufend überarbeitet. Derartige Tabellen dienen als Grundlage für die zusammenfassende Analyse (s. Anhang 10.4).

Im Laufe der axialen Kodierung wurde das Kategoriensystem nach und nach weiter entwickelt und ausdifferenziert. Die folgende Tabelle gibt einen Einblick, das gesamte Kategoriensystem befindet sich im Anhang (10.3).

Komplexität erfassen
Kontexte beschreiben
Fehlende Daten
Fehlende Perspektiven
Mehr Erhebungsaufwand
Anpassung der Methodik
Bestehende Daten nutzen
Monitoringdaten nutzen
Standardisiert erheben
Komplexität über Zahlen fassbar machen Messen; Operationalisieren; Indexbildung; Mit Proxys arbeiten; quantifizieren; nivellieren; Faktoren durch andere Faktoren ausdrücken; Kompromisse machen
offen (qualitativ) herangehen
Nicht voll standardisierte Gespräche
Offene Fragen in schriftlicher Befragung
offen erheben
Erkenntnis durch Nähe erreichen
beobachten
breites Spektrum offen einfangen
Methodenmix
Auf Andere abstützen
Mit Akteuren im Feld partizipieren Erhebungen koordinieren
Auf subjektive Einschätzungen bauen
Priorisierung den Untersuchten überlassen
Personen in Gruppen befragen
Zielgruppen befragen
Schlüsselpersonen befragen
stellvertretend befragen
Zufällig zu Befragende auswählen
Partizipation an Evaluation kontrollieren
Retrospektiv fragen
Prospektiv fragen
Hypothetisch fragen - Potenziale erfragen
persönlich befragen (face to face)
nach Bewertungen fragen

Tab. 6 Auszug aus dem Kategoriensystem

Frühe Emergenz einer Kernkategorie

Die Grounded Theory geht davon aus, dass sich im Analyseprozess eine Kernkategorie herauskristallisiert, welche zum Mittelpunkt der entstehenden Theorie wird:

The core category represents the central phenomenon of the study. It is identified by asking questions such as: What is the main analytic idea presented by this research? If I had to conceptualize my findings in a few sentences, what would I say? (Corbin & Strauss, 1990, 424).

Die im selektiven Kodieren herausgearbeiteten und verfestigten Beziehungen dieser Kernkategorie zu den anderen Kategorien bilden das Gerüst der entstehenden Theorie (Strauss & Corbin, 1996). In der vorliegenden Untersuchung kristallisierte sich bereits früh ein zentrales Thema heraus, welches sich durch alle Kategorien hindurch zieht. Von der Grenzziehung über die Erfassung und die Verarbeitung von Komplexität bis zur eigentlichen Wirkungsbeurteilung drehte sich alles um das Thema Wirkungskonstruktion. Obwohl bereits im Theorieteil thematisiert und mit Blick auf konstruktivistisch orientierte Erkenntnis- und Wissenschaftstheorie eine Selbstverständlichkeit, war die Entdeckung dieses zentralen Themas in seiner Deutlichkeit ein Aha-Erlebnis. Nicht primär der Umstand, dass die Evaluation Wirkungen nicht nachweist (obwohl immer davon gesprochen wird), sondern eben konstruiert, ist verblüffend, sondern die Vielfalt von Elementen der Wirkungskonstruktion, der grosse Spielraum der Evaluation bei der Wirkungskonstruktion, die vielen Unsicherheiten, die sich zeigten und die klaren Gren-

zen der Wirkungskonstruktion. Diese Wirkungskonstruktion im Detail zu untersuchen und zu reflektieren, erschien mehr und mehr lohnenswert, und Wirkungskonstruktion festigte sich immer mehr als übergeordnetes Thema und etablierte sich als Kernkategorie. Durch die verschiedenen Phasen der Evaluation, von der ersten Grenzziehung bis hin zur Wirkungsbeurteilung, wurden die zahlreichen Facetten des Umgangs mit Komplexität als Elemente einer systematischen Wirkungskonstruktion analysiert.

Konfrontation mit der Literatur

In einem weiteren Kapitel wurde der aktuelle Stand der empirischen Analyse in dreierlei Hinsicht in einen theoretischen Bezug gebracht. Zum einen mit den in Kapitel 4 herausgearbeiteten Erkenntnissen aus der Komplexitätstheorie, zum zweiten mit den in Kapitel 4.3 besprochenen Prinzipien der Gesundheitsförderung und drittens und mit aktuellen Ansätzen der Evaluationsliteratur. Dazu wurde für jede Evaluation in einem zusätzlichen Analyseschritt einzeln zusammengefasst, wie sie mit Komplexität, also den Merkmalen komplexer sozialer Systeme bei der Wirkungsbeurteilung umgeht (8.1). In einem weiteren Schritt (8.2-8.6) wurden diese Ergebnisse zusammen mit den Erkenntnissen aus Kapitel 7 und vor dem Hintergrund der aktuellen Diskussionen in der Evaluationsliteratur reflektiert und dann zu einer Synthese verdichtet (Kap. 8.7). Es ging darum, die Analyseergebnisse vor dem Hintergrund der theoretischen Diskussionen zu reflektieren, sie kritisch zu befragen und die Analyse auf dieser Basis zu vertiefen und zu schärfen. Dort wo die Gegenüberstellung der empirischen Ergebnisse mit der Theorie neue Fragen aufwarf, galt es, noch einmal zu den Daten zurück zu kehren und nach Antworten zu suchen.

In Zusammenhang mit kontextsensiblen Evaluationsansätzen (Pawson & Tilley, 1997) ist beispielsweise die Frage aufgekommen, wie differenziert das wissenschaftliche Wissen über Ursache-Wirkungszusammenhänge ist, das in die Evaluationen eingeflossen ist. Eine Überprüfung an den Daten hat gezeigt, dass herbeigezogenes wissenschaftliches Wissen in der Regel einen hohen Generalisierungsgrad aufweist und nur in Ausnahmefällen zielgruppen- oder kontextspezifisch differenzierte Erkenntnisse enthält.

Ergebnisdarstellung

Die Ergebnisse der empirischen Untersuchung wurden jeweils entlang der induktiv aus den Texten heraus entwickelten Kategorien dargestellt. Im Vordergrund stand eine zusammenfassende Beschreibung und Analyse der Befunde. Dabei wurden die beiden Themen „Umgang mit Komplexität“ und „Beurteilung von Wirkungen“ fokussiert, der Umgang der Evaluation mit Komplexität in den Kontext der Wirkungsbeurteilung gestellt. Die Ergebnisse der empirischen Analyse sind jeweils mit Referenzen zu entsprechenden Textstellen versehen (Identifikationsnummer der Evaluation und Seitenzahl). Diese Angaben dienen einerseits der Nachvollziehbarkeit der Ergebnisse durch Dritte und andererseits der Qualitätssicherung: die einzelnen Befunde wurden konsequent mit konkreten Textstellen belegt. Damit sollte vorschnellen und ungenügend abgestützten Ergebnissen vorgebeugt werden.

Die wörtlich wiedergegebenen Zitate dienen in erster Linie der Illustration und inhaltlichen Nachvollziehbarkeit der Befunde. Sie sollen dem Leser die Ergebnisse verständlicher machen, indem sie die zum Teil abstrakten Befunde mit konkreten Beispielen veranschaulichen.

Mittels eigenen Abbildungen und Tabellen wurde versucht, wesentliche Schritte, Faktoren und Zusammenhänge der Wirkungskonstruktion für den Leser oder die Leserin besser nachvollziehen zu können. Diese unterliegen natürlich selber einer starken und auf jeden Fall diskussionswürdigen Reduktion von Komplexität. Die in den Grafiken dargestellten Sachverhalte widerspiegeln die Analyseergebnisse in groben Zügen, indem sie konsequent auf den Hauptkategorien der Analyse aufbauen.

6.3 Potenzial und Grenzen der Untersuchung

Bei der vorliegenden Untersuchung handelt es sich um eine qualitative Fallstudie. Die klassischen Gütekriterien empirischer Forschung, namentlich Objektivität, Validität und Reliabilität, sind in einem objektivistischen Sinne, wie sie in der Testtheorie entwickelt wurden und Verwendung finden, mehrheitlich nicht als Referenz für die Überprüfung der Güte qualitativer Untersuchungen, die sich einem konstruktivistischen Ansatz verpflichten, geeignet. Vielmehr orientiert sich die vorliegende Arbeit an Gütekriterien, wie sie für qualitativ orientierte Forschung entwickelt wurden. Trotz einer Vielzahl unterschiedlicher Kriterienlisten, die in Abgrenzung zur klassischen Trias aus der Testtheorie für qualitative Forschung entstanden sind, lassen sich einige Kriterien oder Leitideen herauskristallisieren, die sich erstens in leicht adaptierter Form in unterschiedlichen dieser Kriteriensets wiederfinden und die sich zweitens - zumindest teilweise - an die klassischen Gütekriterien anlehnen. Der Forschungsprozess sowie die Aussagekraft der Ergebnisse der vorliegenden Studie sollen anhand dieser ausgewählten Kriterien und auch mit Blick auf die Beurteilungskriterien für eine Grounded Theory (nach Corbin & Strauss, 1990) reflektiert werden. Dabei werden sowohl Stärken als auch Limitierungen diskutiert:

Intersubjektive Nachvollziehbarkeit: Das Kriterium der intersubjektiven Nachvollziehbarkeit wurde in Abgrenzung zum Kriterium der Objektivität formuliert. In konstruktivistisch orientierter qualitativer Forschung ist strenge Objektivität keine sinnvolle Referenz. Das Gütekriterium der intersubjektiven Nachvollziehbarkeit fordert dazu auf, die Schritte des Erkenntnisprozesses transparent zu machen, damit andere u.a. beurteilen können, inwiefern der Prozess regelgeleitet geführt wurde, und um subjektive Einflüsse erkennbar machen zu können. Der Forschungsprozess, welcher unzählige Stunden Detailarbeit umfasst, kann nicht im Detail beschrieben werden. Im Kapitel 6.2 wurde er skizziert und der Weg vom Text über Konzepte bis zu den Kategorien aufgezeigt. In der Beschreibung der empirischen Analyse (Kapitel 7) finden sich zahlreiche zitierte Textstellen, welche die entwickelten Konzepte und Kategorien repräsentieren. Zusätzlich wurden die Ergebnisse der Analyse konsequent mit Hinweisen auf meist entsprechende Textstellen untermauert. Da es sich bei den untersuchten Datenquellen um veröffentlichte Evaluationsberichte handelt, sind die Rohdaten für Dritte gut zugänglich und die Ergebnisse lassen sich zumindest teilweise nachprüfen. Da die vorliegende Untersuchung nicht in einer Forschergruppe gemacht werden konnte, sondern in Einzelarbeit durchgeführt werden musste (ein Mangel dieser Untersuchung), wird der systematischen Referenzierung der entsprechenden Textstellen, auch im Sinne der internen Qualitätssicherung, besonderes Gewicht beigemessen.

Beispiel: Es haben sich in der Analyse zahlreiche Plausibilisierungsstrategien herauskristallisiert, mit denen Wirkungsbeurteilungen in Evaluationen überprüft und untermauert werden, so z.B. die Strategie der Plausibilisierung durch unterschiedliche methodische Zugänge. Zu jeder der Strategien wurden einzelne Textstellen angegeben, welche eine Überprüfung ermöglichen. Im genannten Beispiel führt die Referenz EB, 12 zu folgender Textstelle im entsprechenden Evaluationsbericht: „Diese Schätzung kann als robust angesehen werden, da sie durch die Verwendung von zwei unterschiedlichen ökonomischen Modellen bestätigt wurde.“ (Wieser et al., 2010, 12)

Plausibilität der Ergebnisse: Wenn davon ausgegangen wird, dass objektive Realität nicht unmittelbar zugänglich ist, kann auch nicht überprüft werden, ob sie exakt abgebildet wurde. Interne Validität im Sinne des traditionellen Gütekriteriums der Testtheorie kann entsprechend nicht Richtschnur sein zur Beurteilung konstruktivistisch orientierter Forschung (Guba & Lincoln, 1989). In Anlehnung an die konstruktivistische Reformulierung dieses Gütekriteriums bei Guba („credibility“, Guba & Lincoln, 1989, 236) und an das Kriterium der empirischen Verankerung bei Steinke (2007), erscheint Plausibilität der Ergebnisse den dort formulierten Sachverhalt gut zu treffen.

Wesentliches Element, welches die Plausibilität dieser Studie fördert, ist die hohe Dichte der Analyse resp. die Vielzahl der akribisch untersuchten Textstellen. Der für die Grounded Theory charakteristische iterative Analyseprozess, der auf ständigem Vergleichen, Interpretieren und wieder empirischem Überprüfen beruht, bildet ein feines Netz an kontinuierlicher Hinterfragung und Vergewisserung, auf welchem die Plausibilität der Ergebnisse aufbaut. Ein weiteres zentrales Element der Plausibilisierung

ist der Abgleich der Ergebnisse mit der Evaluationsliteratur, der im Kapitel 8 geleistet wird. Was in dieser Studie nicht geleistet wird, ist die Plausibilisierung über methodische Triangulation. Aus Ressourcengründen ist im Laufe des Forschungsprozesses entschieden worden, auf zusätzliche weitere Datenquellen resp. auf eigene Erhebungen, zum Beispiel Einzelinterviews mit den Evaluierenden zu verzichten und das theoretische Sampling auf die Evaluationsberichte zu beschränken. Eine weitere Limitierung liegt, wie bereits weiter oben angesprochen, in der gewählten Datenquelle selbst begründet. Evaluationsberichte sind Produkte von Evaluationen. Als solche gewähren sie nur begrenzt Einblicke in die vielen Interaktionen zwischen den an der Evaluation Beteiligten. Die entsprechenden Aushandlungsprozesse zu kennen, würde die Erkenntnisse auf jeden Fall bereichern.

Beispiel: Die Schwierigkeit, klare Wirkungsaussagen zu machen, hat sich in den Evaluationen bald einmal als übergreifendes Thema gezeigt. So gibt es denn auch zahlreiche Stellen, an denen sich diese Unsicherheit deutlich ausdrückt (es werden z.B. „mögliche“ (E9, 13), „wahrscheinliche“ (E1, 42; E3, 5), und „naheliegende“ Wirkungen (E1, 36; E9, 13) proklamiert. Es wird „ein Eindruck von Wirkung“ (E4, 47) festgestellt oder der Interventionsgruppe werden „eher“ positive Veränderungen nachgesagt (E1, 38; E8, 80). Wirkungen scheinen „plausibel“ (E2, 81) oder werden „mit grosser Sicherheit“ (E5, 30) nachgewiesen. Gleichzeitig gibt es aber in allen Evaluationen auch klare Wirkungsaussagen, wie z.B. „Direkte kleinere Veränderungen in der Infrastruktur wurden vorgenommen, die eine grosse Wirkung hatten.“ (E2, 51). Da sich kein Muster erkennen liess, unter welchen Bedingungen es zu unsicheren oder eben klaren Wirkungsaussagen kommt, wurde vermutet, dass nicht kontextuelle, gegenstandsbezogene oder methodische Elemente zu mehr oder weniger Klarheit führen, sondern dass die Deutlichkeit der Wirkungsaussagen stark von den jeweiligen Autorinnen und Autoren abhängt. Was im einen Fall zu einer unsicheren Wirkungsaussage führt, wird in einem vergleichbaren Fall als deutlicher Wirkungsnachweis ausgegeben. Diese Vermutung hat sich bei nochmaliger Überprüfung der Textstellen zu einem grossen Teil erhärtet – während die einen Evaluationen sehr zurückhaltend mit Wirkungszuschreibungen sind (E3), sind andere, obwohl methodisch nicht erkennbar besser abgestützt, diesbezüglich freizügiger (E2, E4, E8).

Generalisierbarkeit der Ergebnisse: Auch Generalisierbarkeit resp. externe Validität erhält im Lichte konstruktivistischer Forschung eine besondere Bedeutung, da jede Erkenntnis Konstruktion ist und damit nur eine unter verschiedenen möglichen Konstruktionen. Erkenntnis ist vor diesem Hintergrund immer kontextbezogen und Verallgemeinerungen bedürfen prinzipiell der Überprüfung in dem Kontext, auf den Erkenntnis übertragen wird. Die in der Analyse gewonnenen Erkenntnisse zum Umgang der Evaluation mit Komplexität treffen zunächst auf diejenigen Evaluationen zu, aus denen sie gewonnen wurden. Der Anspruch der Grounded Theory geht aber weit darüber hinaus. Durch die Strategie des theoretischen Samplings mit möglichst grosser Varianz der untersuchten Fälle und die Fortführung des iterativen Analyseprozesses bis hin zur theoretischen Sättigung sollen Erkenntnisse gewonnen werden, die für möglichst viele Kontexte Gültigkeit haben. In den untersuchten Evaluationen lassen sich fast unzählige methodische Schritte identifizieren, welche den Umgang mit Komplexität aufzeigen. Auch wenn sich einzelne Evaluationen bezüglich Forschungsdesign unterscheiden, so verwenden sie doch in ähnlichen Kontexten immer wieder vergleichbare Strategien im Umgang mit Komplexität. Es sind Erkenntnisse über diese kontextgebundenen Strategien, die sich über die vielen iterativen Analysenprozesse gefestigt haben, welche über die untersuchten Fälle hinaus verallgemeinert werden können. Durch die Auswahl von z.T. stark kontrastierenden Evaluationsdesigns und durch die sich dennoch abgezeichnete Sättigung kann davon ausgegangen werden, dass ein breites und gehaltvolles Spektrum an Formen des Umgangs mit Komplexität durch die Analyse abgedeckt werden konnte. Weitere Evaluationen mit anderen methodischen Ansätzen mögen andere Wege gehen, um zur Wirkungsbeurteilung zu gelangen. Sie dürften sich aber zu einem grossen Teil auf die hier herausgearbeiteten Strategien abstützen.

Beispiel: In den untersuchten Evaluationen zeigen sich drei unterschiedliche Strategien, auf welchen Wirkungszuschreibungen basieren: systematische Vergleiche, detaillierte Rekonstruktion der Wirkungszusammenhänge und Übernahme von Wirkungszuschreibungen von Befragten. In den neun

Evaluationen finden sich insgesamt viele unterschiedliche methodische Zugänge, die zu Hunderten von Wirkungsaussagen führen. Da beim Sampling auf eine möglichst grosse Varietät der methodischen Zugänge geachtet wird und die Anzahl von Wirkungsevaluationen in der Gesundheitsförderung und Prävention in der Schweiz überschaubar ist, kann davon ausgegangen werden, dass die Untersuchung ein grosses Spektrum aktueller Praxis der Wirkungsevaluation abdeckt. Da sich die gefundenen Wirkungsaussagen letztlich alle auf eine der drei Strategien oder eine Kombination davon zurückführen lassen, dürften diese drei Strategien für die Wirkungsevaluation in der Gesundheitsförderung und Prävention – zumindest in der Schweiz – qualitativ repräsentativ sein. Dass sich in der Evaluationsliteratur keine Hinweise auf weitere Strategien gezeigt haben, stützt diese Generalisierung.

Angemessenheit des Forschungsprozesses: Reliabilität im Sinne einer exakten Reproduzierbarkeit der gewonnenen Erkenntnisse durch wiederholte Studien kann ebenfalls nicht Kriterium für die Beurteilung konstruktivistisch orientierter qualitativer Forschung sein. Der in einer Untersuchung gemachte Erkenntnisgewinn und der Kompetenzzuwachs würden selbst bei methodisch gleichem Vorgehen – wenn dies auch möglich wäre – nicht zu denselben Ergebnissen führen. Angemessenheit des Forschungsprozesses (entspricht der Indikation des Forschungsprozesses bei Steinke (1999)), aufbauend auf der intersubjektiven Nachvollziehbarkeit, scheint grundsätzlich gut als Kriterium zur Beurteilung der (Forschungs-)Prozessqualität geeignet zu sein, wenn es auch einen grossen Ermessensspielraum bietet. Damit kann auch das mitgefasst werden, was Guba und Lincoln (1989) als Pendant zur Reliabilität unter Verlässlichkeit (dependability) verstehen, nämlich die Nachvollziehbarkeit von laufenden Entwicklungen im Forschungsprozess. Der eingeschlagene Weg bei der empirischen Untersuchung, die qualitative Inhaltsanalyse in Anlehnung an die Grounded Theory, hat sich bewährt. Die Analyse war zwar sehr aufwändig, hat aber sehr facetten- und erkenntnisreich die Prozesse der Wirkungskonstruktion von Evaluationen freigelegt.

Beispiel: Mit der primär induktiven Herangehensweise konnte eine eindrucksvolle Vielfalt von Unschärfen und Verzerrungen identifiziert werden, die den Prozess der Wirkungsevaluation begleiten. Es konnte nicht nur aufgezeigt werden, wie sich Wirkungsnachweise dadurch stark relativieren, sondern auch, welche Strategien Evaluationen anwenden, um Unsicherheiten in den Wirkungsaussagen zu reduzieren und so zu besser abgestützten Beurteilungen zu gelangen.

Subjektive Reflexivität

Natürlich handelt es sich bei dieser Forschungsarbeit ebenfalls um einen Konstruktionsprozess, welcher u.a. gespeist wird durch meine akademische Vorbildung, berufliche Erfahrung, viele Gespräche mit anderen Fachpersonen und viele weitere Einflüsse. Vor allem aber ist er getrieben durch eine Verunsicherung, welche dem in Kapitel 2.3 beschriebenen Evidenzdilemma entspringt und durch den ständigen Handlungsdruck als Verantwortlicher für Evaluationen bei Gesundheitsförderung Schweiz und als selbständiger Berater an Brisanz gewann.

Mein sozialwissenschaftlicher Hintergrund¹⁰¹ hat mich zu Beginn meiner Tätigkeit als Verantwortlicher für Evaluation und Qualität bei Gesundheitsförderung Schweiz mit Ansprüchen der Evidence Based Medicine in Konfrontation gebracht. Der interdisziplinäre und systemische Blick, den ich aus meinem Studium und meinen ersten Berufsjahren in der Suchthilfe mitbrachte, und der sich inhaltlich gut mit den Prinzipien der Gesundheitsförderung verträgt, wollte nicht zu dem stark reduktionistischen Forschungsverständnis passen, das zuweilen mit Nachdruck an die Evaluation von Programmen und Projekten der Gesundheitsförderung und Prävention und an mich als Verantwortlichen für entsprechende Evaluationsmandate herangetragen wurde. Wäre mein Hintergrund ein medizinischer gewesen, wären entsprechende Irritationen womöglich ausgeblieben und hätten vielleicht nicht zu dieser Forschungsarbeit geführt. Mein sozialwissenschaftlicher Hintergrund und mein Interesse an der Kom-

¹⁰¹ Studium der Sozialarbeit, Ethnologie und Religionswissenschaften sowie mehrere Jahre wissenschaftlicher Mitarbeiter und Dozent an der Fachhochschule Nordwestschweiz im Bereich Sozialer Arbeit.

plexität sozialer Phänomene waren aber nicht nur wichtiger mitbedingender Faktor für diese Arbeit und für das Herauskristallisieren der dazugehörigen Forschungsfrage, sondern sie beeinflussten den gesamten Such- und Findungsprozess entscheidend mit: Die Wahl der Komplexitäts- und der Systemtheorie als theoretische Referenz und der Grounded Theory als methodischen Ansatz für die empirische Analyse sind ein klares erkenntnistheoretisches Statement im Sinne einer Verortung in einem sozialkonstruktivistischen Paradigma und eine Weichenstellung mit Folgen für die Konstruktion der Ergebnisse. Die ständige Suche nach einer Auflösung der Spannung zwischen den komplexen sozialwissenschaftlichen Fragestellungen, die die Evaluation von Gesundheitsförderung und Prävention mitbringt, und den als unangemessen empfundenen vereinfachenden Evidenzansprüchen der Evidence Based Medicine hat mich herausgefordert, immer wieder die Seite zu wechseln und das Spannungsfeld aus der ‚anderen‘ Perspektive zu beleuchten und zu verstehen. Ich habe diese Seitenwechsel und die vielen bewusst gesuchten Auseinandersetzungen mit der Literatur der Evidence Based Medicine (z.B. Cochrane, 1972; Sackett et al., 1996; Rossoth, Gay, & Lin, 2007) und der (quasi-) experimentellen Wirksamkeitsevaluation (z.B. Khandker, Koolwal, & Samad, 2010; Caspari, 2009; Gertler et al., 2011; Shadish, Cook, & Campbell, 2002) sowie Gespräche mit entsprechenden Exponentinnen und Exponenten bewusst als kontrastierende Elemente genutzt und geschätzt. Es sind nicht zuletzt diese Auseinandersetzungen mit der anderen Seite, die mich in der Ausarbeitung und Festigung meiner Positionen weitergebracht haben. Als zusätzliches wichtiges Korrektiv, das mich immer wieder auf die Verkürzungen einer einseitigen Denkweise (z.B. im Sinne von „je mehr Komplexität, desto besser“) hinwies und mich dadurch vor ‚unangemessenen‘ und ebenfalls reduktionistischen Positionierungen bewahrte, habe ich die ausgleichenden Interventionen meines Doktorvaters, Prof. Dr. Manfred Max Bergman des soziologischen Instituts der Universität Basel, erlebt.

Leider waren meine Möglichkeiten begrenzt, den Konstruktionsprozess systematisch zusammen mit anderen Fachleuten zu gestalten, da diese Forschungsarbeit aus Eigeninitiative und in Eigenarbeit entstand und nicht in einem übergeordneten Forschungsprogramm mit entsprechenden Organisationsstrukturen und gemeinsamen Entwicklungsprozessen eingebettet war. Dies ist sicherlich eine der Schwächen dieser Forschungsarbeit. Plausibilisierungen der sich abzeichnenden Erkenntnisse müssen daher über die Konfrontation mit einer breiten wissenschaftlichen Literatur und über fachliche Diskussionen in unterschiedlichen Bezügen gesucht werden: In Workshops an nationalen und internationalen Tagungen, in Weiterbildungsmodulen, in Forschungskolloquien, in Beratungsmandaten sowie in unzähligen Gesprächen konnte ich Erkenntnisse aus der Komplexitätstheorie in die Fachdiskussion einbringen, reflektieren und weiter entwickeln¹⁰². Das Schreiben von Fachartikeln, Buchbeiträgen und wissenschaftlichen Artikeln (Spencer et al., 2008; Ackermann et al., 2009; Ackermann & Pfyl, 2010; Kolip et al., 2012; Ackermann & Conrad Zschaber, 2011; Broesskamp-Stone, 2008; Ackermann et al., 2015), die sich ebenfalls mit diesen Themen auseinandersetzen, hat mich dazu gezwungen, Verunsicherungen auf den Punkt zu bringen, im wissenschaftlichen Diskurs Stellung zu beziehen und eigene Ansätze für eine Weiterentwicklung der Evaluation der Gesundheitsförderung und Prävention immer wieder zu hinterfragen, zu festigen und explizit zu machen. Immer wieder galt es, von bisher entwickelten Vorstellungen Abschied zu nehmen und neue Wege zu prüfen. Entsprechend habe ich die Erarbeitung dieser Dissertation selber als komplexen Prozess erlebt mit begrenzter Planbarkeit, unvorhersehbaren nicht-linearen Verläufen und ständig neuen emergenten Erkenntnissen, die Bestehendes in Frage gestellt und zum Teil zu radikalem Umdenken gezwungen haben. Der Prozess war sowohl von Zeiten der Euphorie als auch von Zeiten der völligen Verunsicherung geprägt. Auslöser für Ersteres waren beispielsweise Literatur, die mich weiter gebracht hat, anregende Fachdiskussionen, eine fertige Publikation oder das Erkennen von Anschlussmöglichkeiten eigener Ideen an den wissenschaftlichen Diskurs. Wenn eigene Gedanken im wissenschaftlichen Diskurs Bestätigung fanden, konnte das ebenso beflügelnd wirken wie das Bewusstwerden eigener innovativer Positionen, die vom wissenschaftlichen Mainstream abzuweichen schienen. Auslöser für Verunsicherung waren primär

¹⁰² Eine Übersicht über Auftritte, Weiterbildungsmodulen und Mandate findet sich auf meiner Webseite unter den entsprechenden Rubriken: www.guenterackermann.ch.

Ängste, mich in der Komplexität der Thematik selber zu verlieren, verbunden mit hohen eigenen Ansprüchen an die zu erwartenden Erkenntnisse. In Momenten derartiger Verunsicherung fühlte ich mich nicht selten selber als Opfer der Komplexität bzw. wurde die Geister nicht mehr los, die ich gerufen hatte.

Neben den klärenden Gesprächen in diesen Situationen war es aber auch die Literatur selber, die mich aus diesen Tiefen wieder heraus geführt hat. Sehr bewusst habe ich die Konfrontation mit Theorie gesucht und immer wieder eindrucksvoll erlebt, wie mich genau dies zurück zu einer Fokussierung auf Wesentliches gebracht hat und mich nicht in der Vielfalt von Unbedeutendem verlieren liess. Es waren neben den Strategien der theoretischen Sensibilität, wie sie die Grounded Theory propagiert (Strauss & Corbin, 1996), auch die expliziten theoretischen Bezüge in einem nachgelagerten analytischen Prozess, die mich vom rein Deskriptiven ins Analytische zu führen vermochten. Ich durfte einmal mehr den unschätzbaren Wert von Theorie in wissenschaftlichen Konstruktionsprozessen erleben, und durch die sich über mehrere Jahre erstreckende Dissertation war ein Rahmen gegeben, diese Quelle auch gebührend zu nutzen – etwas, das in Evaluationen, die meist unter starken finanziellen und zeitlichen Beschränkungen durchgeführt werden müssen, oft nur unzureichend gegeben ist.

7 Wirkungskonstruktion in Evaluationsberichten

In diesem Kapitel wird anhand der Analyse von Evaluationsberichten der Frage nachgegangen, wie Evaluation bei der Beurteilung von Wirkungen mit Komplexität umgeht. Insbesondere soll untersucht werden, wie mit Vielschichtigkeit, Dynamik, Multidimensionalität und Nichtlinearität von Ursache-Wirkungsbeziehungen, Emergenz und Unvorhersehbarkeit umgegangen wird und welche Konsequenzen sich daraus für eine Beurteilung der Wirkungen und der Wirksamkeit eines Projekts ergeben. Dabei werden in einem iterativen Prozess, der sich an die Grounded Theory anlehnt, Evaluationsberichte analysiert und Schritt für Schritt ein Modell des Umgangs mit Komplexität entwickelt. Als Kernkategorie der Analyse hat sich das Thema der Wirkungskonstruktion herauskristallisiert. Es lässt sich zeigen, dass Evaluation in komplexem Kontext Wirkungen nicht ‚nachweist‘, sondern über einen mehrstufigen sozialen Aushandlungsprozess Wirkungen systematisch konstruiert. Wie in Kapitel 5.4 dargelegt, sind Wirkungen - als eine Sonderform der Kausalität - selbst nicht beobachtbar. Es lassen sich lediglich Einzelphänomene beobachten, welche mehr oder weniger plausibel zu Ursache-Wirkungszusammenhängen verknüpft werden. Diese Verknüpfungen werden durch die Evaluation geleistet, mit Rückgriff auf unterschiedliche Quellen. Die verschiedenen Prozesse, die letztlich zu den Verknüpfungen resp. Wirkungskonstruktionen führen, haben sich zu vier Stufen verdichten lassen: Komplexität eingrenzen, Komplexität erfassen, Komplexität verarbeiten und Wirkungen beurteilen. Obwohl sich im Rahmen einer Evaluation die verschiedenen Stufen zum Teil wiederholen, überlappen und nicht immer linear verlaufen, zeigen sie sich doch idealtypisch in dieser Reihenfolge:

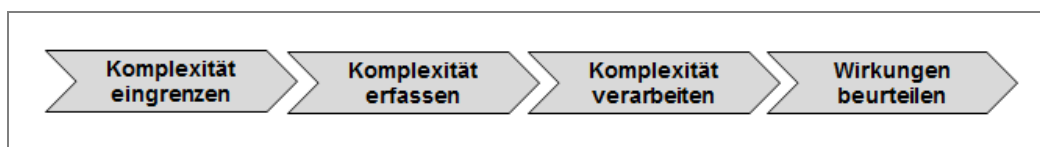


Abb. 12 Stufen der Wirkungskonstruktion in komplexen Feldern

In den folgenden Kapiteln werden Elemente der Wirkungskonstruktion entlang dieser vier Stufen beschrieben. Dabei werden Fragen beleuchtet wie:

- Wer konstruiert Wirkungen und auf welcher Grundlage?
- Was fließt in die Wirkungskonstruktion ein und was nicht?
- Mit welchen Methoden werden Wirkungen konstruiert?
- Als wie gesichert gelten die Wirkungskonstruktionen?

In der Beschreibung der ersten drei Stufen wird die Grundlage gelegt für das Verständnis der vierten Stufe, der Wirkungsbeurteilung. Es werden Schritt für Schritt die Grundlagen erörtert, die letztlich in der Wirkungsbeurteilung gipfeln.

Es werden in dieser Phase nicht die einzelnen benutzten Evaluationsdesigns kritisch beleuchtet, sondern der analytische Blick geht zunächst tiefer. Im Kapitel 8 (Diskussion) werden die Ergebnisse dann vor dem Hintergrund der aktuellen Diskussion über angemessene Forschungsdesigns in der Evaluation reflektiert. Diese Reflexion wird im Kontext der Komplexitätsdiskussion geführt und geht auf die Besonderheiten der Gesundheitsförderung und Prävention ein.

In der Analyse und Diskussion werden auf zwei Arten Bezüge zu einzelnen Evaluationen gemacht:

- Einerseits werden jeweils einige Stellen angegeben, an denen sich beschriebene Sachverhalte deutlich zeigen. Diese sind nur exemplarisch und in den allermeisten Fällen liessen sich weitere in derselben oder in anderen Evaluationen zeigen. Die als Beispiele herangezogenen Textstellen werden mit Angabe der ID der entsprechenden Evaluation und mit der Seitenzahl referenziert, z.B.: (E2, 33; E4, 62). Charakterisiert eine gemachte Aussage die Evaluation als Ganzes, wird auf die

Angabe einer Seitenzahl verzichtet. Anhand der ID's und Seitenangaben können die Befunde überprüft werden.

- Andererseits werden exemplarisch Textstellen wörtlich zitiert, welche den beschriebenen Sachverhalt gut illustrieren. Diese dienen der direkten Nachvollziehbarkeit durch den Leser und erleichtern so das Verständnis der Analyse. Wörtliche Zitate aus den untersuchten Evaluationen sind ebenfalls mit ID und Seitenzahl versehen.

7.1 Komplexität eingrenzen

In komplexen sozialen Systemen sind Ursache-Wirkungsbeziehungen vielfältig. Interventionen können in unvorhersehbarer Weise auf unterschiedlichsten Ebenen Wirkungen entfalten und die anvisierten Wirkungsindikatoren werden durch unzählige Faktoren, mit oder ohne Bezug zur Intervention, beeinflusst. Da es unmöglich ist, alle Ursache-Wirkungsbeziehungen in einer Evaluation zu berücksichtigen, muss die Evaluation die Grenzen des Evaluationssystems definieren:

Defining boundaries is an essential part of systems thinking. A boundary differentiates between what is ‚in‘ and what is ‚out‘, what is deemed ‚relevant‘ and ‚irrelevant‘, what is important and what is unimportant, who ‚benefits‘ and who is ‚disadvantaged‘. (...) It is also a political activity. (Reynolds, 2014, 83)

Es braucht eine Komplexitätsreduktion im Sinne einer Eingrenzung der zu berücksichtigenden Themen, Einflussfaktoren, Perspektiven, Datenquellen etc. Diese Grenzziehung nimmt die Evaluation immer vor, auch wenn sie dies nicht explizit tut. Indem sie bestimmte Fragen stellt und andere nicht, bestimmte Personen befragt und andere nicht berücksichtigt, ausgesuchte Kontextfaktoren einbezieht und wiederum andere aussen vor lässt, werden die Grenzen gezogen. Diese Grenzziehung kann mehr oder weniger explizit sein und mehr oder weniger gut begründet. In den untersuchten Evaluationen lässt sich zeigen, wie die Grenzziehung zunächst von der Perspektive der Anspruchsgruppen, vom herangezogenen wissenschaftlichen Wissen und von Einschränkungen der Praxis beeinflusst wird. Die Grenzziehung beeinflusst die methodische Ausrichtung, wobei eine offene, explorative Methodik die Grenzen noch einmal neu definiert.

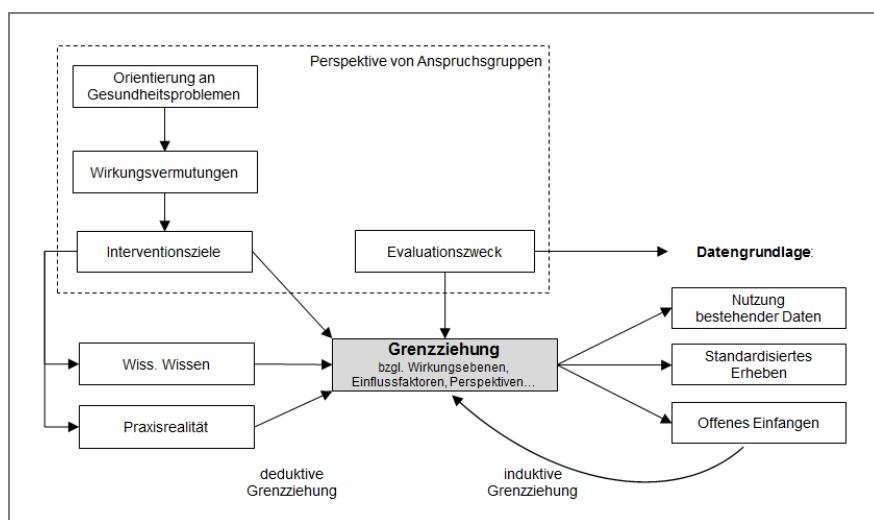


Abb. 13 Komplexität eingrenzen

Die Perspektive der Anspruchsgruppen als Ausgangspunkt der Wirkungskonstruktion

Durch das Auftragsverhältnis gegeben bildet das Erkenntnisinteresse von Trägern und/oder Finanzgebern von Interventionen, die meist auch die Auftraggeber der Evaluation sind, den Ausgangspunkt für die Wirkungskonstruktion. Zur Verbesserung der Gesundheit von bestimmten Bevölkerungsgruppen werden über Wirkungsvermutungen Ziele für die Intervention abgeleitet, welche für die Grenzziehung leitend sind. In den untersuchten Evaluationen lässt sich schön zeigen, wie sich explizite (E3, 13) oder implizite Wirkungsvermutungen der Träger konkret in Zielen niederschlagen¹⁰³ (E3, 57; E6, Z-3), die jeweils mehrere Wirkungsebenen¹⁰⁴ abdecken:

Die bfu ging bei der Entwicklung der Kampagne davon aus, dass Schneesportler/-innen auf den Helm verzichten, weil ihr Gefahrenbewusstsein ungenügend ist (fehlendes Gefahrenbewusstsein), weil sie glauben, dass sie aufgrund ihres Fahrstils oder ihrer Exposition keinem Verletzungsrisiko ausgesetzt sind (fehlende wahrgenommene Vulnerabilität) oder weil sie den Helm als unbequem einschätzen (fehlende Bequemlichkeit). Aus diesem Grund wollte die bfu den Anteil der Schneesportler/-innen reduzieren, die wegen diesen drei Faktoren auf das Tragen eines Helms verzichten. (E3, 13)

Bei den Interventionszielen kann es sich um Leistungs- oder um Wirkungsziele handeln. Für eine summative Wirkungsbeurteilung sind die Wirkungsziele die Referenz. Leistungsziele können aber für die Beurteilung von Wirkungspotenzialen und Wirkungsmechanismen wichtig sein. Die Wirkungsziele sind zum Teil sehr konkret formuliert und über Indikatoren und Sollwerte operationalisiert (E1, 6; E3, 33), zum Teil sind sie aber auch recht vage gehalten (E4, 10, E7, 3ff). Die untersuchten Evaluationen können bezüglich der Ziele der von ihnen evaluierten Interventionen in folgende Kategorien aufgeteilt werden:

¹⁰³ In zwei der untersuchten Evaluationen (E5, E8) fehlen explizite Interventionsziele, da es sich nicht um klassische Projektevaluationen handelt. Diese wurden für die Evaluation im Nachhinein rekonstruiert.

¹⁰⁴ Für das hier behandelte Thema sind insbesondere Wirkungsziele von Bedeutung (und weniger Leistungs- und Qualitätsziele).

Kategorien	Evaluationen	Beispiele
1) Operationalisierte Wirkungsziele mit klarem Anspruchsniveau	E1, E3	„Reduktion der Trinkgelegenheiten mit Bingetrinken im letzten Monat um durchschnittlich 30 % bei der Zielgruppe (Gruppe R/H) sechs Monate nach der Intervention.“ (E1, 32) „Maximal 10 % der Schneesportler/-innen geben an, sich noch nicht mit dem Gedanken «Helm» befasst zu haben.“ (E3, 57)
2) Vage gehaltene Wirkungsziele ohne klar definiertes Anspruchsniveau	E2, E4, E6, E9 E5, E8 (nachträglich rekonstruiert)	„Die Arbeitsprozesse sind gesünder und effizienter gestaltet. Die Gesundheit, das Wohlbefinden und die Arbeitszufriedenheit der einzelnen Mitarbeitenden sind gesteigert.“ (E2, 18) „Die Mädchen lernen, mit schwierigen Situationen konstruktiv umzugehen. Die Mädchen lernen, Fremde zu tolerieren und zu akzeptieren. Die Mädchen bauen ihre Gewaltbereitschaft ab.“ (E4, 10) „Erstere Fragestellung wurde durch eine Schätzung der Auswirkungen der Präventionsmassnahmen auf das Gesundheitsergebnis in Form vermiedener Krankheitsfälle, Verletzungen und Todesfälle beantwortet.“ (E8, 8)
3) Keine ausformulierten Ziele, sondern lediglich Wirkungsdimensionen	E7	„Das Kernstück der vorliegenden Studie ist ein Modell der Systemdynamik der Midnight Basketball Projekte und ihrer Wirkungsweise im Hinblick auf die von Midnight Basketball ausgewiesenen Zieldimensionen „Chancengleichheit“, „Gesundheitsförderung“, „Gewaltprävention“ und „soziale Integration“.“ (E7, 28)

Tab. 7 Kategorien von Wirkungszielen

Eine genauere Betrachtung der Ziele und ihrer Evaluierung bestätigt die naheliegende Vermutung, dass je offener die Ziele formuliert vorliegen, desto offener auch die Untersuchung der Zielerreichung erfolgt. Operationalisierte Wirkungsziele mit quantitativen Zielvorgaben legen eine Überprüfung der Zielerreichung mit statistischen Verfahren nahe. In den meisten Evaluationen (nicht in E5 und E8) werden zusätzlich qualitative Verfahren genutzt, um die quantitativen Ergebnisse zu ergänzen und angemessen interpretieren zu können.

Ungeachtet des Operationalisierungsgrades der Ziele sind diese bezüglich Kontext und Zielgruppen immer sehr allgemein gehalten. Es werden zum Beispiel Ziele für ‚Mitarbeitende‘, ‚Mädchen‘ oder ‚Eltern‘ formuliert und das für Settings wie ‚Betriebe‘, ‚Schulen‘ oder ‚Gemeinden‘. Mit diesen Zielformulierungen ist die Referenz für Wirkungsbeurteilungen auf einer sehr allgemeinen Ebene angesiedelt und es wird implizit unterstellt, dass für die Personen in diesen Zielgruppen und für die unterschiedlichen Settings ähnliche Wirkungsverhältnisse herrschen.

Es wird in den untersuchten Evaluationen grösstenteils nicht ersichtlich, welche Prozesse zu den Interventionszielen geführt haben, wer auf welche Art und Weise an diesen Prozessen beteiligt war und inwiefern diese bereits wissenschaftlich abgestützt sind¹⁰⁵. Die Perspektive von Anspruchsgruppen dürfte aber im Wesentlichen eine Expertensicht widerspiegeln, da es sich bei den Leitenden von grösseren Projekten und Programmen in der Regel um ausgebildete und/oder erfahrene Gesundheitsförderungs- und Präventionsfachleute handelt¹⁰⁶.

Da sich die Evaluationen in ihrer Beurteilung stark an den Interventionszielen orientieren, fallen fehlende oder ungenügend gut formulierte Ziele entsprechend ins Gewicht und können die Aussagekraft von Evaluationen beeinträchtigen (E3, 46). Der Evaluationszweck, welcher entweder stärker der Legitimation (Rechenschaftslegung, Entscheidungsgrundlage über eine Weiterführung oder Ausweitung) oder aber stärker dem Lernen (Steuerung, Optimierung, Argumentarien für Fachleute) verpflichtet sein

¹⁰⁵ Zum Teil wird dazu aber auf entsprechende weiterführende Dokumente verwiesen (z.B. in E3, 43).

¹⁰⁶ Vergleiche dazu die Befragung von Public Health Organisationen in der Schweiz (Frank, Weihofen, Duetz Schmucki, Nocera, and Paccaud (2013)).

kann, beeinflusst die Grenzziehung nicht im zunächst vermuteten Sinne, dass die legitimatorische Perspektive zwingend zu einer engen Fokussierung auf wenige Wirkungsindikatoren führt, die dann quantitativ untersucht werden. In E1 und E3 lässt sich eine entsprechende Tendenz zwar zeigen, E2 ist aber in seiner quantitativen Wirkungsevaluation sehr breit angelegt und E6 stützt seine Bilanzierung gar ausschliesslich auf eine ausgedehnte qualitative Analyse:

Entsprechend dem Evaluationszweck wählten wir für die Evaluation in Absprache mit der Auftraggeberin den methodischen Ansatz der vergleichenden Fallstudienanalyse. Dies gibt die Möglichkeit, in den ausgewählten Gemeinden vertieft und genau hinzuschauen und festzuhalten, was konkret und gemeindespezifisch erreicht wurde. Die Analysetiefe steht im Vordergrund, da es um eine zuverlässige Wirkungsanalyse geht. (E6, 9)

Im Spannungsfeld zwischen der Perspektive von Anspruchsgruppen, wissenschaftlichem Wissen und Praxisrealität

Die Perspektive von Anspruchsgruppen, die primär über die Interventionsziele in die Wirkungskonstruktion einfließt, ist allerdings nur eine der entscheidenden Referenzen für die Wirkungskonstruktion. In den untersuchten Evaluationen lässt sich vielmehr ein Spannungsfeld zwischen drei Referenzen ausmachen: die Perspektive von Anspruchsgruppen, wissenschaftliches Wissen aus der Literatur und die Praxisrealität. Neben der Perspektive von Anspruchsgruppen bildet die wissenschaftliche Literatur die zweite tragende Säule für die Wirkungskonstruktion. Interventionsziele und Wirkungsvermutungen werden zu Beginn durch die Evaluation nur selten kritisch hinterfragt¹⁰⁷. Der Evaluationsblick wird in die entsprechende Richtung gelenkt und läuft dadurch Gefahr, eingeschränkt zu sein. Wissenschaftliches Wissen kann hier als Korrektiv dienen und beispielsweise im Rahmen von Konzeptevaluationen systematisch genutzt werden:

Die Überprüfung des Schneesportkampagnenkonzepts zeigte, dass auf der Ebene der Impacts und der Outcomes geeignete Indikatoren zur Messung der Wirkungen und des Erfolgs enthalten sind, die die definierten Ziele gut abbilden. (...) Dabei wurde geprüft, ob die Kampagne so angelegt ist, dass sie nach evaluationswissenschaftlichen Kriterien zielwirksam sein kann. (E3, 43)

Alle untersuchten Evaluationen stützen sich auf wissenschaftliches Wissen ab, wobei die Interventionsziele den Fokus bestimmen. Die Praxisrealität tritt als beschränkendes Element hinzu, indem nicht alles, was gewünscht oder wissenschaftlich angebracht wäre, in der Praxis auch umgesetzt werden kann. Die verschiedenen Ansprüche, die in Evaluationen kombiniert werden sollen, zeigen sich beispielsweise in folgendem Zitat aus einer der Evaluationen:

Ziel war es, mit diesem Projekt fundierte Argumente und praxistaugliche Vorgehensweisen zu entwickeln, welche zur Verbreitung der Umsetzung von Stresspräventionsmassnahmen in der Schweiz beitragen. (E2, 4)

Wissenschaftliches Wissen

Was das wissenschaftliche Wissen angeht, so können in den untersuchten Evaluationen und in Anlehnung an das Best Practice Konzept von Gesundheitsförderung Schweiz (Broesskamp-Stone, 2009) Aussagen zu folgenden Gegenständen unterschieden werden:

- Wissen zur Verteilung von Gesundheit und Gesundheitsdeterminanten in der Bevölkerung (E1, 1; E3, 5)

¹⁰⁷ Eine nachträgliche kritische Hinterfragung aufgrund der Evaluationsergebnisse ist hingegen häufiger (E1, 49ff, und pointiert E6, 86f).

- Wissen über das Zusammenspiel von Gesundheitsdeterminanten im Sinne von Wirkfaktoren (E2, 23; E5, 4)
- Wissen zur Wirksamkeit von Interventionen (E1, 2; E5, 4)

Mehr oder weniger alle Evaluationen führen Wissensnachweise zu allen drei Wissensgegenständen auf, allerdings in sehr unterschiedlichem Umfang, in wenigen Fällen bieten sie eine ansatzweise systematische Übersicht über die Evidenzlage (z.B. in E1, 2; E2, 22f), mehrheitlich ist die Verwendung eher punktuell/exemplarisch. Für die Grenzziehung erweisen sich vor allem die Aussagen aus dem zweiten und dritten Bereich als bedeutsam¹⁰⁸, weil hier definiert wird, welche Faktoren für die Wirkungsentstehung entscheidend und daher in die Betrachtung einzubeziehen sind. Die Wirkungszusammenhänge werden dabei entweder in Form von Studienergebnissen angeführt (E4, 11; E5, 4) oder aber in Form von theoretischen Modellen (E2, 10; E5, 28; E6; 79), wobei beim Rückgriff auf andere Studienergebnisse nur teilweise transparent reflektiert wird, inwiefern die Ergebnisse aufgrund ihres jeweiligen Kontextes tatsächlich auf die aktuell untersuchte Evaluation übertragbar sind (z.B. in E1, 1).

Die Auswertung möglicher Produktivitätsgewinne beschränkt sich deshalb auf den Absentismus aufgrund von Krankheit und die Arbeitsleistung am Arbeitsplatz, also auf zwei Grössen, die oft im Mittelpunkt der betrieblichen Gesundheitsförderung stehen. Verschiedene Studien haben auch gezeigt, dass die Arbeitsleistung oft wichtiger Erfolgsfaktor der BGF ist (siehe z.B. Mills, Kessler, Cooper & Sullivan, 2007). (E2, 72)

Auch wird nur teilweise klar, inwiefern die zitierten Studien den aktuellen Forschungsstand gut repräsentieren oder gezielt zur Untermauerung eigener Argumente herausgegriffen wurden. Wenn nicht bloss einzelne Studien, sondern passende systematische Reviews, Metaanalysen oder mehrere Studien verfügbar sind und angeführt werden (E2, 9; E5, 5), erhöht dies die Aussagekraft, und das Potenzial derartiger Bezüge kann besser eingeschätzt werden. Auf der anderen Seite kann sich Evaluation natürlich nur dort auf wissenschaftliches Wissen abstützen, wo solches auch tatsächlich existiert. Gerade wenn im Rahmen von Pilotprojekten neue Interventionsansätze getestet werden, ist Evidenz oft fehlend (E1, 2; E4, 11). Zusätzlich lassen sich Studienerkenntnisse nur bedingt auf die jeweiligen Evaluationskontexte übertragen (E7, 5). Liegen, wie bei hoher Komplexität nicht selten, unsichere oder gar widersprüchliche Evidenzen vor (E3, 3), bedarf es einer entsprechenden Einschätzung durch die Evaluation. Zudem beeinträchtigen verschiedene Verzerrungen (wie z.B. der publication bias) die Evidenzbasis.

Theoretisches und empirisches Wissen über Ursache-Wirkungszusammenhänge ist, wie wir später noch sehen werden, von praktisch allen Evaluationen explizit in die Entwicklung ihrer Wirkungs- und Analysemodelle aufgenommen worden und/oder als Referenz für Beurteilungen von Konzeptqualität (E3) oder Umsetzungsergebnissen (E1, 48f, E6) genutzt worden. Allerdings handelt es sich in den meisten Fällen um stark generalisiertes Wissen und kaum um differenziertes, kontext- und zielgruppenspezifisches Wissen über Ursache-Wirkungszusammenhänge.

Rahmenbedingungen der Praxis

Nicht alles was aus Sicht der Anspruchsgruppen oder aus fachlicher Sicht wünschbar ist, lässt sich in der Praxis auch wirklich realisieren, nicht alle gewünschten Ursache-Wirkungszusammenhänge lassen sich untersuchen. Einen entscheidenden Einfluss auf die Grenzziehung hat grundsätzlich der finanzielle Umfang der Evaluation, da der Breite und Tiefe der Untersuchungen aus Ressourcengründen Grenzen gesetzt sind. Oder umgekehrt formuliert: Wenn der Evaluationsgegenstand und entsprechende Ursache-Wirkungsbeziehungen als sehr komplex konzipiert und entsprechend umfassend in

¹⁰⁸ Die Argumente aus dem ersten Bereich fliessen über die Interventionsziele ein, jene zum dritten Bereich sind vor allem dann von Interesse, wenn sie Aussagen über Wirkungszusammenhänge enthalten, die dann aber ebenfalls dem zweiten Bereich zugeordnet werden können.

die Evaluation aufgenommen werden (E2, E7), resultieren ressourcenintensive Evaluationen, in denen mittels unterschiedlicher Methoden und aus mehreren Perspektiven eine Vielzahl von Wirkungsindikatoren und Einflussfaktoren untersucht werden können. Fehlen diese Mittel, müssen entsprechende Abstriche gemacht werden. Die Rahmenbedingungen einer Evaluation spannen einen Möglichkeitsraum auf, welcher in den Evaluationsberichten primär als beschränkend auftaucht. Es werden prioritär die folgenden Beschränkungen sichtbar:

- begrenzte finanzielle resp. personelle Ressourcen resp. zu hoher Aufwand für die Evaluation (E3, 42; E4, 14)
- begrenzte Verfügbarkeit oder Teilnahmebereitschaft von relevanten Informationsträgern (E2, 19/39; E3, 82)
- begrenzte Interventions- und/oder Beobachtungsdauer (E2, 82; E9, 88)
- begrenzte Verfügbarkeit von Daten (E2, 35; E7, 8)
- sich verändernde Kontexte, welche Anpassungen der Intervention (E3, 44) und/oder der Evaluation (E1, 42) erfordern

Die Konsequenz dieser verschiedenen Beschränkungen durch die Praxisrealität ist insbesondere eine verminderte Aussagekraft der Ergebnisse. Sie kann zu Verzerrungen führen (E1, 42), die Analysetiefe einschränken (E2, 45; E5, 12) oder aber grundsätzlicher Natur sein:

„Ohne Vergleichsgruppen bleibt aber immer die Frage im Raum, ob die Veränderung nicht andere Ursachen hat.“ (E2, 87)

Anpassungen an die Praxisrealität – mit all ihren Einschränkungen – werden aber nicht nur als Nachteil gesehen. Evaluationsergebnisse, die im Kontext einschränkender Praxis zustande kommen, mögen zwar interne Validität einbüßen, sie zeichnen sich potenziell aber durch eine höhere externe Validität aus und sind dadurch für die Praxis von besonderem Nutzen:

Sie (die Evaluation. Anm. d. Verf.) prüft somit die direkte Umsetzbarkeit einer präventiven Massnahme in der Situation, in der diese Massnahme auch wirklich umgesetzt werden soll. Sogenannte „Efficacy-Trials“ haben zwar in der Regel eine höhere interne Validität, da sie Situationen, die die direkte Überprüfbarkeit der Wirkung der Intervention gefährden, ausschliessen (...). Sie sind aber aufgrund dieser „experimentellen“ Situation (...) häufig nur schwer auf reale Situationen generalisierbar. Aus diesem Grund wurde ein „Effectiveness-Trial“ konzipiert, d.h., dass die entsprechenden Ergebnisse direkt für die Praxis umsetzbar sind. (E1, 3)

Grenzziehung als Aushandlungsprozess

Die Grenzziehung der Evaluation kann als Aushandlungsprozess zwischen den Ansprüchen der Nutzenden, der Wissenschaft, der Praxis und natürlich der Evaluierenden selbst verstanden werden, in welchem die zu berücksichtigenden Wirkungsfaktoren und Perspektiven festgelegt, die Erhebungsinstrumente definiert, die Stichproben eingegrenzt und weitere Elemente bestimmt werden. In den untersuchten Evaluationen gibt es kaum Hinweise darauf, wie entsprechende Aushandlungsprozesse konkret abgelaufen sind und wer in welcher Weise darin involviert war (E4, 12; E6, 29), meist werden lediglich die Ergebnisse festgehalten. Den Evaluierenden, welche die Evaluation fachlich verantworten, kommt eine vermittelnde Expertenrolle und auch abschliessende Entscheidungskompetenz zu.

Diese bestimmen, wie weit sie die verschiedenen Perspektiven und Ansprüche aufnehmen und verantworten können und letztlich auch darüber, wo fachlich, ethisch, finanziell etc. die Grenzen des Machbaren sind, im Rahmen welcher eine Evaluation also überhaupt durchgeführt werden kann.

Enge und weite Grenzen

Bei allen untersuchten Evaluationen finden mehrere Wirkungsindikatoren, mehrere Perspektiven resp. Datenquellen sowie mehrere potenzielle Einflussfaktoren Eingang in die Wirkungsbeurteilung, wobei sich einzelne Evaluationen im Umfang der berücksichtigten Elemente stark unterscheiden. Beispiele für einen engeren Fokus sind E1 und E3, Beispiele für sehr weite Grenzen liefern E2 und E7.

Die Interviews kreisten um die Lebenswelt der Jugendlichen, also Schule, Eltern, Gemeinde, um ihren Lebensstil, ihre Erwartungen, Haltungen und Perspektiven, ihre Probleme, um ihr Gesundheits- und Suchtverhalten sowie um das Midnight Basketball Projekt, wie sie dieses Angebot erleben und welchen Einfluss es auf sie in ihrer eigenen Wahrnehmung hat. Ein besonderes Augenmerk galt dem interkulturellen Zusammenleben, Konflikten sowie der Geschlechterperspektive. (E7, 10)

Gründe für diese Unterschiede scheinen auf den ersten Blick in der unterschiedlichen Komplexität der Interventionen zu liegen, sind doch die ersten beiden Evaluationen Beispiele für eher begrenzte verhaltensorientierte Interventionen, während es sich bei den beiden anderen um komplexere settingorientierte Ansätze handelt. Bei genauerer Betrachtung und der Hinzunahme weiterer Evaluationen lässt sich diese Vermutung von der Tendenz her zwar bestätigen, allerdings nicht mit Ausschliesslichkeit. In folgender Tabelle wurden die untersuchten Evaluationen in Bezug auf die in die Wirkungsbeurteilung einflussenden Faktoren (engere versus weitere Grenzen), die Interventionsart (Verhaltens- versus Settingorientierung), die Komplexität der Intervention und die methodische Orientierung (quantitativ versus qualitativ) eingeschätzt.

	eher engere Grenzen	eher weitere Grenzen
starke Verhaltensorientierung	E1, E3, E9	E4
starke Settingorientierung	E5	E2, E6, E7
nicht beurteilbar	E8	
wenig komplexe Intervention	E1, E3, E5, E9	
höher komplexe Intervention	E8	E2, E4, E6, E7
stärker exploratives Design	E1, E3, E5, E8, E9	E2
stärker hypothesenprüfende Design		E2, E4, E6, E7

Tab. 8 *Einschätzung der untersuchten Evaluationen*

Die kleine Analyse lässt sich zu folgenden Aussagen verdichten:

- wenn die Interventionen verhaltensorientiert und/oder wenig komplex sind, zieht die Evaluation immer eher enge Grenzen
- bei starker Settingorientierung und/oder komplexer Intervention resultieren mehrheitlich weitere Grenzen, aber auch eine enge Grenzziehung ist möglich
- bei wenig komplexen Interventionen werden die Grenzen nie weit gezogen
- Settingorientierte, komplexe Interventionen werden tendenziell, aber nicht ausschliesslich, mit explorativen Methoden untersucht. Dies dürfte der Grund für eine weite Grenzziehung sein

Trotz der sich zeigenden Tendenz deuten die Erkenntnisse aufgrund der beobachteten Ausnahmen eher darauf hin, dass nicht die Art der Intervention die Grenzziehung bestimmt, sondern die Grenzziehung eine Leistung der Evaluation ist. Diesbezüglich bemerkenswert ist E8, wo anhand weniger Faktoren zusammenfassend die Wirkung von jahrzehntelangen vielfältigen Präventionsbemühungen be-

urteilt wird. Die Frage, inwiefern es zur Beurteilung komplexer Interventionen komplexer Forschungsdesigns bedarf, wird im Kapitel 8 erörtert.

Vielfältige Perspektiven

Was die berücksichtigten Perspektiven anbelangt, so werden in allen Evaluationen, welche Personen befragen, primär Vertreterinnen und Vertreter der Trägerschaft, des Teams der Umsetzenden sowie der Zielgruppen berücksichtigt. Ihre Perspektiven fliessen in die Wirkungskonstruktion ein. In wenigen Evaluationen werden zusätzliche Sichtweisen aus dem weiteren Umfeld der Interventionen mit berücksichtigt (E6, 15; E7, 8). Bei der Selektion der Perspektiven zeigen sich zwei Grundprinzipien: einerseits eine möglichst breite Verteilung und gleichzeitig eine Mindestzahl berücksichtigter Perspektiven – zusammenfassend eine ausgewogene Verteilung. Dies lässt sich sowohl bei qualitativ als auch bei quantitativ orientierten Erhebungen beobachten. Im ersten Fall sollen beispielsweise aus mehreren Gruppen jeweils mehrere Personen befragt werden (E1, 45; E7, 9f) oder aber die Fallstudien ausgewogen zusammengestellt werden (E6, 10). Im zweiten Fall wird bei der Zusammenstellung der Stichprobe darauf geachtet, dass die als relevant definierten Kriterien wie Alter, Geschlecht oder soziale Herkunft möglichst gleichmässig verteilt sind (E1, 4; E9, 42), um unterschiedliche Perspektiven vergleichen zu können. Auch bei Beobachtungen (quantitativ oder qualitativ orientiert) wird darauf geachtet, dass die Auswahl nicht einseitig ist (E7, 10; E3, 40). Sind die Evaluationen quantitativ angelegt, werden die Grenzen im Hinblick auf eine statistische Repräsentativität möglichst weit gezogen, es wird also versucht, mit den vorhandenen Ressourcen eine möglichst grosse Stichprobe zu erreichen (E3, 40ff), oder es werden gar Vollerhebungen angestrebt (E2, 29). Kriterium für die Grenzziehung sind dabei u.a. statistische Überlegungen hinsichtlich der Aussagekraft resp. der statistischen Repräsentativität der Ergebnisse (E1, 16).

Die Vielfalt von Perspektiven beschränkt sich aber nicht auf zu befragende Personen, es können auch verschiedene fachliche Perspektiven berücksichtigt werden:

Gleichsam die bisherigen Ergebnisse zusammenfassend werden die lokalen Alkoholpolitiken im Folgenden aus vier unterschiedlichen Blickwinkeln beurteilt: aus der Sicht der Politikwissenschaft, der Wirkungsziele des Projekts, der tatsächlichen Veränderungen in den Gemeinden sowie aus der Sicht des Vollzugs des Jugendschutzes. (E6, 78)

Mit der Wahl der Stichproben wird definiert, wer in die Wirkungskonstruktion einbezogen wird, und wessen Perspektiven unberücksichtigt bleiben. Eine entsprechende Priorisierung respektive Selektion ist aber unumgänglich, nicht zuletzt aus Kostengründen (E4, 14), auch wenn diese Form der Komplexitätsreduktion immer mit Erkenntnisverlust verbunden ist:

Eine Begrenzung liegt schliesslich im Einbezug der Auskunftspersonen. Interessant könnte ein erweiterter Kreis von Auskunftspersonen sein, das wären nebst Trägerschaft, Leitung, Projektausführenden und Mädchen auch betroffene Kinder und Jugendliche, Eltern, Lehrpersonen, andere Jugendorganisationen, Polizei und Jugendanwaltschaft etc. (E4, 19)

Kontext

Die drei settingorientierten Evaluationen mit weiten Grenzen (E2, E6, E7) verstehen die untersuchten Interventionssysteme als stark vom Kontext geprägt und sie untersuchen explizit Einflüsse von aussen auf die Wirkungszusammenhänge. In den genannten Evaluationen werden diese Aspekte bereits in den Fragestellungen sichtbar, die Analysen erfolgen kontextspezifisch und die Kontexteinflüsse spielen auch bei der Wirkungsbeurteilung eine wichtige Rolle:

Die Wirkungen hängen zweitens von den im Midnight Basketball konkret realisierten Prozessen ab und von deren Weiterverarbeitung in den psychischen Systemen der Jugendlichen. Und die-

se Weiterverarbeitung wiederum ist drittens abhängig von Prozessen, die weit ausserhalb des Projekts ablaufen, z. B. in der Familie, in der Schule, in den Medien etc. Dass die Lebenswelt der Jugendlichen durch Midnight Basketball bereichert wird, ist aufgrund unserer Analysen kaum zu bestreiten. Ob diese Bereicherung aber ausreicht, Benachteiligungen in anderen Bereichen auszugleichen, ist durchaus fraglich. Die Untersuchung zeigt auch, dass beispielsweise Stigmatisierungsprozesse bzw. Ausschlussprozesse, die in anderen Kontexten entstehen (z. B. der Schule), im Midnight Basketball fortgesetzt werden. Dasselbe gilt auch für die gewalt- und suchtpreventive Dimension. (E7, 51)

Bei diesen drei Evaluationen handelt es sich um fallvergleichende Studien, bei denen die Wirkungsweisen in den jeweiligen Settings gesondert, also in Bezug zum jeweiligen Kontext, analysiert werden. In weiteren Evaluationen scheinen einzelne Kontextfaktoren am Rande als unabhängige Variablen in den Berechnungen auf, anhand welcher beispielsweise der Einfluss einer Sprachregion, des Migrationshintergrundes etc. auf gemessene Veränderungen geschätzt wird (E3, E9), oder aber externe Einflüsse werden primär als verzerrende Faktoren wahrgenommen, die zu eliminieren sind (E8, 27). Externe Einflüsse erscheinen in diesen Fällen in pauschalisierender Weise und es wird darauf verzichtet, sie detailliert zu untersuchen. Auf der einen Seite gibt es also Evaluationen, welche äussere mitbedingende Faktoren explizit in die Betrachtung mit aufnehmen. Der grössere Teil der untersuchten Evaluationen kommt allerdings ganz oder zumindest fast ganz ohne Kontextbezug aus. Es werden Wirkungen der Intervention auf bestimmte Zielgruppen untersucht, während sowohl die Intervention als auch die Zielgruppen mehr oder weniger isoliert betrachtet werden, resp. nur die Beziehung zwischen den beiden relevant erscheint. Die folgende Textstelle zeigt eindrucksvoll, wie das Suchen nach unverfälschter Wirksamkeit dazu verleiten kann, jegliche Kontextaspekte auszublenden:

Die Wirkung der Präventionsmassnahmen wurde also von allen anderen Einflussfaktoren isoliert, was sicher der schwierigste Teil der Untersuchung war. (E8, 8)

Deduktive versus induktive Grenzziehung

Zum einen orientieren sich Evaluationen also stark an Ansprüchen, Wirkungshypothesen, Zielen und Vermutungen von Auftraggebern und Projektträgern. Zusätzlich liefert die Literatur zentrale Hinweise auf Ursache-Wirkungsbeziehungen, die in die Evaluationen einfließen und so die Grenzen des Evaluationssystems abstecken. Diese Grenzziehung, welche sich aus externen Quellen ableitet, wird hier als deduktive Grenzziehung beschrieben. Wenn sich die Evaluation zu stark an den Interventionszielen orientiert, ist die Gefahr gross, dass die Grenzziehung zu eng wird und andere als die beabsichtigten Wirkungen nicht thematisiert und untersucht werden. Wenn die Wirkungsvermutungen direkt in standardisierte Erhebungsinstrumente abgegossen werden und wenn entsprechend fokussiert untersucht wird, kann viel verloren gehen. Es werden nur diejenigen Wirkungen gefunden, nach denen auch gesucht wird. Der deduktiven Grenzziehung steht die induktive gegenüber, welche aus explorativen methodischen Zugängen („offenes Einfangen“) hervorgeht (E2, 47f; E3, 42; E4, 63). Offenes Fragen (oder Beobachten) richtet den Blick weit über den eigenen Horizont hinaus:

Den am Projekt beteiligten Fachfrauen stellt die Evaluation die Frage nach der Wirkung für die Mädchen folgendermassen: Welche positiven Veränderungen und welche negativen Wirkungen können bei den Mädchen festgestellt werden? (E4, 63)

Dadurch können Aspekte sichtbar werden und in die Wirkungskonstruktion einfließen, die zuvor nicht auf dem Radar der Intervenierenden oder der Evaluierenden waren.

Sie (die Jugendlichen. Anm. d. Verf.) erwerben unmittelbar im Handeln Qualifikationen, die zum Teil auch noch als Ausweis bei der Suche nach einer Lehrstelle genutzt werden. Dies ist eine im Konzept von Midnight Basketball nicht angelegte Dimension, die von den Jugendlichen in das Projekt eingebracht wird. (E7, 50)

Hervorzuheben ist auch der folgende Hinweis, welcher die Bedeutung induktiver Grenzziehung sowohl durch die Intervention als entsprechend auch durch die Evaluation - vor dem Hintergrund der Komplexitätstheorie - unterstreicht:

Das Wirkungspotenzial ist natürlich jenseits dessen einzuordnen, dass Jugendliche am Freitag- oder Samstagabend auf Drogen verzichten und keine Straftaten begehen. Weiter reichende Wirkungen bei den Jugendlichen sind jedoch durch die Projekte weder direkt planbar noch erzielbar, sondern ergeben sich quasi indirekt durch das bereitgestellte Lernfeld. (E7, iv)

Nicht wie bei standardisierten Erhebungen bestimmt dort die Grenzziehung der Evaluation den Fokus, sondern umgekehrt, die Ergebnisse der Erhebungen bestimmen bzw. erweitern in der Regel die Grenzziehung. Rein induktive Grenzziehungen werden allerdings in keinem Fall beobachtet. In allen Evaluationen, welche auch induktive Grenzziehungen beinhalten, gehen deduktive Grenzziehungen voraus. Es werden Wirkungsfelder oder Wirkungsdimensionen definiert, innerhalb welcher offen nach Wirkungszusammenhängen gesucht wird. Wie bereits angesprochen, sind die Zielsetzungen der Interventionen auch hierbei grundsätzlich leitend (E7, 28). Aber auch umgekehrt können induktive Grenzziehungen Ausgangslage für weitere Erhebungen sein:

Nach je zwei Beobachtungsdurchgängen an jedem Standort kristallisierten sich erste („generative“) Fragen heraus, die zu einer stärker strukturierten Beobachtung führte. Entsprechend konzentrierten wir uns auf die Rollenträger, die sozialen Teilräume und die Cliquesbildung. (E7, 28)

Im Vergleich zwischen den deduktiven und den induktiven Grenzziehungen zeigt sich, dass deduktive eher zu Faktoren allgemeinerer Natur führen (E1, E2, E3,...) und induktive Grenzziehungen auch sehr kontext- und zielgruppenspezifische Aspekte in den Fokus rücken (E4, E7) und zu entsprechend differenzierteren Modellen führen können (E7, 72).

Nebenwirkungen sind auch Wirkungen

In den untersuchten Evaluationen werden potenzielle Nebenwirkungen nur selten thematisiert. In der einen Evaluation wird bereits aus der Perspektive von Anspruchsgruppen auf konkrete potenzielle (negative) Nebenwirkungen aufmerksam gemacht. Diese antizipierten Nebenwirkungen werden denn auch gezielt untersucht und beurteilt:

Die Gefahr einer negativen Etikettierung (Labeling) durch die Teilnahme an einer Kurzinterventionsgruppe hielten die Befragten für sehr gering. (E1,47)

Anderorts wird offener nach unbeabsichtigten Nebenwirkungen gefragt (E1, 36; E6, Z-3) und es werden auch entsprechende Wirkungen, negativer wie positiver Art, dokumentiert (E6, 22). Da (unbeabsichtigte) Nebenwirkungen nur da entstehen können, wo Wirkungen im Vorfeld bereits geplant bzw. beabsichtigt wurden, wird in Evaluationen, denen keine klaren Zielvorstellungen zugrunde liegen, der Begriff der Nebenwirkungen nicht sichtbar. In derartigen Evaluationen wird ganz allgemein und offen nach Wirkungen gesucht. In den beiden Evaluationen, welche Wirkungen aus einer übergeordneten gesellschaftlichen Perspektive untersuchen (E5, E8), sind Nebenwirkungen oder negative Auswirkungen der Präventionsbemühungen auch kein Thema. Insgesamt wird kaum explizit nach Nebenwirkungen oder negativen Auswirkungen von Massnahmen gefragt. Auch dort, wo Bezüge zur wissenschaftlichen Literatur gemacht werden, sind unbeabsichtigte und negative Wirkungen kein Thema. Dies mag mit dem engen Forschungsblick auf beabsichtigte Wirkungen zusammenhängen, die nach Bestätigung suchen, andererseits mag dies aber auch eine Konsequenz verschiedener ‚publication biases‘ sein¹⁰⁹.

¹⁰⁹ Einerseits werden in wissenschaftlichen Zeitschriften im Gesundheitsbereich experimentelle und quasi-experimentelle Studien mit entsprechend engem Forschungsdesign bevorzugt behandelt, andererseits gibt es die

Fazit

Die untersuchten Evaluationen gehen unterschiedlich mit der Grenzziehung und Identifikation von relevanten Wirkungsfaktoren um. Wenn die Grenzen eher eng und starr sind, die Grenzziehung deduktiv erfolgt und die Evaluation stark generalisierte Wirkungsfaktoren berücksichtigt, dann haben wir es mit starker Komplexitätsreduktion zu tun, die Interventionssysteme werden also wenig komplex rekonstruiert. Im Gegensatz dazu liegen Rekonstruktionen von hoher Komplexität vor, wenn versucht wird, möglichst viele und spezifische Einflussfaktoren zu berücksichtigen, wenn die Grenzziehung induktiv erfolgt und flexibel gehandhabt wird:

Rekonstruktionen von geringer Komplexität	Rekonstruktionen von hoher Komplexität
enge Grenzen, wenige Einflussfaktoren	weite Grenzen, viele Einflussfaktoren
starre Grenzen	flexible Grenzen
deduktive Grenzziehung, Fokus auf vordefinierte Wirkungsziele	induktive Grenzziehung, Berücksichtigung von nicht intendierten Nebenwirkungen
stark generalisierte Wirkungsfaktoren	kontext- und zielgruppenspezifische Wirkungsfaktoren

Tab. 9 Handlungsspielraum bei der Eingrenzung von Komplexität

Bei der Identifikation von wirkungsrelevanten Faktoren wird einerseits auf wissenschaftliches Wissen, andererseits auf das Fachwissen und auf Wirkungsvermutungen von Stakeholdern abgestützt. Die Evaluationen versuchen auf unterschiedliche Weise verschiedene Perspektiven in die Wirkungskonstruktion einzubinden und diese damit möglichst breit und ausgewogen abzustützen. Die Evaluation vermittelt diese Perspektiven und trägt die fachliche Verantwortung für diese Wahl. Die zur Verfügung stehenden Ressourcen wirken dabei stark begrenzend mit. Über Wirkungsziele und Wirkungsvermutungen sowie über die zur Verfügung gestellten Ressourcen nehmen die Stakeholder von Beginn an grossen Einfluss auf die Grenzziehung und damit letztlich auch auf die Wirkungskonstruktion. Operationalisierte Wirkungsziele bieten der Evaluation eine klare Orientierung für die Beurteilung der untersuchten Interventionen und bestimmen die Grenzziehung massgeblich mit. Sie verleiten aber dazu, den Evaluationsblick auf die antizipierten Veränderungen hin zu verengen. Je offener die angestrebten Wirkungen formuliert sind, desto eher wird ein breites Suchen nach Wirkungen begünstigt. Die Evaluation steht dabei im Spannungsfeld zwischen einerseits vordefinierten Elementen wie Interventionszielen oder aus der Theorie bekannten Wirkfaktoren und andererseits einer potenziell unlimitierten kontext- und zielgruppenspezifischen Komplexität.

Es haben sich in diesem Spannungsfeld zwei Grundmuster der Grenzziehung herauskristallisiert, welche oft kombiniert werden: die deduktive und die induktive Grenzziehung. Die deduktive Grenzziehung leitet die relevanten Faktoren einerseits von Wirkungsvermutungen der Stakeholder ab und andererseits von wissenschaftlichem Wissen über Wirkungszusammenhänge, das sie aus theoretischen Modellen und aus Wirksamkeitsstudien bezieht. Die induktive Grenzziehung hingegen gründet auf explorativen empirischen Untersuchungen der Evaluation und stützt sich mehrheitlich auf Interviews mit Akteuren des Interventionssystems und auf Beobachtungen in den entsprechenden Settings ab. Während deduktive Grenzziehungen eher zu Faktoren allgemeiner Natur führen – u.a. weil es zu sehr spezifischen Fragen meist kaum passende Evidenzen gibt – lenken induktive Grenzziehungen den Fokus eher auf situative, zielgruppen- und kontextspezifische Aspekte. Deduktive und induktive Grenzziehungen werden in vielen der Evaluationen kombiniert, wobei in der Regel aber eine der beiden Strategien im Vordergrund steht und der Evaluation ihr Gesicht gibt.

Wie sich an den Beispielen zeigt, scheint die Art der Interventionen die Grenzziehung zu beeinflussen. Komplexere Interventionen werden eher mit explorativen Methoden untersucht, was zu induktiver Grenzziehung und tendenziell weiteren Grenzen führt. Aber auch das Gegenteil ist beobachtbar: in zwei Evaluationen werden vielfältige settingorientierte Ansätze zusammenfassend anhand weniger vordefinierter generalisierter Indikatoren auf ihre Wirkungen hin überprüft. Hier zeigt sich das ganze Spannungsfeld des Umgangs mit Komplexität und weist auf die Kardinalfrage hin: wie viel Komplexität ist erforderlich, um eine fundierte Wirkungsbeurteilung zu ermöglichen?

7.2 Komplexität erfassen

In diesem Kapitel wird reflektiert, welche Art von Daten in die Wirkungskonstruktion einfließt und wie und mit welchen Konsequenzen die Evaluation selber Daten erhebt. Dabei soll und kann hier nicht eine vertiefte methodologische Diskussion geführt werden, sondern es werden eng am Material einige zentrale Aspekte herausgearbeitet, die für die Wirkungsbeurteilung in komplexem Umfeld bedeutsam erscheinen.

Die Wirkungskonstruktionen bauen auf unterschiedlichen Daten auf, wobei alle Evaluationen mehrere Datenquellen kombinieren. Ausgehend von den im letzten Kapitel beschriebenen Grenzziehungen wurden erstens gezielt bestehende Daten genutzt, zweitens standardisiert Daten erhoben und drittens offen bzw. explorativ Daten erhoben. Folgende Tabelle gibt einen Überblick:

Kategorien	Datenquellen
Bestehende Daten nutzen	Konzepte (E3, 43), Medienberichte (E3, 40), Projektberichte und Protokolle (E4, 14), Statistiken unterschiedlichster Art ¹¹⁰ (E5, 7), Ergebnisse aus anderen Studien (E8, 41), Daten aus Erhebungen Dritter (E2, 31)
Standardisiertes Erheben	Schriftliche Befragungen (E1, 4), strukturierte Interviews (E2, 16), stark standardisierte Beobachtungen ¹¹¹ (E3, 73), Messungen ¹¹² (E9, 12), Erfassung Betriebskennzahlen (E2, 67), Zeiterfassungen (E2, 31)
Offenes Einfangen	Offene Fragen in Fragebogen (E1, 36), Persönliche (E6, 15) und telefonische (E2, 31) Einzelinterviews, Gruppendiskussionen (E2, 31), teilnehmende Beobachtungen (E7, 9f), Betriebsbegehungen (E2, 31), Fotointerviews (E4, 40)

Tab. 10 Unterschiedliche Datenquellen

In allen untersuchten Evaluationen werden verschiedene Methoden und Datenquellen genutzt. Primär geschieht dies, weil die Vielfalt an Wirkungsebenen und Einflussfaktoren unterschiedliche Erhebungsmethoden nahelegt. Jede Erhebungsmethode und mehr noch jede Datenquelle hat ihre klaren Grenzen und innerhalb dieser spezifische Stärken und Schwächen. Mit der Wahl entsprechender Datenquellen und Erhebungsmethoden, unter Berücksichtigung der zur Verfügung stehenden Ressourcen, der Interessen von Anspruchsgruppen und Praxisrealitäten, stellt die Evaluation bedeutende Weichen für die Wirkungskonstruktion. Wie das vorangegangene Kapitel gezeigt hat, werden mit stark strukturierten und standardisierten Verfahren tendenziell engere Grenzen gezogen und mit explorativen Verfahren wird eine grössere Vielfalt an Faktoren, die zum Teil vorher noch nicht als relevant bekannt waren, in die Wirkungskonstruktion aufgenommen.

¹¹⁰ Zum Beispiel Krankenhausstatistiken (E5), Verkehrsunfall- und Todesfallstatistiken (E8) etc.

¹¹¹ Z.B. Erhebung von Helmtragquoten in Skigebieten (E3) oder systematische Tätigkeitsbeobachtungen (E2).

¹¹² Z.B. Messung von Bewegungsverhalten und Körpermassen (E9).

Bestehende Daten nutzen

Der Grossteil der untersuchten Evaluationen stützt sich u.a. auf bereits bestehende Daten ab. Dabei besteht allerdings die Gefahr, dass die Evaluation den Blick zu stark von den verfügbaren Daten leiten lässt und dass jene Faktoren bevorzugt in die Wirkungskonstruktion aufgenommen werden, zu denen auch tatsächlich Daten verfügbar sind. Zudem lassen sich der Umfang und die Qualität der Daten im Nachhinein nicht mehr steuern. In Ausnahmefällen können Lücken in den Daten nachträglich noch geschlossen werden:

Eine Vielzahl von Datenquellen wurde in der Studie eingesetzt. Das Team aus Neuenburg führte zum Beispiel eine Umfrage auf Kantonebene durch, um die verfügbaren Informationen zu den Präventionsmassnahmen und -kosten zu ergänzen (...). (E8, 10)

Bereits bestehende Daten werden zu ganz unterschiedlichen Zwecken gesammelt und ausgewertet. In Zusammenhang mit der Wirkungskonstruktion beispielsweise

- zur Abschätzung des Wirkungspotenzials von Interventionen (z.B. durch Analyse von Konzepten) (E3, 40/43)
- als Hauptquelle für quantitative Wirkungsanalysen (z.B. durch Nutzung epidemiologischer Statistiken) (E5, E8)
- als Hintergrundinformationen zur Einschätzung von Wirkungsmechanismen (z.B. durch Analyse von Protokollen) (E4, 14)

Wenn Daten in guter Qualität bereits vorhanden sind, ist dies ein Glücksfall für die Evaluation, welche dadurch Erhebungsaufwand sparen kann. Zum Teil liegen Daten gar in einem Umfang oder einer Qualität vor, die für Eigenerhebungen unerreichbar wären, für die Wirkungsbeurteilung aber entscheidend sind (am folgenden Beispiel über Trendanalysen und Kontrollgruppenvergleiche):

Das laufende und beinahe vollständige Monitoring aller Hospitalisierungen in der Schweiz ermöglicht die Beobachtung von zeitlichen Veränderungen bezüglich der anfallenden alkoholbedingten Aufnahmen in Spitälern. (...). Wenn sich die Zahl der wegen Alkohol-Intoxikation hospitalisierten Fälle im Kanton Genf („Experimentalgruppe“) nach der Intervention anders entwickelt als in den anderen Schweizer Kantonen („Kontrollgruppe“), kann dieser Effekt als Hinweis auf eine Wirkung der eingeschränkten Erhältlichkeit („Intervention“) interpretiert werden. (E5, 1)

Sitzungsprotokolle, Berichte, interne und externe Statistiken usw. können ergänzend zu neu erhobenen Daten wichtige Hinweise zu Wirkungsprozessen liefern. Ein Blick auf die untersuchten Evaluationen, respektive auf den begrenzten Einsatz derartiger Quellen, lässt vermuten, dass hier noch ungenutztes Potenzial besteht.

Standardisiertes Erheben

In allen Evaluationen, in welchen Datenerhebungen gemacht werden, wird auch standardisiert erhoben, wobei der Grad der Standardisierung sehr unterschiedlich ist. Das Spektrum reicht von geschlossenen Fragen in Fragebogen über Messungen bis hin zu (teil-) strukturierten Beobachtungen und Interviews (s. Tabelle oben). Die Evaluation stützt sich dabei auf jene Faktoren ab, die im Rahmen der Grenzziehung (vorhergehendes Kapitel) als bedeutsam identifiziert worden sind. Dem standardisierten Erheben ist eigen, dass Wirkungsvermutungen bereits vorher vorliegen müssen, weil die Erhebungsinstrumente darauf aufbauen. Selbst wenn diese Wirkungsvermutungen nicht immer explizit gemacht sind, lassen sie sich doch in den Fragestellungen deutlich erkennen:

Wenn man Ihnen eine Arbeit mit weniger Stress vorschlagen würde, wären Sie bereit, weniger zu verdienen? (E2, 68)

Wie im vorangegangenen Kapitel ausgeführt, orientiert sich die Evaluation hinsichtlich der Grenzziehung und Selektion sowohl an den Interventionszielen als auch an wissenschaftlichem Wissen. Wirkungsziele, aber auch Einflussfaktoren liegen selten in sehr konkreter Form vor. Schwer fassbare Konzepte wie Gesundheit, Lebensqualität oder politische Verankerung müssen über ausgewählte Indikatoren fassbar und beurteilbar gemacht werden. Über diese Operationalisierung unternimmt die Evaluation eine starke Komplexitätsreduktion und leistet einen gewichtigen Teil der Wirkungskonstruktion bereits vor der Erhebung (E2, 72ff, E6, 70ff, E9, 35). Die Evaluation kann sich bei der Operationalisierung auf wissenschaftliches Wissen abstützen, sie kann sich mit den Auftraggebern absprechen, sie kann sich auf eigene Erfahrungen abstützen und sie könnte¹¹³ die Perspektive der Zielgruppen der Interventionen dazu einholen. Auf jeden Fall hat die Evaluation hier grossen Ermessensspielraum.

Die Nutzung wissenschaftlichen Wissens bei der Operationalisierung ist in den untersuchten Evaluationen an vielen Stellen gut sichtbar. Häufig basieren die Erhebungsinstrumente auf theoretischen Modellen, über welche die Ziele operationalisiert werden. Diese Modelle fliessen beispielsweise in Interviewleitfäden (E2, 28, E6, 8ff) oder auch in Fragebogen ein:

Stufen der Verhaltensänderung (Precontemplation, Contemplation, Preparation, Action, Maintenance) wurden gemäss dem Transtheoretischen Model mit fünf Fragen erhoben. (E1, 18)

Standardisierten, geschlossenen Erhebungsinstrumenten ist eigen, dass den Befragten enge Schemen vorgegeben sind, in welche sie ihre Antworten einpassen müssen. Es bleibt dabei wenig Raum für Eigeninterpretation und Zwischentöne. Während diese Komplexitätsreduktion in gewissen Fällen unproblematisch ist (z.B. bei einfachen quantitativen Einschätzungen), wird in anderen Fällen der Möglichkeitsraum stark eingeschränkt, und Antworten bzw. Wirkungszusammenhänge, welche nicht in das vorgegebene Schema passen, bleiben ungehört:

- A = Ich denke nicht darüber nach, etwas an meinem Alkoholkonsum zu ändern
- B = Ich denke darüber nach, innerhalb des nächsten halben Jahres etwas an meinem Alkoholkonsum zu ändern
- C = Ich denke darüber nach, innerhalb des nächsten vier Wochen etwas an meinem Alkoholkonsum zu ändern
- D = Ich habe gerade begonnen, etwas an meinem Alkoholkonsum zu ändern
- E = Ich habe meinen Alkoholkonsum im letzten halben Jahr reduziert (E1, 26)

Es gibt unterschiedliche Wege, mit der grossen Vielfalt an Facetten und Faktoren umzugehen, die in schwer messbaren Wirkungsdimensionen enthalten sind. Um sie handhabbar zu machen und sie später in Wirkungsbeurteilungen einfließen lassen zu können, werden sie zum Teil in komplizierte Indexbildungen aufgenommen (E2, 12), andererseits werden besser operationalisierbare Indikatoren mit zusammenfassendem Charakter gesucht:

Angesichts der grossen Anzahl Indikatoren, die zur Berechnung eines Ressourcen-Belastungs-Index verwendet werden, wäre es wünschenswert, ein Metakriterium zur Verfügung zu haben, das optimalerweise mit geringerem Befragungsaufwand dasselbe erfasst. Das arbeitsbezogene Kohärenzgefühl (kurz: Work-SoC) könnte dafür in Frage kommen. (E2, 36)

Zum Teil werden dabei Ersatzindikatoren verwendet, die zu diskussionswürdigen Komplexitätsreduktionen führen:

Diese in den Wirtschaftswissenschaften breit angewandte Methode (kontingente Bewertung, Anm. d. Verf.) ermittelt den Geldwert von auf den ersten Blick nicht messbaren Grössen anhand der maximalen Zahlungsbereitschaft für diese Grössen. Der so ermittelte Geldwert der Grösse

¹¹³ „Könnte“ deshalb, weil dies in den untersuchten Evaluationen in keinem Fall sichtbar wird.

kann als Nutzen dieser Grösse interpretiert werden, also zum Beispiel als Nutzen der Heilung einer Krankheit, als Nutzen einer Reduktion der Luftverschmutzung – oder als Nutzen der höheren Lebensqualität dank verringerter Stressbelastung. (E2, 69)

Standardisierungen sind mit Blick auf zusammenfassende Auswertungen und vor allem hinsichtlich Vergleichen wichtig, auf denen die Evaluation ihre Wirkungsbeurteilungen aufbaut. Quantitative Wirkungsanalysen sind auf standardisierte Erhebungen, welche in numerische Werte transformiert werden können, angewiesen. Standardisierte Erhebungen können die Komplexität der untersuchten Phänomene aber nur sehr begrenzt erfassen. In Kapitel 7.3 wird näher beleuchtet, wie der Prozess der Wirkungskonstruktion auf dieser Basis weiter geführt wird.

Offenes Einfangen

Neben dem standardisierten Erheben werden Erhebungen gemacht, die von grösserer Offenheit geprägt sind und in der Analyse mit „Offenes Einfangen“ umschrieben sind. Antworten sind nicht bereits vorgedacht und Wirkungszusammenhänge noch nicht fest in Modellen abgebildet und in Form von Hypothesen in die Erhebungsinstrumente eingeflossen.

Die Herangehensweise, die für diese Art der sozialwissenschaftlichen Forschung charakteristisch ist, kann am besten als „explorativ“ bezeichnet werden. Das heisst, dass eine Annäherung an den Gegenstand (...) ohne vorgefasste Modellannahmen stattfindet. (E7, 28)

Offenes Einfangen ist nicht an bestimmte Erhebungsmethoden oder -instrumente gebunden, sondern kann in Fragebogen, in Interviews, in Beobachtungen etc. enthalten sein (vgl. Tab. oben). Offenes Einfangen liefert im Gegensatz zu standardisiertem Erheben vertiefte und differenzierte Informationen (E3, 42), ein breites Spektrum an Meinungen (E3, 42), Hintergründe (E4, 15), Zwischentöne (E2, 39). Zudem dient es dazu, ein Untersuchungsfeld kennen zu lernen (E7, 10), Prozesse zu beschreiben (E2, 12), Erlebniswelten zu rekonstruieren (E7, 10), Organisationen in ihrer Eigenlogik zu verstehen (E2, 12), Wirkungsweisen detailliert aufzuschlüsseln (E3, 39) und Wirkungen differenziert zu beurteilen (E3, 42).

Es geht nicht darum, isolierte Wirkungsfaktoren abzufragen, um sie in Berechnungen aufnehmen zu können, sondern darum, Wirkungsweisen in ihrer Komplexität verstehen zu lernen. Entsprechend weit und unbestimmt können Fragedimensionen sein (E7, 10):

Die Interviews kreisten um die Lebenswelt der Jugendlichen, also Schule, Eltern, Gemeinde, um ihren Lebensstil, ihre Erwartungen, Haltungen und Perspektiven, ihre Probleme, um ihr Gesundheits- und Suchtverhalten sowie um das Midnight Basketball Projekt, wie sie dieses Angebot erleben und welchen Einfluss es auf sie in ihrer eigenen Wahrnehmung hat. Ein besonderes Augenmerk galt dem interkulturellen Zusammenleben, Konflikten sowie der Geschlechterperspektive. (E7, 10)

Durch Nähe und gegenseitige Offenheit soll es gelingen, für die Wirkungsweise entscheidende kontextspezifische Feinheiten zu erkennen, die sonst verborgen bleiben würden (E2, 38; E6, 16):

Diese kleinräumigen und vielfältigen Massnahmen sind von „ausser“ für die Forschenden und oft auch für Beratende nicht mehr sichtbar, aber entscheidend für die Gesamtwirkung von SWiNG.“ (E2, 38) und erläuternd dazu in einer Fusszeile: „Organisationsforscher und -berater verwenden für solche Studien oft den Begriff ‚Shadowing‘ und beobachten die Betriebe wie Ethnologen fremde Völker untersuchen.

Methoden- und Datentriangulation

In allen untersuchten Evaluationen werden verschiedene Erhebungsmethoden im Evaluationsdesign kombiniert. Allerdings stehen diese mehrheitlich nebeneinander; Ergebnisse der einen Erhebung fließ-

sen – abgesehen von wenigen Ausnahmen - nicht in die Entwicklung anderer Erhebungen ein (E7, 8). Erst bei der Ergebnisdarstellung werden die Befunde aus den verschiedenen Datenquellen miteinander in Verbindung gebracht. Dabei wird vor allem der sich ergänzende Charakter der verschiedenen Methoden, gerade auch im Hinblick auf die Wirkungsbeurteilung betont:

Die mündlichen Informationen erweitern die schriftliche Materialbasis. Sie erlauben das Einholen von Einschätzungen zum Projekt sowie Angaben zu Veränderungen und Wirkungen bei den Mädchen und den Institutionen. (E4, 14)

Dabei werden u.a. auch deutlich die Grenzen quantitativ orientierter Methoden angesprochen, dies obwohl üblicherweise für Wirkungsnachweise primär auf quantitative Methoden gesetzt wird (Khandker et al., 2010):

Grundsätzlich sollte gesagt werden, dass man nur das sieht, was man misst (oder messen will). Es mag Dinge geben, die sich durch SWiNG verändert haben, für die es aber keine metrischen Skalen gibt. Darum sind neben den statistischen Wirksamkeitsanalysen die Interviews und Diskussionen zur Beurteilung des Projekts mindestens so wichtig. (E2, 33)

Methoden- und Datentriangulationen werden in Zusammenhang mit Wirkungsbeurteilungen vor allem genutzt:

- um ein umfassendes Gesamtbild zu erhalten (E9, 35; E2, 39)
- um Grenzen der einen Methode durch andere zu kompensieren (E3, 41)
- zur Plausibilisierung und gegenseitigen Abstützung von Ergebnissen (E1, 47; E2, 67)

In Kombination mit den anderen Diagnose-Instrumenten und den vielen Gesprächen im Betrieb, mit denen auch die organisationalen Rahmenbedingungen und die menschlichen Zwischentöne erfasst wurden, konnten die Beratenden in ihren Rückmeldungen ein gutes und umfassendes Bild des Betriebs zeichnen. (E2, 39)

Die meisten der untersuchten Evaluationen sind entweder klar hypothesenprüfend oder klar explorativ ausgerichtet, zum Teil mit Ergänzungen aus dem anderen Bereich. Lediglich eine der Evaluationen (E2) kombiniert diese beiden Ansätze etwa gleichwertig. Wie sich gezeigt hat, kombinieren alle der untersuchten Evaluationen mehrere Datenerhebungs- und Auswertungsmethoden, um ein umfassenderes Gesamtbild über die Wirkungsentstehung zu erhalten, um Grenzen der einen Methode durch andere Methoden zu kompensieren oder um Ergebnisse zu plausibilisieren:

In Kombination mit den anderen Diagnose-Instrumenten und den vielen Gesprächen im Betrieb, mit denen auch die organisationalen Rahmenbedingungen und die menschlichen Zwischentöne erfasst wurden, konnten die Beratenden in ihren Rückmeldungen ein gutes und umfassendes Bild des Betriebs zeichnen. In den meisten Betrieben setzte sich das Management mit diesem Bild dann auch ernsthaft auseinander. (E2, 39)

Es hat sich in der empirischen Untersuchung zwar die Tendenz gezeigt, dass settingorientierte Interventionen eher explorativ und mit höherer Komplexität evaluiert werden (A) und dass bei eher verhaltensorientierten Interventionen die Wirkungsbeurteilung tendenziell auf hypothesenprüfende Methoden und geringere Komplexität abstützt (B). Die Ausnahmen (C) zeigen jedoch, dass dies auch ganz anders sein kann. Es gibt verhaltensorientierte Interventionen, die explorativ-qualitativ untersucht werden (E4) und andererseits settingorientierte Interventionen, die mit einem rein quantitativen Ansatz evaluiert werden, ohne dass auf die Komplexität der beurteilten Interventionen eingegangen würde (E5, E8):

		Verhaltens-orientiert	Setting-orientiert	primär hypo-thesenprüfend	primär explorativ	geringe wahr-genommene Komplexität	hohe wahr-genommene Komplexität
A	E1						
	E3						
	E9						
B	E2						
	E6						
	E7						
C	E4						
	E5						
	E8						

Tab. 11 Kategorisierung der untersuchten Evaluationen

Diese drei Dimensionen lassen sich wie folgt in einer Grafik zusammenfassend darstellen, wobei die Dimension Setting- versus Verhaltensorientierung farblich unterschieden wird (settingorientierte Ansätze grau) integriert ist:

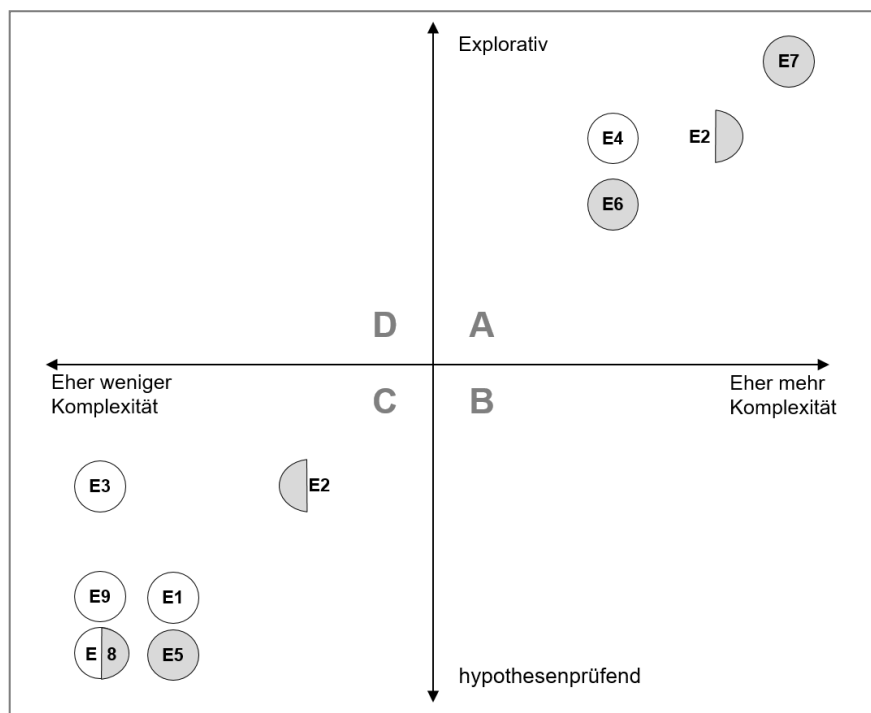


Abb. 14 Komplexität und Design der untersuchten Evaluationen

Die Evaluation E2 verbindet etwa zu gleichen Teilen quantitative wie auch explorativ-qualitative Methodik zur Wirkungsbeurteilung, weshalb sie in der obigen Darstellung in zwei Teilen erscheint, die Evaluation E8 berücksichtigt sowohl setting- als auch verhaltensorientierte Interventionen.

Es fällt auf, dass sich alle Evaluationen in den Feldern A oder C befinden. Die Berücksichtigung höherer Komplexität geht mit explorativ-qualitativer Methodik einher, geringere Komplexität mit quantitativ-prüfenden Verfahren, wobei zunächst nicht klar ist, ob die gewählte Methodik dazu führt, mehr oder weniger Komplexität zu berücksichtigen oder umgekehrt, dass das Mass an wahrgenommener Komplexität nach entsprechender Methodik verlangt. Dass settingorientierte Interventionen eher explorativ-qualitativ untersucht werden, mag ein Hinweis darauf sein, dass die wahrgenommene Komplexität die Methodik bestimmt. Die Ausnahmen zeigen aber, dass dies nicht so sein muss und die Evaluation Spielraum hat und auch nutzt, und über die Methodenwahl ein Mehr oder Weniger an Komplexität in

die Evaluation einbezieht. Diese Ausnahmen sind besonders interessant und bedürfen genauerer Betrachtung:

Bei E4 handelt es sich um eine verhaltensorientierte Massnahme, die sich an eine Gruppe von Personen richtet (gewaltbereite Mädchen). Diese Intervention könnte auch gut mit einem Kontrollgruppendesign untersucht werden, wobei die Intervention die unabhängige und das Gewalthandeln die abhängige Variable darstellen würde. Über die Gründe, weshalb die Evaluation nicht diesen Weg gewählt hat, kann nur gemutmasst werden. Ein Kontrollgruppendesign würde verschiedene voneinander unabhängige Settings bedingen, die entweder als Interventions- oder Kontrollgruppen untersucht werden würden. Dazu bedürfte es zunächst überhaupt Interventionen in mehreren Settings (dies scheint im vorliegenden Fall nicht gegeben), und ein viel aufwändigeres Evaluationsdesign wäre die Folge. Die Evaluation hat sich entschieden, einen anderen Weg zu gehen. Sie geht von der einen Intervention aus und untersucht diese explorativ-qualitativ und ohne Vergleichsgruppe. Die Vergleiche, auf welchen die Beurteilung basiert, sind Vorher-Nachher Vergleiche, zwar aus unterschiedlichen Perspektiven, aber ohne den Vergleich zur Situation ohne die Intervention („counterfactual“). Die Wirkungsbeurteilung erfolgt primär durch die befragten Zielgruppen- und Institutionsvertreterinnen und –vertreter und wird von der Evaluation übernommen. Die Evaluation gewinnt durch den explorativ-qualitativen Ansatz vertiefte Einblicke in die Wirkungsmechanismen, sie beansprucht im Gegensatz zu anderen Evaluationen aber keine ‚harten‘ Fakten, womit in diesem Falle wohl Zahlen gemeint sind:

Die Grenzen der Evaluation sind mehrfach. Es liegen keine Resultate vor, die in einer «harten» Einheit eine Reduktion oder eine Zunahme der Gewalt und ihrer Auswirkungen angeben könnten. (E4, 19)

Auf der anderen Seite stehen die beiden Evaluationen E5 und E8, welche zwar settingorientiert sind (E5) oder zumindest settingorientierte Ansätze beinhalten (E8), aber die damit verbundene hohe Komplexität in keiner Art und Weise zum Gegenstand machen. Diese beiden Evaluationen vergleichen die Entwicklung des Interventionssystems anhand einzelner Impact-Indikatoren mit Entwicklungen anderer ‚ähnlicher‘ Systeme (in diesem Fall Kantone) ohne vergleichbare Intervention:

Wenn sich die Zahl der wegen Alkohol-Intoxikation hospitalisierten Fälle im Kanton Genf („Experimentalgruppe“) nach der Intervention anders entwickelt als in den übrigen Schweizer Kantonen („Kontrollgruppe“), kann dieser Effekt als Hinweis auf eine Wirkung der Intervention interpretiert werden. (E5, 26)

Da die Mechanismen der Wirkungsentstehung allerdings nicht Gegenstand der Untersuchungen sind, bleiben viele offene Fragen bei der Interpretation der Ergebnisse.

Einfluss mentaler Modelle auf die Wirkungsbeurteilung

Anhand der untersuchten Evaluationen lässt sich zeigen, wie die Perspektiven der Evaluation und der Anspruchsgruppen über die Art der Datenerhebung unterschiedlich in die Wirkungsbeurteilung einfließen. Ausgangslage im Erhebungsprozess sind die mentalen Modelle¹¹⁴ der Evaluation, also deren Konzeptualisierung der Wirkungsweise der untersuchten Intervention. Diese Modelle fließen in die Entwicklung der Erhebungsinstrumente ein. Bei standardisierten Erhebungen sind die impliziten oder expliziten Modelle der Evaluation stark bestimmend, bei offenen Erhebungsmethoden ist der Einfluss kleiner. Auf der anderen Seite sind es die mentalen Modelle der Anspruchsgruppen, welche über den Prozess der Datenerhebung selber einfließen. Bei offener Erhebungsmethodik fließen die mentalen Modelle der Anspruchsgruppen dominant ein, bei standardisierten Erhebungen ist ihr Einfluss schwächer. Bei der Nutzung bereits bestehender Daten kann der Einfluss der mentalen Modelle der An-

¹¹⁴ Mentale Modelle sind modellhafte Repräsentationen der Wirklichkeit in den Köpfen der Menschen (Repräsentationssysteme) (Johnson-Laird (1983)).

spruchsgruppen je nach Datenquelle (Interventionskonzepte, Protokolle, Statistiken,...) stark variieren oder gar inexistent sein (bei Datenquellen Dritter). Über die Datenauswertung sind es noch einmal die Modelle der Evaluation, welche prägen, bevor es zur eigentlichen Wirkungsbeurteilung im Sinne einer Synthese kommt. Bei der Delegation der Wirkungskonstruktion nehmen die Befragten direkt die Wirkungsbeurteilung vor. Zumindest in den untersuchten Evaluationen fließen diese, ohne noch einmal durch die Evaluation kritisch reflektiert zu werden, direkt in die Evaluationsergebnisse ein.

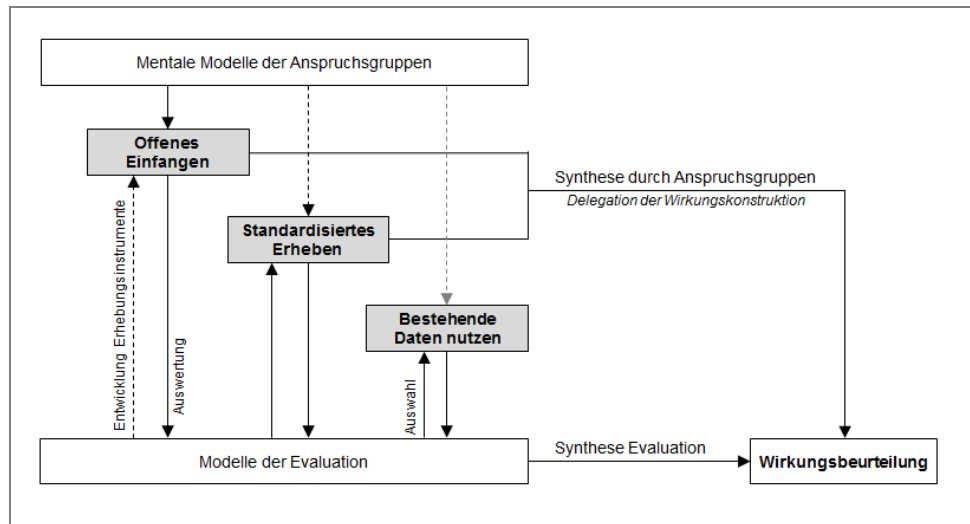


Abb. 15 Einfluss von Modellen auf die Wirkungskonstruktion

Bei Selbstevaluationen (die bei Wirkungsevaluationen allerdings eher selten sind), überlagern sich die Modelle der Anspruchsgruppen und der Evaluation.

Delegation der Wirkungskonstruktion

Eine extreme Form des Einflusses der mentalen Modelle der Anspruchsgruppen auf die Wirkungskonstruktion wird dort sichtbar, wo Anspruchsgruppen direkt um Wirkungsbeurteilungen gebeten werden. Diese Form wurde in der Analyse häufig und in unterschiedlichen Variationen gefunden und mit „Delegation der Wirkungskonstruktion“ betitelt. Zum Teil wird sehr umfassend und offen nach Wirkungen gefragt:

Die Frage, was das Projekt MuM bewirkt hat, in welcher Weise es für die Mädchen Wirkungen zeigte, wurde in der Evaluation auch den Nutzerinnen selbst – den Mädchen – gestellt. (E4, 61)

Aber auch geschlossene Frageformen sind möglich, hier in differenzierter Art und auf der Grundlage entsprechender Wirkungsvorstellungen der Evaluation:

Die (...) Fragen konnten auf 7 Stufen (von 1 = „nein, gar nicht“ bis 7 = „ja, sehr“) beantwortet werden:

Hatten diese Aktivitäten einen positiven Einfluss auf den Betrieb?

Hatten diese Aktivitäten einen positiven Einfluss auf Sie selbst?

Hat sich durch SWiNG Ihr persönlicher Umgang mit Stress verändert?

Wird durch SWiNG das Thema Stress in Ihrem Bereich aktiver angesprochen als früher?

Haben sich durch SWiNG Veränderungen im Verhalten Ihrer Vorgesetzten ergeben? (E2, 36)

Hier zeigt sich, wie das Problem der Attribution von Wirkungen zu Massnahmen, welches die Evaluation in der Regel vor schier unlösbare Aufgaben stellt, gekonnt umgangen, respektive delegiert wird - zumindest zunächst. Die Zuschreibungen im zitierten Beispiel dürften allerdings auch für die Anspruchsgruppen alles andere als leicht zu machen sein (vgl. insbesondere das letzte Item).

Anspruchsgruppen sind sich teilweise durchaus der Schwierigkeit einer Wirkungszuschreibung bewusst und weisen in Rückmeldungen – insofern offene Befragungsformen derartige Relativierungen zulassen - darauf hin:

Die Projektausführende ist vorsichtig im Benennen konkreter, positiver Auswirkungen. Sie begründet ihre Zurückhaltung zum einen damit, dass es bei ihrer Arbeit schwierig sei, zu beurteilen, worauf eine Veränderung genau zurückzuführen sei, ob auf das Projekt MuM oder andere Einflüsse wie Mädonna, Peerumfeld, Schule, Eltern usw. (E4, 63)

Ähnlich kritisches Hinterfragen subjektiver Wirkungseinschätzungen findet sich in den untersuchten Evaluationen nicht. Trotz der Schwierigkeiten entsprechender Zuschreibungen wird der subjektiven Wirkungseinschätzung durch Anspruchsgruppen angesichts der grossen Komplexität von Wirkungszusammenhängen teilweise sogar vorrangige Bedeutung zugesprochen:

Dieser Vorgehensweise liegt auch die Überzeugung zugrunde, dass kein objektives Kriterium besser in der Lage ist, die Komplexität der Gesamtwirkung von SWiNG – mit all den direkten und indirekten, bewussten und unbewussten, fassbaren und nicht fassbaren Komponenten – besser abzubilden, als die subjektive Einschätzung der Mitarbeitenden der SWiNG-Betriebe, ob bei ihnen SWiNG „gewirkt“ hat. (E2, 36)

In diesem Zitat und vor dem Hintergrund der umfassenden Datenerhebungen der betreffenden Evaluation zeigt sich deutlich das Dilemma, in welchem sich die Evaluation befinden kann, wenn es um Wirkungsbeurteilungen geht. Einerseits wird versucht, über umfassende Modelle und ausgeklügelte Analysen der Komplexität möglichst gerecht zu werden – um dann aber gleich wieder davor zu kapitulieren. Ähnlich verlockend sind denn auch zusammenfassende Wirkungsbeurteilungen in folgender, in Befragungen häufig anzutreffender Form:

Würden Sie SWiNG weiterempfehlen? Würden Sie SWiNG einführen, wenn Sie an einem anderen Ort Chef des Betriebs wären? (E2, 83)

Antworten auf solche Fragestellungen wären sehr schwierig zu interpretieren wenn nicht Raum für Diskussion gegeben würde. In der vorliegenden Evaluation war dieser Raum gegeben, die Frage diente als Einstieg in eine Gruppendiskussion, welche nach den Gründen für oder gegen eine Empfehlung fragte. Eine Empfehlung kann in Zusammenhang mit der beabsichtigten Projektwirkung stehen, muss aber nicht zwingend. Es können z.B. Nebeneffekte sein, welche zu einer Empfehlung führen. Dieses Beispiel zeigt die Wichtigkeit vertiefter Erläuterung pauschaler Wirkungsbeurteilungen. Wertvolle Hinweise liefern Rückmeldungen, die sich explizit auf bestimmte Wirkfaktoren beziehen. Sie ergänzen pauschale Wirkungsbeurteilungen und liefern wichtige Hinweise darauf, wie Wirkungsmechanismen gesehen werden und wie die Wirksamkeit aus Sicht der Befragten verbessert werden könnte:

Um Betroffenheit auszulösen, hätte die Kampagne ihrer Ansicht nach vermehrt schockierende Bilder vermitteln müssen. «Es sind ja nur Helis, die vorbeifliegen. Das reicht nicht, es passiert zu wenig.» (E3, 52)

Auf andere abstützen

Die Evaluation stützt sich bei der Wirkungskonstruktion also stark auf die Perspektiven verschiedener Anspruchsgruppen ab. Zunächst bei der Übernahme der Wirkungsvermutungen und Wirkungsziele

und in einem zweiten Schritt über die Datenerhebungen. Bei Befragungen beispielsweise können die gegebenen Antworten in ihrer Komplexität stark variieren. Während es bei einfachen Angaben oder Schätzungen in Zahlen um eher einfache Sachverhalte geht, die weitgehend überprüfbar sind und wenig Interpretationsspielraum lassen, kann es sich bei Einschätzungen, Begründungen und Beurteilungen um hoch komplexe Sachverhalte handeln:

Antwortkategorien	Beispiele
Einfache Angaben	Einfache soziodemographische Angaben wie Alter oder Geschlecht, Anzahl Veranstaltungsteilnahmen (E2,30)
Schätzungen in Zahlen	Monatlicher Alkoholkonsum (E1,18), Alter der beobachteten Personen (E3,40), Bewegungs- und Ernährungsverhalten (E9,35), Investitionen und Personalaufwände (E8,34)
Erfahrungen	Eigenes Gewalthandeln (E4,52), Erleben von Angeboten (E4,10), Gesundheitsverhalten (E4,10)
Einstellungen	Wichtigkeit von Veränderung (E1,18), Akzeptanz von Massnahmen (E3,52), Erwartungen (E7,10)
Einschätzungen	Konsequenzen des eigenen Alkoholkonsums (E1,18), Einfluss von Kontextbedingungen (E6,15), Unbeabsichtigte Wirkungen von Interventionen (E6,15), Ursachen für Gewalthandeln (E4,54), Einschätzung Arbeitsaufwand (E4,66)
Begründungen	Begründungen für präventives Verhalten (E3,58), für mangelnde Interventionswirkungen (E3,52)
Beurteilungen	Beurteilung von Massnahmen (E2,29) oder von Wirkungen (E2,36; E4,66), von Strukturen und Prozessen (E3,49), Beurteilung der eigenen Arbeitsleistung (E2,68),
Projektionen	Projektionen mit zeitlicher und räumlicher Distanz: Beurteilungen in die Zukunft (E1,47), Zuversicht etwas verändern zu können (E1,18), Bereitschaft weniger zu verdienen (E2,35), Einschätzung der Interventionswirkung bei anderen (E4,63)

tendenziell steigende Komplexität



Tab. 12 Unterschiedliche Komplexität von Befragungsdaten

Auch bei Messungen, Beobachtungen und anderen Erhebungsarten kann sich die Komplexität der abgebildeten Sachverhalte erheblich unterscheiden. Erhobene Daten sind von der Zeit und der Art der Erhebung, vom Erhebungssetting, von der momentanen Befindlichkeit und vielen anderen situativen Einflüssen geprägt. Einstellungen, Einschätzungen, Begründungen oder Beurteilungen können sich laufend verändern. Was in die Wirkungskonstruktion einfließt, ist also nicht objektiv und unveränderlich, sondern in den allermeisten Fällen eine Momentaufnahme unter gegebenen Umständen. Entsprechend fließen Verzerrungen in die Wirkungskonstruktion mit ein, wie beispielsweise sozial erwünschte Antworten oder andere Antworttendenzen (response biases). Entsprechend ist auch eine darauf aufbauende Wirkungskonstruktion als eine Momentaufnahme unter den gegebenen Kontextbedingungen zu verstehen. Sie ist stark durch die Anspruchsgruppen geprägt.

In diesem Zusammenhang ist ein kritischer Blick auf den Begriff des Messens angezeigt. Einerseits wird er im herkömmlichen Sinne¹¹⁵ verwendet für die Messung z.B. von Körpermassen wie Grösse, Gewicht oder Blutdruck. Auf der anderen Seite wird er in den Sozialwissenschaften aber auch breiter verwendet, u.a. für die Messung von Einstellungen, Kompetenzen oder auch Wirkungen. In den untersuchten Evaluationen werden beispielsweise Stufen der Verhaltensänderung gemessen (E1,34), Zieler-

¹¹⁵ Messen wird definiert als Zuordnung von Zahlen zu Objekten oder Ereignissen, sofern diese Zuordnung eine homomorphe (eindeutige) Abbildung eines empirischen Relativs in ein numerisches Relativ ist.

reichung (E6, 71), Sensibilisierung (E3, 58), Nutzen (E2, 16) Erfolg (E3, 89), Nachhaltigkeit (E6, 9) und gar Wirkungen (E1, 48; E8, 44).

Ein wichtiges Anliegen der bfu ist, Kampagnen objektiv messbar zu konzipieren und deren Wirkung belegen zu können. (E3, 5)

Die Verwendung des Begriffs Messen mag eine Präzision und Eindeutigkeit vortäuschen, die angesichts der beschriebenen subjektiven und situativen Einflüsse so nicht gegeben ist.

Grenzen der Erfassung

Da die Wirkungskonstruktion der Evaluation auf den erhobenen Daten aufbaut, lohnt sich ein Blick auf weitere Faktoren, welche die Datenqualität und dadurch auch die Aussagekraft der Ergebnisse beeinträchtigen können. In den untersuchten Evaluationen lassen sich folgende Schwierigkeiten in verschiedenen Phasen des Prozesses der Datenerhebung ausmachen:

Bei der Methoden- und Instrumentenentwicklung: Unterschiedliche Begriffsverständnisse (E1, 46), Unschärfen bei der Definition von Kategorien (E4, 68; E5, 7), begrenzte Messbarkeit (E2, 33/39/43), Limitierungen und Verzerrungen durch Operationalisierungen und Verwendung von Proxys (E5, 2), zu wenig feine Skalen (E2, 59), Verlust an Tiefenschärfe durch nivellierende Fragestellungen (E1, 18, E2, 29), arbiträre Grenzziehungen/'cutoffs' (E3, 63), Beeinflussung durch die Art der Frageformulierung (E1, 43), Qualität der Übersetzungen der Erhebungsinstrumente (E2, 116) etc.

So kann es schwierig sein den Wert von weniger Stress genau anzugeben, da ein solches Gut nicht auf dem Markt gekauft werden kann. Ebenso führt die Methode dazu, dass der Stress von Personen mit höherem Einkommen stärker bewertet wird, was aus Gerechtigkeitsgründen unerwünscht sein kann. Trotzdem erlaubt es diese Methode Nutzen zu bewerten, die sonst quantitativ gar nicht erfassbar wären. (E2, 69)

Hier wird klar, dass „eine Teilnahme“ sehr unterschiedliche Ausprägungen haben kann. Bei Projekten dieser Grösse ist es beinahe unmöglich, objektiv für jede Person festzuhalten, wann sie welchen Kurs wie intensiv und mit welchen Inhalten besucht hat und diese Information anhand einer Identifikationsnummer mit der S-Tool-Befragung zu verknüpfen. (E2, 45)

Bezüglich Stichprobe: Begrenzte Perspektiven (E2, 66; E3, 55; E4, 19), Selektionseffekte (E2, 43), mangelnde Teilnahmebereitschaft resp. begrenzter Rücklauf (E2, 73; E9, 98), zu kleine Fallzahlen (E2, 54; E5, 16), Verzicht auf Erhebungen wegen hohem Aufwand (E3, 44), Ausschluss von Fällen aufgrund Unvollständigkeit der Angaben (E1, 21/29), Doppelerfassungen (E2, 74; E8, 53ff), Dropouts/Panelverluste (E2, 116; E2, 19; E1, 29) etc. Das folgende Zitat zeigt nicht nur den Verlust von Befragungsteilnehmenden, sondern lässt über die ungleiche Verteilung der Dropouts auf Interventions- und Kontrollgruppe auch das Verzerrungspotenzial erahnen:

Tabelle 19 zeigt wiederum (vgl. auch Tabelle 2), dass vermehrt Lernende in den Kontrollgruppen und dies insbesondere bei den männlichen Jugendlichen verloren gegangen sind. Besonders stark fällt dies ins Gewicht bei der Gruppe R, wo in den Interventionsklassen „nur“ 18.1% (n=15) verloren gegangen sind, jedoch 38.8% in den entsprechenden Kontrollklassen (n=31). (E1, 29)

Bei den als randomisierte kontrollierte Studien konzipierten Evaluationen zeigt ein Blick in die Stichproben und in die Randomisierung, dass hier die Qualitätsansprüche an ein solches Design kaum eingehalten wurden. In E1 erfolgte weniger als die Hälfte aller Gruppenzuteilungen wirklich randomisiert:

Neun Berufsschulen nahmen an der Studie teil, 3 Schulen jedoch nur unter der Voraussetzung, dass alle an der Studie teilnehmenden Klassen auch von der Intervention profitieren könnten. Zwei Schulen nahmen nur unter der Voraussetzung teil, dass der Klassenverband nur durch die Befragungszeit mit den Fragebögen belastet werden würde. Die Klassen dieser fünf Schulen wurden gesamthaft entweder zur Interventions- oder Kontrollgruppe zugeordnet. (E1, 16)

In E9 war die Anzahl der randomisierten Kindergärten letztlich derart klein (acht Interventionsgruppen versus sieben Kontrollgruppen im Ernährungsmodul und 17 Interventionsgruppen und 20 Kontrollgruppen im Bewegungsmodul), dass bedeutende signifikante Ergebnisse weder erwartet noch gezeigt werden konnten (auf Berechnungen zur erforderlichen Stichprobengröße wird im Bericht nicht eingegangen). Das Verhältnis zwischen Aufwand und Erkenntnissen kann entsprechend kritisch hinterfragt werden.

Im Datenkorpus: Fehlende oder unspezifische Daten (E2, 16; E2, 38), die mit Annahmen und Schätzungen kompensiert werden müssen (E3, 40; 8, 53), Verfälschungen durch das Erhebungssetting (E1, 47; E1, 50), ‚reporting bias‘ (E1, 50; E2, 43/64; E9, 58), allgemein ungenügende Datenqualität (E2, 81), Daten schwierig zu interpretieren (E1, 36), Unschärfen und Kompromisse bei Zuordnungen zu Kategorien (E2, 30), Veränderungen der Bedingungen während der Erhebungszeit (E5, 13) etc.

Abgesehen von fehlenden qualitativen und zu unspezifischen quantitativen Daten auf Einheiten-Team- und Hotspot-Ebene war dieses Vorgehen aufgrund der undurchführbaren Vergleichsgruppenbildung unmöglich, die zentral ist für diesen Teil der statistischen Wirksamkeitsanalyse. (E2, 30)

Dort, wo unterschiedliche Methoden zur Erhebung desselben Sachverhalts eingesetzt werden (z.B. Befragung und Messung), weichen die Ergebnisse oft deutlich voneinander ab (E2, 45; E9, 13ff). In einigen Evaluationen finden sich Formulierungen, welche darauf hindeuten, dass sich die Evaluation der Grenzen der Erfassungen und Lücken in den Daten durchaus bewusst ist (E1, 47):

Im Rahmen der Evaluation von SWiNG wurden die verschiedenen Nutzenarten so gut wie möglich mit dem S-Tool, einem Fragebogen zu den betrieblichen Kennzahlen und einem Fragenbogen zu den Gesundheitskosten gemessen. (E2, 67)

sowie

Bei Abschluss von SWiNG zum Zeitpunkt t2 konnten einige Wirkungen nur unvollständig gemessen werden, da einige Interventionen erst kurz zuvor durchgeführt wurden. (E2, 82)

Zudem finden sich in der einen Evaluation Hinweise, dass sich auch die Befragten selber zur Qualität der Erfassungen kritisch äussern:

In den Spitälern wurden die abgefragten Indikatoren als zu wenig passend empfunden. Ebenso wurden die Fragen hinsichtlich der Vorgesetztenbeurteilung nicht überall als präzise genug empfunden, was auch abhängig von den jeweiligen Führungsebenen und ihrer Unterschiedlichkeit ist. (E2, 39)

und

Allerdings gab es auch Kritik an der Befragung. So waren die Unterschiede in den Ergebnissen der Befragungswellen schwer zu interpretieren und einige Fragen für ein Spital teils wenig passend. Sie waren schwierig zu beantworten und wurden auch als schlecht übersetzt empfunden. So war dann auch der Rücklauf der letzten Befragungswelle sehr tief. (E2, 116)

Es fällt auf, dass Schwierigkeiten und Grenzen bei der Erfassung in nur wenigen der untersuchten Evaluationen explizit thematisiert und daraus folgende potenzielle Verzerrungen transparent gemacht

werden. Insbesondere E2 reflektiert entsprechende Sachverhalte transparent und selbstkritisch und auch in E8, wo zu einem grossen Teil auf Schätzungen aufgebaut wird, werden diese transparent behandelt. In anderen Evaluationen sind entsprechende Hinweise und kritische Reflexionen selten oder fehlen gänzlich.

Fazit

Bei der Erfassung von Komplexität zeigt sich der Handlungsspielraum der Evaluation wiederum deutlich. Die Evaluation kann das Interventionssystem aus Distanz, mit hoch standardisierten Instrumenten und hypothesenprüfendem Blick untersuchen oder aber aus grösserer Nähe, mit explorativen Forschungsmethoden, um Raum für differenzierte Interpretationen und neue Gesichtspunkte zu geben:

Rekonstruktionen von geringer Komplexität	Rekonstruktionen von hoher Komplexität
wenige Perspektiven	Vielfalt von Perspektiven
hypothesenprüfende Zugänge mit hoch standardisierten Erhebungsinstrumenten	explorative Zugänge, offenes Einfangen
nur eine oder sehr wenige Methoden	Kombination verschiedener Methoden
distanzierte Beobachtung	Differenziertheit durch Nähe
wenig Raum für Erklärungen, Hintergründe, Zwischentöne	Raum für Erklärungen, Hintergründe, Zwischentöne
Erfassen weniger Momentaufnahmen	Erfassen dynamischer Entwicklungen
strikte durchgezogenes Evaluationsdesign	emergentes Evaluationsdesign

Tab. 13 Handlungsspielraum bei der Erfassung von Komplexität

Bei der Verwendung von stark strukturierten und standardisierten Erhebungsmethoden erfolgt die Wirkungskonstruktion weitgehend bereits vor der Datenerhebung. Die Wirkungskonstruktion basiert primär auf wissenschaftlichem Wissen und auf allgemeinen Wirkungsvermutungen von Stakeholdern. Beim Einsatz explorativer Forschungsmethoden werden die Wirkungszusammenhänge im Gegensatz dazu nach und nach konstruiert. Explorative Ansätze und die Nähe zum Forschungsgegenstand, in den untersuchten Evaluationen über Beobachtungen und Interviews realisiert, führen zu kontextspezifisch differenzierten Wirkungskonstruktionen. Sie erweisen sich als gut geeignet, den Blick für potenzielle unbeabsichtigte Wirkungen, positiver wie negativer Art, zu öffnen. Eine Sonderform der Wirkungskonstruktion wird dort sichtbar, wo Befragte explizit darum gebeten werden, Wirkungszuschreibungen zu machen: die Evaluation delegiert die Wirkungskonstruktion zumindest teilweise an die Befragten.

Der Grossteil der untersuchten Evaluationen bindet unterschiedliche Perspektiven in die Wirkungskonstruktion ein und versucht diese möglichst breit und ausgewogen abzustützen. Dies sowohl bei hypothesenprüfenden als auch explorativ ausgerichteten Evaluationen. Im Rahmen der Datenerhebung (und –auswertung) wird die Wirkungskonstruktion stark von den mentalen Modellen der Befragten geprägt. Je komplexer (gerade auch im Sinne von Kontingenz), kontextabhängiger und persönlicher die erfragten Sachverhalte sind, umso wahrscheinlicher scheint es, dass momentane Befindlichkeiten sowie zeit- und kontextspezifische Einflüsse über die Antworten in die Wirkungsbeurteilung einfließen.

Angesichts des grossen Spielraums der Evaluation bei der Datenerhebung und der unzähligen Verzerrungen und Unschärfen, welche die Datenerhebung (und –auswertung) begleiten, wird eindrucksvoll sichtbar, dass es sich bei Wirkungskonstruktionen nicht um ein Abbild objektiv gegebener Wirklichkeit handelt, sondern um personen-, zeit- und kontextgebundene soziale Konstruktionen. Je nachdem, wie der vorhandene Spielraum genutzt wird, entstehen potenziell andere Wirkungskonstruktio-

nen. Methoden- und Datentriangulationen werden in allen untersuchten Evaluationen verwendet, um die Wirkungskonstruktionen breiter abzustützen und so zu plausibilisieren.

Die untersuchten Evaluationen verwenden weitgehend starre Evaluationsdesigns, die sich nicht den Entwicklungen der Interventionssysteme anpassen. Die Dynamiken in den Interventionssystemen scheinen sich nicht auf das einmal festgelegte Konzept zur Wirkungsbeurteilung auszuwirken. Sowohl die Wirkungsfragen als auch der Weg zu deren Beantwortung scheinen davon weitgehend unbeeinflusst.

7.3 Komplexität verarbeiten

Strategien zur Wirkungsbeurteilung

Im Hinblick auf Wirkungsbeurteilungen lassen sich vier Kategorien von Strategien zur Verarbeitung von Komplexität identifizieren, die in Kombination genutzt werden: Berechnen, Beschreiben, Vergleichen, und dahinterliegend, Strukturieren:

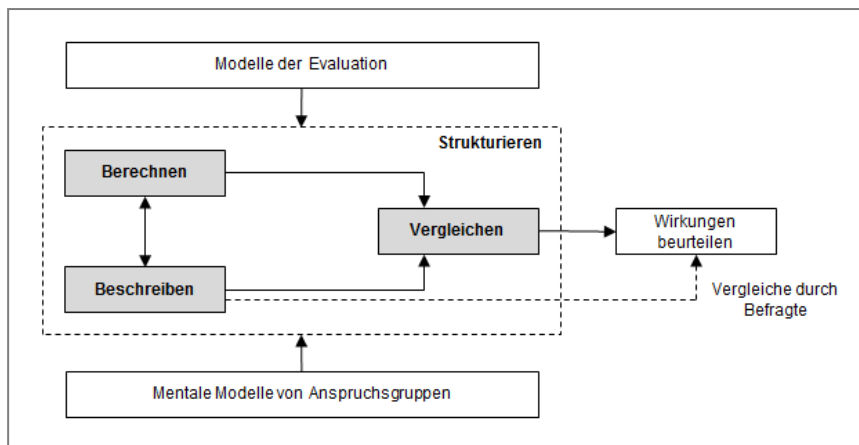


Abb. 16 Strategien der Verarbeitung von Komplexität

Strukturieren

Bei der Wirkungskonstruktion ist die Evaluation laufend gefordert, eine Vielzahl von Informationen angemessen zu berücksichtigen, Grenzen zu ziehen und die relevanten Informationen zwecks Übersichtlichkeit und Handhabbarkeit zu ordnen. Dieses Strukturieren beginnt gleich zu Beginn des Evaluationsprozesses, zieht sich durch alle Evaluationsphasen hindurch und spiegelt sich in verschiedenen Dokumenten wieder, die die Evaluation erstellt: in Evaluationskonzepten, Planungshilfen, Erhebungsinstrumenten und letztlich auch in den Evaluationsberichten. In den untersuchten Evaluationsberichten werden unterschiedliche Strukturierungsformen sichtbar: Die Evaluation klassifiziert (E4, 45; E5, 8), sortiert (E3, 54/66), fokussiert (E4,40; E2, 69), priorisiert (E3, 61/12), ordnet zu (E8, 75), betont (E2, 48; E3, 56), wertet (E3, 14; E4, 62), filtert (E2, 138), kombiniert (E8, 74), verdichtet (E2, 29; E3, 63), schliesst aus (E3, 65; E8,75) und modelliert auf vielfältige Weise (E2, 52/32; E7, 28).

Dem Strukturieren kommt eine bedeutende Funktion im Prozess der Wirkungskonstruktion zu: hier wird am Gerüst der Wirkungskonstruktion gebaut, indem einzelnen Elementen tragende Rollen zugewiesen werden und andere aussortiert werden, Elemente miteinander verknüpft werden, Komplexität für die Wirkungsbeurteilung überschaubar gemacht und reduziert wird:

Klassifizieren dient beispielsweise dazu, ungleiche Verteilungen zu erkennen (E4, 50; E3, 54), Wirkungsfelder (E7, 30; E4, 64f), Wirkungspotenziale (E3, 62; E6, 69) und Wirkungsstufen (E1, 26; E3, 43) zu unterscheiden, Wirkungsfaktoren zu bündeln (E2, 11; E8, 35), Analyseraster zu entwickeln (E6, 17; E7, 29), aber auch Ergebnisse verständlich darzustellen (E3, 62; E6, 68). Klassifizieren ist eine analytische Form der Komplexitätsreduktion. Das Zusammenfassen von Einzelementen zu Kategorien gibt Struktur und ermöglicht systematische Vergleiche:

Im Vergleich der Gemeinden zeigt sich eine Fokussierung auf vier Ausprägungen: Sensibilisierung/Information, Verpflichtung, Kontrolle und Vollzugsmaßnahmen. Es erstaunt nicht, dass Sensibilisierung und Information in allen Gemeinden betrieben werden, handelt es sich doch um die einfachste Massnahme, die wenig Widerstand hervorruft. (E6, 69)

Kategorien können dabei entweder induktiv entwickelt (E7, 29) oder aus der Literatur (E6, 17) oder anderen Quellen abgeleitet werden, wobei auch induktiv entwickelte Kategorien sich nicht einfach aus dem Nichts, wie von allein, ergeben, sondern durch mehr oder weniger starke und mehr oder weniger explizite theoretische Modelle derjenigen geprägt sind, welche die Kategorien entwickeln (Strübing, 2008). Klassifikationen sind nie perfekt und immer starke Vereinfachungen von Komplexität. Es sind soziale Konstruktionen, die kontextgebunden und deshalb relativ, also nie absolut oder perfekt sind (Bowker & Leigh Star, 1999). Wie detailliert oder grob Kategorien gebildet werden, muss kontextspezifisch abgewogen werden, entsprechend ist der zu erwartende Erkenntnisgewinn. Werden Kategorien differenzierter; werden beispielsweise feinere Wirkungsmechanismen sichtbar, werden Kategorien grober, z.B. durch Verdichtungen, geschieht dies um den Preis eines Verlustes an Tiefe:

Die Evaluation verdichtet diese Variablen entlang dem zentralen Organisationsmodell und dem JobDemands-Resources-Modell (...), ihrem Evaluationsdesign (...), aufgrund qualitativer Informationen und methodischer Überlegungen zu vier Faktoren: Ressourcen, Belastungen, positive und negative Gesundheit. Dadurch reduziert sich die Komplexität der Analysen erheblich und es kann ein Modell verwendet werden, das für alle (Gross-) Betriebe gleichermaßen Gültigkeit hat, d.h. die Ergebnisse der Evaluation lassen sich auf andere vergleichbare Betriebe übertragen. (E2, 32)

Fokussieren, Priorisieren oder Bewerten sind zusätzliche Strategien der Komplexitätsreduktion, mit welchen Grenzen gezogen (E2, 66; E4, 40) und Ergebnisse gewichtet werden (E2, 81; E2, 58f; E3, 57; E6, 88). Diese Ordnungsstrategien orientieren sich teilweise an den Interessen der Auftraggeber (E4, 40; E2, 65), an der Sicht der Befragten (E3, 61; E9, 59), an der wissenschaftlichen Literatur (E2, 72; E5, 4), an der Praxisrealität (E2, 17; E3, 12) oder aber die Referenz wird nicht direkt sichtbar (E2, 26; E9, 100). Da die genannten Strategien stark strukturierend in die Komplexität eingreifen und die Wirkungskonstruktion massgeblich beeinflussen, dient die Offenlegung der Kriterien, welche das Strukturieren anleiten, der Nachvollziehbarkeit der Unterscheidungen.

Während die settingorientierten Evaluationen die Interventionssysteme in differenzierter Weise rekonstruieren (insbesondere E2, E4, E7), gehen andere, primär in ihren Berechnungsmodellen, von sehr einfachen Ursache-Wirkungszusammenhängen aus (E1, E3, E9) oder verzichten ganz darauf, Wirkungsmechanismen zu beachten (E5, E8), indem sie sich aus einer übergeordneten gesellschaftlichen Perspektive lediglich an pauschalen „Gesamtwirkungen“ aus vielen Interventionen und über viele Jahre interessieren:

Die Ergebnisse zeigen, dass Informations- und Schulungsmassnahmen wesentlich zur Abnahme der Strassenverkehrsunfälle, des Tabakkonsums und des übermässigen Alkoholkonsums beigetragen haben. Die Studie zeigt nicht, dass jede einzelne der getroffenen Massnahmen wirksam war, sondern dass sie gesamthaft wirksam waren. (E8, 15)

Eine besondere Form des Strukturierens, die in Zusammenhang mit Wirkungsbeurteilungen von besonderem Interesse ist, ist die Arbeit mit Wirkungsmodellen:

Wirkungsmodelle – Gerüste der Wirkungskonstruktion

In Wirkungsmodellen werden Ursache- Wirkungszusammenhänge hergestellt und in der Regel grafisch dargestellt. In den untersuchten Evaluationen wird sehr unterschiedlich mit Wirkungsmodellen umgegangen. In keiner der untersuchten Evaluationen liegt aus der Projektkonzeption bereits ein ausgearbeitetes Wirkungsmodell vor, so dass die Evaluation entsprechende implizite Modelle rekonstruiert (E3, 38, E8, 27), Modelle aus der Literatur übernimmt (E1, 7; E6, 8), Modelle induktiv entwickelt (E7, 28; E2, 32) oder gänzlich auf explizite Wirkungsmodelle verzichtet (E4, E5, E9).

Die Wirkungsmodelle können sich hinsichtlich verschiedener Dimensionen unterscheiden:

Dimensionen (dichotom oder graduell)	Erläuterungen mit Beispielen
explizit – implizit	Modelle können transparent gemacht werden (E7, 72; E2, 27) oder aber sie lassen sich nur über entsprechende Hinweise erahnen (E4, 65)
oberflächlich – differenziert	Wirkungsmodelle können entweder nur grobe Kategorien miteinander verbinden (E1, 7; E3, 38) oder aber sehr detaillierte Wirkungsbeziehungen enthalten (E7,72; E2, 32)
allgemein - spezifisch	Entweder bestehen Wirkungsmodelle aus allgemein formulierten Kategorien, die projekt- und kontextübergreifend als Rahmenmodelle genutzt werden können (E6, 8; E8, 28) oder aber sie sind themen- und/oder kontextspezifisch ausgearbeitet (E7, 72; E3, 38)
deduktiv – induktiv genutzt	Zum Teil werden Wirkungsmodelle aus der Literatur übernommen und für die Evaluation nutzbar gemacht (E6, 8), zum Teil werden sie aus den Daten heraus generiert (E7, 72; E2, 32)

Tab. 14 Dimensionen von Wirkungsmodellen

Explizite Wirkungsmodelle machen Wirkungsvermutungen transparent, legen das Gerüst der Wirkungskonstruktion offen und ermöglichen dadurch die Reflexion von Wirkungszusammenhängen, aber auch von Möglichkeiten und Grenzen des Wirkungsnachweises. Explizite Wirkungsmodelle machen sichtbar, in welcher Komplexität die untersuchten Systeme rekonstruiert werden, also bis zu welchem Grad Komplexität reduziert wird. Dabei zeigt sich einmal mehr deutlich, dass induktive Grenzziehungen über explorative Verfahren zu differenzierteren Wirkungskonstruktionen führen. Bezüglich der Differenziertheit eröffnet sich ein Spannungsfeld: einfachere Modelle sind besser verstehbar, büßen aber an Genauigkeit ein, während differenziertere Modelle die beobachteten Systeme potenziell passender rekonstruieren, dafür aber schwerer versteh- und handhabbar sind. Werden allgemeine Rahmenmodelle verwendet, dienen diese in der Regel als Referenz für die Analyse, an welcher sich die Evaluation orientiert. In drei Fällen gehen die Wirkungsmodelle vom konkreten Untersuchungsgegenstand aus und es wird auf drei unterschiedliche Arten versucht, diesen oder Teile davon im Modell abzubilden. Einmal wird die Programmlogik aus dem Projektkonzept abgebildet (E3), einmal wird das Interventionssystem induktiv rekonstruiert (E7) und einmal werden Teilaspekte des Evaluationsgegenstandes auf Basis statistischer Berechnungen modelliert (E2, *Strukturgleichungsmodell*). In den anderen Verwendungsarten wird ein allgemein gehaltenes Rahmenmodell als Reflexions- und Berechnungsgrundlage verwendet, ohne aber die konkreten Wirkungszusammenhänge der spezifischen Intervention abzubilden. Die verwendeten Wirkungsmodelle erfüllen letztlich ganz unterschiedliche Funktionen. Sie dienen

- der Rekonstruktion der Wirkungslogik vorliegender Interventionskonzepte (E3)
- als Rahmenmodelle für die Reflexion der untersuchten sozialen Systeme und Interventionen (E1, E2, E6)
- als Grundlage für quantitative Analysen (E1, E2, E8)
- der Darstellung komplexer sozialer Systeme und Wirkungsmechanismen, um diese verstehen zu lernen (E2, E7)

- und in allen Fällen, aber auf unterschiedliche Weise, als Gerüst der Wirkungskonstruktion und damit als Grundlage für die Wirkungsbeurteilung.

Wenn versucht wird, sich der Komplexität sozialer Systeme anzunähern und entsprechend detailliert Ursache-Wirkungsbeziehungen abzubilden, resultieren rasch sehr komplexe Modelle. Das nachfolgend dargestellte Modell zeigt eindrucksvoll die verschiedenen Wirkungsebenen und vielfältigen Einflussfaktoren, auch wenn die einzelnen Elemente in dieser Darstellungsform nicht entziffert werden können:

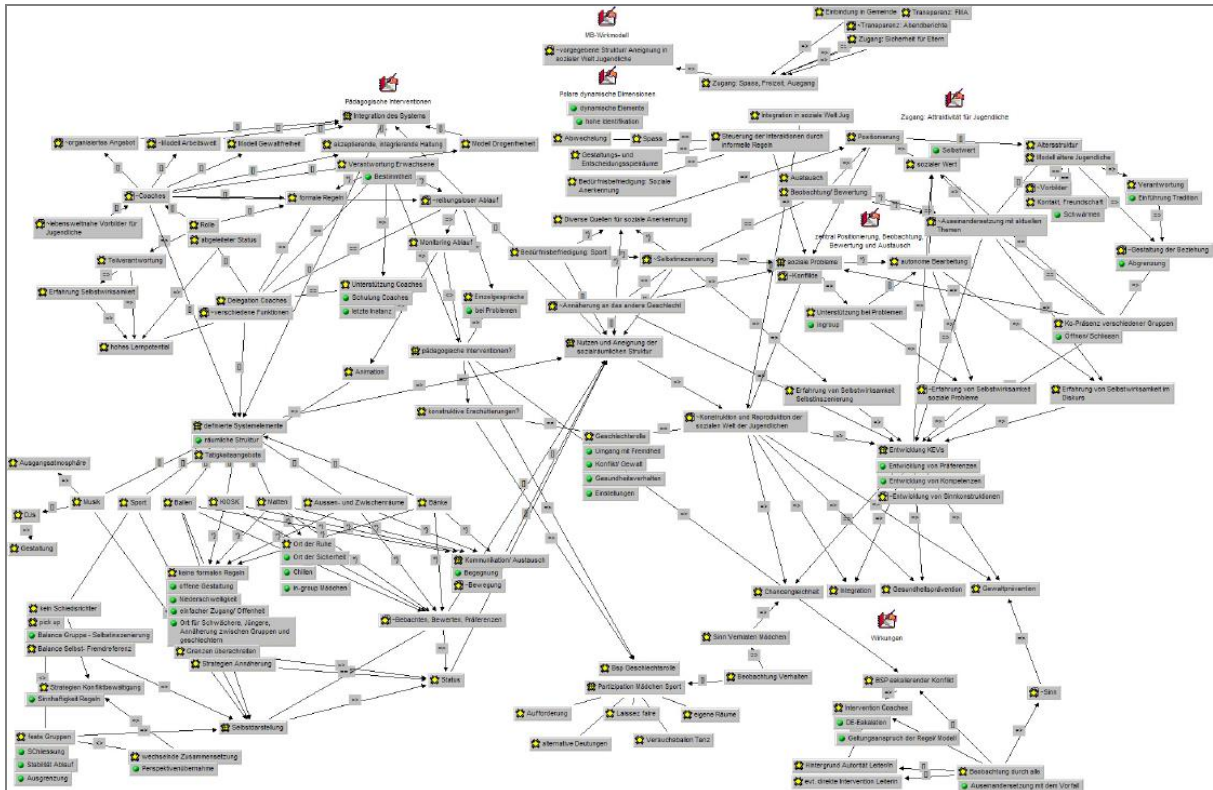


Abb. 17 Umriss eines induktiv entwickelten, sehr differenzierten Wirkungsmodells (Sommerfeld et al. 2006 (E7), 72)

Das Modell vermittelt einen Eindruck davon, in welchen komplexen Zusammenhängen und wie begrenzt durchschaubar sich Wirkungen in Projekten der Gesundheitsförderung und Prävention entfalten.

Mit der Komplexität des Wirkungsmodells wird deutlich, wo und wie die erwartbaren Wirkungen entstehen. Es wird zugleich deutlich, dass es sich um eine anspruchsvolle Aufgabe handelt, dieses System zu leiten. Das Bild ähnelt einem Mobile: Wenn man an einer Stelle zieht, dann verändert sich die gesamte Konstellation und mit ihr auch die Wirkungen, die tatsächlich realisiert werden können. (E7, 32)

Wirkungsmodelle können die Realität aber auch viel einfacher abbilden:

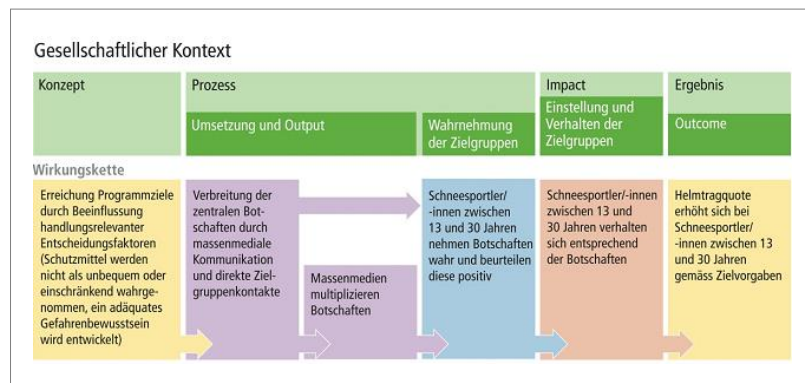


Abb. 18 Einfaches Wirkungsmodell einer Präventionskampagne (Furrer und Balthasar 2011 (E3), 66)

Dieses Wirkungsmodell beschränkt sich auf wenige Wirkfaktoren, die linear miteinander verbunden sind. Die Evaluation, die darauf aufbaut, konzentriert sich auf die Beurteilung dieser Faktoren:



Abb. 19 Evaluationsfragen zum Wirkungsmodell der Präventionskampagne (E3,66)

Bei sehr differenzierten Wirkungsmodellen besteht die Gefahr, dass sie nicht mehr übersichtlich sind und die Anschlussfähigkeit an die Praxis verlieren. Einfache Wirkungsmodelle laufen Gefahr, relevante Aspekte nicht zu berücksichtigen, da diese in den fokussierten Kategorien nicht enthalten sind. Wenn entscheidende Einflüsse unentdeckt bleiben, resultieren Wirkungskonstruktionen, welche die Realität unangemessen abbilden. Wirkungsmodelle bieten sowohl der Intervention als auch der Evaluation eine Orientierungshilfe. Sie dienen dazu, implizite Wirkungsvermutungen offenzulegen, die Interventionslogik darauf abzustützen und der Evaluation einen Rahmen zu geben:

Dazu kommt, dass das Projekt offenbar keine ausformulierte oder explizite Programmtheorie in eigener Sache hatte. Diese hätte zeigen können, welche Schritte im Einzelnen nötig sind und welche Hindernisse zu überwinden sind, um das Projekt zu konzipieren, zu gestalten und umzusetzen. (E6, 87)

und entsprechend die Empfehlung der Evaluation:

Das Bundesamt für Gesundheit verlangt bei Projektgesuchen konsequent eine formulierte Programmtheorie für das Projekt. Diese zeigt auf, wie die angestrebten Wirkungen des Projekts erreicht werden sollen. (E6, 89)

Wirkungsmodelle scheinen in der Evaluation von Gesundheitsförderung und Prävention zwar gebräuchlich zu sein, die stark unterschiedliche Praxis mag aber ein Hinweis auf eine gewisse Unsicherheit im Umgang mit Wirkungsmodellen sein.

Vergleiche als Grundlage für Wirkungsbeurteilungen

Evaluieren bedeutet immer auch Beurteilen. Bei der Beurteilung ist die Evaluation auf eine Referenz angewiesen, an welcher der festgestellte Zustand gemessen wird. Dies geschieht über Vergleiche. Wirkungsbeurteilungen bauen entsprechend auf Vergleichen auf. Die Vergleiche können sowohl quantitativer Art (auf Basis von Berechnungen), als auch qualitativer Natur (auf Basis von Beschreibungen) sein. In den untersuchten Evaluationen konnten unterschiedliche Vergleiche identifiziert werden, die oft auch kombiniert auftreten:

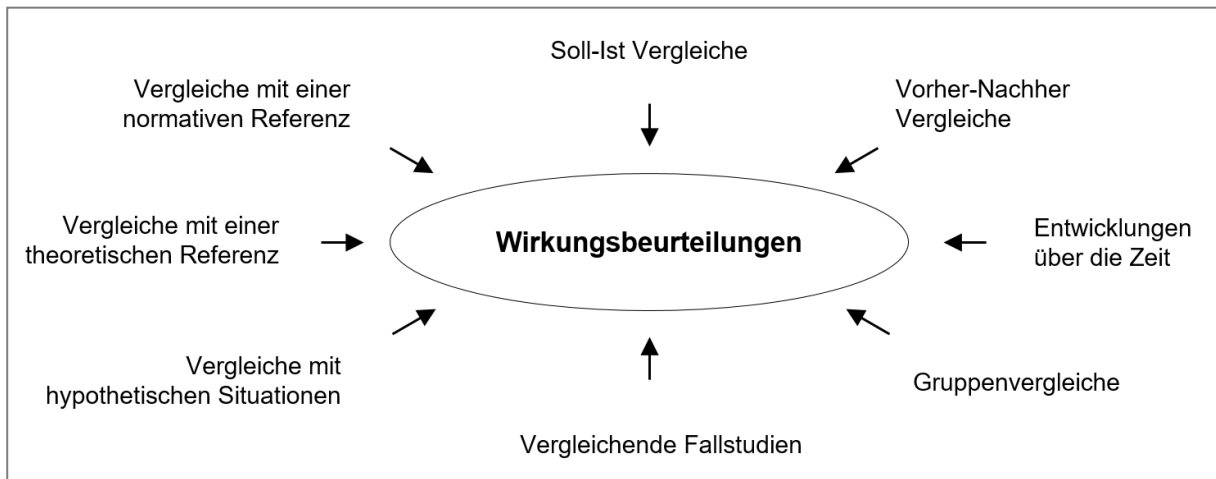


Abb. 20 Unterschiedliche Vergleiche als Basis für Wirkungsbeurteilungen

Soll-Ist Vergleiche: Bei diesem ersten Typus von Vergleichen wird der Entwicklungsstand nach einer bestimmten Zeit mit einem geplanten Soll-Zustand, in der Regel in Form von Interventionszielen, verglichen. Der Vergleich wird meist zum Ende der Interventionsdauer gemacht, er ist aber auch früher (Zwischenevaluation) oder später (Beurteilung der Nachhaltigkeit) möglich. Wirkungsziele werden sowohl quantitativ (E1, 32f; E3, 59f) als auch qualitativ beurteilt (E4, 61f; E6, 81f). Da oft keine quantitativen Zielwerte vorliegen, sondern nur Zielrichtungen (E2, 18; E4, 10; E9, 34), können die Ziele bereits dann als erreicht beurteilt werden, wenn die Ergebnisse in die gewünschte Richtung zeigen. In lediglich zwei Fällen liegen konkrete Zielwerte vor (E1, 6; E3, 33), die überprüft und auch teilweise erreicht werden (E1, 32f; E3, 59/66). Dass in vielen Fällen keine Zielwerte vorliegen, mag mit der Komplexität der Interventionssysteme und der schwierigen Vorhersage der Entwicklungen zusammenhängen. Werden Zielwerte vorgegeben, kann es sein, dass diese in der gegebenen Zeit keinesfalls oder auch bereits viel früher erreicht werden:

Zwei der drei im 2006 gesteckten Impact-Ziele (Vulnerabilität und Gefahrenbewusstsein) konnten erreicht werden. Allerdings entsprachen die Werte für die fehlende wahrgenommene Vulnerabilität bereits 2007 und für das fehlende Gefahrenbewusstsein bereits 2008 der festgelegten Zielsetzung. (E3, 59)

Wie die untersuchten Evaluationen zeigen, sind quantitativ konkretisierte Wirkungsziele aufgrund der hohen Komplexität der untersuchten sozialen Systeme nicht nur schwierig festzulegen, sie sind für Wirkungsbeurteilungen auf den ersten Blick auch keine zwingende Voraussetzung. In statistischen Tests können bei entsprechender Stichprobengröße bereits sehr kleine Unterschiede zwischen Interventions- und Kontrollgruppen signifikant werden und auf Wirkungen hinweisen. Da statistische Signifikanz aber nicht gleichbedeutend ist mit inhaltlicher Bedeutsamkeit, macht es durchaus Sinn, sich darüber zu einigen, ab welchem Wert von einer bedeutsamen Wirkung gesprochen werden soll, wie dies in einer der untersuchten Evaluationen auch getan wird (E9, 58). Im Sinne von Mindestwerten für Wirksamkeit können Zielwerte also durchaus sinnvoll sein.

Vorher-Nachher-Vergleiche: Vorher-Nachher-Vergleiche sind für Wirkungsbeurteilungen entscheidend. In der Regel beziehen sich das Vorher und das Nachher auf eine klar abgegrenzte Intervention (E1, 5; E9, 37), wobei zum Teil auch andere Zeitpunkte für Vergleiche gewählt werden, zum Beispiel vor und nach Ende der Intervention (E9, 37; E6, 17). Aus den Vorher-Nachher Vergleichen wird auf Interventionseffekte geschlossen. Vorher-Nachher-Vergleiche werden nicht nur quantitativ durchgeführt, sondern auch qualitativ, beschreibend (E3, E6; E2, 122).

Das Bewusstsein konnte geschärft werden. Mitarbeitende spüren, dass sich das Arbeitsklima verbessert hat. Sie achten vermehrt auf einen besseren Umgang mit Stress. Es gibt eine offener Kommunikation darüber, und Stress wird nicht mehr als Tabu gesehen. Je nach Abteilung konnten unterschiedliche kleinere oder grössere Verbesserungen umgesetzt werden. (E2, 122)

Bei Vorher-Nachher Vergleichen werden Prä- und Post-Messungen gemacht (E1, E9), zum Teil werden Dokumente (E6, 15) oder statistische Daten Dritter (E5, 7; E8, 10) aus der Vorphase als Referenz genutzt und zum Teil wird auf retrospektive Einschätzungen abgestützt (E6). Die Vorher-Nachher Vergleiche umfassen zeitliche Spannweiten von einigen Monaten (E1, 17; E9, 37) bis zu mehreren Jahren (E5, 1; E8, 8), letzteres bei der Nutzung von länger zurückreichenden Monitoringdaten. Zum Teil werden zu den Postmessungen Nachbefragungen durchgeführt, um die Nachhaltigkeit der beobachteten Veränderungen beurteilen zu können (E9, 37).

Bei Vorher-Nachher Erhebungen müssen Stichproben und die Erhebungsinstrumente möglichst konstant gehalten werden, damit die Vergleichbarkeit gewährleistet bleibt. Vorher-Nachher Vergleiche bergen das Risiko, dass festgestellte Veränderungen vorschnell der untersuchten Intervention zugeschrieben und fälschlicherweise als Wirkungen ausgegeben werden. Die Kombination mit Kontrollgruppenvergleichen (E1, 4; E5, 9; E9, 36), aber auch die Plausibilisierung über die Analyse von Wirkungsbedingungen qualitativer oder quantitativer Art (E4, E6) können hier ergänzen.

Vergleiche von Entwicklungen über die Zeit: Zum Teil werden nicht einfach einzelne Erhebungszeitpunkte miteinander verglichen, sondern es werden wirkungsrelevante Entwicklungen über die Zeit beschrieben (s. auch das Kapitel ‚Beschreiben‘):

Der Vollzug der Alkoholpolitik verlief harzig. Einige Teilprojekte wurden wieder eingestellt. Dies lag vor allem auch daran, dass mit dem Abgang des Projektleiters die bestehende Arbeitsgruppe aufgelöst wurde und es zur Bildung einer kleineren, "schlankeren", stärker operativ tätigen Arbeitsgruppe kam. Mit dem Projekt "Werdenberg handelt!" konnte eine Sensibilisierung der Gastwirte und Verkaufsstellen erreicht werden, was die guten Resultate der Testkäufe 2005 zeigten. Bei den Festanlässen ist gemäss Aussagen von Mitgliedern der Projektgruppe ebenfalls eine Verhaltensänderung zu beobachten. Bei den jährlichen Grossanlässen (z.B. Buchserfest, Erlenfest, WIGA) sei ein weniger exzessiver Alkoholkonsum feststellbar. Die Atmosphäre am Fest selbst habe sich verändert und es gebe weniger negative Folgeerscheinungen (z.B. Vandalismus). Gemäss einem Veranstalter des jährlich stattfindenden Erlenfestes seien die Anforderungen heute strenger geworden als im Vergleich zu früher (z.B. "Bändelsystem", Alterskontrollen am Eingang). (E6,30)

Für die Wirkungsbeurteilung in komplexen Feldern würden sich kontinuierliche und kleinräumige Betrachtungen anbieten, um Wirkungsmechanismen und Wirkungspotenziale besser verstehen und Wirkungszuschreibungen plausibilisieren zu können:

Es benötigt daher entweder häufige Befragungen zur Trenddarstellung oder kleinräumige Auswertungen auf Teamebene, verbunden mit Informationen aus Gesprächen im Betrieb, um auf diese Weise Veränderungen mit hoher Plausibilität auf die Wirkung einer Intervention wie SWiNG zurückführen zu können. Ohne Vergleichsgruppen bleibt aber immer die Frage im Raum, ob die Veränderung nicht andere Ursachen hat. (E2, 87)

Nicht nur der vertiefte Blick in die Interventionsphase, sondern auch der Blick über diese hinaus lässt sich in den untersuchten Evaluationen ausmachen. Wenn entsprechende Daten vorliegen, können Veränderungen während der Interventionszeit in einen grösseren zeitlichen Horizont verortet werden. Damit lassen sich gemessene Verbesserungen gegebenenfalls relativieren, und die Gefahr wird vermindert, dass positive Veränderungen vorschnell als Interventionswirkungen ausgegeben werden:

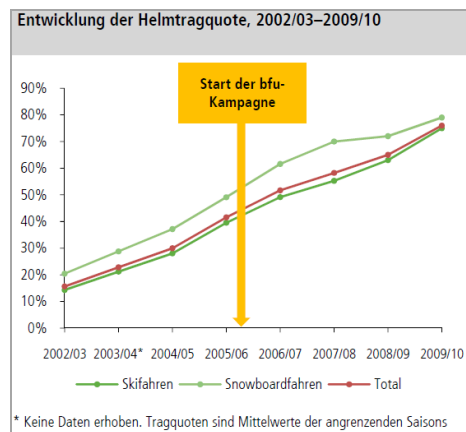


Abb. 21 Betrachtung der Entwicklung über die Interventionszeit hinaus
(Furrer und Balthasar 2011 (E3), 60)

Gruppenvergleiche: Vergleiche zwischen verschiedenen Gruppen sind in den untersuchten Evaluationen sehr häufig. Quantitative Wirkungsbeurteilungen bauen auf Vorher-Nachher-Vergleichen in Kombination mit Gruppenvergleichen auf, wobei sehr oft Interventions- mit Kontrollgruppen verglichen werden (ausser in E4 und E7), teilweise werden aber auch andere Gruppenvergleiche angewendet, zum Beispiel innerhalb der Interventionsgruppen aufgeteilt nach Altersgruppen (E3, 51), Sprachregionen (E3, 51), Berufsgruppen (E2, 40), Risikogruppen (E1, 49,), Gruppen mit höherer versus tieferer Wirkungseinschätzung (E2, 58) etc.

Referenz für statistische Vergleiche zwischen Interventions- und Kontrollgruppen bilden in der Regel explizite Wirkungshypothesen:

Die Hypothese, dass Personen, die die bfu-Kampagne «1000 Unfälle pro Tag – Schütz dich mit einem Helm» kennen, häufiger einen Helm tragen, wird durch diese Analyse bestätigt. (E3)

In zwei der untersuchten Evaluationen (E1, E9) wird mit randomisierten Interventions- und Kontrollgruppen gearbeitet, wobei in beiden Fällen nicht Personen, sondern Klassen zufällig zugeteilt werden (Clusterrandomisierung). Die Randomisierung gilt als entscheidende Voraussetzung für Wirkungsnachweise schlechthin, indem andere Einflussfaktoren als die Intervention durch die Randomisierung ‚neutralisiert‘ werden:

Durch ein kontrolliert-randomisiertes Design wird der Zufall und der Einfluss von Fremdeinflüssen als Ursache für beobachtete Veränderungen so weit wie möglich ausgegrenzt. (E9, 78)

Bei der Umsetzung dieses Designs ergeben sich zum Teil aber Anpassungen an die Praxisrealität, welche dieses entscheidende Merkmal gleich wieder opfern¹¹⁶.

Während Vergleiche zwischen Interventions- und Kontrollgruppen dazu verwendet werden, um Interventionseffekte als solche zu identifizieren, werden Vergleiche zwischen einzelnen Untergruppen (Al-

¹¹⁶ So wird in der einen Evaluation auf die Wünsche einiger Klassen eingegangen, die Randomisierung durchbrochen und teilweise durch ein Matching ersetzt (E1, 16).

ter, Geschlecht,...) dazu genutzt, um zu überprüfen, bei wem die Intervention mehr (bzw. weniger) gewirkt hat und um damit Wirkungsbedingungen identifizieren zu können (E1, 49f; E3, 52; E9, 83). Einen interessanten Ansatz zur Wirkungsbeurteilung ohne Kontrollgruppe bildet der Vergleich zwischen Personen mit höherer bzw. tieferer Wirkungseinschätzung der Intervention (E2, 36). Dieses Verfahren vergleicht die eigene Wirkungseinschätzung mit gemessenen Veränderungen und ermöglicht so eine Plausibilisierung der Wirkungszuschreibung.

Vergleichende Fallstudien: Zum Teil werden nicht ganze Gruppen, sondern lediglich einzelne Fälle qualitativ miteinander verglichen (E2, E6, E7). Durch die detaillierte vergleichende Beschreibung sollen fallspezifische Wirkungsmechanismen sichtbar werden und auf diese Weise Wirkungspotenziale ersichtlich und Wirkungsbeurteilungen möglich werden:

Entsprechend dem Evaluationszweck wählten wir für die Evaluation in Absprache mit der Auftraggeberin den methodischen Ansatz der vergleichenden Fallstudienanalyse. Dies gibt die Möglichkeit, in den ausgewählten Gemeinden vertieft und genau hinzuschauen und festzuhalten, was konkret und gemeindespezifisch erreicht wurde. Die Analysetiefe steht im Vordergrund, da es um eine zuverlässige Wirkungsanalyse geht. (E6, 9)

Im Unterschied zum Wirkungsnachweis über Kontrollgruppendesigns wird hier die Attribution der Wirkung zur Intervention nicht über den Vergleich zwischen Interventions- und Kontrollgruppe (u.a. mittels Randomisierung) gemacht, sondern durch eine vertiefte Analyse der Verbindungen zwischen Intervention und Veränderungen.

Vergleiche mit hypothetischen Situationen: Eine weitere Möglichkeit zur Wirkungsbeurteilung bietet der Vergleich mit einer hypothetischen Situation. Bei Fehlen einer Vergleichssituation kann diese simuliert werden. In E8 wird dieses Verfahren angewendet, wobei – aufgrund der hypothetischen Situation – auf Schätzungen abgestützt wird:

Die Gesundheitsergebnisse im Zeitraum, in dem die Präventionsmassnahmen durchgeführt wurden, wurden mit den Gesundheitsergebnissen in einer hypothetischen Situation ohne diese Präventionsmassnahmen verglichen. Ein Beispiel ist die Senkung der Blutalkoholgrenze für Fahrzeuglenker von 0,8 auf 0,5 Promille im Jahr 2005. Der Effekt dieser Verringerung wird geschätzt, indem die tatsächliche Zahl der Strassenverkehrsoffer im Zeitraum, in dem die 0,5 Limite in Kraft war, mit der hypothetischen Zahl von Strassenverkehrsoffern verglichen wird, zu denen es bei der alten Limite von 0,8 gekommen wäre. (E8, 8)

Vergleiche mit einer theoretischen Referenz: Um beispielsweise festgestellte Veränderungen oder Interventionswirkungen auf ihre Bedeutsamkeit hin zu untersuchen, wird u.a. auf wissenschaftliche Literatur Bezug genommen. Wirkungen werden so nicht nachgewiesen, sondern in ihrer Qualität bewertet (E6, 79f):

Auf der Ebene der Politikformulierung und der Politikimplementierung vermochten die entstandenen Massnahmen oder sogenannten lokalen Alkoholpolitiken den politikwissenschaftlichen Kriterien nur ansatzweise zu genügen. Was in den untersuchten Gemeinden erarbeitet wurde, kann kaum als Politik bezeichnet werden. (E6, 80)

Vergleiche mit einer normativen Referenz: Anstelle einer theoretischen Referenz kann auch eine normative Referenz wie in folgendem Beispiel die Gesetzgebung (E6, 83f), für die Beurteilung von Wirkungen genutzt werden. Dabei wird beurteilt, ob eine festgestellte Veränderung auch tatsächlich der Absicht des Gesetzgebers entspricht:

Fazit: Der konsequente Vollzug des Jugendschutzes ist in keiner der untersuchten Gemeinden nachhaltig gesichert. (E6, 84).

In den untersuchten Evaluationen konnte eine Vielzahl von Vergleichen identifiziert werden, mit je spezifischen Potenzialen und Grenzen. Die verschiedenen Evaluationen gelangen dabei über ganz unterschiedliche Vergleiche zu Wirkungsurteilen: quantitativ oder qualitativ, mit oder ohne Vergleichsgruppen, mit oder ohne Randomisierung. Was die Häufigkeit betrifft, so gehen in den untersuchten Evaluationen die meisten klaren Wirkungsaussagen, welche auf Vergleichen beruhen, auf statistische Vergleiche zurück, die Zeitpunkte und Gruppen kombiniert vergleichen (Kontrollgruppendesigns).

Für eine effektive Wirkungsbeurteilung kristallisieren sich quer durch die verschiedenen Vergleiche einige zentrale Aspekte heraus:

- Vergleiche müssen (theoretisch) gut begründet sein, Wirkungsmodelle können eine gute Grundlage dafür liefern. Vergleichsgruppen müssen sorgfältig gewählt werden und beispielsweise angemessen differenziert sein (damit tatsächliche Wirkungen nicht durch Nivellierung verschwinden).
- Beobachtete Veränderungen müssen in einen Bezug zur untersuchten Intervention gebracht werden. Dazu zeigen sich in den Evaluationen unterschiedliche Wege.
- Vergleiche bauen grundsätzlich immer auf starken Komplexitätsreduktionen auf und es muss auch hier immer abgewogen werden zwischen Erkenntnisgewinn durch Zulassen versus Reduktion von Komplexität. Einmal festgelegte Vergleichskategorien bergen das Risiko, nicht mehr flexibel auf neue Entwicklungen reagieren zu können und dadurch Wichtiges zu verpassen.
- Sowohl aus quantitativ orientierten Vergleichen wie auch aus qualitativen Analysen lassen sich wertvolle Wirkungsbeurteilungen ableiten. Keine der beiden Formen scheint der anderen grundsätzlich überlegen und eine Kombination erscheint besonders ergiebig
- Während quantitative Vergleiche eher darauf ausgelegt sind, zu untersuchen, *ob* eine Intervention Wirkung erzeugt hat, gehen qualitative Vergleiche, z.B. in vergleichenden Fallstudien, stärker in die Tiefe und machen Aussagen zu kontextspezifischen Wirkungsmechanismen und Wirkungspotenzialen.
- Wirkungen müssen im Kontext ihrer Entstehung beurteilt und Kontextfaktoren entsprechend berücksichtigt werden (z.B. unterschiedliche Settings, saisonale Schwankungen). Die Entwicklungen vor der Interventionsphase als auch danach sind ergänzend sehr aufschlussreich.

Berechnen

Vergleiche, die auf Berechnungen aufbauen, sind in den meisten der untersuchten Wirkungsevaluationen eine prioritäre Grundlage der Wirkungsbeurteilung (E1, E2, E3, E5, E8, E9) und ein Grossteil der zentralen Wirkungsaussagen baut auf diesen Berechnungen auf. Die Berechnungen und die dabei verwendeten Methoden und Modelle sind sehr unterschiedlich, zum Teil sehr umfassend und auch nicht im Detail in den vorliegenden Berichten aufgeführt (E9, 35; E8, 35/100; E2, 32). Berechnungen liegt in der Regel eine Wirkungskonstruktion zu Grunde, auf Basis welcher die Variablen bestimmt werden, die in die Berechnungsmodelle einfließen. Die Berechnungen dienen in erster Linie dazu, statistische Vergleiche anzustellen, um über diesen Weg Wirkungen beurteilen zu können. Berechnungen ohne weiterführende Vergleiche werden nur in Ausnahmefällen zur Wirkungsbeurteilung genutzt, und zwar primär um das Wirkungspotenzial von Interventionen in der Breite abzuschätzen (E7, 48).

Die untersuchten Evaluationen verwenden sehr unterschiedliche Berechnungsmodelle, die letztlich aber alle in den beschriebenen Vergleichen münden. Die einzelnen Berechnungsmodelle werden in dieser Arbeit nicht detailliert beschrieben und beurteilt, jedoch lohnt wiederum ein Blick auf den Umgang mit Komplexität: In die Berechnungsmodelle der meisten Evaluationen fließen nur wenige, überschaubare unabhängige und abhängige Variablen ein. Die Evaluationen mit randomisiertem Kontrollgruppendesigns (E1, E9) sowie die Evaluation mit dem Ein-Gruppen-Design (E3) thematisieren zwar durchaus auch komplexere Wirkungszusammenhänge (E1, 8; E3, 47; E9, 81f):

Der Prozess der Veränderung wird im TTM als spiralförmiges Geschehen verstanden, in dem Personen nach einem Rückfall auf eine frühere Veränderungsstufe zurückfallen können. Das

Spiralmodell zeigt, dass Klientinnen und Klienten mit Rückfällen nicht bei Null beginnen müssen, sondern durch den Misserfolg lernen. (E1, 8)

allerdings ohne dass diese (z.B. hier die spiralförmigen Entwicklungen) entsprechend in den Berechnungsmodellen berücksichtigt würden: Das Design, welches den Berechnungen zugrundliegt ist einfach aufgebaut und vergleicht in Vorher-Nachher Messungen Unterschiede in Interventions- und Kontrollgruppen bzw. Untergruppen in Bezug auf wenige Zielindikatoren und unter Berücksichtigung einiger weniger beeinflussender Faktoren. Dynamische Entwicklungen zwischen den Messzeitpunkten, komplexere Wirkungszusammenhänge oder kontextspezifische Ausprägungen fließen nicht in die Berechnungen ein. Auch in den retrospektiven Evaluationen mit gesamtgesellschaftlichem Blick (E5, E8), welche zwar komplizierte Berechnungsmodelle verwenden, ist die Situation vergleichbar. Nichtlineare komplexe Ursache-Wirkungszusammenhänge und dynamische Veränderungen werden durch die Modelle nicht erfasst.

In die Berechnungen zur Wirksamkeit der betrieblichen Gesundheitsförderung in E2 fließen zwar viele einzelne Einflussfaktoren ein:

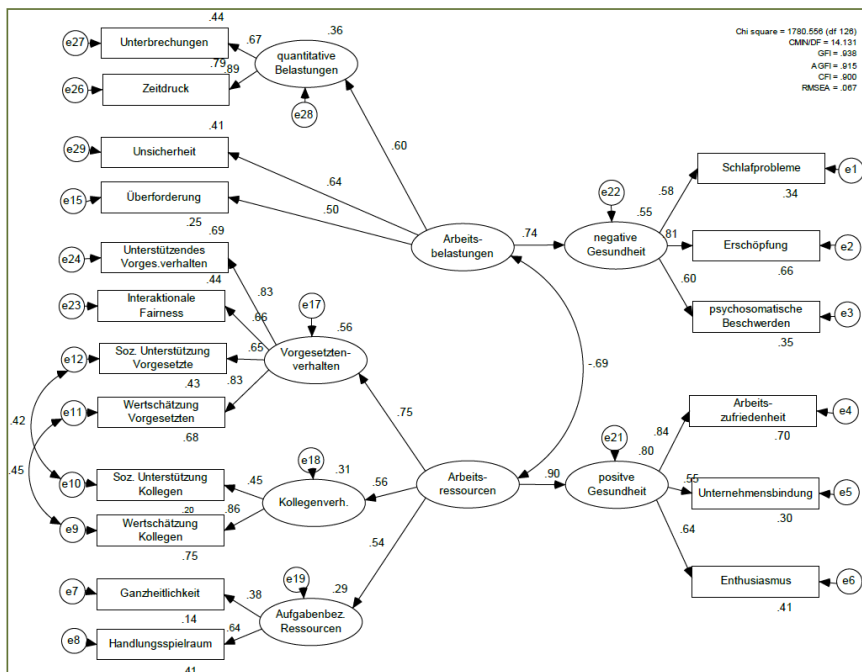


Abb. 22 Strukturgleichungsmodell (Jenny et al. 2011 (E2), 32)

Allerdings bilden auch diese Faktoren nur einen Ausschnitt aus dem Gesamtmodell (E2, 27) und sie werden zur Wirkungsbeurteilung gleich wieder zu wenigen Indikatoren verdichtet, um den Preis des Verlustes an Differenziertheit:

Es ist für jeden einzelnen Belastungs-, Ressourcen- oder Gesundheitsindikator wenig darüber bekannt, ob er sich langsam über die Zeit verändert oder sprunghaft, ob die Zusammenhänge zwischen den Indikatoren linear oder kurvenförmig sind, ob die Veränderung anhält oder schnell wieder abflacht. Für die Wirksamkeitsanalyse verdichtet die Evaluation alle Indikatoren zu vier Faktoren: Belastungen, Ressourcen, negative und positive Gesundheit (siehe Kapitel 3.7). Dadurch verlieren solche Eigenarten der Indikatoren im Gesamtfaktor ihren Einfluss. (E2, 29)

Mit den in E2 verwendeten drei Messzeitpunkten bleiben dynamische Entwicklungen ebenfalls weitgehend unberücksichtigt – zumindest was die Berechnungen betrifft.

Berechnungen resp. numerische Ergebnisse suggerieren in der Regel eine hohe Genauigkeit und Verlässlichkeit. In den untersuchten Evaluationen zeigt sich allerdings eine Vielzahl von Hinweisen, welche diese Genauigkeit relativieren:

- Viele Berechnungen basieren auf Schätzungen (E3, 40; E8, 77; E9, 63). Diese sind teilweise bewusst konservativ, so dass Wirkungen der Gesundheitsförderung und Prävention tendenziell unterschätzt werden (E8, 58; E2, 82).
- Die Daten, welche in Berechnungen einfließen, sind in vielen Fällen lückenhaft oder nicht eindeutig (E2, 16; E5, 7; E7, 70) (vgl. auch Kapitel ‚Grenzen der Erfassung‘)
- Daten werden korrigiert, Verzerrungen bereinigt (E8, 75; E9, 58)
- Bei Ergebnissen statistischer Tests handelt es sich nicht um eindeutige Nachweise, sondern um statistische Wahrscheinlichkeiten mit einer gewissen Irrtumswahrscheinlichkeit (E1, 15; E3, 41; E8, 79)
- In den Berechnungen werden verschiedene Verfahren verwendet (z.B. Gewichtungen,...), die potenziell zu Verzerrungen führen können (E8, 45; E3, 50; E5, 13/26) oder explizit davon ausgehen, dass es einen Spielraum für Berechnungsergebnisse gibt (Sensitivitätsanalysen) (E8, 51)
- Es werden zum Teil Hochrechnungen verwendet, um von Stichproben auf Gesamtpopulationen zu schliessen (E2, 76), in anderen Fällen fließen hypothetische Annahmen in Berechnungen ein (E2, 79)
- Zum Teil werden mangelhafte Methoden mangels besserer Alternativen eingesetzt (E2, 69).
- Das Zusammenlegen von Indikatoren zu übergeordneten Faktoren geht mit einem Verlust an Genauigkeit einher (E2, 29).
- Zum Teil muss zwischen verschiedenen (komplexen) Berechnungsansätzen oder -methoden gewählt werden, die wohl selten zu exakt denselben Ergebnissen führen dürften (E5, 12; E8, 73).

Zeitreihenanalysen sind per se verhältnismässig komplex. In den oben beschriebenen Modellen werden die Zeitreihenanalysen durch das Interventionsdesign und die nicht äquivalenten Kontrollgruppen zusätzlich komplexer. Da keine zufällige Einteilung der Kantone in Experimental- und Kontrollgruppen erfolgen konnte, werden die „anderen Schweizer Kantone (exkl. Genf)“ als nicht äquivalente Kontrollgruppe bezeichnet. Wir sind zwar der Meinung, hier mit dem Differenz- und Interaktionsmodell zwei verschiedene, korrekte Ansätze zu verwenden. Dennoch finden sich in der Literatur auch andere Ansätze (Alternative A und Alternative B), welche im Folgenden beschrieben werden: (...) (E5, 12)

Beschreiben

Evaluationsberichte bestehen naturgemäss vor allem aus Beschreibungen. Es werden die Ausgangslage, die untersuchte Intervention, die Fragestellungen, die Methodik und vor allem die Ergebnisse beschrieben, die Evaluationsfragen beantwortet, Schlussfolgerungen abgeleitet und Empfehlungen abgegeben. In Zusammenhang mit dem Thema dieser Untersuchung liegt der Fokus darauf, inwiefern sich die Komplexität sozialer Systeme in den Beschreibungen abbildet und wie Beschreibungen im Kontext der Wirkungsbeurteilung genutzt werden. Mit Blick auf die Komplexität sozialer Systeme können sich Beschreibungen hinsichtlich verschiedener Dimensionen unterscheiden:

Dimensionen (dichotom oder graduell)	Erläuterungen mit Beispielen
ausführlich - knapp	Beschreibungen können Sachverhalte sehr differenziert und ausführlich beschreiben (E7, 33f; E2, 63f) oder aber nur knapp zusammenfassen (E1, v; E6, 18)
eindeutig - unsicher	Die gemachten Aussagen erscheinen entweder als klar und eindeutig (E2, 7; E6, 67) oder aber sie sind sehr vorsichtig (z.B. als Vermutungen) formuliert (E5, 24; E9, 13)
homogen – heterogen	Zum Teil zeigen die Beschreibungen unterschiedliche Sichtweisen auf (E4, 73), meistens aber sind sie homogen, aus der Sicht der Evaluation zusammenfassend (E2, 4; E7, 49)
kontextspezifisch - kontextübergreifend	Es können einzelne Kontexte/Fälle separat beschrieben werden (E2 87f; E6, 17f; E7, 43f) oder aber die Beschreibungen fassen mehrere Kontexte zusammen (E2, 83f; E6, 64f)

Tab. 15 Dimensionen von Beschreibungen

In Beschreibungen werden bezüglich des untersuchten Evaluationsgegenstandes vor allem:

- Kontexte beschrieben (E4, 9; E7, 11f)
- die Interventionen selber beschrieben (E4, 10; E6, Z-2) und Interventionsprozesse nachgezeichnet (E4, 42f; E2, 38f)
- erreichte Veränderungen präsentiert (E2, 51; E9, 53f)
- Wirkungspotenziale beschrieben (E2, 14; E4, 74f)
- Wirkungsmechanismen analysiert (E2, 15; E7, 33f;)
- Berechnungsergebnisse erläutert (E2, 87; E5, 20)
- Befunde diskutiert (E1, 48f; E5, 26) und *plausibilisiert* (E2, 72; E5, 27)
- die Interventionen beurteilt (E6, 78f; E9, 56)
- auf Nebenwirkungen hingewiesen (E4, 64; E6,22)
- *Ergebnisse mit anderen Studien verglichen* (E2, 82; E5, 5ff)
- Beurteilungen zusammengefasst (E2, 83f; E5, 30)

Die Optionen, die sich aus der Verbindung zwischen den Dimensionen und den Beschreibungsinhalten ergeben, zeigen den grossen Spielraum, welche Evaluationen bei der Konstruktion von Wirkungen haben und auch nutzen. In einigen Evaluationen, welche in ihrer Wirkungsbeurteilung vor allem auf quantitative Analysen abstützen (E1, E3, E5, E8, E9), sind die Beschreibungen eher knapp und dienen primär dazu, die Berechnungsergebnisse zu beschreiben und zu erläutern. Ergänzend dazu werden Hinweise, oft aus qualitativen Interviews, zum Ablauf der Interventionen und zu Gelingensbedingungen gegeben (E1, 45f; E3, 66; E9, 53f). Die Reflexion von Wirkungspotenzialen und Wirkungsmechanismen konzentriert sich in diesen Evaluationen auf wenige in den quantitativen Analysen berücksichtigte unabhängige Variablen (E1, 58; E5, 16). In den stärker oder ausschliesslich qualitativ ausgerichteten Evaluationen (E2, E4, E6, E7) hingegen nehmen ausführliche Beschreibungen von Wirkungspotenzialen und Wirkungsmechanismen einen grossen Platz ein:

In diesen kommunikativen Prozessen, zu denen auch die Selbstdarstellung gehört, wird ständig beobachtet und bewertet. Der Status der einzelnen wird verhandelt, worüber sich Identität und Habitus ausbilden. In diesem Sinne nutzen die Jugendlichen Midnight Basketball und eignen sich diese sozialräumliche Struktur an. Midnight Basketball ist ein wertvoller Ort für die Konstruktion und die Reproduktion der sozialen Welt der Jugendlichen, die durch die im Projekt eingelassenen Prinzipien und Regeln bereichert und orientiert wird. Durch diesen Zusatz und die diversen Erfahrungsräume für Selbstwirksamkeit wird die Entwicklung der individuellen Kognitions-Emotions-Verhaltensmuster erweitert und in ihrer Entwicklung stimuliert. Über diesen Weg kann Midnight Basketball tendenziell Wirkungen im Hinblick auf Chancengleichheit, Gesundheitsförderung und Gewaltprävention entfalten. (E7, 42)

In fallvergleichenden ausführlichen Beschreibungen von Kontexten, Interventionsprozessen, Wirkungspotenzialen und Wirkungsmechanismen und unter Berücksichtigung verschiedener Perspektiven wird versucht, Interventionssysteme detailliert in ihrer Wirkungsweise zu verstehen, und die Wirkungsbeurteilung baut wesentlich oder ausschliesslich auf diesen Beschreibungen auf. Dem Einfluss von Kontextbedingungen auf die Intervention wird dabei unterschiedlich grosse Bedeutung beigemessen, entsprechende Einflüsse werden zum Teil sehr ausführlich beschrieben (E2, 47f; E6, 64f), eher knapp erwähnt (E3, 67; E9, 78) oder sie werden – trotz stark qualitativer Orientierung - kaum thematisiert (E4).

Eine Reihe individueller und organisationaler Rahmenbedingungen begünstigen den Umsetzungsprozess, welche dieser Faktoren für einen bestimmten Betrieb entscheidend sind, zeigt sich aber oft erst während dem Umsetzungsprozesses. Aufgrund von Projekten wie SWiNG wird aber deutlich, worauf die BGF-Beratenden und Betriebsverantwortlichen ein waches Auge haben sollten, um frühzeitig zu reagieren oder proaktiv Einfluss zu nehmen. Im Projekt SWiNG wurden die Beratenden in dieser Hinsicht doppelt gefordert, da die Umwelt der Betriebe sich während des Projektes stark veränderte (wirtschaftliche Unsicherheit) und so die Wahrnehmung und Durchführung des Projektes beeinflusste. (E2, 48)

In einzelnen Evaluationen (E2, E6 und E7) werden die verschiedenen Umsetzungen derselben Intervention einzeln beschrieben und die Evaluationsfragen teilweise (E6) für jede Umsetzung kontextspezifisch beantwortet, bevor dann eine Synthese der Befunde auf einer übergeordneten Ebene gemacht wird.

Im Gegensatz zu den Berechnungen finden komplexe Sachverhalte an vielen Stellen Eingang in die Beschreibungen. Hier bietet sich die Möglichkeit, die im Feld angetroffenen vielfältigen Interaktionen, dynamischen Prozesse, feinen Wirkungsmechanismen und unvorhergesehenen Entwicklungen zu thematisieren. In allen untersuchten Evaluationen werden komplexe Wirkungszusammenhänge beschrieben, allerdings in sehr unterschiedlichem Umfang. In vier der Evaluationen (E2, E4, E6, E7) wird dies sehr ausführlich gemacht und diese Evaluationen bauen die Wirkungsbeurteilung zu einem grossen Teil (E2) oder gänzlich auf diesen Beschreibungen auf. Im Kapitel 8 wird anhand von Beispielen gezeigt, auf welche Weise die Merkmale komplexer sozialer Systeme in den Evaluationen sichtbar werden.

Fazit

Auch bei der Verarbeitung von Komplexität zeigt sich das Spektrum zwischen Rekonstruktionen geringer und hoher Komplexität anhand konkreter Aspekte deutlich. So kann die Evaluation die untersuchten Interventionssysteme mittels Modellen mit einem hohen Abstraktionsgrad oder aber mit sehr differenzierten und kontextspezifischen Modellen, die dann meist induktiv entwickelt werden, rekonstruieren. Die Analysen können eher allgemein, wenig differenziert und monoperspektivisch erfolgen oder aber sehr differenziert und kontextspezifisch sein und unterschiedliche Sichtweisen einarbeiten:

Rekonstruktionen von geringer Komplexität	Rekonstruktionen von hoher Komplexität
einfache Mechanismen und Modelle, lineares Wirkungsverständnis, wenig Analysetiefe	differenzierte Mechanismen und Modelle, Betonung von Interaktion, Wechselwirkungen und Nicht-Linearität, mehr Analysetiefe
deduktiv genutzte Modelle	induktiv entwickelte Modelle
Überprüfen externer Wirkungskonstruktion	auf Verstehen der Eigenlogik der Systeme ausgelegt
allgemeine Analyse	kontextspezifische Analyse
Vergleiche weniger Variablen und weniger Zeitpunkte	differenzierte Vergleiche, Zeitverläufe
monoperspektivisch	Diskussion verschiedener Perspektiven
unkritischer Methodeneinsatz	Reflexion des Potentials und der Grenzen der eingesetzten Auswertungsmethoden

Tab. 16 Handlungsspielraum bei der Verarbeitung von Komplexität

Die Evaluation fällt über vielfältige Arten des Strukturierens laufend Entscheide, welche die Wirkungskonstruktion massgeblich beeinflussen. Auch in Bezug auf die Referenzen, an welcher sich das Strukturieren orientiert, nutzt die Evaluation einen grossen Spielraum. Je nachdem, auf welche Modelle sie sich abstützt, treten andere Wirkungszusammenhänge in den Fokus. Psychologische Referenzmodelle dürften beispielsweise zu anderen Wirkungsbeurteilungen führen als soziologische oder politikwissenschaftliche. Die meisten der untersuchten Evaluationen machen die unterstellte Wirkungslogik der untersuchten Interventionssysteme über Programmtheorien oder Wirkungsmodelle transparent. In diesen konkretisiert sich die Wirkungskonstruktion der Evaluation. Während deduktiv aus der Literatur oder aus Wirkungsvermutungen von Stakeholdern abgeleitete Wirkungskonstruktionen tendenziell zu einfachen und oberflächlichen Modellen führen, führen induktive Vorgehen zu differenzierten und teilweise sehr unüberschaubaren Konstruktionen. In den untersuchten Evaluationen gibt es jeweils mehrere Beispiele für beide Arten.

Alle untersuchten Evaluationen nutzen Vergleiche als zentrales Analyseelement, um zu Wirkungsbeurteilungen zu gelangen. Vergleiche werden nicht nur quantitativ, sondern auch qualitativ realisiert und sie sind sehr unterschiedlich aufgebaut. Auch hier zeigt sich wieder das grosse Spektrum, welches die Komplexität des Untersuchungsgegenstandes bietet: während die einen Evaluationen sich darauf beschränken, wenige Variablen vor und nach einer Intervention zu vergleichen, sind andere Vergleiche sehr differenziert. Wirkungsaussagen, die auf Berechnungen basieren, verwenden in der Regel nur wenige abhängige und unabhängige Variablen und zugrundeliegende Modelle sind wenig differenziert. Dabei wird oft eine Genauigkeit suggeriert, die sich bei näherer Betrachtung nicht halten lässt. Qualitative Analysen zeigen ein grosses Potenzial für differenzierte kontextspezifische Wirkungsbeurteilungen. Sie erlauben es, komplexe Zusammenhänge zu beleuchten und Wirkungsmechanismen differenziert zu untersuchen. Quantitative Vergleiche werden deshalb meist um qualitative Erhebungen ergänzt, zwecks Interpretation der errechneten Ergebnisse.

7.4 Wirkungen beurteilen

Die untersuchten Evaluationen sind sehr unterschiedlich. Im Sample befinden sich randomisierte Kontrollstudien (E1, E9), qualitative Fallstudien (E4, E6, E7), retrospektive Studien mit nachträglich gebildeten Kontrollgruppen (E5, E8), eine quantitativ angelegte Studie ohne Kontrollgruppe (E3) sowie eine Fallstudie mit einem Multi-Methoden-Ansatz (E2).

	Forschungsdesign/methodische Schwerpunkte	Hypothesenprüfend	Explorativ
E1	Randomisierte kontrollierte Studie mit ergänzenden Interviews	x	(x)
E2	Fallstudie mit einem vielfältigen Multi-Methoden-Ansatz	x	x
E3	Quantitativ angelegte Ein-Gruppen Studie (ohne Kontrollgruppe) mit ergänzenden Interviews	x	(x)
E4	Qualitative Fallstudie		x
E5	Retrospektive quantitative Fallstudie mit Vergleichsgruppen	x	
E6	Qualitative retrospektive Fallstudie		x
E7	Qualitative Fallstudie	(x)	x
E8	Retrospektive quantitative Fallstudie mit hypothetischer Kontrollsituation	x	
E9	Randomisierte kontrollierte Studie mit ergänzenden Interviews	x	(x)

Tab. 17 Forschungsdesign der untersuchten Evaluationen

Die Evaluationen nutzen unterschiedliche Erhebungs- und Auswertungsmethoden sowie Kombinationen davon, sie stützen sich auf unterschiedliche Vergleiche ab, über die sie zu Wirkungsaussagen kommen, sie verwenden zum Teil sehr differenzierte, zum Teil sehr einfache und zum Teil gar keine Wirkungsmodelle, sie untersuchen in sehr unterschiedlichem Ausmass Wirkungspotenziale und Wirkungsmechanismen, und sie stützen sich auf verschiedene Datenquellen und Perspektiven ab. Alle Evaluationen machen dabei Wirkungsaussagen, die zum Teil sehr klar und gesichert formuliert sind, zum Teil aber auch sehr vorsichtig und vage.

Strategien der Wirkungszuschreibung (Attribution)

Von Wirkung wird gesprochen, wenn eine Kausalbeziehung zwischen einer Intervention und einem beobachteten Phänomen hergestellt werden kann (vgl. Kapitel 5.4). In den untersuchten Evaluationen geschieht dies auf unterschiedliche Art und Weise. Aus der Analyse aller Wirkungsaussagen in den untersuchten Evaluationen konnten drei zentrale Strategien herauskristallisiert werden, über welche Evaluation zu Wirkungsaussagen kommt. Obwohl in den untersuchten Evaluationen jeweils einzelne Strategien priorisiert werden¹¹⁷, werden meist mehrere kombiniert. Da die einzelnen Strategien jeweils mehr oder weniger stark ausgeprägt sein können, lassen sich die drei Strategien als konzeptuelle Dimensionen darstellen, die einen Möglichkeitsraum aufspannen:

¹¹⁷ Dimension 1 (Kontrafaktische Vergleiche): E1, E3, E5, E8, E9; Dimension 2 (Rekonstruktion): E6, E7; Dimension 3 (Wirkungsbeurteilung durch Befragte (E4)). Die Evaluation E2 stützt sich stark auf alle drei Dimensionen ab.

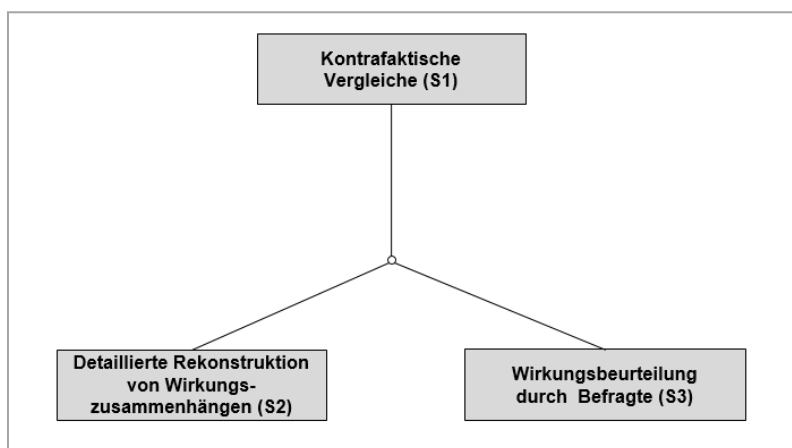


Abb. 23 Strategien der Wirkungszuschreibung

S1 - Kontrafaktische Vergleiche: Beziehungen zwischen beobachteten Phänomenen und Interventionen können einerseits über mehr oder weniger aussagekräftige Vergleiche hergestellt werden, wobei dies in der Regel erst durch Vergleichskombinationen erreicht wird. Soll-Ist Vergleiche, Vorher-Nachher Vergleiche oder Analysen über Entwicklungen über die Zeit geben beispielsweise zunächst nur Hinweise darauf, ob und in welchem Ausmass Veränderungen in die gewünschte Richtung eingetreten sind, sie lassen aber keine Schlüsse darauf zu, inwiefern diese Veränderungen mit einer durchgeführten Intervention in Verbindung stehen oder durch äussere Faktoren bedingt sind. So ist es möglich, dass ein Interventionsziel erreicht wird, obwohl der Intervention keine entsprechende Wirkung nachgewiesen werden kann. Zum Beispiel, wenn eine gewünschte Veränderung sowohl in der Interventions- als auch in der Kontrollgruppe stattgefunden hat (E1, 66). Dies zeigt, dass Ziele immer mit Bezug zur entsprechenden Intervention formuliert oder zumindest gedacht werden müssen und dass grosse Vorsicht angezeigt ist bei der Interpretation beobachteter Veränderungen. Ziele werden teilweise durch die Evaluation inhaltlich hinterfragt, was dazu führen kann, dass trotz Zielerreichung eine schlechte Wirksamkeit attestiert wird:

Das Ziel des Projekts, mindestens eine Massnahme langfristig zu verankern, wurde in den untersuchten Gemeinden erreicht. Allerdings sind es Massnahmen mit geringer Reichweite, un-spezifischer Zielsetzung oder ohne Kontrollen. Ihre Wirkung beurteilen wir deshalb als gering. (E6, Z8)

Auch bei Vorher-Nachher-Vergleichen ist ein Bezug der beobachteten Veränderung zur Intervention zunächst nicht gegeben und muss hergestellt werden, was zum Attributionsproblem führt. Üblicherweise und vor allem in quantitativ angelegten Untersuchungen werden Vorher-Nachher Vergleiche mit Gruppenvergleichen kombiniert, um Wirkungszuschreibungen (Attribution) zu erreichen. Dies kann mittels randomisierter kontrollierter Studien geschehen, denen höchste Aussagekraft bei der Herstellung von Kausalität zugesprochen wird oder aber mittels Vergleichen von Interventionsgruppen mit nicht zufällig gebildeten Vergleichsgruppen. Die Vielfalt an Möglichkeiten ist sehr gross, mit jeweils unterschiedlichem Potenzial und unterschiedlichen Grenzen. Der Zusammenhang zwischen einer festgestellten Veränderung und der Intervention wird über statistische Tests hergestellt und in seiner Bedeutsamkeit beurteilt. Sowohl in E1 als auch in E9 basieren Wirkungsaussagen primär auf derartigen statistischen Wahrscheinlichkeiten aus randomisierten kontrollierten Studien. Weitere Evaluationen basieren ihre Wirkungsaussagen auf statistische Vorher-Nachher-Vergleiche in Kombination mit Untergruppenvergleichen (E2, E3) oder auf Zeitreihen- oder Vorher-Nachher-Vergleiche zwischen Kantonen (E5, E8).

Tabelle 28 belegt, dass - kontrolliert für Alter und Geschlecht - ein signifikanter Interventionseffekt, also eine Reduktion von etwa 3.5 Gläsern pro Woche im Vergleich zur Kontrollgruppe, für die Risikogruppe R gefunden werden kann. (E1, 38)

In der jüngsten Altersgruppe der 10- bis 15-Jährigen sanken im Kanton Genf die Intoxikationsraten ab dem Beginn der Intervention, während sie in den anderen Kantonen weiter anstiegen. Dieser Effekt wurde in beiden Zeitreihen-Modellen (ARIMA) signifikant. (E5, 1)

In qualitativ angelegten Untersuchungen (z.B. in Fallstudien) wird in Vorher-Nachher Vergleichen oder Vergleichen über die Zeit die Wirkungszuschreibung nicht über gleichzeitige Kontrollgruppenvergleiche statistisch plausibilisiert, sondern Bezüge zwischen Veränderung und Intervention werden durch die Analyse von Wirkungspotenzialen und –mechanismen hergestellt und plausibilisiert.

Insgesamt gesehen hat die Durchführung von "Die Gemeinden handeln!" in den Gemeinden mit abgeschlossener Alkoholpolitik eher wenig bewirkt; wir beurteilen die Veränderungen als klein bis mittel. Die Massnahmen haben oft eine geringe Reichweite, und zudem sind klare Ziele und Massnahmenpläne, eben Politiken, kaum vorhanden. In den Gemeinden mit Alkoholpolitik in Entwicklung sind die Veränderungen bisher gering. Allerdings ist aus heutiger Sicht fraglich, ob bis Abschluss des Projekts Ende dieses Jahres Ziele und Strukturen beschlossen werden, die zu deutlicheren Veränderungen führen. (E6, 82)

S2 - Detaillierte Rekonstruktion von Wirkungszusammenhängen: In dieser Strategie stützen sich Wirkungsaussagen auf eine differenzierte Analyse und Rekonstruktion des Interventionssystems und seiner Wirkungsketten ab. Dabei wird die vermutete Wirkungsbeziehung zwischen der Intervention und der beobachteten Veränderung in einzelne Ursache-Wirkungsbeziehungen mit kleinerer zeitlicher und räumlicher Distanz zerlegt. Bei diesem Ansatz wird im Wesentlichen von der Annahme ausgegangen, dass durch die grössere zeitliche und räumliche Nähe Wirkungszusammenhänge plausibler konstruiert werden können, um sie dann zu einem besser abgestützten Gesamtbild zusammenfügen zu können. Durch diese detaillierte Betrachtung wird versucht, das Interventionssystem und die Wirkungslogik im Detail zu verstehen und über den Weg der Analyse einzelner intermediärer Wirkungsdikatoren und Wirkungsketten den Zusammenhang zwischen der Intervention und der beobachteten Veränderung zu plausibilisieren.

Der Trick, wenn man das so bezeichnen will, besteht nun darin, dass die Jugendlichen sich mit der vorgegebenen Struktur auseinandersetzen müssen, wenn sie diesen sozialen Raum für sich auf Dauer nutzen wollen, weil die verantwortlichen Erwachsenen, unterstützt durch die Jugendlichen, die in die Coach-Rolle schlüpfen, die Geltung der Modelle, Prinzipien und Regeln durchsetzen. Dadurch entsteht ein prinzipiell fruchtbarer Mix aus symmetrischen und asymmetrischen sozialen Strukturierungen, der das Lernfeld kennzeichnet und sich in diverse polare Dimensionen ausdifferenziert. Die Wirkung bei den Jugendlichen im Hinblick auf die vom Projekt verfolgten Zieldimensionen entsteht also quasi indirekt. Indem Midnight Basketball den Jugendlichen den Raum und seine Struktur zur autonomen Nutzung zur Verfügung stellt, entstehen selbstorganisierte Erfahrungsräume in der sozialen Welt der Jugendlichen, die durch den sozialen Sinn angereichert sind, der von Midnight Basketball repräsentiert und von den beteiligten Erwachsenen und Coaches verkörpert wird. (E7, 29)

Insbesondere bei der Richterswiler Beachparty, die jährlich Tausende von meist jugendlichen Besucherinnen und Besuchern anlockt, konnten Verbesserungen erzielt werden. Mit den Betreibern der Beachparty wurde mehrmals das Gespräch gesucht und Auflagen wurden gemacht, um die Einhaltung der Jugendschutzbestimmungen zu verbessern (Verbot von Alcopops). (E6, 21)

Die detaillierte Rekonstruktion des Interventionssystems erfolgt über explorative qualitative Zugänge, mittels welchen auch zunächst unbekannte Faktoren und Zusammenhänge identifiziert werden können (E4, 68; E7, 28). Erkenntnis wird dabei durch Nähe der Forschenden zum Untersuchungsgegenstand und durch eine offene Herangehensweise erreicht resp. erleichtert:

Die Herangehensweise, die für diese Art der sozialwissenschaftlichen Forschung charakteristisch ist, kann am besten als „explorativ“ bezeichnet werden. Das heisst, dass eine Annäherung an den Gegenstand (hier Midnight Basketball) ohne vorgefasste Modellannahmen stattfindet. Entsprechend wurde die Beobachtung als erster Schritt der Studie zunächst offen und unstrukturiert konzipiert (Girtler 2001). (E7, 28)

S3 - Wirkungsbeurteilung durch Befragte: Zum Teil werden die Wirkungsbeurteilungen befragter Personen von der Evaluation übernommen und direkt als Interventionswirkungen ausgegeben (E4, E6). Diese Wirkungen können mehr oder weniger gut abgestützt sein, wobei das „gut“ verschiedene Dimensionen umfassen kann, so zum Beispiel die eigene Klarheit bezüglich des gemachten Wirkungsurteils:

Beim Projekt „Stadtquartier“ wird Midnight Basketball von der interviewten Person ebenfalls eine hohe Wirksamkeit zugesprochen. Es werden in diesem Zusammenhang die Junior Coaches, deren Entwicklung gefördert werde, erwähnt. Dies sei im Alltag zu beobachten und mache sich in ihrem Verhalten in Bezug auf Abfallentsorgung bemerkbar. (E7, 47)

versus:

Beim Projekt „Agglomeration“ hat die Gemeindevertretung den Eindruck, dass Midnight Basketball eine positive Wirkung für die Gemeinde und auch für deren Bild entfaltet. (E7, 47)

Die Evaluation weist sowohl für die Mädchen selbst wie auch für die Trägerinstitutionen des Projekts positive Wirkungen und Veränderungen nach. Die Mädchen äussern, dass sie Regeln kennengelernt haben, wie sie sich verhalten sollen, und dass reden besser ist als schlagen. Sie schränken ein, dass sie dieses Wissen nicht immer in die Tat umsetzen, finden aber dennoch, MuM hätte schon «ein bisschen» geholfen, «wenn diejenigen, die Streit haben, Hilfe wollen und etwas verändern wollen – dann kann MuM helfen» – so äussert ein Mädchen. (E4, 73)

Eine andere Dimension der Abstützung betrifft den Umfang der berücksichtigten Perspektiven. Bei den Wirkungsbeurteilungen kann es sich um die Meinung einer einzelnen Person handeln, oder aber die Urteile mehrerer Personen deuten in dieselbe Richtung (E6, A1). Zudem werden die Wirkungsbeurteilungen mitunter durch Erkenntnisse aus anderen Erhebungen der Evaluation (Beobachtungen, Dokumentenanalysen) (E6, 15) oder Daten aus dem Feld gestützt:

Durch die wiederholten Testkäufe konnten Wirte und Leiter von Verkaufsstellen für die Problematik Jugendschutz sensibilisiert werden; dies zeigten die verbesserten Ergebnisse der Testkäufe. (E6, 41):

Manchmal werden derartige Wirkungsbeurteilungen nicht weiter ausgeführt, manchmal werden sie aber ausführlich umschrieben und so begründet und nachvollziehbar gemacht. Auf den Umstand, dass sich die Befragten der Schwierigkeit der Wirkungszuschreibung zum Teil selber sehr wohl bewusst sind, wurde bereits hingewiesen.

Kombination S1-S2: Sehr häufig werden die Strategien „Kontrafaktische Vergleiche“ und „Rekonstruktion von Wirkungszusammenhängen“ kombiniert und zwar u.a. in Kombination mit quantitativer mit qualitativer Methodik. Zum Beispiel dann, wenn in Studien mit Kontrollgruppendesign qualitative Verfahren eingesetzt werden, um Befunde aus quantitativen Vergleichen hinsichtlich Steuerungs- und Optimierungswissen anzureichern (E1, 47; E3, 52) und/oder zu plausibilisieren (E1, 47; E9, 13). Eine Kombina-

tion ist aber auch in anderer Hinsicht möglich, nämlich die Verwendung von Gruppenvergleichen, qualitativ oder quantitativ, zur Unterstützung der Rekonstruktion des Interventionssystems (E2, E6, E7):

Es wurde eine vergleichende Fallstudienanalyse durchgeführt (insgesamt 13 Fallstudien). Neben Gemeinden, die das Projekt abgeschlossen haben bzw. im Projekt arbeiten, wurden Gemeinden untersucht, die sich einmal dafür interessierten und Kontakt mit dem Projekt hatten, dann aber von einer Teilnahme absahen. Zudem wurden im Sinne einer Kontrollgruppe Gemeinden studiert, die sich nie für das Projekt interessierten und am Projekt nicht teilnahmen. (E6, Z-5)

Kombination S1-S3: Immer dann, wenn Wirkungsurteile von Befragten in systematische Vergleiche einfließen, liegt eine Kombination zwischen der Strategie „Kontrafaktische Vergleiche“ und der Strategie „Beurteilung durch Befragte“ vor.

Eine interessante Kombinationsmöglichkeit liefert E2. Dort wird die Wirkungseinschätzung der Befragten als Grundlage für Untergruppenvergleiche genutzt. Die Gesamtwirkung ergibt sich aus der Differenz im Zielindikator zwischen der Gruppe mit hoher und der Gruppe mit mittlerer/tiefer Wirkungseinschätzung:

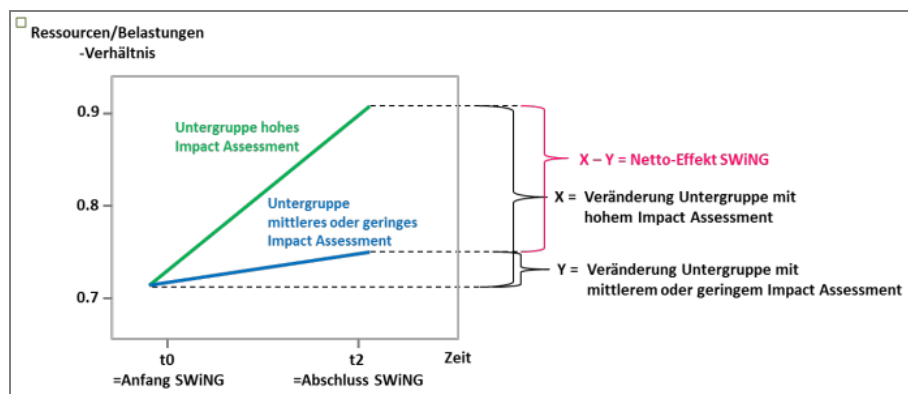


Abb. 24 Berechnung der Projektwirkung
(Jenny et al. 2011 (E2), 73)

Dieser Vergleich geht davon aus, dass alle Personen in der untersuchten Zeit eine Entwicklung machen und sie lassen diese Veränderung bei Personen, die der Intervention wenig Wirkung zuschreiben, nicht als Wirkung gelten. Da angenommen wird, dass alle Personen diese Veränderung auch ohne Intervention gemacht hätten, wird nur der Zugewinn an Veränderung in der Gruppe mit hoher Wirkungseinschätzung als Wirkung deklariert.

Kombination S2-S3: Und letztlich werden natürlich auch Wirkungsbeurteilungen von Befragten zur Rekonstruktion des Interventionssystems verwendet, nämlich indem nicht nur danach gefragt wird, ob die Intervention gewirkt habe, sondern insbesondere auch wie:

Die Frage, was das Projekt MuM bewirkt hat, in welcher Weise es für die Mädchen Wirkungen zeigte, wurde in der Evaluation auch den Nutzerinnen selbst – den Mädchen – gestellt. Sie nehmen gegenüber dieser Frage eine ambivalente Haltung ein. Einerseits finden sie, MuM hätte bei ihnen etwas verändert und bewirkt, dass sie weniger (körperliche und verbale) Gewalt anwenden. Sie hätten gelernt, dass es besser sei, zu reden, und zum Beweis werden die von der Projektausführenden gelernten und diskutierten Verhaltensmöglichkeiten aufgezählt. (E4, 61)

Grenzen der Wirkungsnachweise

Trotz der verschiedenen Wege, die eingeschlagen werden, um beobachtete Veränderungen den untersuchten Interventionen zuzuschreiben (Attribution) und sie dadurch als Wirkungen zu deklarieren, bleibt diese Zuschreibung grundsätzlich eine grosse Herausforderung (E4, 63/64; E5, 30). Es bleibt immer die Frage, ob die beobachteten Veränderungen nicht auch durch externe Einflüsse hätten zustande kommen können:

Es ist deutlich schwieriger nachzuweisen, dass eine Verbesserung bei den Nutzenarten tatsächlich auf die BGF-Massnahmen zurückzuführen ist. Eine Verbesserung der Lebensqualität der Mitarbeitenden könnte auch auf andere Ereignisse, wie einen neuen Abteilungsleiter, eine Verbesserung der allgemeinen Wirtschaftslage oder reinen Zufall zurückzuführen sein. (E2, 67)

Gerade wenn grössere Umwälzungen mit potenziell grossem Einfluss auf die Zielindikatoren stattfinden, wird die Interpretation von Veränderungen schwierig (E2, 39). In vielen Formulierungen zeigen sich die Unsicherheiten, mit denen diese Wirkungszuschreibungen verbunden sind. Es werden „mögliche“ (E9, 13), „wahrscheinliche“ (E1, 42; E3, 5), und „naheliegende“ Wirkungen (E1, 36; E9, 13) proklamiert. Es wird „ein Eindruck von Wirkung“ (E4, 47) festgestellt oder der Interventionsgruppe werden „eher“ positive Veränderungen nachgesagt (E1, 38; E8, 80). Wirkungen scheinen „plausibel“ (E2, 81) oder werden „mit grosser Sicherheit“ (E5, 30) nachgewiesen. Zuweilen wird auch vor einer eindeutigen Wirkungszuschreibung gewarnt (E5, 27). Klarere Aussagen werden zum Beispiel dadurch verhindert, dass zu wenig von der geplanten Intervention realisiert werden konnte (E6, 87), die Messkriterien zu wenig sensibel resp. die Veränderungen zu klein waren (E2, 37), viele Annahmen getroffen werden mussten (E8, 53), die Datenqualität (E2, 16/81) oder die Datenmenge ungenügend war (E2, 59; E5, 9), nicht alle Faktoren berücksichtigt werden konnten (E2, 81), die Ergebnisse schwer zu interpretieren sind (E2, 59/116) etc. (vgl. dazu auch das Kapitel ‚Grenzen der Erfassung‘). Die Schwierigkeit der Wirkungszuschreibung und damit klarer Wirkungsnachweise betrifft quantitativ angelegte Forschung:

Aus den realisierten statistischen Analysen lässt sich auch ein knapp signifikanter Zusammenhang zwischen der Bekanntheit der Kampagne und dem Helmtragen erkennen. Dennoch kann die Tatsache, dass sich die Helmtragquote in den letzten Jahren im Schneesport deutlich erhöht hat, nicht in erster Linie der Kampagne zugeschrieben werden. (E3, 67)

Aber ebenso wie qualitativ orientierte:

Dies mache es schwierig, darüber zu befinden, ob nun MuM oder etwas Anderes oder verschiedene Einflüsse gemeinsam den Ausschlag für positive Veränderungen geben. Unter Berücksichtigung dieser Vorbehalte äussert die Projektausführende dennoch, dass sie bei den Mädchen erfreuliche Veränderungen bemerkt hat. (E4, 64)

Auch die Wirkungsweise einzelner Teilelemente im gesamten Wirkungsmechanismus ist sehr schwierig einzuschätzen:

Wie die Ressourcen genau im Veränderungsprozess wirken und wie sie durch individuelle Kompetenzen oder organisationale Kultur unterstützt werden, ist sehr schwierig zu beurteilen: Die Evaluation kann nur Puzzle-Teile liefern, nicht aber die Komplexität des ganzen Systems abbilden. (E2, 63)

Es kommt zudem vor, dass festgestellte Korrelationen fälschlicherweise als Wirkungen ausgegeben werden (E3, 65ff). In folgendem Beispiel werden mittels Regressionsanalysen Korrelationen überprüft und zunächst richtig interpretiert:

Die Hypothese, dass Personen, die die bfu-Kampagne «1000 Unfälle pro Tag – Schütz dich mit einem Helm» kennen, häufiger einen Helm tragen, wird durch diese Analyse bestätigt. (E3, 65)

Um sie dann aber im Fazit als Wirkungen auszugeben:

Schliesslich lässt sich mit den durchgeführten Analysen auch ein signifikanter Einfluss der Kampagne auf die Helmtragquote nachweisen. (E3, 66)

Wenn der Einfluss externer Faktoren auf die Wirkungsindikatoren mit in Betracht gezogen wird, dann ist oft von einem Beitrag der Intervention oder einzelner Wirkfaktoren an die beobachtete Veränderung die Rede (Kontribution), oder es wird von einem Mitnahmeeffekt gesprochen (E6, 84)¹¹⁸. Das gilt sowohl in qualitativ (E7, iv), wie auch in quantitativ orientierten Studien (E8, 14):

Die Ergebnisse zeigen, dass Informations- und Schulungsmassnahmen wesentlich zur Abnahme der Strassenverkehrsunfälle, des Tabakkonsums und des übermässigen Alkoholkonsums beigetragen haben. (E8, 15)

Dabei wird dieser Wirkungsbeitrag auch in seinem Umfang qualifiziert, so z.B. als „bescheiden“ (E3, 66), „wesentlich“ (E8, 65), oder „hoch effektiv“ (E7, 50) deklariert und in seltenen Fällen auch quantitativ bestimmt respektive geschätzt:

Dies entspricht einer Abnahme von etwa 55'000 Personen mit einem schädigenden Alkoholkonsum. Unter Berücksichtigung des Preiseffekts ist fast die Hälfte der Abnahme (47,3% mit einer Spanne von 22,2% bis 60,2%) auf die Prävention zurückzuführen. (E8, 14)

Gleichwohl werden aber in allen Evaluationen und damit über alle unterschiedlichen Forschungsdesigns und Wirkungskonstruktionen klare Wirkungszuschreibungen gemacht (E8, 12/66; E9, 84) bzw. alle Evaluationen machen eindeutige Aussagen über Wirkungen (E5, 22; E8, 12), Teilwirkungen (E2, 60; E4, 63), Nichtwirkungen (E1, 41; E9, 14), Entwicklungen in die gewünschte Richtung (E3, 14; E5, 25) oder über einzelne Wirkungszusammenhänge (E1, 26; E4, 75). Klare Wirkungsaussagen gründen sowohl auf Berechnungen (E1, 42; E2, 61) als auch auf qualitativen Analysen (E2, 63; E6, Z-9).

Es liess sich zunächst kein klares Muster erkennen, in welchen Fällen die Evaluation zu klaren Wirkungsaussagen kommt und in welchen die Wirkungsaussagen unsicher sind. Wie angesprochen läuft die Trennlinie nicht entlang unterschiedlicher Forschungsdesigns- oder Methoden und es lag die Vermutung nahe, dass es zu einem grossen Teil den Evaluierenden obliegt, wie eindeutig sie Wirkungsaussagen einschätzen und ausformulieren. Bei einer weiteren Prüfung der Daten erhärtete sich diese Vermutung. Während einzelne Evaluationen sehr vorsichtig mit Wirkungszuschreibungen umgehen (E3), sind andere, die sich methodisch aber nicht erkennbar besser abstützen, freizügiger mit Wirkungszuschreibungen (E2, E4, E8).

Wirkungen plausibilisieren

Es lässt sich anhand der untersuchten Evaluationen deutlich zeigen, dass die Evaluationen bei ihrer Wirkungskonstruktion mit unterschiedlichsten Herausforderungen konfrontiert sind, welche Entscheide verlangen, die man auch anders fällen könnte und die immer ein gewisses Mass an Unsicherheit und eine Limitierung der Aussagekraft der Ergebnisse mit sich bringen. Die Evaluation begegnet diesem Umstand teilweise dadurch, dass sie das Mass an Unsicherheit durch zusätzliche Belege zu reduzieren sucht. Es konnten die folgenden Formen der Plausibilisierung identifiziert werden:

Interne Vergleiche

- Vergleiche verschiedener ähnlich gelagerter Indikatoren (E2, 36/75; E3, 16; E5, 27)
- eine Abstützung auf viele Textstellen in qualitativen Studien (E9, 54f)
- nachträgliche zusätzliche Erhebungen im Rahmen der Evaluation (E1, 50)

¹¹⁸ „Ein wichtiges Kriterium für die Wirkung oder den Effekt eines Projekts ist die Frage, ob das, was gemacht wurde, auch ohne das Projekt gemacht worden wäre. Ist das der Fall, spricht man von Mitnahmeeffekt.“ (E6, 84)

- die Weiterführung und erneute Evaluation der Intervention (E4, 77)
- die Bewertung der Qualität der Konzeption und damit des Wirkungspotenzials der Intervention (E2, 78; E3; E6)

Externe Vergleiche

- Vergleiche mit anderen Interventionen (E9, 77)
- Daten anderer Quellen zum untersuchten Feld (E5, 28)
- Überprüfung an der Praxis (E5, 29)

methodische Erweiterungen

- gleichzeitiger Einsatz verschiedener Methoden (andere Methode, mehr Tiefe, Vergleichsgruppen, andere Berechnungsmodelle ...) (E2, 87; E8, 12; E4, 19; E7, 47)
- Einbettung der Ergebnisse in einen weiteren (zeitlichen) Kontext (E3, 14; E8, 75)
- Einbezug anderer/weiterer Perspektiven (E4, 56; E9, 13)

fachliche/wissenschaftliche Abstützung

- Theorie/Kriterien aus der Literatur (E2, 42; E6, Z-9)
- Ergebnisse anderer Studien (E5, 2; E5, 6; E8, 65)
- Expertenmeinungen (E5, 4; E8, 39)

Es werden dabei sowohl Berechnungsergebnisse mittels qualitativer Erhebungen plausibilisiert (E1, 47) als auch beispielsweise subjektive Einschätzungen mittels objektiver Messungen (E9, 13ff).

Das Standard-Bewegungsprogramm führte nach Aussagen der Eltern zu mehr körperlicher Aktivität der Kinder und weniger, vor dem Fernseher verbrachter Zeit. Bei den objektiven Messungen ließ sich eine Änderung des Bewegungsverhaltens allerdings nicht beobachten. (E9, 13ff)

Zum Teil werden unterschiedliche Szenarien für Wirkungszuschreibungen diskutiert und dem Leser die Wahl gelassen, welche Lesart ihm plausibler erscheint (E1, 50; E5, 2). Meistens werden aber gezielt einzelne Belege angeführt, die eine Wirkungsaussage bestärken, ohne dass dabei beurteilt werden kann, ob es nicht auch andere gegeben hätte, wenn man danach gesucht hätte.

Wirkungen verallgemeinern

In einigen der untersuchten Evaluationen (E1, E2, E3, E6, E9) wird ein klarer Anspruch an eine Verallgemeinerbarkeit der Ergebnisse resp. der Interventionswirkungen über die untersuchte Stichprobe oder über den untersuchten Evaluationsgegenstand hinaus erhoben, in einer der Evaluationen nur andeutungsweise¹¹⁹ (E7). Dabei wird in zwei Evaluationen statistische Repräsentativität der Ergebnisse angestrebt (E3, 75; E9, 97), in den anderen Fällen geht es um eine inhaltliche Übertragbarkeit der Ergebnisse über die untersuchten Fälle hinaus. In Bezug auf die Verallgemeinerbarkeit von Wirkungsaussagen über die untersuchten Fälle hinaus zeigen sich in den untersuchten Evaluationen drei Spannungsfelder:

Breite und Tiefe: Zum einen erscheint sowohl hinsichtlich statistischer als auch qualitativer Repräsentativität die Menge und angemessene Verteilung der untersuchten Fälle ein zentrales Kriterium zu sein, um Repräsentativität geltend machen zu können. Im ersten Fall (E3, E9) rechtfertigt eine statistisch repräsentative Stichprobe allein die Übertragbarkeit der Ergebnisse auf die Grundgesamtheit. Im Falle einer qualitativen Repräsentativität sind nicht statistische Kriterien leitend. Durch eine inhaltlich möglichst breite Abdeckung von unterschiedlichen Fällen (E6, Z-6; E2, 25) und durch eine vertiefte qualitative

¹¹⁹ „Abschliessend bleibt festzuhalten, dass wir die prinzipielle Wirkungsweise und einige Faktoren, die auf die Wirkungstiefe einen Einfluss haben, modellhaft beschreiben können.“ (E7, 52)

Analyse soll die Thematik derart umfassend ausgeleuchtet werden, dass die wesentlichen Aspekte und Zusammenhänge, die potentiell an anderen Orten auftreten können, in den untersuchten Fällen bereits sichtbar werden. Es wird davon ausgegangen, dass die Aussagen zur Wirksamkeit einer Intervention deshalb argumentativ auf andere Orte übertragbar sind.

Es wird also nicht quantitative Repräsentativität angestrebt. Angestrebt wird qualitative Repräsentativität oder die Repräsentativität des Inhalts. Damit ist gemeint, dass die ganze inhaltliche Breite des Untersuchungsgegenstandes erfasst wird.“ (E6, 16) „Das heisst, dass von den Ergebnissen der Fallstudien in den acht Gemeinden nicht im statistischen Sinne auf die anderen, nicht untersuchten Gemeinden geschlossen werden kann. Sie sind aber argumentativ verallgemeinerbar. (E6, Z-6)

Während in E6 die Anzahl der untersuchten Fälle mit 13 relativ gross ist, baut eine andere Evaluation (E7) auf lediglich drei untersuchten Standorten auf. Die Evaluation geht in ihrer Analyse aber tiefer und analysiert in vielen Beobachtungen und Gesprächen eine grosse Menge und Vielfalt an Situationen (und damit Fällen), die sie zu einem Modell verdichtet, das über die drei Standorte hinaus „modellhafte“ Gültigkeit haben soll.

Abschliessend bleibt festzuhalten, dass wir die prinzipielle Wirkungsweise und einige Faktoren, die auf die Wirkungstiefe einen Einfluss haben, modellhaft beschreiben können. Dadurch wird deutlich, dass Midnight Basketball ein komplexes System ist, das für die Jugendlichen eine wichtige Ressource ist. (E7, 52)

Auf der einen Seite ist es also die Breite der untersuchten Fälle und Situationen, welche Verallgemeinerbarkeit begründet, auf der anderen Seite ist es die Tiefe der Analyse.

Kontextspezifisch versus allgemein: Ein weiteres Spannungsfeld zeigt sich hinsichtlich der Kontextspezifität der Untersuchungen und Ergebnisse. Während auf der einen Seite einzelne Interventionen möglichst detailliert und spezifisch untersucht werden, um Wirkungspotenziale und Wirkungsmechanismen im Detail kennen und verstehen zu lernen (E6, E7), wird auf der anderen Seite argumentiert, dass allzu spezifische Untersuchungen und Ergebnisse einer Verallgemeinerbarkeit im Wege stehen:

Somit erarbeitete die Evaluation Ergebnisse, die für die breite Belegschaft von Grossbetrieben gültig sind. Dies steht im Gegensatz zu Studien mit fokussierten Risikogruppen mit sehr spezifischen Problemen und Interventionen. (E2, 17)

Durch Verdichtung werden kontextspezifische Besonderheiten (die Umsetzungen unterscheiden sich zum Teil deutlich (E2, 79)) ausnivelliert und damit eine breitere Gültigkeit des verwendeten Modells und der Ergebnisse deklariert:

Die Evaluation verdichtet diese Variablen entlang dem zentralen Organisationsmodell und dem JobDemands-Resources-Modell (...), ihrem Evaluationsdesign (...), aufgrund qualitativer Informationen und methodischer Überlegungen zu vier Faktoren: Ressourcen, Belastungen, positive und negative Gesundheit. Dadurch reduziert sich die Komplexität der Analysen erheblich und es kann ein Modell verwendet werden, das für alle (Gross-) Betriebe gleichermassen Gültigkeit hat, d.h. die Ergebnisse der Evaluation lassen sich auf andere vergleichbare Betriebe übertragen. (E2, 32)

In einzelnen Evaluationen werden Ergebnisse resp. Wirkungsaussagen sowohl für jedes untersuchte Setting einzeln ausgewiesen (E2, E6, E7), als auch zu allgemeinen Aussagen verdichtet.

Interne versus externe Validität: In der einen Evaluation (E1) wird die Frage der Übertragbarkeit von Ergebnissen aus einer anderen Perspektive beleuchtet. Sie fokussiert auf die Übertragbarkeit der Ergebnisse auf reale Situationen und grenzt sich damit ab von kontrollierten Studien, die zwar eine

hohe interne Validität der Ergebnisse erreichen, aber kaum über das kontrollierte Setting hinaus Aussagekraft haben. Die Studie nimmt dabei bewusst einen Verlust an interner Validität in Kauf, um ein höheres Mass an externer Validität zu erreichen.

Sie prüft somit die direkte Umsetzbarkeit einer präventiven Massnahme in der Situation, in der diese Massnahme auch wirklich umgesetzt werden soll. Sogenannte „Efficacy-Trials“ haben zwar in der Regel eine höhere interne Validität, da sie Situationen, die die direkte Überprüfbarkeit der Wirkung der Intervention gefährden, ausschliessen (ungleiche Kontrollgruppen, Panelverluste etc.). Sie sind aber aufgrund dieser „experimentellen“ Situation (z.B. Bezahlung von Probanden; Randomisieren von Teilnehmenden in „unnatürliche Gruppen“ (z.B. ausserhalb des Klassenverbandes) häufig nur schwer auf reale Situationen generalisierbar. Aus diesem Grund wurde ein „Effectiveness-Trial“ konzipiert, d.h., dass die entsprechenden Ergebnisse direkt für die Praxis umsetzbar sind. (E1, 3)

Fazit

Da die Wirkungsbeurteilungen auf der Eingrenzung, Erfassung und Verarbeitung von Komplexität aufbauen, zeigt sich auch hier dasselbe Spektrum zwischen geringer und hoher Komplexität. Während die einen Evaluationen auf intendierte Wirkungen fokussieren, nehmen andere auch nicht-intendierte Wirkungen systematisch in die Suche und Beurteilung auf. Auf der einen Seite werden eindeutige Wirkungsnachweise proklamiert und Wirkungen verallgemeinert, auf der anderen Seite werden differenzierte und kontextspezifische Wirkungsbilder gezeichnet und Wirkungszuschreibungen sehr vorsichtig formuliert und verallgemeinert:

Rekonstruktionen von geringer Komplexität	Rekonstruktionen von hoher Komplexität
Fokus auf intendierte Wirkungen	Berücksichtigung nicht-intendierter Wirkungen
starke Verdichtung	Belassen von Vielfalt
eindeutige Wirkungsnachweise	Diskussion der Grenzen von Attribution
Wirkung als Ergebnis der durchgeführten Intervention verstanden	Wirkung als Co-Produktion und gemeinsamer Lern- und Entwicklungsprozess verstanden
allgemeingültige Wirkungsaussagen und unkritische Generalisierungen	kontextspezifische Wirkungsbeurteilungen und vorsichtige Generalisierung
keine Plausibilisierungen der Wirkungsaussagen	Nutzung verschiedener Formen der Plausibilisierung der Wirkungsaussagen
fehlende kritische Reflexion der Wirkungszuschreibungen und des Einflusses der Evaluation	selbstkritische Reflexion der Wirkungszuschreibungen und des Einflusses der Evaluation

Tab. 18 Handlungsspielraum bei der Beurteilung von Wirkungen

Wirkungsaussagen in den untersuchten Evaluationen basieren jeweils entweder auf kontrafaktischen Vergleichen, auf der detaillierten Rekonstruktion von Wirkungszusammenhängen oder aber auf Wirkungszuschreibungen von Befragten. In den meisten Evaluationen werden diese Strategien kombiniert verwendet. Die Zuschreibung von beobachteten Veränderungen zu Interventionen wird weitgehend als grosse Herausforderung anerkannt und entsprechend fallen viele Wirkungsaussagen zurückhaltend aus. Trotzdem werden in den meisten Evaluationen auch ganz klare Wirkungsaussagen gemacht. Weil sich nicht erkennbar auf methodische oder andere Unterschiede zurückführen lässt, wann Wirkungsaussagen als gesichert dargestellt werden und wann nicht, wird vermutet, dass dies stark von den Evaluierenden selber abhängt – die einen sind freizügiger mit Wirkungszuschreibungen, andere sind vorsichtiger. Um Unsicherheit in den Wirkungszuschreibungen zu reduzieren, werden Wirkungsaussagen auf unterschiedliche Weise plausibilisiert.

Die Evaluationen zeigen sehr schön die Spannungsfelder, die sich daraus ergeben, dass in komplexen sozialen Systemen Wirkungsweisen immer sehr spezifisch sind und andererseits Ergebnisse gewünscht werden, die kontextübergreifend gültig und verwendbar sind.

7.5 Synthese

Die letzten Kapitel sind der Frage nachgegangen, wie Evaluationen konkret vorgehen, um in komplexen sozialen Systemen Wirkungen zu beurteilen. Wirkungen zeigen sich dabei keineswegs als objektiv gegebene Sachverhalte, die mittels ‚richtiger‘ Methodik lediglich identifiziert und gemessen zu werden brauchen. Vielmehr sind Wirkungen das Ergebnis eines differenzierten Konstruktionsprozesses, in welchen unterschiedliche Perspektiven einfließen und in denen viele Entscheide gefällt werden müssen, die sich auf das auswirken, was letztlich als Wirkung ausgegeben wird. Dabei konnte gezeigt werden, dass Wirkungsfragen mit vielen Ungenauigkeiten, Limitierungen und offenen Fragen verbunden sind. Der Evaluation kommt die Rolle zu, den Wirkungskonstruktionsprozess zu steuern, um letztlich die Evaluationsergebnisse zu verantworten. In Abstimmung mit den Auftrag Gebenden und im Rahmen verfügbarer Ressourcen kann sie enge Grenzen ziehen oder mehr Offenheit zulassen, viele oder nur wenige Perspektiven einbeziehen, eher auf standardisierte Erhebungsverfahren setzen oder aber explorativ vorgehen, sich auf statistische Kalkulationen abstützen oder Wirkungen explorativ rekonstruieren, sich bei der Beurteilung von Wirkungen mehr oder weniger auf das Urteil anderer abstützen etc. Die unterschiedlichen beobachteten Wirkungskonstruktionen lassen sich in folgendem Schema einordnen, welches sich über den gesamten Prozess der Wirkungskonstruktion erstreckt:

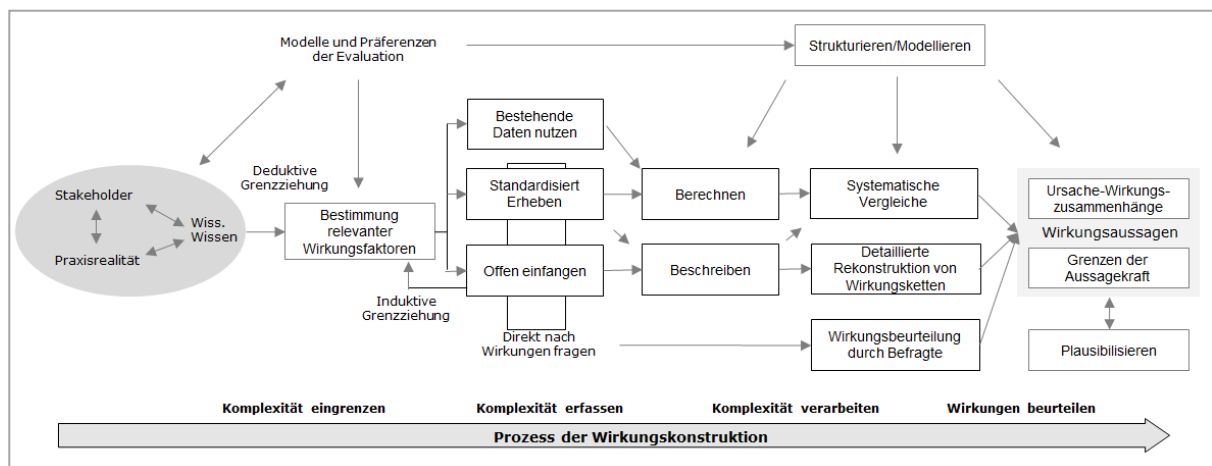


Abb. 25 Der Prozess der Wirkungskonstruktion

Am Anfang der Wirkungskonstruktion steht das Spannungsfeld zwischen Stakeholderperspektiven, Praxisrealität und wissenschaftlichem Wissen. Der Evaluation, welche zusätzlich eigene Modelle und Präferenzen in die Diskussion einbringt, kommt in Absprache mit den Auftraggebern die Aufgabe zu, aus den Perspektiven, die dieses Spannungsfeld prägen, die für die Wirkungsbeurteilung relevanten Faktoren abzuleiten und darauf aufbauend das Evaluationsdesign und die Evaluationsmethoden zu definieren (‚deduktive Grenzziehung‘). Zum Teil wird auf bestehende Daten zurückgegriffen, zum Teil werden Daten standardisiert erhoben und zum Teil wird explorativ an den Untersuchungsgegenstand herangegangen. Bei offener Herangehensweise ergeben sich in der Regel neue Wirkungsfaktoren, die in die Betrachtung einfließen (‚induktive Grenzziehung‘). Je nach methodischer Ausrichtung werden die Daten quantitativ (‚Berechnen‘) oder qualitativ (‚Beschreiben‘) verarbeitet, wobei die Daten in diesem Analyseprozess auf unterschiedliche Weise strukturiert oder gar modelliert werden. Zu einer Wirkungsbeurteilung gelangen die Evaluationen über drei Wege, die oft kombiniert werden: über kont-

rafaktische Vergleiche, über die detaillierte Rekonstruktion von Wirkungszusammenhängen und/oder über das Abstützen auf Wirkungsbeurteilungen von Befragten. Es resultieren Wirkungsaussagen, die zum Teil sehr klar, zum Teil aber auch sehr unsicher und vorsichtig formuliert sind. Sie sind eingebettet in Ursache-Wirkungszusammenhänge, die wiederum stark von der Art der Strukturierung und Modellierung durch die Evaluation geprägt sind. Um die Wirkungsaussagen abzustützen, wird auf unterschiedliche Arten der Plausibilisierung zurückgegriffen.

Die empirische Analyse hat gezeigt, wie unterschiedlich detailliert die Interventionssysteme in den untersuchten Evaluationen rekonstruiert sind. Die Evaluationen haben einen grossen Spielraum und in jeder Stufe der Wirkungskonstruktion gibt es vielfältige Möglichkeiten, Komplexität zuzulassen oder zu reduzieren. Die folgende Tabelle ist entlang der vier Stufen der Wirkungskonstruktion aufgebaut, die in der empirischen Analyse herausgearbeitet wurden (erste Spalte). Spalten zwei und drei stecken den Möglichkeitsraum ab, den Evaluationen nutzen können, um die untersuchten Interventionssysteme zu rekonstruieren. Die einzelnen Aspekte wurden induktiv aus dem Material herausgearbeitet.

Stufen der Wirkungskonstruktion	Rekonstruktionen von geringer Komplexität	Rekonstruktionen von hoher Komplexität
Komplexität eingrenzen	enge Grenzen, wenige Einflussfaktoren	weite Grenzen, viele Einflussfaktoren
	starre Grenzen	flexible Grenzen
	deduktive Grenzziehung, Fokus auf vordefinierte Wirkungsziele, unkritische Übernahme der Ziele	induktive Grenzziehung, Berücksichtigung von nicht intendierten Nebenwirkungen, kritische Reflexion der Wirkungsziele
	stark generalisierte Wirkungsfaktoren	kontext- und zielgruppenspezifische Wirkungsfaktoren
Komplexität erfassen	wenige Perspektiven	Vielfalt von Perspektiven
	hypothesenprüfende Zugänge mit standardisierten Erhebungsinstrumenten	explorative Zugänge, offenes Einfangen
	nur eine oder sehr wenige Methoden	Kombination verschiedener Methoden
	Distanzierte Beobachtung	Differenziertheit durch Nähe
	wenig Raum für Erklärungen, Hintergründe, Zwischentöne	Raum für Erklärungen, Hintergründe, Zwischentöne
	Erfassen weniger Momentaufnahmen	Erfassen dynamischer Entwicklungen
	strikte durchgezogenes Evaluationsdesign	emergentes Evaluationsdesign
Komplexität verarbeiten	einfache Mechanismen und Modelle, lineares Wirkungsverständnis, wenig Analysetiefe	differenzierte Mechanismen und Modelle, Betonung von Interaktion, Wechselwirkungen und Nicht-Linearität, mehr Analysetiefe
	deduktiv genutzte Modelle	induktiv entwickelte Modelle
	Überprüfen externer Wirkungskonstruktion	auf Verstehen der Eigenlogik der Systeme ausgelegt
	allgemeine Analyse	kontextspezifische Analyse
	Vergleiche weniger Variablen und weniger Zeitpunkte	differenzierte Vergleiche, Zeitverläufe
	monoperspektivisch	Diskussion verschiedener Perspektiven
	unkritischer Methodeneinsatz	Reflexion des Potentials und der Grenzen der eingesetzten Auswertungsmethoden
Wirkungen beurteilen	Fokus auf intendierte Wirkungen	Berücksichtigung nicht-intendierter Wirkungen
	starke Verdichtung	Belassen von Vielfalt
	eindeutige Wirkungsnachweise	Diskussion der Grenzen von Attribution
	Wirkung als Ergebnis der durchgeführten Intervention verstanden	Wirkung als Co-Produktion und gemeinsamer Lern- und Entwicklungsprozess verstanden
	allgemeingültige Wirkungsaussagen und unkritische Generalisierungen	kontextspezifische Wirkungsbeurteilungen und vorsichtige Generalisierung
	keine Plausibilisierungen der Wirkungsaussagen	Nutzung verschiedener Formen der Plausibilisierung der Wirkungsaussagen
	fehlende kritische Reflexion der Wirkungszuschreibungen und des Einflusses der Evaluation	selbstkritische Reflexion der Wirkungszuschreibungen und des Einflusses der Evaluation

Tab. 19 Handlungsspielraum der Evaluation bei der Wirkungskonstruktion

Grob können die untersuchten Evaluationen in zwei Gruppen aufgeteilt werden. Auf der einen Seite die Evaluationen, welche ein explorativ-qualitatives Forschungsdesign verwenden, um Wirkungen zu beurteilen und auf der anderen Seite Evaluationen, welche ein hypothesenprüfend-quantitatives Design verwenden. Dort wo die Referenzsysteme für die Wirkungsmessung (Individuen, soziale Systeme) nur in kleinen Fallzahlen vorliegen, werden explorative Designs verwendet, wo die Referenzsysteme in grossen Fallzahlen vorliegen, primär hypothesenprüfend. In explorativ orientierten Evaluationen werden die untersuchten Systeme mit vergleichsweise hoher Komplexität rekonstruiert und der Wirkungsnachweis wird primär über die detaillierte Rekonstruktion der Wirkungszusammenhänge erbracht. In hypothesenprüfend verfahrenen Evaluationen werden die untersuchten Systeme mit vergleichsweise geringer Komplexität rekonstruiert, also ein hoher Grad an Komplexitätsreduktion verwendet, und der Wirkungsnachweis wird primär über kontrafaktische Vergleiche (unterschiedlicher Art) erbracht. In den hypothesenprüfend ausgerichteten Evaluationen werden die quantitativen Wirkungsmessungen zum grossen Teil mit explorativen Elementen ergänzt. Die Erkenntnisse dieser Verfahren (Interviews oder Beobachtungen) fliessen nicht direkt in die Wirkungsmessungen ein, sie dienen aber zur Plausibilisierung und Interpretation der Ergebnisse der Messungen. Wirkungsbeurteilungen von Befragten fliessen teilweise sowohl in explorativ sowie auch in hypothesenprüfend ausgerichtete Evaluationen ein:

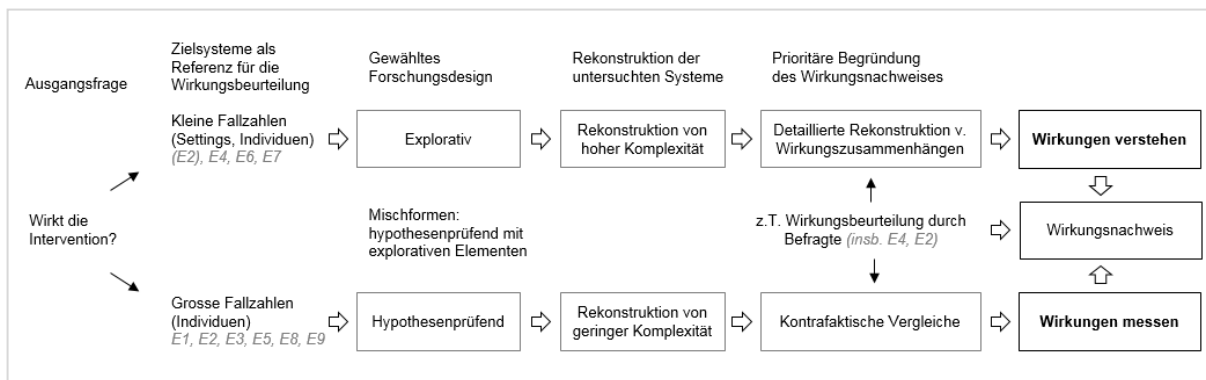


Abb. 26 Vorgehensweisen zur Wirkungsbeurteilung

Vor dem Hintergrund der Theorie komplexer Systeme und den Ergebnissen der empirischen Analyse scheint es auf der einen Seite wünschbar, Interventionssysteme differenziert zu untersuchen, um Wirkungen verstehen und beurteilen zu können. Auf der anderen Seite wäre es vermessen, soziale Systeme in ihrer gesamten Komplexität abbilden und beurteilen zu wollen. Jede Evaluation, auch wenn sie noch so umfassend ist, ist eine starke Komplexitätsreduktion. Ob Komplexität reduziert werden soll, ist vor diesem Hintergrund keine sinnvolle Fragestellung. Vielmehr interessiert die Frage, wie Evaluationen angemessen mit Komplexität umgehen können, wo sich welche Formen der Komplexitätsreduktion anbieten und was Evaluationen mit Rekonstruktionen geringer – oder eben hoher - Komplexität gewinnen und verlieren können. Die in der empirischen Analyse herausgearbeiteten Erkenntnisse werden im nächsten Kapitel vor dem Hintergrund der Merkmale komplexer sozialer Systeme, der Evaluationsliteratur und der Prinzipien der Gesundheitsförderung reflektiert, um daraus Ansatzpunkte für eine Weiterentwicklung der Evaluationspraxis in der Gesundheitsförderung und Prävention ableiten zu können.

8 Umgang der Evaluationen mit Komplexität

Komplexität ist ein Wesensmerkmal sozialer Systeme, und Interventionen der Gesundheitsförderung und Prävention sind von besonderer Komplexität geprägt, wie in Kapitel 4.3 gezeigt wurde. Dieses Kapitel widmet sich der Forschungsfrage und analysiert mit Rückgriff auf die im Theorieteil herausgearbeiteten Merkmale komplexer sozialer Systeme (Kap. 4.2), wie Evaluation in der Gesundheitsförderung und Prävention bei der Beurteilung von Wirkungen mit Komplexität umgeht.

Dazu wird in einem ersten Schritt jede der neun in dieser Studie berücksichtigten Evaluationen im Hinblick auf ihren Umgang mit Komplexität reflektiert. In den folgenden Unterkapiteln (8.2-8.6) werden die Erkenntnisse aus den bisherigen Analysen zusammengefasst, vor dem Hintergrund des Theorieteils und aktueller Diskussionen und Ansätze der Evaluationsliteratur reflektiert¹²⁰ und zu einer Synthese verdichtet.

8.1 Evaluationsspezifische Analyse

In diesem ersten Unterkapitel wird beleuchtet, wie die einzelnen Evaluationen mit Komplexität umgehen. Für jede Evaluation wird ein Überblick über die Ziele, die Massnahmen, das Evaluationsdesign und die in der jeweiligen Evaluation berücksichtigten Faktoren gegeben. Das Raster zur Darstellung der Faktoren basiert auf einer noch unveröffentlichten systemtheoretischen Modelllogik von Martin Hafen der Hochschule Luzern. In einem zweiten Schritt werden für jede Evaluation zentrale Erkenntnisse zum Umgang mit Komplexität aufgelistet. Grundlage für die Analyse bildet die deduktiv vorgenommene Kodierung der Evaluationsberichte entlang der zehn Merkmale komplexer sozialer Systeme. Bei dieser Analyse wird bewusst auf Wirkungsbeurteilungen fokussiert. Indikatoren und Aspekte aus prozessevaluativen Teilen werden nur dann berücksichtigt, wenn sie zur Wirkungsbeurteilung genutzt werden. Mittels Seitenzahlen in Klammern wird auf die entsprechenden Stellen in den Evaluationen verwiesen.

¹²⁰ Die Herausforderungen, die sich der Evaluation komplexer sozialer Systeme stellen, werden erst seit wenigen Jahren in der Evaluationsliteratur explizit und ausführlich thematisiert (Patton 2011; Robert Koch Institut (2012); Forss, Marra, and Schwartz (2011); Williams and Hummelbrunner (2009); DeGEval (2013)).

E1: Kurzinterventionen bei risikoreich Alkohol konsumierenden Berufs- und Mittelschülerinnen und -schülern im Kanton Zürich (Gmel et al., 2010)

Interventionsziele

Mit den Kurzinterventionen sollen die folgenden Ziele erreicht werden (5ff):

- Reduktion der Trinkgelegenheiten mit ‚Bingetrinken‘ im letzten Monat um durchschnittlich 30 % bei der Zielgruppe sechs Monate nach der Intervention.
- Erhöhung der Motivationsstufe zur Verhaltensveränderung um durchschnittlich 0.5 Stufen bei der Zielgruppe sechs Monate nach der Intervention.
- Erhöhung der telefonischen und persönlichen Kontakte zu Hilfsangeboten bei 5% der Zielgruppe in den sechs Monaten nach der Intervention.
- Höchstens minimale negative Etikettierung durch Schulleitung, Lehrpersonen und Mitlernende im Zusammenhang mit der Teilnahme an den Gruppen R/H in den sechs Monaten nach der Intervention

Überblick über die untersuchten Systeme und Faktoren

Intervention	Systeme im Fokus der Intervention	Einflussfaktoren	Prioritäre Ergebnisindikatoren	Folgen
<u>Gruppen-Kurzinterventionen</u> nach folgendem Ablauf: - Setting klären - Trinkverhalten anschauen - Ambivalenz fördern - Diskrepanzen suchen - Ratschlag geben - Abschluss	Soziale Systeme - Schulklassen (Interventionsort, keine Systemveränderung vorgesehen)	- <u>Kontaktaufnahme zu Hilfsangeboten</u> - Beteiligung der Jugendlichen an der Intervention - Gruppendynamik (zur Interpretation der Ergebnisse hinzugezogen) - <u>Negative Etikettierung (Folge d. Intervention)</u>		- Zank oder Streit, Balgerei oder Kampf, Probleme mit den Eltern, Freunden oder Lehrpersonen, schlechte Leistung in der Schule, Probleme mit der Polizei, bereuter Sex
	Psychische Systeme - Jugendliche	- <u>Motivation (Motivationsstufen im transtheoretischen Modell)</u> - Wichtigkeit der Veränderung - Zuversicht hinsichtlich einer Veränderung - Offenheit im Interventionsprozess		
	Physische Systeme - Jugendliche	-	- <u>Trinkverhalten (Konsumhäufigkeit und -menge)</u>	- Unfälle, Verletzungen, Einweisung ins Krankenhaus... - Verlust von Geld/Wertgegenständen, Beschä-

				digung des Eigentums, Opfer von Diebstahl oder Raub, Sex ohne Kondom
--	--	--	--	--

Evaluationsdesign: Als randomisierte kontrollierte Studie konzipiert, wobei die Randomisierung auf Ebene der Schulklassen erfolgte (Clusterrandomisierung). Quantitative Befragung der Interventions- und Kontrollgruppen mit Vor- und Nachbefragung. Ergänzende Prozessevaluation (Interviews mit beteiligten Lehr- und Präventionsfachpersonen), in welcher wirkungsförderliche und wirkungshemmende Aspekte angesprochen werden (z.B. Offenheit und engagierte Beteiligung der Jugendlichen). Diese fließen nicht in die Wirkungsmessung ein, werden aber punktuell zur Interpretation und Plausibilisierung der Messergebnisse genutzt.

Umgang mit Komplexität bei der Wirkungsbeurteilung

Interaktionsvielfalt *Konnektivität und Interdependenz, Offenheit*

- Es wird auf sehr wenige individuelle Indikatoren (18) und Interaktionen fokussiert (32f), die sich aus den Interventionszielen ableiten.
- Umweltfaktoren fließen kaum in die Betrachtungen und gar nicht in die Wirkungsmessung ein. Beobachtet wird vor allem ein Faktor, welcher die Motivation und das Trinkverhalten beeinflusst, nämlich die Gruppenkurzintervention.
- Die Selektion (und damit die Grenzziehung) wird einerseits durch das transtheoretische Modell begründet (7ff), andererseits durch das gewählte Studiendesign, in welchem Umweltfaktoren ausgeblendet werden, um die ‚pure‘ Wirkung extrahieren zu können.

Komplexe Wirkungsmechanismen *Feedback, Multidimensionalität/Nicht-Linearität/Zirkularität, Dynamik*

- Ursache-Wirkungszusammenhänge werden für die Wirkungsbeurteilung auf die Wirkungskette Gruppenkurzintervention → Motivationsstufen → Trinkverhalten → negative Konsequenzen des Alkoholkonsums reduziert (32f).
- Es werden keine Feedbacks und Rückkopplungen zwischen verschiedenen Systemen oder Systemebenen untersucht.
- Gruppendynamiken werden im Rahmen methodischer Fragen und möglicher Nebenwirkungen angesprochen (47), aber nicht untersucht. Weitere Dynamiken kommen nicht zur Sprache.
- Die Entwicklungen entlang der Motivationsstufen des transtheoretischen Modells werden als spiralförmig beschrieben (8), die Wirkungsmessung berücksichtigt aber nur Übergänge von der einen in die andere Stufe zwischen zwei Messzeitpunkten (34ff). Dynamische, spiralförmige Prozesse können so nicht beurteilt werden.

Selbstorganisation *Adaption und Co-Evolution, Reflexivität, Selbstorganisation/operative Geschlossenheit*

- Selbstorganisationsprozesse werden als Grundlage für die Intervention genannt (Eigenständigkeit, selbstgesteuerte Verhaltensänderung auf Ebene Individuum, Gruppenprozesse auf Ebene Interventionsgruppen) (1), aber in der Evaluation nicht untersucht.
- Die Gruppendynamik wird als möglicher Erklärungsgrund für erstaunliche Messergebnisse angeführt, aber nicht weiter beleuchtet (47).

Unvorhersehbarkeit *Emergenz, Unvorhersehbarkeit/Ungewissheit*

- Systementwicklungen und emergente Systemeigenschaften werden nicht untersucht, aber punktuell angeführt, z.B. wenn auf Gruppendynamiken verwiesen wird.
- Nebenwirkungen der Intervention auf die Jugendlichen werden in der Nachbefragung offen abgefragt (36ff).
- Im Rahmen der Prozessevaluation werden einzelne unvorhersehbare Entwicklungen thematisiert, welche die Durchführung erschwert haben (17, 42).
- Zum Teil mussten (oder sollten) kleinere Abweichungen vom Konzept der Intervention oder der Evaluation gemacht werden, wenn es die Situation erforderte (46, 49).

Wirkungsbeurteilung

- Die Wirkungsbeurteilung basiert auf dem kontrafaktischen Vergleich zwischen Interventions- und Kontrollgruppe und fokussiert auf das Trinkverhalten, den Wechsel zwischen den Motivationsstufen und die Kontakte zu Hilfsangeboten (12f).
- Es wird Wert darauf gelegt, dass die Studie unter ‚realen Bedingungen‘ durchgeführt wird (2ff), um eine höhere externe Validität (auf Kosten der internen Validität) zu erreichen.
- Bei den Angaben zu den Wirkungsindikatoren (Trinkverhalten, Inanspruchnahme von Hilfsangeboten, Konsequenzen des Alkoholkonsums etc.) handelt es sich um Selbsteinschätzungen der Jugendlichen (18). Die Jugendlichen werden gefragt, ob die genannten potenziellen Folgeprobleme kausal dem Alkoholkonsum zuzurechnen sind (Wirkungsbeurteilung durch Befragte) (18).
- Die Ergebnisse der Wirksamkeitsevaluation werden mit Hilfe der Literatur und mit einzelnen Bezügen zu den Ergebnissen der Prozessevaluation plausibilisiert (48-51).
- Die Ergebnisse werden insofern generalisiert, als die untersuchten Interventionen (für Risikokonsumierende) unter Berücksichtigung von Optimierungen empfohlen werden (51).

E2: Projekt SWiNG (Jenny et al., 2011)

Interventionsziele

Mit dem Projekt SWiNG sind die folgenden Ziele verbunden (18):

- Die Arbeitsprozesse sind gesünder und effizienter gestaltet.
- Die Gesundheit, das Wohlbefinden und die Arbeitszufriedenheit der einzelnen Mitarbeitenden sind gesteigert.
- Die Leistungsfähigkeit, die Leistungsbereitschaft, die Motivation und die individuellen Coping-Fähigkeiten der Mitarbeitenden sind gesteigert.
- Die Unternehmenskultur ist verbessert.
- Das Image des Betriebs (intern und extern) ist verbessert und leistet einen Beitrag zu einer verbesserten Wettbewerbsfähigkeit und zu einem Standortvorteil.

Überblick über die untersuchten Systeme und Faktoren

Intervention	Systeme im Fokus der Intervention	Einflussfaktoren (Auszug)	Prioritäre Ergebnisindikatoren	Folgen
<p><u>Umfassende Stressprävention</u> in Betrieben, als betriebsspezifische Entwicklungsprozesse konzipiert:</p> <ul style="list-style-type: none"> - Einstiegsworkshop mit der Geschäftsleitung - Umfassende Analysephase (quantitative und qualitative Erhebungen) - Führungsschulungen - Gesundheitszirkel - Stressmanagement-Kurse - betriebsspezifische Massnahmen 	<p>Soziale Systeme</p> <ul style="list-style-type: none"> - Betrieb 	<p>Organisationale Struktur</p> <ul style="list-style-type: none"> - Vorhandene Gefässe und Kompetenzen für Kurse und Projekt-/Steuergruppen - Ressourcenbereitstellung für Kurse, Massnahmen und Kommunikation - Handlungsspielraum in der Massnahmenumsetzung und dem Ressourceneinsatz - Stellenprozente für Projekt-/Themenverantwortliche - Vernetzung und Verankerung der Projekt-/Steuergruppen - Vernetzung wichtiger Stellen und Personen - Kontinuität der verantwortlichen Personen - Wahrgenommene Kommunikationskanäle, Face-to-Face-Kommunikation - Stabilität der organisationalen Strukturen <p>Organisationale Strategie</p>	<ul style="list-style-type: none"> - Sensibilisierung - strukturelle Veränderungen - Führungskultur - Austausch - etc. 	<p>Ökonomischer Nutzen</p>

		<ul style="list-style-type: none"> - Commitment und Durchsetzung durch das Management (Veränderungsbereitschaft) - Vorbildcharakter des Managements, Verbindlichkeit für die Führung - Transparenz der Projektziele (strategische und ethische Ziele) - Kongruenz mit den Unternehmenszielen - Gemeinsames Bild und klarer Fokus des Projekts (Umfang und Zielgruppe der Veränderung) - Verortung und Integration des Themas (Personalabteilung, Arbeitssicherheit, u.a.) - Lösungsorientierung und Belohnung (vs. Bestrafung und Sanktion) <p>Organisationale Kultur</p> <ul style="list-style-type: none"> - Vertrauen und Fairness - Gemeinschaftlichkeit (vs. autoritärer Umgang) - Kultur des Austausches und Zusammenarbeit zwischen/in den Teams - Umgang mit Diversität (Vielfalt als Chance und Voraussetzung für das Funktionieren) - Bewusstsein für und Stellenwert von psychosozialen Faktoren - Gemeinsame Sprache zum Thema <p>Organisationale Umwelt</p> <ul style="list-style-type: none"> - Wirtschaftliche Sicherheit (Auftragslage, Existenzsicherung, Holding-Strukturen) - Politische Sicherheit (Parteilinien, Gesetze und Regelungen) <p><u>Arbeitsbelastungen</u></p> <ul style="list-style-type: none"> - Unterbrechungen - Zeitdruck - Unsicherheit - Überforderung <p><u>Arbeitsressourcen</u></p>		
--	--	---	--	--

		<ul style="list-style-type: none"> - Unterstützendes Vorgesetztenverhalten - Interaktionale Fairness - Soziale Unterstützung Vorgesetzte - Wertschätzung Vorgesetzte - Soziale Unterstützung Kollegen - Ganzheitlichkeit - Handlungsspielraum 		
	Psychische Systeme <ul style="list-style-type: none"> - Mitarbeitende - Führungspersonen 	Individuelle Kompetenz <ul style="list-style-type: none"> - Führungskompetenzen (Personenführung, Fachführung) - Sprachliche Kompetenzen (Befragung, Auseinandersetzung mit dem Thema) Individuelle Motivation <ul style="list-style-type: none"> - Teilnahmebereitschaft an Massnahmen - Bestehendes gesundheitsorientiertes Verhalten - Wahrgenommene Vereinbarkeit/Passung zwischen Projekt und Betrieb Individuelle Identität <ul style="list-style-type: none"> - Stellenwert von Gesundheit und Haltung gegenüber dem Stressbegriff - Führungsbilder, Menschenbilder - Stellenwert der Arbeit („Nine-to-Five“) Individuelle Umwelt <ul style="list-style-type: none"> - Balance zwischen Arbeit und Freizeit - Arbeitsbezogenes Kohärenzgefühl 	<u>Verhältnis zwischen Ressourcen und Belastungen</u> (Index aus einer Vielzahl von organisationalen und individuellen Faktoren, vgl. Strukturgleichungsmodell E2, 32)	<u>Positive Gesundheit</u> <ul style="list-style-type: none"> - Arbeitszufriedenheit - Unternehmensbindung - Enthusiasmus
	Physische Systeme <ul style="list-style-type: none"> - Mitarbeitende 			<u>Negative Gesundheit</u> <ul style="list-style-type: none"> - Schlafprobleme - Erschöpfung - psychosomatische Erkrankungen
	Relationen zwischen den Systemen	Passung zwischen der Intervention, den Organisationen und den Individuen		

Evaluationsdesign: Das Evaluationsdesign ist vielfältig und kombiniert hypothesenprüfende Verfahren mit explorativen Untersuchungen etwa zu gleichen Teilen. Es werden 14 unterschiedliche Erhebungsmethoden aufgeführt. Der hypothesenprüfende Teil besteht aus einer Wirksamkeitsanalyse, die auf einer umfassenden standardisierten Befragung der Mitarbeitenden zu drei Erhebungszeitpunkten gründet. Der explorative Teil stützt sich auf Daten aus Leitfaden-Interviews und Gesprächen auf unterschiedlichen Ebenen (Anbieter, Management, Gesundheitszirkel,...) und aus systematische Beobachtungen. Die Ergebnisse aus den explorativen Elementen fließen prominent in die Wirkungsbeurteilungen ein, wenn auch nicht in die Wirksamkeitsmessung selber.

Umgang mit Komplexität bei der Wirkungsbeurteilung

Interaktionsvielfalt *Konnektivität und Interdependenz, Offenheit*

- Die untersuchten Interventions- und Zielsysteme werden als komplexe soziale Systeme konzipiert und entsprechend vielfältig sind die untersuchten Ebenen und Faktoren (14, 32).
- Die in der Wirksamkeitsmessung berücksichtigten Faktoren werden aus der Literatur abgeleitet (23, 26, 38f). Viele weitere Faktoren werden im Rahmen der explorativen Elemente identifiziert (14).
- Die Beziehungen zwischen Intervention, organisationaler und individueller Ebene werden sowohl im hypothesenprüfenden (32f) als auch im explorativen Teil (38f) intensiv untersucht.
- Als Ergebnisse der Intervention werden sowohl Veränderungen auf individueller Ebene wie auch auf Ebene sozialer Systeme (Betriebe) untersucht (15), die Intervention verbindet verhaltens- und verhältnisorientierte Ansätze.
- Die Analyse von System-Umwelt-Bezügen ist ein tragendes Element der Evaluation. Es wird eine Vielzahl von organisationalen und individuellen Umweltfaktoren aus der Literatur als relevant identifiziert und untersucht (38f). Weitere Umweltfaktoren emergieren aus den explorativen Erhebungen (14).
- Es werden unterschiedliche Strategien angewandt, um die Vielfalt gegen Ende wieder zu verdichten und zu reduzieren (32, 36, 37).
- Trotz der Vielfalt der Faktoren, die berücksichtigt wurde, und trotz der differenzierten Wirkungsbeurteilung spitzt sich die Evaluation auf klare Wirkungsaussagen zu: „Mit der Unterstützung von SWiNG konnten 25% der Belegschaft ihre Ressourcen und Belastungen optimieren und dadurch ihre Gesundheit erhalten oder steigern.“ und „Die Mitarbeitenden, die von SWiNG profitieren sind produktiver, gesünder und weisen weniger Absenzen auf.“ (7)

Komplexe Wirkungsmechanismen *Feedback, Multidimensionalität/Nicht-Linearität/Zirkularität, Dynamik*

- Sowohl die Wirksamkeitsanalyse als auch die explorativen Analysen zeichnen sich durch eine hohe Komplexität aus.
- Insbesondere in den explorativen Wirkungsbeurteilungen wird ein grosses und dichtes Netz an Einflussfaktoren aufgespannt (38f), die vielfältig miteinander verbunden werden.
- Dynamiken (53) und Feedbackprozesse (39, 43, 51, 53) und Rückkoppelungen (10, 26) sind ein wichtiger Bestandteil der explorativen Analyse und fließen in die Wirkungsbeurteilung mit ein.
- Prozessfaktoren werden als wichtig für die Wirkungsbeurteilung erachtet (9) und umfassend in die Analyse integriert (38f).
- Es wird ein überschaubares Rahmen-Wirkungsmodell verwendet, welches die Vielfalt an in der Evaluation berücksichtigter Einflussfaktoren stark reduziert (27).
- Für die Wirksamkeitsmessung wird auf einen Teil der Gesamteinflussfaktoren fokussiert (32) und diese werden zu wenigen Indizes verdichtet (Belastungen und Ressourcen,

positive und negative Gesundheit).

Selbstorganisation *Adaption und Co-Evolution, Reflexivität, Selbstorganisation/operative Geschlossenheit*

- Die Evaluation basiert auf einem systemtheoretischen Verständnis und verwendet systemische Organisationsmodelle, die u.a. psychologische, soziologische und ökonomische Aspekte verbinden (10).
- Die Betriebe werden als selbstreferenzielle Systeme konzipiert und untersucht (23, 41, 85). Die Evaluation versucht, die Betriebe in ihrer Eigenlogik zu verstehen (25). Wirkungsentfaltung wird als Eigenleistungen der Zielsysteme verstanden (17, 41, 85), die sich in kleinräumigen Schritten und in ständigem Austausch mit umgebenden Systemen vollziehen. Innerbetriebliche Reflexionen werden dabei als zentrales Element beleuchtet (17, 50, 53, 85ff).
- Die untersuchten Systeme werden als adaptive, co-evolutive Systeme verstanden (47), die Passung zwischen der Intervention und dem Zielsystem wird als zentrales Element herausgearbeitet (13).

Unvorhersehbarkeit *Emergenz, Unvorhersehbarkeit/Ungewissheit*

- Standardisiert sind einige zentrale Aspekte der Intervention (Analysephase, Führungsschulungen...), die konkreten Entwicklungsprozesse in den Betrieben und die dabei entstehenden Massnahmen sind unvorhersehbar (49).
- Die Dynamiken in den Betrieben (86) wie auch die Entwicklungen im Umfeld (116) werden als unvorhersehbar verstanden. Viele relevante Faktoren und wichtige Wechselwirkungen tauchen erst im Laufe der (explorativen) Untersuchungen auf (13, 53).
- In den explorativen Erhebungen werden unvorhergesehene Nebenwirkungen sichtbar (39)
- Evaluationsschritte werden teilweise den Entwicklungen in den Betrieben angepasst (28), dies aber nur in den explorativen Elementen.

Wirkungsbeurteilung

- Die Wirkungsbeurteilung besteht aus zwei Teilen: den Ergebnissen der Wirksamkeitsmessungen und den Ergebnissen der explorativen Analysen (83f).
- Die Wirksamkeitsmessungen werden unabhängig von den explorativen Elementen konzipiert und durchgeführt. Die Messergebnisse werden als eigenständig dargestellt (51-64) und ‚für alle grösseren Betriebe‘ generalisiert (6). Der Unterschiedlichkeit der Interventionen in den einzelnen Betrieben wird im Rahmen der Wirksamkeitsanalyse entsprechend nicht Rechnung getragen (im Gegensatz zur Wirkungsbeurteilung auf Basis der explorativen Elemente).
- Für die Wirksamkeitsanalyse wird eine Vielzahl von Einflussfaktoren zu einem Ressourcen-Belastungsindex verdichtet (32f) und mit der subjektiven Wirkungseinschätzung der Mitarbeitenden (Skala Impact-Assessment, 36f) in Beziehung gebracht (Wirkungsbeurteilung durch Befragte).
- Der Wirksamkeitsnachweis basiert auf einem kontrafaktischen Vergleich. Dieser erfolgt nicht zwischen Gruppen mit oder ohne Intervention, sondern zwischen Gruppen mit einer hohen versus einer tiefen subjektiven Wirkungseinschätzung (58).
- Die explorativ hervorgebrachten Ergebnisse bilden den zweiten, sehr differenzierten Teil der Wirkungsbeurteilung. Dieser Teil ist stark geprägt von der Komplexität sozialer Systeme und beruht auf einer differenzierten Rekonstruktion der Wirkungszusammenhänge (38f).
- Neben dieser differenzierten Darstellung werden die Ergebnisse der quantitativen Analyse auf wenige pointierte und klare Wirkungsaussagen verdichtet (7).
- Die Ergebnisse der Wirksamkeitsanalyse und des explorativen Teils werden in der Gesamtwirkungsbeurteilung zusammengeführt diskutiert (83f). Sie werden auf unterschiedliche Weise plausibilisiert, vor allem im Rahmen der Methodentriangulation innerhalb der Evaluation (57, 61, 67), mit Bezug zu anderen Studien (33, 68, 72),

E3: Evaluation der bfu-Präventionskampagne im Schneesport (Furrer & Balthasar, 2011)

Interventionsziele

Die Präventionskampagne verfolgt die nachstehenden Ziele (33):

- Erhöhung des Gefahrenbewusstseins: Maximal 10 % der Schneesportler/-innen geben an, sich noch nie mit dem Gedanken Helm befasst zu haben.
- Erhöhung der wahrgenommenen Vulnerabilität: Maximal 20 % der Schneesportler/-innen geben an, keinen Helm zu benötigen, weil sie aufgrund ihres Fahrstils oder ihrer Exposition keinem Verletzungsrisiko ausgesetzt sind.
- Senkung der fehlenden Bequemlichkeit: Maximal 3 % der Schneesportler/-innen geben an, dass ein Helm unbequem ist oder ihr Freiheitsgefühl dadurch eingeschränkt wird.
- Erhöhung der Helmtragquote.
- Übergeordneter Outcome: Reduktion von Kopfverletzungen (wird in der Evaluation nicht beurteilt).

Überblick über die untersuchten Systeme und Faktoren

Intervention	Systeme im Fokus der Intervention	Einflussfaktoren	Prioritäre Ergebnisindikatoren	Folgen
Präventionskampagne mit einem Mix von Kommunikationsmassnahmen: - TV-, Kino- und Radiospots - Inserate - Plakate - Publireportagen - Banner auf Internetportalen - Helmtesttage - etc.	Soziale Systeme - Intervention - Schneesportgebiete	- Wirkungspotenzial - Umsetzungsqualität - Zusammenarbeit mit den Partnern - Medienpräsenz - Umweltfaktor Markt: Verfügbarkeit und Preis von Helmen - Sprachregion		
	Psychische Systeme - Schneesportler/-innen	- <u>Bekanntheit</u> und Akzeptanz <u>der Kampagne</u> bei den Zielgruppen - Gründe für das Tragen resp. Nicht-Tragen eines Helms - Gefahrenbewusstsein - Wahrgenommene Vulnerabilität - Bequemlichkeit	-	-
	Physische Systeme - Schneesportler/-innen	- Art des ausgeführten Skisports - Fahrniveau - Geschlecht - Alter	- <u>Helmtragquote</u>	- Kopfverletzungen (nicht gemessener Indikator zum Outcome „Reduktion von Kopfverletzungen“)

Evaluationsdesign: In einer Konzeptevaluation wurde das Kampagnenkonzept auf Basis der wissenschaftlichen Literatur überprüft (Dokumentenanalysen). Eine Prozessevaluation befasste sich mit der Qualität der Umsetzung, der Resonanz in den Medien und der Zielgruppenerreichung (Interviews, Medienanalysen). In der Impact-Evaluation wurde die Einstellung der Schneesportler/-innen mittels Befragungen in Schneesportgebieten gemessen und in der Outcome-Evaluation wurde die Helmtragquote in Schneesportgebieten mittels Beobachtungen vor, während und nach der Kampagne zu mehreren Messzeitpunkten ermittelt. In Gruppengesprächen wurden zusätzlich Wirkungszusammenhänge, die das Verhalten von Jugendlichen und den Einfluss der Kampagne bestimmen, etwas detaillierter aufgeschlüsselt.

Umgang mit Komplexität bei der Wirkungsbeurteilung

Interaktionsvielfalt *Konnektivität und Interdependenz, Offenheit*

- In der Evaluation wird eine begrenzte Anzahl von Einflussfaktoren auf das Helmtragen untersucht: einerseits die drei vermuteten Gründe für das Tragen resp. Nicht-Tragen eines Helms, andererseits soziodemographische Faktoren, Sportart und Fahrniveau sowie die Bekanntheit der Kampagne (57-67).
- Für die Messung der Wirksamkeit der Kampagne wird allerdings nur auf den Faktor „Bekanntheit der Kampagne“ abgestützt (14, 65ff).
- Als Umweltfaktoren werden Verfügbarkeit und Preise von Helmen untersucht, allerdings ohne einen direkten Zusammenhang zur Wirksamkeit herzustellen (50).
- Die Ergebnismessung fokussiert auf individuelle Faktoren. Veränderungen auf Systemebene (z.B. der erwähnte Kohorteneffekt) werden bei der Interpretation der Ergebnisse angesprochen (67), aber nicht untersucht.

Komplexe Wirkungsmechanismen *Feedback, Multidimensionalität/Nicht-Linearität/Zirkularität, Dynamik*

- Es wird ein einfaches, lineares Wirkungsmodell verwendet, welches die Wirkungskette von der Konzeption der Kampagne über die Umsetzung, Wahrnehmung und Einstellung der Zielgruppen bis zur Verhaltensänderung abdeckt (38).
- Auf dynamische Aspekte wird in der Ergebnisinterpretation punktuell und hypothetisch hingewiesen (15). Dynamische Prozesse oder Wechselwirkungen zwischen verschiedenen Faktoren oder Wirkungsebenen werden hinsichtlich einer Wirkungsbeurteilung nicht untersucht.

Selbstorganisation *Adaption und Co-Evolution, Reflexivität, Selbstorganisation/operative Geschlossenheit*

- Selbstorganisations- oder adaptive Lern- und Entwicklungsprozesse werden nicht thematisiert oder untersucht.
- Der Untersuchungsgegenstand wird nicht systemisch beobachtet.

Unvorhersehbarkeit *Emergenz, Unvorhersehbarkeit/Ungewissheit*

- Ziele und Zielgruppen mussten zum Teil den laufenden Entwicklungen angepasst werden (33, 42, 46, 48).
- Die Evaluation hat stark auf die Messung der Zielerreichung fokussiert, unvorhersehbare oder unerwünschte Entwicklungen wurden lediglich im Rahmen der Prozessevaluation thematisiert (12, 49ff).

- Es werden keine positiven oder negativen Nebenwirkungen der Kampagne untersucht. In den Gruppengesprächen tauchen einzelne neue Erkenntnisse bei der Beurteilung der Kampagnenqualität auf (52f).

Wirkungsbeurteilung

- Die Wirksamkeitsmessung der Kampagne stützt sich primär auf die Untersuchung des Zusammenhangs zwischen dem Kennen der Kampagne und dem Helmtragen (64). Da keine Kontrollgruppen verwendet worden sind und weitere Aspekte gegen eine Wirksamkeit der Kampagne in gemessenem Umfang hindeuten (67), ist die Aussagekraft allerdings begrenzt. Es tragen zwar vergleichsweise mehr Personen, die die Kampagne kennen, einen Helm, aber ein Wirksamkeitsnachweis wird davon nicht abgeleitet (66). Das Ergebnis der Wirksamkeitsmessung wird entsprechend vorsichtig beurteilt (67). Die Interpretation der Ergebnisse, welche die Evaluation abschliesst, ist sehr hypothetisch.
- Die Ergebnisse der Konzept- und der Prozessevaluation (inkl. der gestiegenen Bekanntheit und der hohen Akzeptanz der Kampagne) sprechen für das Wirkungspotenzial der Kampagne (55ff). Andere Aspekte wie die zu frühe Zielerreichung (46), Erkenntnisse aus der Literatur (67) oder der Zusatzaufwand in der Romandie (59), der nicht zu messbaren Erfolgen führte, sprechen eher gegen einen Kausalzusammenhang zwischen der Kampagne und den gemessenen Veränderungen.
- Die Ergebnisse der Evaluation werden nicht generalisiert und es gibt keine entsprechenden Hinweise oder Empfehlungen zur Multiplikation der Kampagne.

E4: Evaluationsbericht zum Projekt «MuM – mutige Mädchen – Integration von gewaltbereiten Mädchen im unteren Kleinbasel» (Gloor & Meier, 2004)

Interventionsziele

Das Projekt MuM verfolgt die folgenden Interventionsziele (10):

- Die Fachfrau baut den Kontakt zu den Mädchen auf, mit der Zeit entsteht zwischen den Mädchen und der Fachfrau eine tragfähige Beziehung.
- Die Fachfrau verbringt mit den Mädchen gemeinsame Zeit, die für Gespräche und soziokulturelle Aktivitäten genutzt wird.
- Die Fachfrau begleitet und stärkt die Mädchen in ihrer Entwicklung.
- Die Mädchen lernen, sich gesund zu verteidigen und sich im öffentlichen Raum sicher zu fühlen.
- Die Mädchen lernen, mit schwierigen Situationen konstruktiv umzugehen.
- Die Mädchen lernen, Fremde zu tolerieren und zu akzeptieren.
- Die Mädchen bauen ihre Gewaltbereitschaft ab.

Überblick über die untersuchten Systeme und Faktoren

Intervention	Systeme im Fokus der Intervention	Einflussfaktoren (induktiv auftauchend)	Prioritäre Ergebnisindikatoren	Folgen
<ul style="list-style-type: none"> - Aufsuchende Jugendarbeit mit gewaltbereiten Mädchen, mit Gruppen- und Einzelkontakten - Durchführen gemeinsamer Anlässe und Ausflüge - Regeln aufstellen - Mediationen - etc. 	<p>Soziale Systeme</p> <ul style="list-style-type: none"> - Mädchengruppen - Institutionen, die mit den Mädchen arbeiten - Fach- und breite Öffentlichkeit 	<ul style="list-style-type: none"> - verbale Gewalt - Regeln - Veränderungen in der Kommunikation - Vorbildrollen - Fachliche Auseinandersetzung - Entlastung - Veränderung von Aufgaben - erweiterter Blick (Aussenblick) - zusätzliche Ressourcen - Zugänglichkeit zu den Mädchen - Sensibilisierung - Einsichten in die Problematik - Verständnis - Fachwissen 		

<p>Psychische Systeme - Mädchen</p>	<ul style="list-style-type: none"> - Beruhigung der Situation - Akzeptanz des Projekts - wahrgenommene Notwendigkeit zur Gewaltausübung - Reflexionserfahrung - Einsicht in Problemverhalten - Bereitschaft, Hilfe anzunehmen - Personale Ressourcen (z.B. Coping-Fähigkeiten) - Sozialkompetenz 		
<p>Physische Systeme - Mädchen</p>		<p><u>körperliche Gewalt</u> (Selbstdeklaration)</p>	

Evaluationsdesign: Die Evaluation ist explorativ angelegt, enthält ergänzend aber quantitative Elemente wie wenn auch einzelne quantitative Elemente integriert sind (soziodemographische Angaben zu den Teilnehmenden, Teilnahmeprotokolle). Die Wirkungsevaluation stützt sich auf Leitfadeninterviews mit den teilnehmenden Mädchen, mit den Projektdurchführenden sowie mit Schlüsselpersonen aus dem Umfeld (14). Um die Lebenswelt der teilnehmenden Mädchen kennen zu lernen, wurden Foto-Interviews durchgeführt (17ff). Diese werden aber nicht mit der Wirkungsevaluation in Verbindung gebracht und deshalb hier auch ausgeblendet.

Umgang mit Komplexität bei der Wirkungsbeurteilung

Interaktionsvielfalt *Konnektivität und Interdependenz, Offenheit*

- In den explorativen Erhebungen werden viele Wirkungsfaktoren zusammengetragen (61-71). Da sie aber nur vereinzelt auftreten und nicht systematisch untersucht werden, ist es schwierig, deren Bedeutung einschätzen zu können.
- Durch die stark induktive Herangehensweise der Intervention erscheinen die untersuchten Faktoren als Folgen resp. Wirkungen der Intervention und nicht primär als Einflussfaktoren auf die Ergebnisindikatoren. In den Interventionszielen werden zwar Wirkungsvermutungen und Einflussfaktoren sichtbar (10), diese werden aber analytisch nicht verwendet, d.h. die Evaluationsergebnisse werden nicht auf diese Faktoren bezogen.
- Das Umfeld wird insofern thematisiert, als die Auswirkungen der Intervention auf umliegende Organisationen untersucht werden (67-69). Der Einfluss des Umfeldes auf das Projekt und auf die Wirkungen wird nicht untersucht.
- Gegen Ende der Evaluation werden die Ergebnisse nicht auf wenige klare Aussagen verdichtet, es bleibt ein differenziertes Wirkungsbild ohne klare und generalisierte Wirkungsaussage.

Komplexe Wirkungsmechanismen *Feedback, Multidimensionalität/Nicht-Linearität/Zirkularität, Dynamik*

- Die zitierten Interviewausschnitte weisen auf eine komplexe Problematik (Gewalthandeln) mit vielen Einflussfaktoren hin (61-64), die aber nicht weiter untersucht werden.
- Zusammenhänge zwischen einzelnen Wirkungsfaktoren und Wirkungsebenen werden ebenfalls nicht untersucht, die verschiedenen Faktoren stehen weitgehend unkommentiert nebeneinander. Es gibt keine Übersicht, keine Zusammenführung oder vergleichende Bewertung dieser Faktoren.

Selbstorganisation *Adaption und Co-Evolution, Reflexivität, Selbstorganisation/operative Geschlossenheit*

- Es gibt lediglich einzelne und isolierte Hinweise auf adaptive Entwicklungs- und Selbstorganisationsprozesse (62ff), die untersuchten Systeme (die Gruppen von Mädchen oder die Institutionen der Mädchenarbeit) werden aber nicht als selbstorganisierend konzipiert.

Unvorhersehbarkeit *Emergenz, Unvorhersehbarkeit/Ungewissheit*

- Da die Evaluation explorativ aufgebaut ist, ist sie nicht durch operationalisierte Ziele oder geschlossene Erhebungsinstrumente eingeschränkt. Sie nimmt offen Aspekte auf, die den Interviewten wichtig sind und stellt sie dar (61f).
- Die Projektausführenden wurden offen nach negativen Entwicklungen gefragt (64).

Wirkungsbeurteilung

- Die (kausale) Wirkungsbeurteilung durch die Evaluation stützt sich ausschliesslich auf die Wirkungseinschätzung der Befragten. Wirkungszuschreibungen werden anekdotisch angeführt und erschweren eine Gesamt-Wirkungseinschätzung. Dies auch deshalb, weil die Datenlage sehr dünn ist (z.B. nur 4 interviewte Mädchen, 16).
- Mit Verweis auf die schwierige Bestimmung von Ursache-Wirkungszusammenhängen (19) wird die Wirkungsbeurteilung vorsichtig formuliert und die Wirkungsbeurteilung bleibt unscharf (73ff).
- Die verschiedenen Perspektiven dienen dazu, die Wirkungsaussagen zu plausibilisieren (61-67).
- Aufgrund seines Wirkungspotenzials (nicht der nachgewiesenen Wirksamkeit) wird das Projekt auch anderen Gemeinden empfohlen (78).
- Es wird weitere Forschung empfohlen, um mehr Wissen über die (längerfristige) Wirksamkeit des Projektes zu gewinnen (77).

E5: Effekt der Einschränkung der Erhältlichkeit von Alkohol auf Alkohol-Intoxikationen im Kanton Genf (Gmel & Wicki, 2010)

Interventionsziele

Die Intervention bezweckt eine Reduktion des Alkoholkonsums, gemessen an den Intoxikationsraten in Spitälern (1).

Überblick über die untersuchten Systeme und Faktoren

Intervention	Systeme im Fokus der Intervention	Einflussfaktoren	Prioritäre Ergebnisindikatoren	Folgen
Einschränkung der Erhältlichkeit von Alkohol nachts, an Tankstellen und in Videotheken.	Soziale Systeme - Gemeinwesen	- <u>Erhältlichkeit von Alkohol</u>		
	Psychische Systeme		-	-
	Physische Systeme - Jugendliche und Erwachsene	- Alter - Geschlecht	- <u>Intoxikationsraten</u> (akute Intoxikation, schädlicher Gebrauch, toxische Wirkung von Alkohol, sowie Unterkategorien)	-

Evaluationsdesign: Es handelt sich um eine vergleichende quantitative Fallstudie mit einem nicht randomisierten kontrafaktischen Vergleich. Die Evaluation vergleicht die Intoxikationsraten aus den medizinischen Statistiken der Krankenhäuser des Kantons Genf mit jenen aller anderen Schweizer Kantone, die quasi als Kontrollgruppe genutzt wurden. Da die Daten zu den Intoxikationen monatlich ausgewiesen werden, können ausführliche Zeitreihenanalysen gemacht werden.

Umgang mit Komplexität bei der Wirkungsbeurteilung

Interaktionsvielfalt *Konnektivität und Interdependenz, Offenheit*

- Die Wirksamkeitsanalyse ist extrem reduktionistisch angelegt. Es wird primär die Beziehung zwischen lediglich zwei Variablen untersucht: der Einfluss von Einschränkungen

der Verfügbarkeit von Alkohol auf die Intoxikationsraten (5). Dieser Zusammenhang wird in Bezug auf Alter und Geschlecht gesondert ausgewertet.

Komplexe Wirkungsmechanismen *Feedback, Multidimensionalität/Nicht-Linearität/Zirkularität, Dynamik*

- Der Untersuchung liegt ein sehr einfaches Ursache-Wirkungsverständnis mit einer Ursache- und einer Wirkungsebene zugrunde. Feedbackprozesse, dynamische Entwicklungen oder Wechselwirkungen zwischen einzelnen Faktoren werden nicht beleuchtet.
- Das Evaluationsdesign ist so angelegt, dass komplexe Entwicklungen für den Wirksamkeitsnachweis als nicht relevant erachtet werden. Entscheidend ist die gemessene Veränderung, die im kontrafaktischen Vergleich der untersuchten Intervention zugeschrieben werden kann.

Selbstorganisation *Adaption und Co-Evolution, Reflexivität, Selbstorganisation/operative Geschlossenheit*

- Die Evaluation untersucht den Zusammenhang zwischen den beiden Hauptvariablen isoliert und verzichtet auf eine systemische Betrachtungsweise. Entsprechend stehen Adaption, Co-Evolution, Reflexivität oder Selbstorganisation von Systemen nicht im Blickfeld der Untersuchung.

Unvorhersehbarkeit *Emergenz, Unvorhersehbarkeit/Ungewissheit*

- Da sich die Evaluation auf die beschriebenen wenigen Variablen beschränkt, auf die Untersuchung von Entwicklungen verzichtet und keine explorativen Elemente enthält, sind auch Unvorhersehbarkeit und Emergenz keine Themen, die interessieren.
- Potenzielle Nebenwirkungen der Einschränkung der Erhältlichkeit von Alkohol (z.B. das Umsteigen auf andere Rauschmittel) werden nicht untersucht.

Wirkungsbeurteilung

- Entwicklungsprozesse, die zu den gemessenen Veränderungen geführt haben, interessieren in der Studie nicht. Die Wirksamkeitsmessung erfolgt unabhängig der Kenntnis der vielfältigen Einflussfaktoren auf Trinkverhalten und Intoxikationen.
- Die Ergebnisse der Wirksamkeitsmessung werden unter Bezugnahme auf andere Studien plausibilisiert (26-30).
- Die Ergebnisse der Wirksamkeitsmessung werden vorsichtig formuliert: Hinweise auf Wirkungen (24), die in Anbetracht der in internationaler Fachliteratur dokumentierten Erfahrungen inhaltlich plausibel erscheinen (30).

E6: Evaluation der Wirkungen lokaler Alkoholpolitiken – eine vergleichende Fallstudienanalyse (Peters & Wapf, 2006)

Interventionsziele

Förderung von lokalen Alkoholpolitiken (2):

- Die Gemeindevertreter/-innen und die lokalen Fachstellen kennen das Projekt sowie die Kriterien einer lokalen Alkoholpolitik.
- Alkoholpolitiken in den Gemeinden führen zu politischer Verbindlichkeit und Akzeptanz sowie zu verbindlichen Grundlagen im Bereich Alkoholprävention als Beitrag an die Reduktion des risikohaften Alkoholkonsums (speziell des episodisch risikohaften Alkoholkonsums bei jungen Leuten).
- Eine geeignete Netzwerkstruktur, welche die Vernetzung sowie den Informations- und Erfahrungsaustausch zwischen den beteiligten Akteuren sicherstellt, ist vorhanden.
- Die finanzielle Unterstützung bildet einen Anreiz für Gemeinden, sich am Projekt zu beteiligen, und ermöglicht die Umsetzung von Aktionen/Projekten in den Gemeinden, die die Entwicklung der Alkoholpolitik unterstützen.

Überblick über die untersuchten Systeme und Faktoren

Intervention	Systeme im Fokus der Intervention	Einflussfaktoren (Auszug, z.T. deduktiv, z.T. induktiv bestimmt)	Prioritäre Ergebnisindikatoren	Folgen (berichtet)
Unterstützung von Gemeinden bei der Entwicklung von kommunalen Alkoholpolitiken	Soziale Systeme - Gemeinden	Bereitschaft zur Entwicklung einer Alkoholpolitik lokale Alkoholpolitiken: Leitung, Steuerung, Struktur - Entstehen div. Arbeitsgruppen Mitglieder - Vertreter aus Gemeinden, Jugendarbeit, Vereinen, Eltern, Lehrerschaft, Gastwirte, Verkaufsstellen, Polizei etc. Fokus, Aktivitäten - Jugendschutz Networking - Vernetzung zwischen Schule, Vereinen, Lehrerschaft, Gemeinderat etc. Programme/Dienste	<u>Auswirkungen der Alkoholpolitiken auf Massnahmenebene:</u> - Verbesserungen im Bereich Jugendschutz - Einführung von Leitfäden, Formularen - Strengere Anforderungen für öffentliche Anlässe - Durchführung von Kontrollen und Sanktionen - Zertifikate - Vernetzungsgefäße - Verbindlichkeiten im Handeln gegen die Alkoholproblematik - Verankerung (geregelt Verantwortlichkeiten) - Nachhaltigkeit	-

		<ul style="list-style-type: none"> - Jugendschutzkurse - Testkäufe - alkoholfreie Bars - Kampagnen - Öffentlichkeitsarbeit - Leitfäden - Präventionskonzepte - Bewilligungsverfahren - Labels, Auditierungen, Zertifizierungen - Hausarztprojekte <p>Politikbeeinflussung</p> <ul style="list-style-type: none"> - Durchsetzung Jugendschutz - Regionale Vernetzung - Aufnahme Thema in Regionalplanung <p>Veränderung der Umwelt</p> <ul style="list-style-type: none"> - Sensibilisierungsaktionen für Jugendliche, Vereine, Gastwirte, Öffentlichkeit, Festwirtschaftsbetreiber - Umdenken bei öffentlichen Anlässen <p>Systemveränderung</p> <ul style="list-style-type: none"> - Aufträge an Dritte - Gründung von Arbeitsgruppen <p>Monitoring, Evaluation</p> <ul style="list-style-type: none"> - Testkäufe - Evaluationsberichte <p>Produkte, Materialien</p> <ul style="list-style-type: none"> - Leitfäden, Formulare - Jugendschutzvereinbarungen <p>Kontextbedingungen</p> <ul style="list-style-type: none"> - Vorhandensein von Arbeitsgruppen - Einbettung in eine integrale Suchtpolitik 	<p>Unbeabsichtigte Nebenwirkungen auf Ebene Gemeinde:</p> <ul style="list-style-type: none"> - Entstehen von neuen Angeboten - Entstehen von Vereinen - positive regionale Ausstrahlung 	
--	--	--	--	--

		<ul style="list-style-type: none"> - Problemdruck - Kleinräumigkeit - Wille und Engagement wichtiger Akteure - Art und Wechsel Projektleitung - Positive Erfahrungen in Nachbargemeinden 		
	Psychische Systeme			Sensibilisierung von Jugendlichen
	Physische Systeme			weniger Vandalismus

Evaluationsdesign: In einer vergleichenden Fallstudienanalyse wurden 13 Gemeinden untersucht. Neben Gemeinden, die am Projekt teilnahmen, wurden Gemeinden untersucht, die interessiert waren aber letztlich nicht teilnahmen und solche, die sich nie dafür interessierten und auch nicht teilnahmen. Bei den teilnehmenden Gemeinden wurde ein Vergleich zwischen der Projekt- und der Nachprojektphase durchgeführt, kombiniert mit einem Soll-Ist-Vergleich, ausgehend von den gesetzten Zielen. Im Rahmen der Evaluation wurden Dokumentenanalysen und problemzentrierte Interviews durchgeführt.

Umgang mit Komplexität bei der Wirkungsbeurteilung

Interaktionsvielfalt *Konnektivität und Interdependenz, Offenheit*

- Ausgehend von einer Reihe von aus der Literatur abgeleiteten Kriterien (8, 9) wurden in den vielen Interviews zahlreiche Faktoren systematisch zusammengetragen und Setting spezifisch analysiert (18-63).
- Der Einfluss von (förderlichen oder hinderlichen) Kontextbedingungen wurde als wirkungsrelevant erachtet und explizit untersucht (15).
- In der Diskussion werden die Ergebnisse kategorisiert und Setting übergreifend zusammengefasst (64-85).
- In den Schlussfolgerungen findet noch einmal eine starke Verdichtung der Ergebnisse zu 9 Hauptpunkten statt (9).
- Im Abstract werden die Ergebnisse noch einmal stark zu einer Hauptwirkungsaussage verdichtet: „Das Projekt hat kaum die Formulierung und Implementierung von lokalen Alkoholpolitiken im politikwissenschaftlichen Sinne ausgelöst.“ (Z-1).

Komplexe Wirkungsmechanismen *Feedback, Multidimensionalität/Nicht-Linearität/Zirkularität, Dynamik*

- Es geht in der Analyse nicht nur darum, ob lokale Alkoholpolitiken etwas bewirken, sondern was sie bewirken (5); im Vordergrund steht die Analysetiefe (9).
- Die lokalen Alkoholpolitiken werden als Resultat von dynamischen Entwicklungsprozessen verstanden (Z-3), welche im Bericht ansatzweise nachgezeichnet werden (18-63). Auf eine detailliertere Beschreibung dieser Prozesse in einem Materialienband (der hier nicht untersucht wurde), wird verwiesen (17).
- Differenzierte Fragestellungen (Z-3) führen zu einer differenzierten Betrachtungsweise (18-63).
- Die Evaluation versteht sich als theoriegestützt und greift zur Beurteilung der Wirksamkeit auf ein differenziertes theoretisches Modell zurück (7ff).

Selbstorganisation *Adaption und Co-Evolution, Reflexivität, Selbstorganisation/operative Geschlossenheit*

- Alkoholprobleme werden als Ergebnis komplexer Interaktionen zwischen verschiedenen Subsystemen verstanden, in welche die Menschen eingebettet sind (2ff).
- Das Zusammenwirken von Akteuren auf Gemeindeebene wird eingehend untersucht (20).
- Selbstorganisation ist ein wichtiges Kriterium, besteht doch das Ziel der Intervention darin, die Gemeinden anzuregen, selber eine lokale Alkoholpolitik umzusetzen (Z-11). Es geht also nicht darum, in den Gemeinden von aussen her Massnahmen einzuführen, sondern die Gemeinden zu unterstützen, selber Massnahmen zu entwickeln (2).
- Die Evaluation untersucht detailliert die Prozesse und Ergebnisse dieser Selbstorganisation: „Wenn eine Alkoholpolitik in der Gemeinde x eingeführt ist (Outputs), was wird damit konkret ausgelöst?“ (15).
- Es werden Interaktionen zwischen Subsystemen auf Settingebene untersucht, aber nicht die Auswirkungen des Settings auf die Individuen (5).
- Pro untersuchte Gemeinde werden mehrere Interviews durchgeführt, um eine differenzierte Innensicht zu erhalten (15).

Unvorhersehbarkeit *Emergenz, Unvorhersehbarkeit/Ungewissheit*

- Die Evaluation ist von ihren Fragestellungen und vom methodischen Vorgehen her darauf ausgerichtet, unvorhersehbare Entwicklungen zu beobachten und emergente Organisationsstrukturen aufzuspüren (15).
- Zudem wird gezielt nach unbeabsichtigten Nebenwirkungen der Alkoholpolitiken gefragt (5).
- Der Fokus Selbstorganisation zeigt sich auch daran, dass zwei konkrete Modelle der Unterstützung von Gemeinden bezüglich ihrer Alkoholpolitik diskutiert werden: die Unterstützung beim Aufbau versus die Unterstützung beim Vollzug und bei der Weiterentwicklung (Z-12). Es wurde im Rahmen der Evaluation aber nicht untersucht, welches der beiden Modelle die Gemeinden stärker in ihrer Selbstorganisation unterstützen würde.
- Es gibt an einer Stelle einen Hinweis, der auf eine leichte Anpassung des Evaluationsvorgehens hindeutet (telefonische Zweitinterviews, Z-19).

Wirkungsbeurteilung

- Im Vordergrund der Wirkungsbeurteilung steht ein Vorher-Nachher Vergleich kombiniert mit einem Soll-Ist-Vergleich (Z-5).
- Es handelt sich um eine qualitative Form eines kontrafaktischen Vergleichs, in welchem die Entwicklungen in den Gemeinden mit Intervention mit dem aktuellen Entwicklungsstand in Gemeinden ohne entsprechende Intervention verglichen werden (9ff).
- Aus der umfassenden qualitativen Analyse wird mit Rückgriff auf das theoretische Modell und seine Kriterien (7ff) ein klares Wirkungsurteil abgeleitet (Z-1), die Erkenntnisse werden entsprechend verdichtet. Ergänzend werden weitere Erkenntnisse für eine differenzierte Auseinandersetzung dargestellt (9).
- Es wird eine qualitative Repräsentativität angestrebt (Z-7), die Ergebnisse werden argumentativ verallgemeinert (16).
- Die Literatur wird dazu herangezogen, das Wirkungsurteil abzustützen und zu plausibilisieren (Z-9).

E7: Evaluation „Midnight Basketball“ Projekte (Sommerfeld et al., 2006)

Interventionsziele

Das in der offenen Jugendarbeit verortete Interventionsprojekt verfolgt Ziele auf unterschiedlichen Ebenen (3ff):

- Prävention von Jugendkriminalität und -gewalt
- Chancengleichheit durch niederschwellige Teilnahme
- Integration von Jugendlichen, Austausch unterschiedlicher Gruppen
- Gesundheitsförderung (u.a. durch die sportliche Betätigung, aber auch durch Kompetenz- und Persönlichkeitsbildung) und Alkohol- und Tabakprävention

Überblick über die untersuchten Systeme und Faktoren

Intervention	Systeme im Fokus der Intervention	Einflussfaktoren <i>(Auszug, primär induktiv bestimmt)</i>	Prioritäre Ergebnisindikatoren	Folgen
Gemeindeturnhallen werden an Wochenendabenden für Jugendliche geöffnet. An den pädagogisch begleiteten Abendveranstaltungen können sich die Jugendlichen treffen und sportlich betätigen (meist wird Basketball oder Fussball gespielt).	Soziale Systeme - Abendveranstaltungen - Gemeinde -	Struktur des sozialen Raums - Regeln und Grundprinzipien (Gewaltfreiheit, Drogenfreiheit, akzeptierende Haltung, Freundlichkeit, Bestimmtheit etc.) - Sport, Musik, Tanz etc. - Rollen (Coaches etc.) - Mitwirkung von Erwachsenen - Zugang Soziale Welt der Jugendlichen - Freizeit und Spass - Integration in die Sozialwelt - Position und Sinn - Inszenierung des Selbst - Erfahrungen und Austausch, Auseinandersetzungen - Annäherung an das andere Geschlecht	- <u>soziale Integration</u> - <u>Chancengleichheit</u>	

		<ul style="list-style-type: none"> - Gruppendynamik Spannungsfelder - Struktur vs. Aneignung - Animation vs. Partizipation - Verantwortung Erwachsene vs. Jugendliche - unspezifische vs. spezifische Auseinandersetzung - Tätigkeit vs. Diskurs - Aktivität vs. ‚chillen‘ - Geschlossenheit vs. Offenheit - Homogenität vs. Heterogenität - Nähe vs. Distanz - Konsum vs. Gestaltung Projekt - Qualität - Unterstützungswürdigkeit - Transparenz - Attraktivität Wirkung auf das Umfeld - Image des Angebots im Umfeld - Lärmbelästigung Umweltfaktoren - Problemdruck - Image des Stadtteils - Erwartungen an das Projekt - elterliche Zustimmung Ebene Gemeinde - Verbesserte Vernetzung der Jugendarbeit - Alternativen zum Angebot - Finanzierung - Verankerung in Konzepten 		
--	--	---	--	--

		<ul style="list-style-type: none"> - Akzeptanz bei Schlüsselakteuren - Neue Inputs für die Jugendarbeit 		
	Psychische Systeme <ul style="list-style-type: none"> - Jugendliche - jugendliche Coaches 	<ul style="list-style-type: none"> - Motivation zur Teilnahme (Spass, Sport, Zugehörigkeit, Anerkennung, Abwechslung etc.) - Identifikation mit Midnight Basketball - Beobachten und Bewerten - Verantwortungsgefühl - Selbstorganisationsfähigkeiten 	<ul style="list-style-type: none"> - <u>Entwicklung von Kompetenzen und Persönlichkeit, Identitätsbildung</u> - <u>Tabak- und Alkoholkonsum</u> 	
	Physische Systeme <ul style="list-style-type: none"> - Jugendliche - Turnhallen 	<ul style="list-style-type: none"> - sportliches Können - Ausstattung der Turnhallen 	<ul style="list-style-type: none"> - <u>Körperliche Bewegung</u> - <u>Vandalismus, Kriminalität</u> 	

Evaluationsdesign: Bei der Evaluation Midnight Basketball handelt es sich um eine rekonstruktive qualitative Fallanalyse in Anlehnung an die Grounded Theory. Es werden Projektumsetzungen an drei Standorten mittels teilnehmender Beobachtung und Interviews (Jugendliche, relevante Erwachsenen aus dem Umfeld, Vertreter/-innen der Projekte und der beteiligten Gemeinden) untersucht. Ergänzend wurden Teilnehmerzahlen ausgewertet, die punktuell in die Wirkungsbeurteilung eingeflossen sind (z.B. Anzahl Jugendliche mit Migrationshintergrund zur Beurteilung des Integrationsziels).

Umgang mit Komplexität bei der Wirkungsbeurteilung

Interaktionsvielfalt *Konnektivität und Interdependenz, Offenheit*

- Ausgehend von den groben Zieldimensionen rekonstruiert die Evaluation die Lebenswelt der Jugendlichen an den Abendveranstaltungen.
- Durch Beobachtung und in Gesprächen werden Wirkungsfaktoren und Wirkungsbeziehungen identifiziert und dann ausführlich beschrieben (7).
- Auf diese Weise spannt die Evaluation ein komplexes, schwer durchschaubares Netz an Ursache-Wirkungszusammenhängen.
- Die erhobenen Faktoren und Beziehungen werden mit Rückgriff auf die wissenschaftliche Literatur geordnet (28-32).
- Einflüsse aus dem Umfeld der Jugendlichen (z.B. Schule, Familie) werden angesprochen (51), aber nicht detailliert untersucht.
- Gegen Berichtsende werden die Ergebnisse auf mehreren Seiten zusammengefasst (48-52), die Wirkungsbeurteilung bleibt allerdings noch immer sehr komplex.

Komplexe Wirkungsmechanismen *Feedback, Multidimensionalität/Nicht-Linearität/Zirkularität, Dynamik*

- Die Evaluation geht von einem systemischen Ansatz aus und versteht Wirkung als komplexes Zusammenspiel unterschiedlichster Faktoren (28f). Individuelle Kompetenz- und Persönlichkeitsbildung wird als Resultat vielfältiger und dynamischer Entwicklungs- und Lernprozesse in sozialen Systemen verstanden (49).
- Die Evaluation geht in ihrer Analyse stark in die Tiefe und beschreibt ausführlich hoch dynamische und komplexe Entwicklungsprozesse, die von vielfältigen Rückkoppelun-

gen geprägt sind (33f).

- Schritt für Schritt wird eine hoch komplexe Programmtheorie aus den Daten heraus entwickelt.
- Die vielfältigen Wirkungspfade werden in einem komplexen Wirkungsmodell grafisch dargestellt (72).
- Der Wirkungsmechanismus, der über viele Seiten entwickelt wurde, wird gegen Ende des Berichts verdichtet (49), bleibt aber komplex.

Selbstorganisation *Adaption und Co-Evolution, Reflexivität, Selbstorganisation/operative Geschlossenheit*

- Der Evaluation geht von Selbstorganisationsprozessen unter den Jugendlichen aus. Die Intervention selber kann nur Anstöße geben, diese selbstorganisierenden Entwicklungsprozesse zu unterstützen (iv). Wirkung entfaltet sich in diesen Selbstorganisationsprozessen (49).
- Die Evaluation untersucht die Entwicklungs- und Lernfelder, die das Projekt den Jugendlichen gibt. Die Wirkmächtigkeit des Projekts wird in den autonomen Prozessen gesehen, welche den Jugendlichen ermöglicht werden (51).
- Es wird deutlich darauf hingewiesen, dass die Ziele, Vorstellungen und Interessen der Erwachsenen lediglich einen Rahmen setzen, in welchem die Jugendlichen ihre Welt selber gestalten (38).
- Das Erleben der Jugendlichen kann nicht von aussen determiniert werden, vielmehr eignen sich die Jugendlichen die soziale Struktur des Angebotes durch Beobachten und Bewerten Schritt für Schritt an (38), wenn es für sie zu relevanter Umwelt wird.
- In den Interviews mit den Jugendlichen fließt deren Perspektive in die Wirkungsbeurteilung mit ein (10).

Unvorhersehbarkeit *Emergenz, Unvorhersehbarkeit/Ungewissheit*

- Es wird darauf hingewiesen, dass Wirkungen nicht direkt planbar sind, sondern sich indirekt ergeben, über das bereitgestellte Lernfeld (iv).
- So kann es dazu kommen, dass sich Prozesse nicht in die gewünschte, sondern in die Gegenrichtung (Nebenwirkungen) entwickeln (51). Durch die induktive Vorgehensweise werden sowohl intendierte wie auch nicht-intendierte Wirkungen sichtbar.
- Die Evaluation zeigt eine Reihe von Spannungsfeldern auf (31), in welchen ein Gleichgewicht gefunden werden muss, damit sich die Wirkung optimal entfalten kann (38ff, 51ff).
- Die Evaluationsinstrumente werden teilweise erst im Laufe der Untersuchung entwickelt. So dienen die ersten, wenig spezifischen Beobachtungen dazu, die Beobachtungsraster zu verfeinern (28) und die Interviewleitfäden zu entwickeln (19). Es kann entsprechend von einem emergenten Evaluationsdesign gesprochen werden.

Wirkungsbeurteilung

- Es wird festgehalten, dass über den gewählten Zugang der Nachweis über die Wirksamkeit nicht eindeutig geführt werden kann (iv).
- Es wird ein Modell entwickelt, welches aufzeigt, wie das Projekt potenziell Wirkung entfaltet. Ob und in welchem Ausmass diese Wirkung tatsächlich eintritt, lässt sich nicht bestimmen oder gar generalisieren (51), da die Dynamiken immer wieder anders sind.
- Die Evaluation geht davon aus, dass das Projekt lediglich unterstützenden Charakter hat und demzufolge nur einen Wirkungsbeitrag leistet (51). Sie kann diesen Beitrag aber mehrheitlich nicht näher bestimmen und schon gar nicht quantifizieren.
- Die Ergebnisse auf Ebene der Jugendlichen, die aus den drei Projektumsetzungen gewonnen wurden, werden standortübergreifend generalisiert (48-52). Es wird herausgearbeitet, welches Potenzial das Projekt zur Wirkungsentfaltung hat. Es geht also nicht um einen Nachweis eingetretener Wirkungen, sondern um die Erwartbarkeit von Wirkungen (29).
- Die Wirkung der Projektumsetzungen auf die Gemeinde wird standortspezifisch untersucht und dargestellt (43-48).

E8: Kosten-Nutzen-Analyse der Verkehrsunfallprävention (Wieser et al., 2010)

Für diese Detailanalyse wurde auf einen der drei untersuchten Präventionsbereiche fokussiert, auf die Verkehrsunfallprävention.

Interventionsziele

Reduktion von Todesfällen und Verletzten im Strassenverkehr. Gewinnung von behinderungsfreien Lebensjahren (8).

Überblick über die untersuchten Systeme und Faktoren

Intervention	Systeme im Fokus der Intervention	Einflussfaktoren (Auszug, primär induktiv bestimmt)	Prioritäre Ergebnisindikatoren	Folgen
Gesamtheit der durch die öffentliche Hand initiierten Massnahmen zur Unfallverhütung im Strassenverkehr zwischen 1975 und 2007 (mit Ausnahme der Investitionen in die Sicherheit der Strasseninfrastruktur). z.B. Verhaltensorientierte Kampagnen, Gesetze und deren Umsetzung, Investitionen in die Sicherheit von Strassen und Fahrzeugen etc.	Soziale Systeme - Kantone	- Verkehrsgesetze und Verordnungen - Allgemeine wirtschaftliche Lage		
	Psychische Systeme			
	Physische Systeme - Verkehrsteilnehmende - Infrastrukturelles Umfeld	- Alkoholkonsum bei Unfall - Gurtentragequote - Motorrad- und Moped-Helmtragequoten - Fahrradhelmtragequoten - Alter - Zahl der betriebenen Fahrzeuge - Unterschiede in der Strasseninfrastruktur und im Grad der Urbanisierung zwischen den Kantonen	- Unfälle - Todesfälle - dauerhafte Behinderungen - schwere, mittlere und leichte Verletzungen - <u>Behinderungsbereinigte Lebensjahre (DALY)</u>	- Kosten

Evaluationsdesign: Kern des Evaluationsdesigns ist ein kontrafaktischer Vergleich der Veränderungen zwischen 1975 und 2007 verglichen mit einer hypothetischen Situation 2007. Die hypothetische Situation entspricht den Präventionsmassnahmen, die zu Beginn der Schätzperiode im Jahr 1975 bereits eingeführt worden waren. Zudem wurden Vergleiche zwischen verschiedenen Kantonen (mit und ohne entsprechende Massnahmen) in die Berechnungen einbezogen, um die Effekte einzelner Bündel von Präventionsmassnahmen abschätzen zu können.

Umgang mit Komplexität bei der Wirkungsbeurteilung

Interaktionsvielfalt *Konnektivität und Interdependenz, Offenheit*

- Die Wirksamkeitsanalyse fokussiert auf die Beziehung zwischen Bündeln von Präventionsmassnahmen und wenigen Ergebnisindikatoren (33-37).
- Zusätzlich wird eine Reihe von erklärenden Variablen in die Berechnungsmodelle einbezogen (35-37, 79). Es wird in diesem Bericht nicht ersichtlich, auf welcher Grundlage diese Variablen ausgewählt wurden.
- Es werden ganz wenige Umweltfaktoren in die Berechnungen mit einbezogen, vor allem um Verzerrungen aufgrund äusserer Bedingungen (z.B. wirtschaftliche Lage oder Grad der Urbanisierung) zu vermeiden (37).
- Alle verwendeten Variablen werden letztlich dazu genutzt, den Einfluss der Präventionsmassnahmen auf die Gewinnung von behinderungsfreien Lebensjahren zu berechnen.
- Die Interventionen setzen zwar teilweise auf struktureller Ebene an, aber es werden keine Systemveränderungen als Interventionsergebnisse untersucht.

Komplexe Wirkungsmechanismen *Feedback, Multidimensionalität/Nicht-Linearität/Zirkularität, Dynamik*

- Die Berechnungsmodelle werden im Evaluationsbericht nicht näher ausgeführt. Aufgrund der berücksichtigten Variablen dürfte es sich aber um komplizierte statistische Modelle handeln.
- Das Grundwirkungsmodell ist aber sehr einfach aufgebaut: Präventionsmassnahmen wirken auf individuelles Verhalten wobei einzelne verzerrende Faktoren mitspielen. Folge sind intermediäre Gesundheitsauswirkungen (27).
- Es wird nicht untersucht, wie einzelne Präventionsmassnahmen Wirkung entfalten, Wirkungsmechanismen werden nicht untersucht.
- Es gibt keine Hinweise darauf, dass Feedbackprozesse oder dynamische Entwicklungen berücksichtigt würden. Vielmehr ist von einem linearen Grundverständnis auszugehen.
- Wechselwirkungen werden gemäss Bericht insofern berücksichtigt, als die Abnahme von z.B. schweren Verletzungen durch Prävention zu einer Zunahme von leichten Verletzungen führen kann (37) – es handelt sich dabei aber eher um Verschiebungen als um Wechselwirkungen.

Selbstorganisation *Adaption und Co-Evolution, Reflexivität, Selbstorganisation/operative Geschlossenheit*

- Die Evaluation hat keinen systemischen Ansatz und es werden entsprechend auch keine Selbstorganisationsprozesse untersucht.
- Die Perspektive der Verkehrsteilnehmenden selber, also die Beweggründe für mehr oder weniger sicheres Verhalten im Strassenverkehr, interessieren in dieser Evaluation nicht.

Unvorhersehbarkeit *Emergenz, Unvorhersehbarkeit/Ungewissheit*

- Unvorhersehbare Entwicklungen haben für diese Art von Evaluation keine Bedeutung. Es wird die Gesamtwirkung der Präventionsmassnahmen ermittelt, ohne Entwicklungsprozesse zu beurteilen.
- Spezifische Wirkungsziele von einzelnen Präventionsmassnahmen werden nicht untersucht und entsprechend auch keine nicht-intendierten Ziele. Im Fokus stehen die wenigen übergeordneten Wirkungsziele. Potenzielle Nebenwirkungen der Gesamtheit der Präventionsmassnahmen werden nicht untersucht.
- Da es sich um eine bilanzierende Ex-Post Evaluation handelt, nehmen die Entwicklungen der Interventionen auch keinen Einfluss auf das Evaluationsdesign.

Wirkungsbeurteilung

- Die Wirkungsbeurteilung wird einerseits als Schätzung deklariert (37), andererseits werden klare Wirkungsaussagen gemacht („die Wirkung der Interventionen war beträcht-

- lich“ (12), „die Ergebnisse zeigen, dass Informations- und Schulungsmassnahmen wesentlich zur Abnahme der Strassenverkehrsunfälle, ... beigetragen haben“ (15)).
- Zu den einzelnen berücksichtigten Variablen scheint es zwar relativ verlässliche Zahlen zu geben (34-37), die grosse Ungenauigkeit im Modell dürfte aber die angenommene hypothetische Situation (2007 ohne Präventionsmassnahmen) sein.
 - Für die Plausibilisierung der Wirksamkeitsschätzung werden erklärende intermediäre Faktoren (z.B. Gurten- oder Helmtragequoten, 35-37), Vergleiche zwischen den Kantonen (resp. Einheiten von Kantonen) und Sensitivitätsanalysen (51) hinzugezogen (35). Weiter wurde zur Plausibilisierung punktuell auf die Literatur (z.B. im Bereich der Tabakprävention, 58, 65) und auf Experteneinschätzungen (39) zurückgegriffen.
 - In der Wirkungsbeurteilung werden Gesamtwirkungen von Bündeln von Präventionsmassnahmen ausgegeben. Über die Wirksamkeit einzelner Präventionsmassnahmen können keine Angaben gemacht werden (15).
 - Mit Hilfe der Berechnungsmodelle ist es möglich, den Wirkungsbeitrag von derartigen Bündeln von Präventionsmassnahmen zu schätzen (vgl. Abbildung auf Seite 38).

E9: Gesundheitsförderung im Kindergarten. Evaluation des Programms „Komm mit in das gesunde Boot“ (De Bock & Fischer, 2011)

Interventionsziele

Das übergeordnete Ziel des Kindergartenprogramms ist es, den Anteil übergewichtiger Kinder bei Schuleintritt durch ausgewogene Ernährung und Bewegung zu vermindern (11). Für die zwei umgesetzte Module gibt es je spezifische Unterziele (34):

- „Das Ernährungsmodul hat zum Ziel, die Handlungskompetenz der Kinder rund um Essen und Trinken zu erhöhen, das kindliche Verhaltensrepertoire zu erweitern und zu vermitteln, dass Essen und Trinken Spaß machen.
- Das Bewegungsprogramm hat zum Ziel, die Kinder spielerisch und durch Einbinden in eine Piratengeschichte altersgerecht in Bewegung zu bringen, in ihnen die Lust an Bewegung und sportlicher Aktivität zu entfachen.“

Überblick über die untersuchten Systeme und Faktoren

Intervention	Systeme im Fokus der Intervention	Einflussfaktoren (primär deduktiv bestimmt, einige explorativ erhoben)	Prioritäre Ergebnisindikatoren	Folgen
<p><u>Ernährungsmodul</u> mit 10 Bausteinen für Kinder im Kindergarten und 5 Bausteinen für Eltern und Kinder (25-27).</p>	<p>Soziale Systeme</p> <ul style="list-style-type: none"> - Soziales Umfeld (Eltern, Erzieherinnen, Fachkräfte etc.) 	<ul style="list-style-type: none"> - Einbindung der Eltern - gesundheitsförderliches Umfeld (u.a. Leadership, Parallelangebote) 		
<p><u>Bewegungsmodul</u> mit zweimal wöchentlich einer Bewegungsstunde (über 20 Wochen), eine Auftakt-Aktionstag und einem bewegten Elternabend (28-30).</p> <p><u>Partizipativer Elternbaustein</u> für einen Teil der teilnehmenden Kindergärten (38-40). Die Bausteine wurden mehr-</p>	<p>Psychische Systeme</p> <ul style="list-style-type: none"> - Kinder - Eltern - Erzieherinnen - Fachkräfte 	<p>Kinder</p> <ul style="list-style-type: none"> - Wahrnehmung von Gesundheitsbedürfnissen - Handlungskompetenz und Verhaltensrepertoire - Freude an der Bewegung und am Essen und Trinken - Selbstbewusstsein - Lebensqualität und allgemeine Gesundheit - Migrationshintergrund - sozioökonomischer Hintergrund 		

<p>heitlich von den Erzieherinnen durchgeführt, begleitet von Fachkräften.</p>		<p>Eltern</p> <ul style="list-style-type: none"> - Sensibilisierung - Freude an Bewegung - Empowerment <p>Erzieherinnen</p> <ul style="list-style-type: none"> - Bewusstsein für die Wichtigkeit von Bewegung - Überforderung und Bedrängung durch externe Angebote <p>Fachkräfte</p> <ul style="list-style-type: none"> - Qualifikation 		
	<p>Physische Systeme</p> <ul style="list-style-type: none"> - Kinder - Umfeld 	<p>Kinder</p> <ul style="list-style-type: none"> - Koordinative und konditionelle Fähigkeiten - Bewegungsaktivität (Zeit im Sitzen, bei mittlerer und hoher Aktivität) - Herzfrequenz - draussen verbrachte Zeit, vor dem TV verbrachte Zeit - Konsum von Obst, Gemüse und gesunden Lebensmitteln - Mithelfen im Haushalt - Schlafqualität - Lebensqualität und allgemeine Gesundheit - Geschlecht <p>Umfeld</p> <ul style="list-style-type: none"> - Jahreszeit 	<ul style="list-style-type: none"> - <u>Essverhalten</u> - <u>Bewegungsverhalten</u> - <u>Übergewicht bei Kindern</u> (BMI, Körperfettprozente und Taillen-Grössenverhältnis) 	

Evaluationsdesign: Es handelt sich um eine Cluster-randomisierte kontrollierte Studie mit einer Interventionsgruppe, die sowohl das Bewegungs- als auch das Ernährungsmodul umsetzen und einer Wartekontrollgruppe. In der Interventionsgruppe erhielt die Hälfte der Kindergärten, ebenfalls randomisiert zugeteilt, zusätzlich einen Elternbaustein. Es gibt vier Messzeitpunkte, um Veränderungen zu messen (6 Monate vor der Intervention, zu Beginn der Intervention, gegen Ende der Intervention und sechs Monate nach Ende der Intervention, 37). Die Interventionen wurden in zwei Wellen durchgeführt, um saisonale Effekte ausgleichen

zu können (78). Durch die zwei Wellen Ergänzend wurden die Erzieher/-innen (explorative Interviews) und die Eltern (standardisierter Fragebogen) zu den Wirkungen der Intervention befragt.

Umgang mit Komplexität bei der Wirkungsbeurteilung

Interaktionsvielfalt *Konnektivität und Interdependenz, Offenheit*

- Die Interaktion wird als Zusammenspiel zwischen Kindern, Eltern, Erzieherinnen und Fachkräften beschrieben, die Erhebungen fokussieren aber lediglich Veränderungen bei den Kindern (34ff). Systemänderungen sind nicht im Fokus der Evaluation.
- Bei der Bestimmung der beeinflussenden Faktoren stützt sich die Evaluation auf die wissenschaftliche Literatur ab (18-22) sowie auf übergeordnete bildungspolitische Konzepte (25).
- Als externe Einflüsse auf die Kinder wird vor allem die Intervention selber und in diesem Rahmen das Verhalten erziehender Erwachsener thematisiert und im Rahmen des Elternbausteins auch untersucht (38-40). Zusätzlich werden sozioökonomische Faktoren der Kinder einbezogen (Migrationshintergrund, sozioökonomischer Status) sowie die Jahreszeit. Andere externe Faktoren wie zum Beispiel Bewegungs- oder Ernährungsangebote im Umfeld sind nicht Gegenstand der Evaluation und erscheinen nicht als relevante Umwelt.

Komplexe Wirkungsmechanismen *Feedback, Multidimensionalität/Nicht-Linearität/Zirkularität, Dynamik*

- Die Ursache-Wirkungszusammenhänge, welche in der Evaluation untersucht werden, sind relativ einfach. Es wird hauptsächlich der Einfluss der Intervention als Ganzes auf die Kinder der Interventionsgruppe (Ernährungs- und Bewegungsverhalten sowie Körpermasse) untersucht.
- Es werden keine wechselseitigen Ursache-Wirkungszusammenhänge oder Rückkopplungen untersucht.
- Für die Intervention liegt kein grafisches Wirkungsmodell vor.

Selbstorganisation *Adaption und Co-Evolution, Reflexivität, Selbstorganisation/operative Geschlossenheit*

- Es wird darauf hingewiesen, dass Projekte ‚lernende Systeme‘ sind, sich also dank neuen Erkenntnissen verbessern können (33).
- Zudem gibt es wenige Hinweise, die darauf hindeuten, den Kindergarten als autonom entscheidendes System zu begreifen („Ein Problem war daher, dass der Elternbaustein von bestimmten Kindergärten als von außen aufgezwungen empfunden wurde, so dass die Erzieherinnen sich die Idee des partizipativen Elternbausteins nicht zu eigen machen wollten“, 49).
- Der partizipative Elternbaustein wird ebenfalls aus einer systemischen Sichtweise begründet (38ff).
- Systemische Aspekte oder Selbstorganisationsprozesse werden nicht weitergehend thematisiert oder untersucht.
- Kontextspezifische Anpassungen sind nicht erwünscht, um das Forschungsdesign nicht zu beeinträchtigen (90) – obwohl festgehalten wird, dass Flexibilität und lokale Anpassbarkeit wichtige Erfolgsfaktoren seien (90). Beim partizipativen Elternbaustein können kontextspezifische Ausformungen nicht verhindert werden, da die Eltern selber entscheiden, auf welche Weise und wie stark sie sich an der Intervention beteiligen.

Unvorhersehbarkeit *Emergenz, Unvorhersehbarkeit/Ungewissheit*

- Die Bewegungs- und Ernährungsmodule wurden weitgehend standardisiert durchgeführt, während der partizipative Elternbaustein – auf die Bereitschaft der Eltern zur Mitwir-

kung angewiesen – sehr unterschiedlich umgesetzt wurde (48-51). Die Unterschiedlichkeit der Umsetzung wurde allerdings in der Evaluation nicht weiter berücksichtigt, die Auswertung basierte auf dem kontrafaktischen Vergleich zwischen Kindergärten mit und ohne Elternbaustein (72ff).

- Es wurden keine Nebenwirkungen der Interventionen untersucht.
- Die Evaluation verfolgt ein vorgegebenes Forschungsdesign, von welchem sie nicht abweicht.

Wirkungsbeurteilung

- Die Wirkungsbeurteilung gründet auf einen Vorher-Nachher-Vergleich, kombiniert mit dem kontrafaktischen Vergleich zwischen Interventions- und Kontrollgruppe.
- Es werden Gesamtwirkungen des Projekts berechnet. Zu den Wirkungen einzelner Massnahmen können keine Aussagen gemacht werden (mit Ausnahme zum separat untersuchten Elternbaustein).
- Die Ergebnisse der Elternbefragungen werden unter dem Vorbehalt der sozialen Erwünschtheit berichtet (58ff).
- Zur Plausibilisierung der Ergebnisse wird auf die Literatur zurückgegriffen (13) und mit den Ergebnissen anderer Projekte verglichen (77). Zudem wird auf den wertvollen Vergleich zwischen objektiv gemessenen Daten mit qualitativ erhobenen Daten hingewiesen (78) und die Ergebnisse werden entsprechend auch mithilfe der explorativ erhobenen Daten interpretiert.
- Durch die grosse Stichprobe wird von einer Generalisierbarkeit der Ergebnisse für andere Kindergärten in der Region ausgegangen (97).

8.2 Interaktionsvielfalt

Komplexe Systeme sind von einer Vielzahl von Systemkomponenten geprägt, die vielfältig und wechselseitig interagieren. Wirkungen haben, der Unendlichkeitsproblematik der Kausalität folgend (Pawson & Tilley, 1997), tendenziell unendlich viele aufeinander aufbauende und sich gegenseitig beeinflussende Ursachen und dasselbe gilt umgekehrt für jede Ursache, die potenziell unendlich viele Wirkungen entfaltet. Gesundheit ist ein Konzept, welches diese vielfältigen und schier unendlichen Ursache-Wirkungsbezüge – nicht nur sozialer, sondern auch psychischer und physischer Natur - in sich trägt. Das Modell der Gesundheitsdeterminanten führt deutlich vor Augen, dass praktisch alle Lebensbereiche direkt oder indirekt gesundheitsrelevant sind. Bildung, Arbeit, Wohnsituation, nachbarschaftliches Umfeld, ökonomische Situation, kulturelle Zugehörigkeit, Sozialpolitik, Mobilität, Verkehrsinfrastruktur etc. wirken sich direkt oder indirekt auf unsere Lebensweise und unsere Gesundheit aus. Sozialräumliche Systeme (Schulen, Betriebe, Stadtteile,...), in denen Interventionen der Gesundheitsförderung- und Prävention ansetzen, sind ihrerseits durch eine Vielzahl von Akteuren und potenziell wirkungsrelevanten Beziehungen geprägt, systemintern und gegen aussen. Diese Vielfalt an Ursache-Wirkungszusammenhängen birgt einerseits ein grosses Potenzial für gesundheitsfördernde und krankheitspräventive Interventionen, andererseits ist sie interventions-, aber auch evaluationsmethodisch eine grosse Herausforderung. Interventionsseitig gilt es diejenigen Faktoren zu identifizieren, die mit vertretbarem Einsatz einen möglichst grossen Gesundheitsnutzen für ganze Bevölkerungsgruppen versprechen und evaluationsseitig besteht die Herausforderung darin, die Wirkungsvermutungen der Intervention und alternative Wirkungsbeziehungen zu prüfen, Wirkungen und Wirkungsweisen zu beurteilen und Anregungen für Weiterentwicklungen zu geben. Jede Evaluation muss vor dem Hintergrund der vielfältigen beeinflussenden Faktoren Komplexität reduzieren und sich mit der Frage auseinandersetzen, welche Einflussfaktoren auf ein untersuchtes System als besonders wirkungsrelevant in die Analyse aufgenommen werden sollen und welche nicht. In Bezug auf die komplexe Herausforderung Gesundheit tendieren die potenziellen Einflussfaktoren gegen unendlich und es geht darum, die bedeutenden von den weniger bedeutenden oder unbedeutenden zu unterscheiden und entsprechende Grenzen zu ziehen.

Wie im empirischen Teil aufgezeigt werden konnte, sind diese Grenzen weder objektiv gegeben, noch werden sie allein durch die Evaluation gezogen. Vielmehr wird die Grenzziehung in Aushandlungsprozessen realisiert, in welche neben der Perspektive der Evaluation jene der Stakeholder und der Wissenschaft einfließen. Zudem muss Rücksicht auf Einschränkungen der Praxis wie zum Beispiel finanzielle Ressourcen, Verfügbarkeit von Akteuren etc. genommen werden. Letzteres bedeutet, dass nicht die fachliche Relevanz allein für die Komplexitätsreduktion massgebend ist, sondern viele andere Kriterien die Wirkungsanalyse und deren Aussagekraft von Beginn an mitbestimmen:

Auch auf Ebene der Führungskräfte gibt es keine direkte Information, ob sie sich selbst durch die Führungsschulungen kompetenter fühlen im Umgang mit ihren Mitarbeitenden und dem Thema Stress, und ob sie ihren Führungsstil angepasst haben. Eine solche Erfassung unmittelbar nach den Kursen erfordert zusätzliche Datenerhebungen und sprengt den Rahmen von Studien, die die allgemeine Wirksamkeit eines Programms beurteilen. (E2, 59)

Das Evaluationssystem muss sich um die Anschlussfähigkeit an die anderen beteiligten Systeme bemühen und Kompromisse suchen, um die Evaluation überhaupt durchführen zu können: sie muss die Ansprüche der auftraggebenden Systeme berücksichtigen, um den Evaluationsauftrag zu erhalten. Sie ist dem Wissenschaftssystem verpflichtet, da sie durch die Anschlussfähigkeit an dieses System fachliche Legitimation erhält und sie muss sich nach dem Interventions- und den Zielsystemen richten, da sie auf deren Mitwirkung angewiesen ist. Der letztliche Entscheid über die Selektion und Grenzziehung liegt aber bei den Evaluierenden selbst, da sie für die Evaluation die Verantwortung tragen. Als selbstreferentielles System kann das Evaluationssystem nicht anders, als auf den eigenen Sys-

temstrukturen aufzubauen. Entsprechend fließen zum Beispiel das eigene Evaluationsverständnis, die methodischen Qualifikationen und Vorlieben, frühere Evaluationserfahrungen etc. massgeblich in die Entscheidung über den Umgang mit Komplexität ein.

In den meisten der untersuchten Fälle werden die Grenzen des Evaluationssystems primär aus den Projektzielen und aus der Literatur abgeleitet. Sie werden von aussen und in einer stark generalisierter Form an die zu untersuchenden Systeme herangetragen. Eine beobachtete Strategie des Umgangs mit Interaktionsvielfalt ist es, sich auf diese vorgegebenen Wirkungsziele zu beschränken, um sie hypothesenprüfend zur Intervention in Bezug zu setzen. In den explorativ ausgerichteten Evaluationen bilden die vorgegebenen Ziele und wissenschaftliches Wissen nur eine grobe Grenzlinie, die durch induktive Vorgehensweise laufend verändert wird. Durch den Einsatz von Beobachtungen und Interviews tauchen laufend neue Aspekte auf, die in die Wirkungsbeurteilung mit aufgenommen werden. Rein induktiv verfahrenende Wirkungsevaluationen, die weder von vermuteten Wirkungszusammenhängen noch von Erkenntnissen aus der Literatur ausgehen, sind grundsätzlich zwar vorstellbar, in den untersuchten Evaluationen aber nicht vertreten. Wenn hypothesenprüfend ausgerichtete Evaluationen explorative Elemente ergänzend einsetzen, so werden die neu auftauchenden Elemente zwar für die Plausibilisierung und Interpretation der Ergebnisse der Wirkungsmessung genutzt, sie fließen aber nicht in die eigentliche Wirkungsmessung ein. Der Wirkungsnachweis wird anhand der definierten Faktoren erbracht. In länger dauernden Evaluationen kann eine kontinuierliche Überprüfung der Grenzziehung wichtig werden, weil Grenzen in komplexen Interventionssystemen nicht starr sind, sondern sich über die Zeit verschieben können:

Complex adaptive systems have fuzzy boundaries. System membership can change, and agents can be members of multiple systems at the same time (15). Complex adaptive systems are open systems characterized by agents interacting and exchanging information with others beyond the system boundaries. (Lessard & Birch, 2010, 10)

Nicht nur in den settingorientierten Interventionsansätzen, die beispielsweise in E2 oder E7 zum Tragen kommen, sondern auch in eher verhaltensorientierten Ansätzen wie in E1 oder E3 wird die Bedeutung sozialer Systeme für die Wirkungsentfaltung deutlich. In diesen Interventionen steht individuelles Verhalten zwar im Fokus der beabsichtigten Veränderung und die Evaluation misst den Erfolg auch primär auf der Verhaltensebene, aber die Einflüsse von umgebenden sozialen Systemen sind offensichtlich und werden auch in diesen Evaluationen immer wieder deutlich sichtbar:

Ein gewisser Kohorteneffekt könnte die höhere Helmtragquote bei den 13- bis 30-Jährigen als auch die überproportionale Steigerung bei den 18- bis 25-Jährigen erklären. Da Kinder heute von Beginn ihrer Schneesportaktivitäten an mit einem Helm «aufwachsen», ist die Wahrscheinlichkeit gross, dass sie auch im Jugend- und im Erwachsenenalter nicht darauf verzichten werden. (E3, 66)

Selektionen und Grenzziehungen in den untersuchten Evaluationen folgen allerdings, mit Ausnahme von E2 und E7, nicht einer Systemlogik. Es werden zwar implizit Systeme sichtbar und Einflussgrenzen gezogen, indem ausgewählte Wirkungsfaktoren aus dem Umfeld ein- und andere ausgeschlossen werden, aber es fehlt eine explizite Bestimmung von Systemen und Systemgrenzen. Die Unterscheidung zwischen Zielsystemen, Interventionssystemen und umliegenden Systemen wird nicht gemacht und ebenso wenig wird geklärt, welche Systemtypen (sozial, psychisch, physisch) im Fokus der Interventionen stehen und wie sie gegebenenfalls kombiniert sind. Kern einer systemischen Herangehensweise sind gemäss dem Arbeitspapier der Deutschen Gesellschaft für Evaluation (DeGEval, 2013) drei generische systemische Konzepte: Beziehungen, Perspektiven und Systemgrenzen. Systemisch ausgerichtete Evaluationen stützen sich in ihren Beurteilungen auf die Identifikation und Abgrenzung von einzelnen Systemen, auf die Untersuchung von Beziehungen zwischen Systemen und Systemkomponenten sowie auf die Nutzung verschiedener Perspektiven zur Herausarbeitung systemrelevanter Aspekte. Lediglich zwei der untersuchten Evaluationen (E2, E7) lassen vor diesem Hinter-

grund eine systemische Ausrichtung erkennen, indem sie explizit Systeme und Systembeziehungen untersuchen und unterschiedliche Perspektiven innerhalb einzelner Systeme herausarbeiten. Nur eine der Evaluationen basiert explizit auf systemtheoretischen Modellen (E2, 26). Systemische Evaluationsansätze wie z.B. die ‚Critical Systems Heuristics‘ (Ulrich & Reynolds, 2010), das ‚Outcome Mapping‘ (Earl, Carden, & Smutylo, 2001) oder das ‚Viable System Model‘ (Hoverstadt, 2010) finden in keiner der untersuchten Evaluationen Verwendung.

Es fällt auf, dass bei der Selektion relevanter Einflussfaktoren die Perspektive der Zielsysteme und eine kontextspezifische Bestimmung der potenziell relevanten Systembezüge in der Regel zunächst fehlen. Im Laufe der Durchführung der Evaluationen wird diese Perspektive teilweise durch den Einsatz explorativer Verfahren wieder hereingeholt: In Interviews kommen Vertreterinnen und Vertreter der Zielsysteme zu Wort und können auf bisher unberücksichtigte Wirkungsfaktoren und Zusammenhänge aufmerksam machen.

In sozialen Systemen sind immer mehrere Personen mit spezifischen Rollen, Anliegen und Sichtweisen vertreten. Diese schätzen Systeme und Wirkungszusammenhänge potenziell unterschiedlich ein. Systemische (Evaluations-) Ansätze nutzen die unterschiedlichen Sichtweisen, um Systeme und Wirkungszusammenhänge aus verschiedenen Perspektiven zu reflektieren, um dadurch ein differenzierteres und besser abgestütztes Bild zu erhalten. Das Gewährwerden von Differenzen soll das untersuchte System zur Klärung und Weiterentwicklung anregen. Die meisten der untersuchten Evaluationen greifen implizit oder explizit auf die Sichtweise von Stakeholdern zurück, zum Beispiel indem sie auf deren Programmtheorien und Zielen aufbauen oder sie zu Wirkungsmechanismen oder Wirkungseinschätzungen befragen¹²¹. Explorative Forschungsmethoden werden genutzt, um vor allem die Sichtweisen der Trägerschaft, der Umsetzenden und der Zielgruppen zu erfragen. Allerdings findet man in keiner der Evaluationen explizite Stakeholder- oder Netzwerkanalysen und die Perspektiven unterschiedlicher Stakeholder werden nur selten detailliert herausgearbeitet und verglichen (z.B. in E4 oder E7). Wenn darauf verzichtet wird, die Perspektiven der verschiedenen Akteure im Interventionssystem zu untersuchen, werden potenzielle Interessenskonflikte und wichtige Wirkungsfaktoren nicht sichtbar. Wenn bestimmte Akteursgruppen beispielsweise der Einführung eines Gesundheitsförderungsprogramms in einem Betrieb kritisch gegenüberstehen, kann dies zu Widerstand in der Umsetzung und zu entscheidenden Konsequenzen für die Wirkungsentfaltung führen. Die unterschiedlichen Perspektiven zu kennen hilft, Wirkungsmechanismen und beobachtete Wirkungen adäquat zu interpretieren und differenzierte Empfehlungen für künftige Umsetzungen zu geben.

Die Beschreibungen komplexer Sachverhalte basieren primär auf Interviews mit Vertretern der Zielgruppen, Projektbeteiligten und Schlüsselpersonen sowie auf Beobachtungen. Durch die Wahl und Einschränkung der Perspektiven, die dabei berücksichtigt werden, wird die Wirkungskonstruktion stark beeinflusst. Des Weiteren wird bei Beschreibungen oft nicht klar, in welchem Umfang die berichteten Aspekte relevant sind. Handelt es sich um Einzelphänomene oder um breit erhobene und abgestützte Phänomene? Für die Abschätzung des Wirkungspotenzials ist dies von grosser Bedeutung.

Kontextbezüge

Soziale Systeme stehen in ständigem Austausch mit ihrer Umwelt. Kontexteinflüsse werden von den meisten der untersuchten Evaluationen in der einen oder anderen Form als wirkungsrelevant eingestuft (E1 ist hier eine Ausnahme). Das folgende Zitat zeigt exemplarisch, wie wichtig diese Einflüsse für die Wirkungsentfaltung in Interventionssystemen sein können:

Betriebe wurden als komplexe soziale Systeme beschrieben (...). Sie operieren in einer Umwelt, in der es andere Systeme gibt, die für die Organisation relevant sind (und solche, die es

¹²¹ Eine aktuelle Übersicht über aktorsbasierte Evaluations- und Monitoringverfahren liefern (van Ongevalle, Huyse, and van Petegem (2014)).

nicht sind). Relevant sind Kapitalgeber, Kunden, die Politik mit ihren regulierenden und kontrollierenden Instanzen, sowie andere öffentliche und privatwirtschaftliche Einrichtungen mit Einfluss auf das Geschehen in der Organisation wie Versicherungen, Arbeitgebende- und Arbeitnehmende-Verbände, Wirtschafts-Dachverbände, Forschungs- und Entwicklungseinrichtungen, Stiftungen, Think Tanks, Weiterbildungsinstitutionen, u.v.m. (E2, 86)

In der Evaluation, aus welcher das Zitat stammt, werden entsprechende Aspekte vor allem in qualitativen Teilen mit berücksichtigt. In die quantitative Wirkungsanalyse fliessen sie nicht ein, da diese kaum zu erfassen und zu berechnen sind. Dasselbe ist beispielsweise auch in E9 beobachtbar. Hier wird mit Blick auf die Literatur auf die Wichtigkeit eines förderlichen Umfeldes für den Erfolg des Ernährungs- und Bewegungsprojektes im Kindergarten hingewiesen und es werden Faktoren wie engagierte Eltern, Fachkräfte, Erzieher/innen und Grosseltern, Transportwege, gemeinsames in die Natur gehen etc. als wirkungsrelevant aufgezählt. In einem Elternbaustein konnten die Eltern „abhängig von den vorhandenen Möglichkeiten, Angeboten und Problemen ein jeweils individuelles „Menü“ zusammenstellen“ (E9, 40), welches die Intervention im Kindergarten ergänzte. In der Wirkungsanalyse wurden diejenigen Kindergärten mit diesem zusätzlichen Elternbaustein mit jenen Kindergärten verglichen, in denen kein Elternbaustein umgesetzt wurde. Die vielfältigen Kontextfaktoren werden auf diese eine Variable Elternbaustein reduziert. Dass sich dahinter unterschiedlichste Konzepte und Umsetzungen verbergen, wird in der Analyse nicht berücksichtigt – obwohl gerade diese spezifischen Ausgestaltungen des Elternbausteines höchst wirkungsrelevant sein dürften. Selbst wenn versucht wird, eine Intervention in mehreren Settings konzeptgetreu auf dieselbe Weise umzusetzen, wird die Wirkungsentfaltung aufgrund der spezifischen Kontexte dennoch sehr unterschiedlich sein, da jedes soziale System grundsätzlich einzigartig ist ¹²².

In den primär hypothesenprüfenden Evaluationen werden durchgängig nur ganz wenige meist sehr allgemeine Kontextfaktoren berücksichtigt und in die Berechnungsmodelle aufgenommen, so zum Beispiel die Sprachregion in E3 oder saisonale und lokale Unterschiede in E9:

Durch Kontrolle (also methodische Berücksichtigung) von Faktoren wie Saison und Lokalisation der einzelnen Kindergärten konnten wir unterscheiden zwischen einem echten und einem vorgetäuschten Effekt: z.B. hängt das Bewegungsausmass von Kindern von der Saison ab. Kinder sind am aktivsten im Frühling, etwa gleich aktiv im Sommer und Herbst und deutlich weniger aktiv im Winter. (E9, 78)

Wirkungsaussagen, die auf hypothesenprüfenden Verfahren gründen, sind in der Folge meist nicht oder nur sehr oberflächlich kontextspezifisch formuliert. In quasi-experimentellen Studien ist eine Analyse nach einzelnen Kontextfaktoren möglich, in randomisierten kontrollierten Studien verhindert die randomisierte Gruppeneinteilung kontextspezifische Auswertungen, es sei denn in Untergruppen, die dann allerdings nicht mehr der Randomisierung unterliegen. In hypothesenprüfenden Verfahren fliessen Kontextfaktoren also kaum in die Wirkungsbeurteilung ein. Sie werden zwar teilweise ergänzend in Interviews thematisiert, nicht aber in die Berechnungen und darauf gründende Wirkungsaussagen aufgenommen.

In den explorativ verfahrenen Evaluationen werden kontextspezifische Aspekte zum Teil sehr breit diskutiert und die Ergebnisse entsprechend spezifisch dargestellt (E2, E6 und E7). Wirkungen werden als stark kontextspezifisch wahrgenommen und es wird darauf hingewiesen, dass Wirksamkeit nicht losgelöst vom Kontext bestimmt und generalisiert werden kann:

Die Erfolge der Midnight Basketball Projekte in der Schweiz und deren Wirkungsweise sind, wie gezeigt, wenig erforscht. Die verschiedenen Studien aus anderen Ländern vermögen diese Lü-

¹²² Ein diesbezüglich eindrucksvolles Beispiel ist die Evaluation der Einführung eines didaktischen Lehrmittels zur Förderung von Bewegung und ausgewogener Ernährung in Kindergärten des Kantons Zug (Ehmig, Dunkel, and Lenares (2009)).

cke nicht zu schliessen, da einerseits die Umsetzung in den konkreten Projekten stark variiert und andererseits unterschiedliche gesellschaftliche, wirtschaftliche oder kulturelle Kontextbedingungen vorherrschen, welche die Wirkungsweise, die Wirkungsbreite und die Wirkungstiefe wesentlich beeinflussen. (E7, 7)

Es wird die Nähe zu den untersuchten Systemen gesucht, um detaillierte kontextspezifische Informationen zu erhalten. Da soziale Systeme von prinzipieller Einzigartigkeit sind und sich hinsichtlich zielgruppen- und kontextspezifischer Merkmale in der Regel stark unterscheiden, diese aber höchst wirkungsrelevant sein können, ist entsprechend differenziertes Wissen für die Beurteilung der Wirksamkeit einer Intervention wichtig. Dies umso mehr, wenn davon ausgegangen wird, dass Gesundheit das Ergebnis vielfältiger Interaktionen zwischen Individuen und ihrer Umwelt ist und dass sich Gesundheit ständig neu entwickelt¹²³. Differenziertes Wissen über System-Umwelt-Beziehungen erlaubt es, besser einschätzen zu können, inwiefern Wirkungen tatsächlich durch die untersuchte Intervention herbeigeführt wurden, und es liefert wertvolle Hinweise für künftige Interventionen. Allerdings wird es mit zunehmender Komplexität schwieriger, den Überblick zu wahren und kurze und prägnante Wirkungsaussagen zu formulieren, was die Anschlussfähigkeit an die Interventions- und Zielsysteme gefährdet. Obwohl sich explorativ verfahrenende Evaluationen darum bemühen, der gegebenen Komplexität gerecht zu werden, bleibt letztlich doch die realistische Erkenntnis, dass dies nie wirklich gelingen kann:

Die Evaluation kann nur Puzzle-Teile liefern, nicht aber die Komplexität des ganzen Systems abbilden. (E2, 15)

Reduktion von Komplexität

Während die eine Herausforderung ist, Komplexität auszuweiten, um die wirkungsrelevanten Faktoren in der Evaluation angemessen berücksichtigen zu können, ist es eine weitere Herausforderung, die zugelassene Komplexität wieder zu reduzieren und zu klaren Wirkungsaussagen zu verdichten. In hypothesenprüfenden Evaluationen, die sich auf die Untersuchung sehr weniger Faktoren beschränken (E1, E3, E5, E8, E9), sind solche Verdichtungen nicht notwendig, oder sie wurden bereits ganz zu Beginn vorgenommen:

Es wird eine Variable eingeführt, die das im Jahr 2005 eingeführte Bündel an neuen Verordnungen repräsentiert, die unter anderem die Einführung eines maximalen Blutalkoholspiegels von 0,5 Promille für alle Fahrzeugführer und die Möglichkeit der Durchführung von Blutalkoholtests bei Fahrern durch die Polizei ohne Hinweis auf vorherigen Alkoholkonsum umfasste. (E8, 36)

Wenn wie in diesem Fall unterschiedliche Massnahmen von Beginn an zusammengefasst werden, sind Aussagen über die Wirksamkeit einzelner Massnahmen allerdings nicht mehr möglich.

In breiter angelegten hypothesenprüfenden Verfahren (E2) sowie in explorativen Untersuchungen, die rasch eine grosse Vielfalt an Faktoren berücksichtigen (ebenfalls E2 sowie E4, E6, E7), braucht es diese Reduktionen, wenn die Wirkungsaussagen zusammengefasst werden sollen. In E2 werden unterschiedliche Strategien der Komplexitätsreduktion für die Berechnungen angewandt, so unter anderem die Zusammenführung verschiedener Indikatoren zu Indizes:

Die Evaluation verdichtet diese Variablen entlang dem zentralen Organisationsmodell und dem JobDemands-Resources-Modell (vgl. Kap. 3.3), ihrem Evaluationsdesign (Kap. 3.5), aufgrund qualitativer Informationen und methodischer Überlegungen zu vier Faktoren: Ressourcen, Belastungen, positive und negative Gesundheit. Dadurch reduziert sich die Komplexität der Analy-

¹²³ Dieses Verständnis ist u.a. im Lebensweisenkonzept grundgelegt, das ab den 1980'er Jahren in der WHO diskutiert wurde und 1986 in der Ottawa-Charta zur Gesundheitsförderung (WHO (1986)) prominenten Ausdruck fand.

sen erheblich und es kann ein Modell verwendet werden, das für alle (Gross-)Betriebe gleichermassen Gültigkeit hat, d.h. die Ergebnisse der Evaluation lassen sich auf andere vergleichbare Betriebe übertragen. (E2, 32)

In explorativen Evaluationen erfolgen diese Verdichtungen ähnlich. Detaillierte Sachverhalte werden entlang von übergeordneten Kategorien eingeschätzt und dann zusammenfassend beurteilt:

Insgesamt gesehen hat die Durchführung von "Die Gemeinden handeln!" in den Gemeinden mit abgeschlossener Alkoholpolitik eher wenig bewirkt; wir beurteilen die Veränderungen als klein bis mittel. Die Massnahmen haben oft eine geringe Reichweite, und zudem sind klare Ziele und Massnahmenpläne, eben Politiken, kaum vorhanden. (E6, 82)

8.3 Komplexe Wirkungsmechanismen

In komplexen sozialen Systemen sind Ursache-Wirkungszusammenhänge nicht eindimensional und nicht linear. Vielmehr sind vielfältige Wechselwirkungen die Regel. Die einzelnen Systeme sind eng miteinander verflochten, beeinflussen sich gegenseitig und entwickeln sich gemeinsam weiter. Dieses Verständnis zeigt sich auch in einigen der untersuchten Evaluationen:

Das Modell der gesundheitsorientierten Organisationsentwicklung versteht Arbeitsprozesse als Interaktion zwischen Organisation und Person: Die Struktur, Strategie und Kultur des Betriebs sowie die Kompetenz, Motivation und Identität der Mitarbeitenden formen und lenken die Arbeitsprozesse. Gleichzeitig wirken die Arbeitsprozesse zurück auf die Organisation und die Personen. (E2, 10)

Veränderungen entwickeln sich in der Regel langsam, in kleinräumigen dynamischen Prozessen:

SWiNG bewirkte schrittweise kleinere Veränderungen am „Gesamtgebäude Betrieb“. Fundamentale Umwälzungen sind in der Regel nicht Gegenstand von BGF. Die durch SWiNG bewirkten Veränderungen sind dafür auf lange Sicht und für Viele wirksam, wenn sie sich in der Struktur, Strategie und Kultur etablieren. SWiNG baut also auf einer kontinuierlichen und nachhaltigen Entwicklung und Selbst-Optimierung der Betriebe auf. Entsprechend sind flexible Beratungsformen und -techniken nötig, die auf den Betrieb abgestimmt sind. (E2, 7)

Die Evaluation muss die vermuteten Wirkungsmechanismen explizit machen und begründen, worauf sie bei der Wirkungsbeurteilung fokussiert. Sie muss entsprechende Potenziale und Grenzen nachvollziehbar dokumentieren.

Modellierung von Wirkungszusammenhängen

Die untersuchten Evaluationen gehen mit der Konzipierung und Überprüfung von Wirkungszusammenhängen sehr unterschiedlich um. Gemeinsam ist allen Evaluationen, dass Vorstellungen über die Wirkungsweise der Interventionen vorliegen. Zum Teil fliessen diese bereits in Form von theoretischen Modellen in die Evaluation ein (z.B. E1, E2, E6), zum Teil werden sie im Laufe der Analyse zu grafischen Wirkungsmodellen entwickelt (z.B. E2, E3, E7, E8) und zum Teil sind sie auch nur implizit ersichtlich (z.B. E4, 65). Die Wirkungsvermutungen für die betreffenden Interventionen sind teilweise sehr differenziert ausgearbeitet, andernorts liegen sie nur in groben Zügen vor. Wenn in Wirkungsevaluationen die Wirkungsweise nicht Gegenstand der Untersuchung ist, sondern die gewünschten Wirkungen nur anhand einzelner Wirkungsindikatoren überprüft werden, spricht man von ‚Black Box‘-Evaluationen (Funnell & Rogers, 2011). Experimentelle Studien (vgl. E1, E9) gehören in diese Kategorie. Deren Grundkonzept ist es, die Wirksamkeit einer Intervention über die Untersuchung der Beziehungen zwischen einzelnen unabhängigen und abhängigen Variablen im Vergleich zu einer Kontrollgruppe zu bestimm-

men. Sowohl die Varietät der umgesetzten Interventionen als auch mit beeinflussende Kontextfaktoren interessieren nicht und werden durch Randomisierung ausgeglichen. In Zusammenhang mit der Evaluation von komplexen Interventionen sind experimentelle Studien in Missgunst geraten. In unterschiedlichen Varianten wird vorgeschlagen, die ‚Black Box‘ von Programmen zu erhellen und entsprechende Programmtheorien¹²⁴ zu explizieren. Eines der Grundlagenwerke dieser Bewegung ist Chens Klassiker „Theory-driven evaluations“ (Chen, 1990). Unter einer Programmtheorie versteht er: „a specification of what must be done to achieve the desired goals, what other important impacts may also be anticipated, and how these goals and impacts would be generated.“ (Chen 1990, 43). Das Ziel einer Programmtheorie ist es, die Logik eines Interventionsprogrammes explizit zu machen und die vermuteten Wirkungszusammenhänge aufzuzeigen. Der Begriff der Theorie verweist in diesem Zusammenhang nicht auf wissenschaftliche Erklärungstheorien, sondern ist allgemeiner gehalten und fragt nach den Wirkungshypothesen, die einem Programm zugrunde liegen (Chen 1990, 40). Mittlerweile gibt es eine grosse Vielfalt von Modellen, welche auf unterschiedliche Weise Programmtheorien und entsprechende grafische Wirkungsmodelle zu konstruieren und abzubilden vorschlagen, so etwa der aus der Entwicklungszusammenarbeit stammende Logical Framework Approach (LFA) (Norad, 1999); das Modell von Funnell (2000) oder der Programmbaum von Univation (Beywl, 2006). Spezifische für die Gesundheitsförderung ausgerichtete Modelle sind etwa das outcome-Modell von Nutbeam (2000), das darauf aufbauende Ergebnismodell von Gesundheitsförderung Schweiz (Spencer et al., 2008), das Proceed Precede-Modell von Gmel et al. (2010) oder der „Health Promotion Framework“ von Saan und de Haes (2005). Die Grundlogik dieser Modelle ist es, Ursache-Wirkungszusammenhänge von den Inputs bis zu den Impacts eines Programmes darzustellen. Wie auch in den empirischen Daten ersichtlich ist, sind Programmtheorien und Wirkungsmodelle soziale Konstruktionen und stellen starke Komplexitätsreduktionen dar, in welche unterschiedliche Perspektiven einfließen.

Kategorisierungen und Klassifizierungen sind demnach keine natürlichen Vorgänge, sondern komplexe soziale, politische und ideologische Vereinbarungen (...). Durch die Zuordnung von Phänomenen in bestimmte Kategorien, die spezifische Benennung und Bezeichnung sowie die kausale Verbindung derselben werden Sinn und Bedeutung kreiert sowie bestimmte Logiken, Hierarchien und Abhängigkeiten konstruiert. (Ackermann et al., 2009, 3)

Programmtheorien und entsprechende Wirkungsmodelle können sehr unterschiedlich ausgestaltet sein¹²⁵. Sie helfen, die unterstellte Wirkungslogik von Interventionen sichtbar und damit reflektier- und überprüfbar zu machen. Durch die Bestimmung relevanter Wirkfaktoren und Wirkungspfade dienen sie dazu, Komplexität bis zu einem gewissen Grad aufzufächern. Von der anderen Seite her gedacht dienen sie dazu, die Komplexität sozialer Systeme systematisch zu reduzieren und damit handhabbar zu machen. Bei der Auswahl der in einem Wirkungsmodell enthaltenen Aspekte muss ein pragmatischer Kompromiss zwischen Komplexität und Vereinfachung gefunden werden. Auf der einen Seite wird vor einer Übersimplifizierung gewarnt, die beispielsweise dann vorliegt, wenn ein hoch komplexes Interventionsprogramm in einem sehr einfachen Modell abgebildet wird, welches grosse Lücken oder Sprünge in den Wirkungsketten aufweist und entscheidende Wirkungszusammenhänge unerklärt lässt (Funnell und Rogers 2011). Auf der anderen Seite wäre es eine Überforderung, wenn versucht würde,

¹²⁴ Es gibt viele verwandte Konzepte mit unterschiedlichen Bezeichnungen rund um logische Modelle, die hauptsächlich aus dem angelsächsischen Sprachraum stammen: program theory, program logic, logic model, outcome model, theory of change, logframe, intervention theory,... (s. dazu Rogers (2008), 30; Funnell and Rogers (2011)). Gemäss Patton wird ein logisches Modell dann zur Programmtheorie, wenn zusätzlich zur Abfolge der Wirkungen die Mechanismen, wie diese Übergänge zustande kommen, erklärt werden. Program logic und program theory würden sich dadurch unterscheiden, dass in einer Programmtheorie auf wissenschaftliche Theorien gebaut wird, was in program logic-Modellen nicht der Fall sein muss (Patton (2008), 336 f, vgl. zu dieser Unterscheidung auch Funnell und Rogers 2011). In Letzterem unterscheidet sich Patton von Chen (1990), welcher theory im Kontext von Programmtheorien nicht explizit auf wissenschaftliche Theorien bezieht.

¹²⁵ Zum Vergleich zwischen einem inhaltlich offenen Modell mit linearer Grundausrichtung und einem inhaltlich spezifischen Modell für die Entwicklung komplexerer Wirkungsmodelle siehe Heinzmann et al. (2008) sowie Ackermann et al. (2009).

alle möglichen Einflussfaktoren und Zusammenhänge in einem Wirkungsmodell unterzubringen. Das Modell würde rasch unübersichtlich und würde seinen Nutzen einbüßen. Das detaillierte Wirkungsmodell in E7 (vgl. Seite 135 dieser Arbeit) dürfte eine diesbezügliche Grenze erreicht, wenn nicht überschritten haben. Die Herausforderung ist es, die für den Erfolg einer Intervention besonders relevanten Faktoren und Zusammenhänge zu identifizieren und zu berücksichtigen (Funnell und Rogers 2011).

Entlang der Unterscheidung zwischen einfachen, komplizierten und komplexen Aspekten von Interventionen¹²⁶ schlagen Funnell und Rogers (2011) die Nutzung entsprechender Evaluationsansätze und Wirkungsmodelle¹²⁷ vor. Die dabei verwendeten Modelle sollten dem Untersuchungsgegenstand in seiner Einfachheit, Kompliziertheit oder Komplexität entsprechen. Während komplizierte oder komplexe Interventionsaspekte nicht auf Basis einfacher logischer Modelle evaluiert werden sollen, gäbe es aber auch durchaus einfache Aspekte, zu deren Evaluation einfache Modelle ausreichend seien. Einfache logische Modelle zeichnen sich durch eine eindeutige, eindimensionale und lineare Ursache-Wirkungsbeziehung in der Regel entlang der Wirkungskette Input-Activity-Output-Outcome-Impact aus. Dabei ist nicht umstritten, dass die Inputs über die geplanten Massnahmen zum geplanten Output führen und damit die eindeutigen Output und Impact erreicht werden können:

Das Wirkungsmodell der Evaluation E3 entspricht dieser Grundlogik¹²⁸:

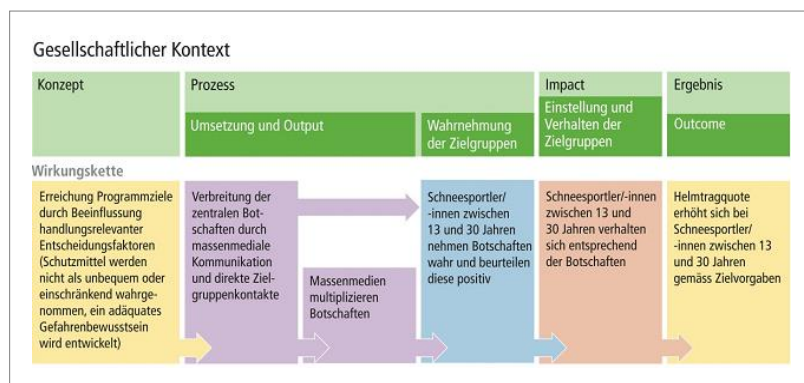


Abb. 27 Einfaches Wirkungsmodell im Kontext einer Präventionskampagne (Furrer und Balthasar 2011 (E3), 60)

Im Sinne eines Rahmenmodells für eine grobe Orientierung kann ein derart lineares Modell durchaus sinnvoll sein. Bei genauerer Betrachtung werden die Limitierungen aber schnell deutlich: Das Modell geht von sehr idealistischen und pauschalen Ursache-Wirkungszusammenhängen aus und ist auf sehr wenige grobe Faktoren reduziert. Wichtige Einflussfaktoren, welche das Helmtragen ebenfalls beeinflussen können (z.B. Gruppendruck, gesellschaftliche Normen, individuelle Gesundheitskompetenzen etc.), sind im Modell nicht enthalten. Auch Fragen, inwiefern andere Akteure ebenfalls ähnliche oder gegenläufige Ziele verfolgen, bleiben unberücksichtigt, obwohl dies für eine Beurteilung der Wirksamkeit von entscheidender Bedeutung wäre. Im Gegensatz zu einfachen Interventionen zeichnen sich komplizierte Interventionen gemäss Rogers (2008) durch das Zusammenspiel verschiedener Akteure resp. die Umsetzung von Interventionsteilen in unterschiedlichen Settings, durch simultan verlaufende Wirkungszusammenhänge sowie durch alternative Wirkungsketten aus. Wirkungsmodelle für komplizierte Interventionen kombinieren entsprechend mehrere Wirkungsketten. Funnell und Rogers vertreten den Standpunkt, dass Programmtheorien und Wirkungsmodelle für einfache und

¹²⁶ Eine Unterscheidung, die hier nicht berücksichtigt wird, da soziale Interventionen nicht anders als komplex sein können (vgl. Kapitel 4.4). Sie können lediglich mehr oder weniger komplex sein.

¹²⁷ Funnell und Rogers verwenden den Begriff logische Modelle.

¹²⁸ In diesem Modell sind die Outcome- und Impact-Ebenen anders definiert als in der vorgestellten Grundkonzeption.

komplizierte soziale Interventionen sehr gut tauglich sind, aber an (überwindbare) Grenzen stossen, was komplexe Interventionen betrifft (Funnell und Rogers 2011). Kybernetische Modelle (Gomez & Probst, 1999; Morecroft, 2010; Vester, 2007) bilden hierzu Alternativen, um komplexe Wirkungsbeziehungen, Rückkopplungen und Wechselwirkungen abzubilden und zu reflektieren. In den untersuchten Evaluationen finden derartige systemische Modelle keine Verwendung. Das in Evaluation E7 entwickelte Wirkungsmodell ist im Ausmass der aufgefächerten Komplexität der Wirkungsbeziehungen aber vergleichbar.

Da im Gegensatz zu Funnell und Rogers in dieser Arbeit Interventionen der Gesundheitsförderung und Prävention generell als komplex verstanden werden, dürfen Modelle mit einfachen, linear verlaufenden Wirkungsketten immer nur als stark limitierte Rahmenmodelle verstanden werden. Vor dem Hintergrund komplexer sozialer Systeme mit ungewisser Wirkungsentfaltung sind derartige Wirkungsmodelle deshalb vermehrt in Kritik geraten (Widmer et al. 2009¹²⁹; Sommerfeld et al., 2006; Patton, 2011). Es wird ihnen u.a. vorgeworfen,

- von vereinfachten Ursache-Wirkungsbeziehungen auszugehen und damit die Komplexität sozialer Realität zu verkennen,
- durch den Fokus auf Geplantes nicht-intendierte Nebenwirkungen zu vernachlässigen,
- Kontextfaktoren nicht angemessen zu berücksichtigen,
- zu starr zu sein und die Dynamik sozialer Systeme nicht erfassen zu können,
- sowie ungeeignet zu sein im Umgang mit Unvorhergesehenem und emergenten Systemeigenschaften.

Es wird argumentiert, dass die Nutzung von Wirkungsmodellen in komplexen Kontexten zu einer allzu reduktionistischen Auffassung der sozialen Realität, zu einseitigen Datenerhebungen und Evaluationsergebnissen, zu Fehlinterpretationen und unangemessenen Schlussfolgerungen führen kann. Tatsächlich verleiten ausformulierte Modelle dazu, sich in der Evaluation nur auf Geplantes zu konzentrieren und nicht intendierte Wirkungen zu übersehen. Eine starke Fokussierung auf Massnahmen und Ziele kann gleichzeitig dazu führen, dass Kontextfaktoren nicht genügend Aufmerksamkeit geschenkt wird. Sind Programmtheorien einmal definiert, ausformuliert und gegebenenfalls visuell ausgearbeitet, besteht die Gefahr, dass die Modelle als starr betrachtet und nicht den Entwicklungen angepasst werden. Es kann schnell geschehen, dass Wirkungsmodelle von der Realität überholt werden und zunehmend weder dem Programm selber, noch der Evaluation dienlich sind. Zudem sind Wirkungsmodelle – als statische Abbilder - nicht dazu geeignet, Dynamiken sozialer Systeme zu erfassen. In der Evaluationsliteratur werden Wirkungsmodelle deshalb vielfach als Rahmenmodelle für weitergehende Analysen gehandelt. Auf Basis einer ‚theory of change‘ und einer grafischen Repräsentation werden einzelne Wirkungsketten identifiziert und zu einer differenzierten ‚theory of action‘ mit z.B. kontextspezifischen Wirkungsannahmen, Erfolgsfaktoren, Risiken, Kontextfaktoren, alternativen Erklärungen etc. ausgearbeitet (Funnell und Rogers 2011; Mayne 2011). Durch diesen Schritt ist es möglich, tiefer in die Komplexität der Wirkungsmechanismen einzudringen. In den untersuchten Evaluationen ist eine derartig systematische Differenzierung der Programmtheorien in keinem Fall beobachtet worden, die Modelle bleiben auf Stufe einer generellen ‚theory of change‘ stehen, wenn auch teilweise explorative Erhebungsmethoden dazu genutzt werden, um kontextspezifische Feinheiten in Ursache-Wirkungsbeziehungen zu reflektieren. Der Grossteil der Wirkungsmodelle bildet soziale Systeme in einer Weise ab, die stark an triviale Maschinen erinnert. Es werden einfache Ursache-Wirkungsbeziehungen gezeichnet, wechselseitige Feedbackprozesse oder gar zirkuläre Beziehungen kommen kaum vor. Die Modelle haben aber den Vorteil, dass sie gut verständlich sind. Folgendes Zitat zeigt das Spannungsfeld, in welchem die Arbeit mit logischen Modellen steckt, schön auf:

Whether it was prepared through reaching consensus in participatory stakeholder workshops or simply consultants in the office, the logical framework necessarily leaves many things out –

¹²⁹ Die in seinem Artikel geäusserte Kritik bezieht sich gezielt auf den Logical Framework Approach.

therein lies its usefulness. At the same time therein also lies its danger, as the things which do not appear in the summary that goes to decision makers tend to be forgotten. (Bakewell & Garbutt, 2005, 13)

Analysetiefe

Die Evaluationen haben viel Gestaltungsspielraum, was den Grad der Komplexitätsreduktion bei der Eingrenzung, Erfassung und Verarbeitung von Informationen zu den untersuchten Systemen betrifft (vgl. Tab. 19). Der Umgang mit Komplexität und die Tiefe der Analyse sind in den untersuchten Evaluationen entsprechend unterschiedlich. Während die hypothesenprüfenden Verfahren den Untersuchungsgegenstand wenig komplex rekonstruieren, führen die explorativ ausgerichteten Evaluationen zu grösserer Komplexität und Analysetiefe, aber auch zu differenzierteren und schwerer überschaubaren Modellen und Ergebnissen. Einige der Evaluationen (*E1, E2, E3, E9*) kombinieren die beiden Ansätze, wobei die hypothesenprüfenden Verfahren im Mittelpunkt stehen und durch die nachgelagert eingesetzten explorativen ergänzt werden. Rekonstruktionen von höherer Komplexität zeichnen sich beispielsweise dadurch aus, dass sie eine vergleichsweise grosse Vielfalt von Einflussfaktoren und Perspektiven berücksichtigen und Wirkungsweisen sehr detailliert und kontextspezifisch untersuchen (vgl. Tab. 19). Durch Offenheit und Nähe können Feinheiten erhoben werden, die für eine Wirkungsbeurteilung äusserst relevant sind, die aber sonst verborgen bleiben würden:

Indessen galt es, sehr gut darauf zu achten, was diesbezüglich von den Mädchen selbst eingebracht wird. Kamen von Seiten der Mädchen bei den Ausführungen zu den Fotos gewalttätige Handlungen, Diskriminierungserfahrungen, sexistische Äusserungen oder rassistische Vorfälle zur Sprache, so fragten die Evaluatorinnen genauer und interessiert nach: Was ist da passiert, wer hat was gemacht, wie ist das konkret vor sich gegangen, wo war das, wann ist das passiert, warum was das so und nicht anders etc.? (*E4, 40*)

Wenn man davon ausgeht, dass sich soziale wie auch psychische Systeme über Sinn konstituieren (Luhmann, 1987), dann erscheint es lohnend, die Bedeutungen zu rekonstruieren, welche sozialer Interaktion und individuellem Gesundheitshandeln zugrunde liegen, um Wirkungsmechanismen besser verstehen und Wirkungsvermutungen überprüfen zu können. Fast alle der untersuchten Evaluationen (*bis auf E5 und E8*) nutzen zumindest ergänzend zu anderen Verfahren explorative Methoden, um die Perspektive von Beteiligten zu erfassen und so die Wirkungsanalysen mit kontextspezifischen Bedeutungen anzureichern. In einigen der hypothesenprüfenden Studien (*z.B. E1, E9*) helfen diese ergänzenden Erkenntnisse, die Ergebnisse aus den Berechnungen zu plausibilisieren und zusätzliche Erkenntnisse über detailliertere Wirkungszusammenhänge zu liefern.

Die Gesprächsleitenden schätzten die Veränderungsmotivation der Jugendlichen bezüglich ihres Alkoholkonsums als tief ein: 75 - 80 % der Jugendlichen wurden der Stufe der Absichtslosigkeit zugeteilt. Diese Einschätzung stimmt mit den Fragebogendaten sehr gut überein. Die Ursache für die tiefe Veränderungsmotivation fanden die Beratenden in den wenigen negativ wahrgenommenen Folgen des Konsums (Erbrechen, Kater, kleine Auseinandersetzungen mit den Eltern). Die grösste Veränderungsmotivation entstand bei den Jugendlichen beim Gedanken, eventuell den Fahrausweis zu verlieren. (*E1, 47*)

Will man den Einfluss sozialer Interventionen auf die Gesundheit differenziert beurteilen, dann kommt man nicht umhin, die strukturellen Kopplungen zwischen sozialen Systemen, psychischen Systemen und physischen Systemen zu untersuchen, wobei der Einfluss von sozialen Systemen auf physische Systeme und umgekehrt über die psychischen Systeme läuft.

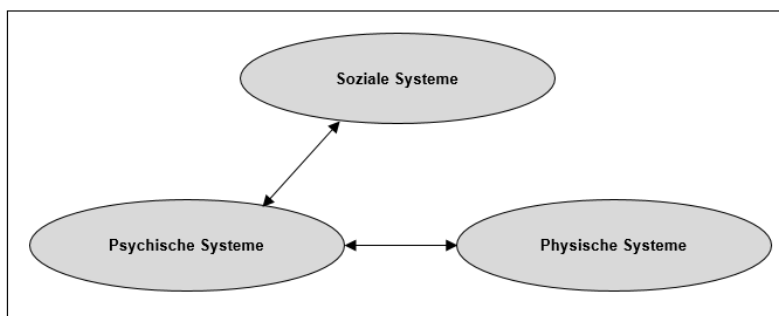


Abb. 28 Strukturelle Kopplung zwischen sozialen, psychischen und physischen Systemen

(Quasi-) experimentelle Verfahren, in der Evaluation von Gesundheitsförderung und Prävention angewendet, vergleichen in der Regel unabhängige Variablen aus dem Bereich der Sozialsysteme (soziale Interventionen) mit abhängigen Variablen entweder aus dem Bereich psychischer Systeme (z.B. Einstellungen) oder aus dem Bereich physischer Systeme (z.B. Body Mass Index), ohne die strukturellen Kopplungen zwischen diesen Systemtypen im Detail zu untersuchen. Mittels explorativer Verfahren können Wirkungsbeziehungen zwischen verschiedenen sozialen Systemen sowie zwischen sozialen und psychischen Systemen tiefer gehend und kontextspezifisch beleuchtet werden:

Die Evaluation arbeitete mit qualitativen Informationen aus Gesprächen und Beobachtungen. Diese Daten bildeten die Grundlage, den Umsetzungsprozess zu beschreiben, die Organisation in ihrer Eigenlogik zu verstehen und die Konsequenzen für die in ihr arbeitenden Personen als auch für das SWiNG-Projekt zu beurteilen. (E2, 28)

Dynamik

Die Mehrheit der untersuchten Evaluationen ist nicht darauf ausgelegt, dynamische Entwicklungen nachzuzeichnen und zu beurteilen. Die Wirkungsbeurteilung beruht von der Grundlogik her – wenn auch methodisch unterschiedlich umgesetzt - primär auf einem Vergleich der Situation vor und nach einer Intervention. Auch wenn in einzelnen Evaluationen dynamische Aspekte angedeutet oder explizit benannt werden (E1, 7ff), fließen diese meist nicht in die Berechnungsmodelle und in die Wirkungsbeurteilungen ein. Drei Evaluationen (E2, E4, E7) rekonstruieren die untersuchten Interventionen als hoch dynamische und facettenreiche Entwicklungsprozesse, allerdings mittels explorativer Forschungsmethodik. Bei der Wirkungsbeurteilung und den Empfehlungen wird entsprechend auch auf diese Dynamiken Bezug genommen:

Im Umsetzungsprozess zeigte sich, dass die Verbindlichkeit des Managements und der Wille der Führungskräfte zentral sind für den Umsetzungserfolg. Auch auf Ebene der Mitarbeitenden wurde deutlich, dass sich dort, wo sie involviert waren und die persönliche Motivation mit der Betriebskultur zusammenpasste, eine positive Wirkungsdynamik entfaltete. Um eine gute Passung zwischen diesen Faktoren zu erreichen, gehört das Thema in die Kategorie der langfristigen Struktur-, Strategie- und Kulturentwicklung. Dann lohnt sich auch die Investition. Die entsprechenden organisationalen und individuellen Faktoren wurden identifiziert und die Evaluation hat Vorschläge gemacht, wie man den Umsetzungsprozess in Zukunft gestalten könnte: Durch Verbindlichkeit und Empowerment der Führungskräfte im Umgang mit der Organisationsanalyse, der anschließenden Reflexion der Ergebnisse im Team und dem gemeinsamen Entscheid zu vertiefter Kompetenzentwicklung in Kursen u.ä. Es wird aber auch klar, dass flexible Beratungsformen und -techniken nötig sind, die auf den Betrieb abgestimmt sind. Ebenso ist der Zeitpunkt der Veränderung wichtig – SWiNG hätte in einigen Betrieben zu einem anderen Zeitpunkt besser gepasst. (E2, 17)

In hypothesenprüfenden Evaluationen scheinen entsprechende Aspekte hin und wieder auf, vor allem in den qualitativen Teilen, werden aber vor allem ergänzend und erklärend genutzt. Feedbackprozesse, welche für Entwicklungen und damit auch für Wirkungsentfaltung in sozialen Systemen von zentraler Bedeutung sind, werden in den vorliegenden Evaluationen kaum thematisiert und in die Wirkungsanalyse aufgenommen. Eine Ausnahme bilden zwei Evaluationen (E2, E7), in welchen derartige Prozesse als wirkungsrelevant untersucht werden, wobei in E7 recht ausführlich:

Die soziale Welt der Jugendlichen bietet also einen Erfahrungsraum, der auf unterschiedlichen Bühnen und in unterschiedlichen Formen Austausch und eigenständige Auseinandersetzung mit anderen ermöglicht und damit dem Individuum erlaubt, sich den sozialen Sinn seiner Umwelt anzueignen, sich daraufhin zu positionieren und Rückmeldungen zu erhalten, die schliesslich die eigene Persönlichkeitsbildung vorantreiben. Der zentrale Rückkopplungsmechanismus besteht in der Erfahrung von Selbstwirksamkeit. Das heisst, dass die Erfahrung gemacht werden kann, dass eigene Äusserungen oder Handlungen in der sozialen Umwelt einen Effekt haben, ob positiv oder negativ. Was einzelne lernen, hängt also unmittelbar von ihrer Umgebung und den Bewertungen und Reaktionen ab, die ihr Verhalten auslöst. (E7, 34)

Wie im Kapitel 7 gezeigt wurde, werden in allen Evaluationen unterschiedliche Handlungs- und Wirkungsebenen angesprochen und Zusammenhänge zwischen diesen mehr oder weniger ausführlich thematisiert. Zum Teil werden diese in Programmtheorien oder Wirkungsmodellen explizit gemacht (vgl. dazu das nächste Kapitel). Allerdings sind die Interaktionsprozesse zwischen den verschiedenen Handlungsebenen nur in einzelnen Fällen (insb. in E2 und E7) explizit Gegenstand systematischer Untersuchungen:

Auch auf Ebene der Mitarbeitenden wurde deutlich, dass sich dort, wo sie involviert waren und die persönliche Motivation mit der Betriebskultur zusammenpasste, eine positive Wirkungsdynamik entfaltete. Um eine gute Passung zwischen diesen Faktoren zu erreichen, gehört das Thema in die Kategorie der langfristigen Struktur-, Strategie- und Kulturentwicklung. Dann lohnt sich auch die Investition. (E2, 17)

In den meisten Evaluationen werden aber vor allem Veränderungen einzelner Ergebnisindikatoren dargestellt und nicht systemübergreifende Interaktions- oder Entwicklungsprozesse nachgezeichnet. Wenn zusätzlich zur Wirkungsbeurteilung Prozessevaluationen durchgeführt werden (E3, E6), dann konzentrieren sich diese primär auf Implementierungsfragen und sind meist ohne direkten Bezug zur Wirkungsbeurteilung. Eine isolierte Betrachtung einzelner, im Voraus festgelegter Kennzahlen zur Evaluation von Gesundheitsförderungs- und Präventionsprogrammen trägt der Komplexität des Untersuchungsgegenstandes jedoch kaum Rechnung und birgt Risiken, wie im Positionspapier „Komplexität und Evaluation“ der DeGEval Gesellschaft für Evaluation festgehalten wird:

Die Konzentration auf wenige ausgewählte Kennzahlen kann der Komplexität der sozialen Realität in den allermeisten Fällen nicht gerecht werden. Wer dennoch – ausschliesslich oder überwiegend – nach Kennzahlen steuert, ohne die Kennzahlenauswahl kontinuierlich zu überprüfen und anzupassen, blendet Komplexität aus. Damit verbunden ist das Risiko der Fehlsteuerung und Ressourcenverschwendung. (DeGEval, 2014, 2)

8.4 Selbstorganisation

Soziale Systeme sind reflexive, adaptive und selbstorganisierende Systeme, die sich in dynamischen Interaktionsprozessen mit umgebenden Systemen entwickeln. Gemäss dem für die Gesundheitsförderung prägenden Konzept der Lebensweisen entsteht Gesundheit in Interdependenz von Lebensbedingungen, kollektiven und individuellen Lebensweisen sowie vorhandenen sozialen und materiellen Ressourcen (Ruckstuhl, 2011). Systemstrukturen entstehen und verändern sich über komplexe Interaktionen zwischen dem System und seiner Umwelt. Dabei wird schnell einsichtig, dass es nicht allein die Qualität eines Interventionsprogrammes ist, die über das Wirkungspotenzial von Interventionen entscheidet, sondern dass Wirkung in wahrscheinlich weit grösserem Ausmass von den Ressourcen abhängt, die einem Zielsystem zur Verfügung stehen. In der Gesundheitsförderungsliteratur werden zurückgehend auf Bourdieu (1986) derartige Ressourcen unter dem Begriff Kapital diskutiert. Ökonomisches, soziales und kulturelles Kapital sind dabei nicht als individuelle Ressourcen zu verstehen, sondern als interagierende gesellschaftliche Ressourcen, die Bevölkerungsgruppen und einzelne Gesellschaftsmitglieder in ihren Möglichkeiten fördern oder eben auch einschränken (Abels, 2007). Die Gesundheitsförderung baut insofern auf diesen Konzepten auf, als sie Interventionen (idealerweise) als partizipative Empowermentprozesse konzipiert, in denen Zielgruppen in ihrem selbstbestimmten Handeln unterstützt und dazu ermächtigt werden, ihre Lebensbedingungen und Lebensweisen selbst gesundheitsförderlicher zu gestalten. Durch das gleichzeitige Ansetzen auf mehreren gesellschaftlichen Handlungsebenen¹³⁰ sollen gesundheitsfördernde Prozesse von verschiedenen Seiten angestossen und genährt sowie entsprechende Synergien gezielt genutzt werden. Die Partizipation ist in der Gesundheitsförderung und Prävention ein Leitprinzip, welches die Wirksamkeit und Nachhaltigkeit von Interventionen generell erhöhen soll, indem es die Anschlussfähigkeit an die Zielsysteme gewährleistet und so die Bewältigung von Komplexität von innen heraus fördert. Ziel der Gesundheitsförderung und Prävention ist es letztlich, die Strukturen in den Zielsystemen so zu beeinflussen, dass sie nachhaltige gesundheitsförderliche Prozesse ermöglichen.

Die Evaluation (insbesondere der Gesundheitsförderung, wo systemische, partizipative und Empowermentansätze gefordert sind) muss darlegen, inwiefern sie Selbstorganisationsprozesse in den Interventionssystemen untersucht. Dies vor dem Hintergrund, dass nicht das Interventionsprogramm ‚Wirkung erzeugt‘, sondern dass Wirkungen das Ergebnis von Selbstorganisationsprozessen in den Zielsystemen sind. Die Evaluation sollte entsprechend untersuchen, inwiefern das intervenierende System wirkungsrelevante Selbstorganisationsprozesse im Zielsystem unterstützt und wie es auf (dynamische, unvorhersehbare,...) Entwicklungen im Zielsystem reagiert. Wirkungen der Gesundheitsförderung und Prävention können sich letztlich nicht anders entfalten als über Selbstorganisationsprozesse der involvierten komplexen (sozialen) Systeme, die von umgebenden Systemen beeinflusst werden. Ein Beispiel dafür liefert folgendes Zitat aus einer der untersuchten Evaluationen:

Die Wirkung bei den Jugendlichen im Hinblick auf die vom Projekt verfolgten Zieldimensionen entsteht also quasi indirekt. Indem Midnight Basketball den Jugendlichen den Raum und seine Struktur zur autonomen Nutzung zur Verfügung stellt, entstehen selbstorganisierte Erfahrungsräume in der sozialen Welt der Jugendlichen, die durch den sozialen Sinn angereichert sind, der von Midnight Basketball repräsentiert und von den beteiligten Erwachsenen und Coaches verkörpert wird. (E7, 49)

Grundsätzlich scheint in keiner der Evaluationen eine deterministische Haltung in der Form vorzuherrschen, dass Interventionen, falls nur richtig durchgeführt, automatisch zu den gewünschten Ergebnissen führen würden. Die anvisierten Zielgruppen und Settings werden grundsätzlich als selbstbestimmend wahrgenommen, d.h. sie können im Sinne der Intervention reagieren, aber auch ganz anders.

¹³⁰ Individuen, Gruppen, Organisationen, Lebenswelten und Gesamtpolitik (WHO 1986).

Zweck der Evaluation ist es zu prüfen, wie diese Reaktion ausfällt. Auch in Evaluationen, die von einfachen Wirkungsmechanismen ausgehen, ist dies klar erkennbar:

Die motivierende Gesprächsführung hat den Vorteil, dass es in Beratungsgesprächen weniger zu Kommunikationsproblemen oder Machtkämpfen kommt, weil es nicht darum geht, wer „objektiv“ Recht hat. Die Sichtweise der Konsumierenden ist grundsätzlich die Richtige, welche die Beratungsperson in kleinen Schritten zu ergänzen versucht, so dass die Konsumierenden die neuen Informationen in ihr Bild einbauen können. (E1, 10)

In den explorativ orientierten Evaluationen, welche settingorientierte Interventionen untersuchen, werden die Interventionssysteme am ehesten und zum Teil explizit als adaptive, lernende Systeme beschrieben und entsprechende Entwicklungsprozesse zwischen dem Zielsystem und dem intervenierenden System nachgezeichnet.

Weiterhin ist im Zusammenhang mit den tatsächlich realisierbaren Effekten noch einmal zu betonen, dass die sozialen Erfahrungsräume dann erwartbare Wirkungen hervorbringen, wenn sie tatsächlich Erfahrungen eröffnen. Wenn beispielsweise keine Erfahrungen mit Gewalt gemacht werden, weil alles immer friedlich ist, und jeder Konflikt sofort unterbunden wird, dann ist die davon ausgehende Wirkmächtigkeit zwar nicht völlig ausgeschlossen, aber sie ist nicht so gross, wie wenn tatsächliche Auseinandersetzungen befriedet werden können. Ebenso gilt im Hinblick auf interkulturellen Austausch, dass Kopräsenz ohne effektiven Austausch eine geringere Wirkung hat, als wenn die verschiedenen Gruppen tatsächlich interagieren. (E7, 52)

In E2 wurde die Passung zwischen dem Interventions- und dem Zielsystem gar als entscheidendes Erfolgskriterium für die Wirkungsentfaltung herausgearbeitet:

Das Fazit zum Umsetzungsprozess ist zuerst einmal sehr allgemein zu formulieren: Eine gute Passung („Fit“) zwischen der Projektarchitektur und der Organisation mit ihren Mitgliedern ist der Schlüssel zum Erfolg. (E2, 13)

Mittels Interviews wird Raum für Reflexion geschaffen und den Beteiligten oder Betroffenen Gelegenheit gegeben, ihre Sicht der Dinge offen zu legen. So werden beispielsweise wirkungsrelevante Einstellungen sichtbar, welche differenzierte Wirkungsbeurteilungen ermöglichen:

Sehr einsichtig verdeutlichen die Beschreibungen der Mädchen, dass sie sich, im Vergleich zu den Buben, an einem deutlich anderen Platz wahrnehmen. Die Sicht der Mädchen: Die Buben nehmen Raum ein und sind aktiv, die Mädchen haben weniger Raum, langweilen sich manchmal, grenzen ihr Aktivitätsfeld – was ein Mädchen tun soll oder darf – ein und nehmen sich teilweise selbst zurück.. (E7, 41)

Lern- und Entwicklungsprozesse werden in etwa der Hälfte der untersuchten Evaluationen beschrieben, wenn auch unterschiedlich detailliert (E2, E4, E6, E7). Eigentliche Selbstorganisationsprozesse und ihr Einfluss auf die Wirkungsentstehung werden aber nur in Ausnahmen (E2, E7) explizit thematisiert und untersucht:

Die Beratenden versuchten, das Wissen und die Kompetenz aufzubauen, um das Thema betriebsintern zu verstehen und zu beobachten, Analyseergebnisse interpretieren zu können und im Team darüber zu sprechen. Was sich darüber hinaus an Wirkungsdynamik zu entfalten vermochte, lag in der Verantwortung des Betriebs, der Führungskräfte und des einzelnen Teams. (E2, 53)

BGF setzt sich zum Ziel, sowohl individuelles Verhalten als auch organisationales Geschehen zu verändern. Dabei verfolgt sie einen salutogenen Ansatz: Man setzt breit an, nicht tief. Man befähigt die Organisation, ihre Gesundheitsrisiken und -chancen wahrzunehmen und sich selbst

zu optimieren. SWiNG hat dazu beigetragen, diesen Optimierungsprozess in Gang zu setzen und nachhaltig zu verankern. (E2, 63)

Indem Midnight Basketball den Jugendlichen den Raum und seine Struktur zur autonomen Nutzung zur Verfügung stellt, entstehen selbstorganisierte Erfahrungsräume in der sozialen Welt der Jugendlichen, die durch den sozialen Sinn angereichert sind, der von Midnight Basketball repräsentiert und von den beteiligten Erwachsenen und Coaches verkörpert wird. (E7, 49)

In den hypothesenprüfenden Evaluationen (E1, E3, E5, E8, E9) werden Selbstorganisationsprozesse nicht untersucht. Zum Teil wird zwar auf entsprechende Konzepte verwiesen:

Diese Ansätze akzeptieren die Jugendlichen als eigenständiges und vollständiges Individuum, ohne ihnen Vorträge zu halten oder Ultimaten zu stellen. Diese Interventionsform schafft somit ein Umfeld für selbstgesteuerte Verhaltensänderung, das Lehrpersonen, Eltern oder andere Autoritätspersonen nur sehr selten zu schaffen vermögen. (E1, 1)

Aber das Forschungsdesign ist nicht darauf ausgelegt, um derartige Prozesse zu analysieren.

Wenn sich Wirkung in komplexen (sozialen) Systemen nicht anders als über Selbstorganisationsprozesse entfalten kann, dann sind diese Selbstorganisationsprozesse für die Wirkungsevaluation von Bedeutung. Wenn Wirkungsmechanismen differenziert beurteilt werden sollen, dann müssen die Selbstorganisationsprozesse in den Interventions- und Zielsystemen sowie die strukturellen Kopplungen dieser Systeme besser verstanden werden. Wenn es das Ziel der Gesundheitsförderung und Prävention ist, Zielsysteme zu befähigen, Lebensbedingungen und Lebensweisen gesundheitsförderlicher zu gestalten, dann müssen die Selbstorganisationsprozesse der Zielsysteme und der Einfluss der Intervention auf diese Prozesse in den Fokus der Evaluation gelangen. Wenn lediglich die Wirkungen der Interventionssysteme untersucht und ausgewiesen werden, kann die Evaluation zwar Aussagen über die Wirksamkeit der Interventionssysteme machen, nicht aber über das Potenzial der Interventionen für nachhaltige Veränderungen in den Zielsystemen. Wenn in einer randomisierten kontrollierten Studie die Wirksamkeit einer Intervention zum Beispiel in einer Schule nachgewiesen wird (wie in E9), dann ist die Wirkungsaussage genau genommen die folgende: Wenn in vielen vergleichbaren Schulen vergleichbare Interventionen durchgeführt werden, kann davon ausgegangen werden, dass im Durchschnitt ein ähnlicher Effekt erzielt werden kann. Da es aber nicht Ziel der Gesundheitsförderung und Prävention sein kann, die Interventionen in den Settings immer und immer wieder auf dieselbe Art durchzuführen, ist ein solches Ergebnis von begrenzter Aussagekraft. Wichtiger wären Aussagen darüber, wie sich die Zielsysteme organisieren und wie sie in Richtung gesundheitsförderlicher Strukturen beeinflusst werden können.

Wenn Zielsysteme als lernende, adaptive, co-evoluierende und selbstorganisierende Systeme verstanden werden, dann stellt sich nicht nur die Frage, inwiefern Gesundheitsförderungsprogramme als partizipative Prozesse konzipiert sein sollen, sondern auch welcher Einfluss und welche Rolle der Evaluation selber zuteil kommen sollen. Soll sie eine unabhängige, externe beurteilende Instanz sein (und kann sie das überhaupt), oder inwieweit soll sie sich in den gesundheitsfördernden Entwicklungs- und Lernprozess einbringen. Bei keiner der untersuchten Evaluationen handelt es sich um eine ausgewiesene partizipative Evaluation. Nicht dass es diese in der Gesundheitsförderung und Prävention nicht geben würde (siehe dazu ausführlich Nitsch et al., 2013). Primär mag dies damit zusammenhängen, dass sich die vorliegende Untersuchung auf Wirkungsevaluationen beschränkt, welche zur Rechenschaftslegung eine ‚neutrale und objektive‘ Sicht verlangen.

„Ein wichtiges Anliegen der bfu ist, Kampagnen objektiv messbar zu konzipieren und deren Wirkung belegen zu können.“ (E3, 5)

Wie die in Kapitel 7 geleistete Rekonstruktion der Wirkungskonstruktion in der Evaluation aufzeigt, kann in Wirkungsevaluationen, welche komplexe Interventionssysteme untersuchen, von einem objek-

tiven und unabhängigen Urteil allerdings nicht die Rede sein. Wirkungen sind vielmehr Ergebnis vielfältiger Interaktionsprozesse zwischen Stakeholdern, Evaluierenden und der Wissenschaft. Durch mehr Involviertheit der an der Intervention Beteiligten kann deren Perspektive in der Wirkungskonstruktion gestärkt werden. Unterschiedliche Perspektiven werden sichtbar und diskutierbar, was kollektives Lernen, gemeinsam getragene Entscheide und innovative Lösungen fördert:

Il s'agit en quelque sorte de parier sur l'intelligence collective au-delà des difficultés que soulève toute confrontation de points de vue, pour coconstruire le problème de décision et imaginer une solution partagée. Il s'agit de mobiliser l'énergie issue de nos différends pour la canaliser vers la création d'une chose qui n'a encore jamais été créée. (Plottu & Plottu, 2009)

Im Extremfall der Selbstevaluation ist die Perspektive der Stakeholder (mit entsprechendem Fachwissen) bestimmend. Wenn auf Partizipation gänzlich verzichtet wird, wird die Wirkungskonstruktion tendenziell der Evaluation allein überlassen. Weder das eine Extrem noch das andere scheint angemessen und zwar aus unterschiedlichen Gründen: Reine Selbstevaluationen sind bei Wirkungsbeurteilungen neben Fragen zur methodischen Qualität vor allem deshalb problematisch, weil sie kaum Akzeptanz in politischen Entscheidungsprozessen finden dürften. Wirkungsevaluationen mit grosser Distanz zum Interventionssystem laufen Gefahr, die Interventionen und Wirkungsweisen nicht angemessen erfassen und interpretieren zu können. De Laats Reflexionen zum Verhältnis zwischen Evaluation, Auftrag Gebenden und Evaluierten (de Laats, 2013) bieten einen Orientierungsrahmen für die Wahl eines dem Evaluationszweck angemessenen Arrangements.

8.5 Unvorhersehbarkeit

Die unzähligen dynamischen Interaktionen und wechselseitigen Rückkoppelungen in komplexen Systemen führen zu Entwicklungen, die sich im Voraus nicht mit Bestimmtheit voraussehen lassen. Interventionen der Gesundheitsförderung und Prävention greifen in derart komplexe Sozialsysteme ein und sind deshalb nur bis zu einem bestimmten Grad planbar. Je mehr dabei auf Freiwilligkeit, Partizipation und Selbstverantwortlichkeit gesetzt wird, umso eher haben die Zielsysteme die Möglichkeit, sich frei zu entfalten und umso schwieriger wird es, Entwicklungen und damit auch Wirkungen vorauszusagen zu können. Es bilden sich Strukturen, Verhaltensmuster und Regeln heraus, welche die geplanten Prozesse unterstützen, aber auch behindern können. Diese emergenten Entwicklungen sind dadurch für die Wirkungsentfaltung grundsätzlich entscheidend.

Die untersuchten Evaluationen gehen sehr unterschiedlich mit emergenten Systemeigenschaften um. Bei den primär verhaltensorientierten Interventionen (E1, E3, E9) liegt der Analysefokus entsprechend nicht auf der Systemebene. Emergente Eigenschaften übergeordneter Systeme werden wenig thematisiert, kommen jedoch partiell vor, so z.B. in E1, wenn Klassendynamiken zur Sprache kommen. Allerdings fließen diese Dynamiken nicht in die Berechnungen und damit in die Wirkungsnachweise ein. E5 und E8 sind interessant, da sie zwar durch ihre Vergleiche auf gesamtgesellschaftlicher Ebene grössere Systeme im Blick haben, Subsysteme und darauf aufbauende Systemdynamiken aber weitgehend ausser Acht lassen. Die Evaluationen mit starker Settingorientierung (E2, E6, E7) verstehen die untersuchten Settings als sozialräumliche Systeme mit spezifischen Eigenschaften und Dynamiken:

Da Organisationen komplexe soziale Systeme mit eigenen Routinen und Regeln sind, wurden u.a. systemtheoretische Modelle entwickelt, um sie zu verstehen und zu verändern. Mit diesen Modellen versucht man die Eigenlogik eines Betriebs zu verstehen und besser auf Wechselwirkungen und unerwartete Nebenwirkungen von Interventionen vorbereitet zu sein. (E2, 25)

Insbesondere in E2 und E7 werden diese Settings detailliert rekonstruiert und Interaktionen zwischen einzelnen Subsystemen sowie ihr Einfluss auf die Gesamtentwicklung und die Wirkungsentfaltung thematisiert.

Mit der Komplexität des Wirkungsmodells wird deutlich, wo und wie die erwartbaren Wirkungen entstehen. Es wird zugleich deutlich, dass es sich um eine anspruchsvolle Aufgabe handelt, dieses System zu leiten. Das Bild ähnelt einem Mobile: Wenn man an einer Stelle zieht, dann verändert sich die gesamte Konstellation und mit ihr auch die Wirkungen, die tatsächlich realisiert werden können. (E7, 32)

Dieses Bild des Mobiles verdeutlicht, dass durch einzelne Interventionen Eigendynamiken im System ausgelöst werden, welche sich nicht vorhersehbar entwickeln und welche sich nicht auf die Eigenheiten einzelner Systemkomponenten zurückführen lassen. In Bezug auf Emergenz ist die Unterscheidung zwischen Evaluationen mit deduktiver und solchen mit induktiver Grenzziehung interessant. Evaluationen, die sich stark an den Wirkungsvermutungen der Träger orientieren und sich auf die Überprüfung von Wirkungshypothesen resp. der Erreichung festgelegter Ziele konzentrieren, laufen Gefahr, durch diesen engen Blick andere Entwicklungen zu übersehen (E1, E3, E5, E8, E9). Zum Teil werden einige wenige antizipierte potenzielle Nebenwirkungen mit berücksichtigt (E1, 36), der Blick bleibt dabei aber stets eng fokussiert. Auf der anderen Seite stehen explorative Evaluationen, welche versuchen, möglichst unvoreingenommen Entwicklungen zu beschreiben:

Die Herangehensweise, die für diese Art der sozialwissenschaftlichen Forschung charakteristisch ist, kann am besten als „explorativ“ bezeichnet werden. Das heisst, dass eine Annäherung an den Gegenstand (hier Midnight Basketball) ohne vorgefasste Modellannahmen stattfindet. (E7, 28)

Durch diese offene Herangehensweise sollen Wirkungsmechanismen sichtbar werden, die nicht bereits zum Vorneherein gedacht wurden:

Gibt es Pfade, die plausibel als Wirkungsrichtungen dargestellt werden können? Welche dynamischen Elemente in Kombination mit welchen inhaltlichen Elementen machen welche Wirkungen im Hinblick auf die genannten Zieldimensionen erwartbar? (E7, 29)

Unvorhersehbarkeit wird dabei nicht als ein Mangel gesehen, sondern als ein konstitutives Element sozialer Intervention:

Zugleich kommt das pädagogische Dilemma daran zum Vorschein. Je reibungsloser der Ablauf organisiert wird, desto geringer sind die Lernchancen, weil es dann kaum Anlässe dafür gibt, die abstrakten Prinzipien der Gewaltfreiheit, der Drogenfreiheit, der Chancengleichheit, der Integration selbst zu erfahren. (E7, 39)

Interventionen von hoher Komplexität zeichnen sich dadurch aus, dass sowohl die Prozesse als auch die Ergebnisse der Interventionen nicht zum Vorneherein bestimmt werden können, sondern diese sich im Laufe der Intervention dynamisch entwickeln. Wenn weder die letztlich beteiligten Akteure im Voraus klar sind noch die Massnahmen, welche ergriffen werden, erübrigt es sich auch, detaillierte Ziele zu formulieren. Bestenfalls können für derartige Interventionen grobe Stossrichtungen resp. Zielbereiche vorgegeben werden. Obwohl der Ansatz der ‚goal-free evaluation‘ (Scriven, 1991a) nicht explizit für die Evaluation von Interventionen ohne Ziele entwickelt wurde¹³¹, eignet er sich durch seine offene Herangehensweise aber auch für die Beurteilung von Interventionen, bei denen aufgrund der grossen Dynamik eine klare Zielformulierung zu Beginn nicht sinnvoll ist. Die Grundidee hinter dem Ansatz der goal-free Evaluation ist, dass sich die Evaluierenden nicht durch die gesetzten Ziele leiten lassen sollen, weil damit die Gefahr verbunden ist, dass sie blind für nicht-intendierte Wirkungen werden:

¹³¹ Das ‚goal-free‘ bezieht sich nicht auf das Interventionssystem, sondern auf die externe Evaluation, welche keine Kenntnis der Ziele der Intervention haben soll.

Finally the staff person (die Evaluator/-innen. Anm. d. Verf.) is likely to have occupational tunnel-vision with respect to the effects of the materials (or methods, etc.,) - that is, a tendency to look mainly in the direction of the announced goals“ (...) „The less the external evaluator hears about the goals of the project, the less tunnel-vision will develop, the more attention will be paid to *looking for actual effects* (rather than *checking on alleged effects*). (Scriven, 1991b, 57)

Die ‚goal-free‘ Evaluation entwickelte Scriven in Abgrenzung zu den gängigen ‚goal-based‘ Ansätzen, welche sich vorrangig auf die Beurteilung der Zielerreichung von Programmen konzentrieren, ohne erstens die Angemessenheit der Ziele zu hinterfragen und zweitens auf nicht-intendierte Wirkungen, vor allem auch auf Nebenwirkungen, zu achten. Programmziele würden oft sehr unrealistisch und/oder sehr vage formuliert und würden sich deshalb als Referenz für den Erfolg eines Programmes nicht eignen. Auch sei es für die Evaluation nicht erheblich, ob Wirkungen intendiert oder nicht intendiert, Ziele oder Nebenwirkungen seien (Scriven 2006). Goal-free Evaluation eignet sich insofern für die Evaluation von komplexen Interventionen, als sie sehr breit nach Wirkungen sucht und eine grosse Offenheit für Wirkungen auf allen möglichen Ebenen hat. Durch seine Breite ist dieser Evaluationsansatz allerdings sehr ressourcenintensiv. ‚Goal-free‘ Evaluation bewegt sich im Spannungsfeld, einerseits durch seine weite Perspektive potenzielle Wirkungsfelder absuchen zu können, andererseits aber auch vergeblich dort zu suchen, wo es nichts zu finden gibt. Eine fokussierte Suche nach intendierten Zielen ist einerseits wesentlich effizienter, läuft aber Gefahr, Wichtiges zu übersehen. Chen (1990) kritisiert am ‚goal-free‘ Ansatz vor allem, dass er mit dem Verzicht der Beachtung der Ziele eine wichtige Dimension eines Programmes vernachlässigt; Ziele seien wichtige Treiber eines Programms. Er schlägt vor, Ziele eines Programmes nicht ausser Acht zu lassen, sie aber nicht unhinterfragt als Referenz zu übernehmen und gleichzeitig explizit nach nicht-intendierten Wirkungen zu fragen. Wenn es die Ressourcen erlauben, kann ein ‚goal-based‘ Ansatz gar mit einer ‚goal-free‘ Evaluation kombiniert werden. Zudem ist es wichtig, die Dynamik eines Programmes im Auge zu behalten und die Änderung von Zielen mit zu verfolgen. Sehr häufig kommt es vor, dass die offiziellen Programmziele zwar unverändert stehen bleiben, innerhalb des Programmes aber laufend neue operative Ziele entstehen und sich wandeln.

Eine genauere Betrachtung der Ziele und ihrer Evaluierung bestätigt die naheliegende Vermutung, dass je offener die Ziele formuliert wurden, desto offener auch die Untersuchung der Zielerreichung erfolgt. Operationalisierte Wirkungsziele mit quantitativen Zielvorgaben legen eine Überprüfung der Zielerreichung mit statistischen Verfahren natürlich nahe. Es besteht dann die Gefahr, dass diese den Blick einengen und unvorhersehbare Entwicklungen und Wirkungen nicht gesehen werden. Ergänzende qualitative Verfahren, wie in den Evaluationen E1 und E3 durchgeführt, können dieser Gefahr zumindest teilweise vorbeugen.

Adaption des Evaluationssystems

Evaluationen in hoch komplexen Interventionssystemen sollen sich den Entwicklungen des Interventionssystems anpassen, neu auftauchende Fragestellungen aufnehmen und mit situationsspezifisch passenden Methoden beantworten, um Lernprozesse flexibel unterstützen zu können. Patton (2011) hat dafür den Begriff des emergenten Evaluationsdesigns geprägt. Das Evaluationsdesign wird ebenso wie die untersuchte Intervention als emergent beschrieben, weil sich Evaluationsfragen nur begrenzt im Voraus bestimmen lassen, sondern während des Verlaufs – immer neu – herauskristallisieren und mit passender Methodik bearbeitet werden müssen:

Emergence as a core complexity construct tells us that innovators can't determine in advance what will happen, so evaluators can't determine in advance what to measure. We have to be watching for whatever emerges. We have to be prepared. We have to expect the unexpected. (Patton 2011, 126)

In den untersuchten Evaluationen finden sich nur vereinzelt Stellen mit Hinweisen auf derartige Anpassungen:

Der Aufwand für die Organisation von Gruppengesprächen mit älteren Zielgruppenmitgliedern erwies sich als zu aufwändig, da diese nicht mehr über Schneesportlager zugänglich sind. In der Folge wurde auf weitere Gruppengespräche verzichtet. (E3, 42)

Die Evaluation stellte sich zu Beginn des Projekts allgemeine Prozess-, Struktur- und Wirkungsfragen. Diese wurden im Verlauf des Projekts angepasst an weiterentwickelte Modelle und Kriterien zur Beurteilung organisationaler Veränderungsprozesse. (E2, 28)

In keiner der Evaluationen wird ein emergentes Evaluationsdesign eingesetzt, welches von Beginn an darauf ausgelegt ist, sich laufend den Entwicklungen des Interventionssystems anzupassen. Die untersuchten Evaluationen verwenden vielmehr starre Evaluationsdesigns, die möglichst konzeptgetreu durchgeführt werden. Dies dürfte mit dem Umstand verbunden sein, dass bereits zu Beginn der Evaluationen die zu messenden Wirkungen festgelegt sind.

Wirkungsmodelle für hoch komplexe Interventionen

Im Kontext hoch komplexer Interventionen, in welchen weder die Akteure noch die (Wirkungs-) Ziele und Massnahmen zu Beginn der Intervention bekannt sind, sondern sich erst im Laufe der Intervention entwickeln, ist es auch nur begrenzt möglich und sinnvoll, konkrete Wirkungsmodelle zu erarbeiten. Die Konstruktion von Wirkungsmodellen ist dann angebracht, wenn zu Beginn einer Intervention bereits realistische Vorstellungen über die Entwicklung und die Wirkungsweise vorliegen. In anderen Fällen kann die Konzeption von Wirkungsmodellen unnützlich, ja sogar kontraproduktiv sein, indem sie dazu verleitet, sich auf Geplantes zu versteifen und indem sie Flexibilität einschränkt (vgl. dazu Patton 2011). Diese ist insbesondere für Gesundheitsförderungsprogramme bedeutend, in denen Ziele und Massnahmen häufig erst in der Umsetzungsphase in Partizipationsprozessen mit den Zielgruppen erarbeitet und laufend weiter entwickelt werden. Für Funnell und Rogers steht ausser Frage, dass Programmtheorien und Wirkungsmodelle auch für hoch komplexe, emergente Interventionen genutzt werden können. Sie nennen fünf verschiedene Anwendungsarten (Funnell und Rogers 2011), aus denen folgende Aspekte besonders bedeutsam oder interessant erscheinen:

- Nutzung eines emergenten Verständnisses von Programmtheorie: Programmtheorien und Wirkungsmodelle werden nicht zu Beginn definiert und bleiben dann stabil, sondern sie entwickeln sich im Laufe der Zeit und müssen ständig adaptiert werden.
- Ausgehen von gewünschten Endpunkten (Wirkungszielen) und kontinuierliche Entwicklung der Programmtheorie resp. des Wirkungsmodells.
- Nutzung eines generischen Rahmenmodells, welches die Interventionslogik grob darstellt („theory of change“) und laufende Entwicklung ausdifferenzierter, kontextspezifischer Modelle („theory of actions“).
- Ergänzende Nutzung von systemischen Konzepten und Methoden (z.B. Netzwerkanalysen), um Interventionssysteme besser verstehen zu können.
- Nutzung von zyklischen Steuerungsmodellen, wie sie in der systemischen Managementberatung verwendet werden (Vester, 2007; Gomez und Probst 1999), da traditionelle Wirkungsmodelle kaum wechselseitige Interaktionen abbilden.
- Abbildung eines Programms in Form von Regeln oder Prinzipien (anstatt in Form eines Wirkungsmodells), welche für die Programmtheorie entscheidend erscheinen.

Die Evaluation E2 war am ehesten mit derart komplexen und ergebnisoffenen Interventionen konfrontiert, in denen weder Ziele noch Massnahmen der einzelnen Betriebe von Beginn an definiert waren. Vielmehr wurden in den Betrieben durch initiiierende Massnahmen wie Gesundheitsbefragungen oder Führungsschulungen Prozesse in Gang gesetzt, welche zu betriebspezifischen Zielen und Mass-

nahmen führten. Die Evaluation hat einerseits auf Basis eines allgemeinen Rahmenmodells generell als gültig erachtete Indikatoren gemessen, andererseits die Entwicklungsprozesse über explorative Verfahren verfolgt und deren Ergebnisse kontextspezifisch, also für jeden Betrieb separat, beurteilt.

8.6 Wirkungsbeurteilung

Die grosse Herausforderung der Wirkungsevaluation ist die Zuschreibung von beobachteten Veränderungen resp. vermuteten Wirkungen zu entsprechenden Interventionen (Attribution) sowie die Ermittlung des Beitrags einer Intervention an eine beobachtete Veränderung (Kontribution). Für viele Forschende sind Wirkungsaussagen nicht anders als über experimentelle oder quasi-experimentelle Studien machbar, da nur diese einen systematischen Vergleich mit einer Situation ohne Intervention ermöglichen (Stockmann, 2010; Caspari, 2009; Khandker et al., 2010).

„Echte“ Wirkungsevaluationen, die analysieren, inwieweit Veränderungen auf Seiten der Zielgruppe einer bestimmten Maßnahme eindeutig zugeschrieben (attribuiert) werden können, müssen der Frage nachgehen, was ohne Programm geschehen wäre, d.h. das Kontrafaktische (counterfactual) berücksichtigen. Dies erfordert härtere/rigorosere Methoden. (Caspari 2009, 188)

Im englischsprachigen Raum hat sich der Begriff ‚Impact Evaluation‘ zur Bezeichnung von Studien etabliert, die sich explizit mit der Frage der Wirksamkeit von Interventionen befassen. Üblicherweise basieren Impact Evaluationen im Kern auf einem quantitativen kontrafaktischen Vergleich zwischen einer Interventions- und einer Kontrollgruppe. Im Vordergrund stehen Interventionsstudien mit experimentellem oder quasiexperimentellem Design (Khandker et al., 2010). In der deutschen Sprache ist die Begrifflichkeit nicht eindeutig. Unter den Begriffen Wirkungs- und Wirksamkeitsevaluation werden teilweise auch nur Studien mit Kontrollgruppendesign verstanden, zum Teil ist die Begriffsverwendung aber auch breiter und umfasst jegliche evaluative Beurteilung von Wirkungen, unabhängig von der gewählten Methodik. Da der Begriff der Wirkungsevaluation hier in einem breiteren Sinne benutzt wird, verwende ich im Folgenden den Begriff Impact Evaluationen als Oberbegriff für quantitative Wirksamkeitsstudien auf Basis kontrafaktischer Vergleiche. Die Zusätze ‚rigoros‘ und ‚echt‘ zu Wirkungsevaluation erscheinen nicht angebracht, da andere Verfahren, die sich ebenfalls mit Wirkungsfragen beschäftigen, weder als nicht rigoros noch als unecht abqualifiziert werden sollen. Dazu Potvin: „Defining rigor as the systematic, lucid and critical appraisal of the relationship between an observer and the empirical world, rather than the blind application of some kind of methodological recipe, ...“ (Potvin & Richard, 2001, 231)

Obwohl quantitative Analysen bei Impact Evaluationen im Zentrum stehen, wird die Kombination von quantitativen mit qualitativen Methoden mehr und mehr als wertvoll erachtet, um ein umfassendes Bild der Wirksamkeit eines Programmes zu erhalten (Khandker et al., 2010; Stern et al., 2012). Qualitative Verfahren eignen sich beispielsweise dazu, Programmtheorien aufzubauen, potenzielle Nebenwirkungen zu identifizieren oder auch um Ergebnisse aus den quantitativen Vergleichen interpretieren zu helfen (Caspari, 2009). Die Kombination unterschiedlicher Methoden und insbesondere die ergänzende Nutzung qualitativer Methoden ist mittlerweile auch in der Impact Evaluation weitgehend unbestritten, wenn auch noch nicht überall in der Evaluationspraxis verbreitet (Khandker et al. 2010; Caspari, 2009; Caspari, 2012).

A good evaluation necessitates the combining of both qualitative and quantitative approaches to data collection and data analysis. Combining both approaches provides quantified results of program impacts as well as explanations of the process and intervening factors that yielded these outcomes. Combined approaches enrich interpretation or explanation (causality) of outcomes measured by the evaluation. (Ezemenari, Rudqvist, & Subbarao, 1999)

Angesichts der Komplexität sozialer Systeme wird die Kombination unterschiedlicher Methoden und insbesondere quantitativer und qualitativer Zugänge in der Evaluationsliteratur zunehmend gefordert und in der Praxis auch weitgehend umgesetzt, um zu einem angemessenen Bild, einem möglichst guten Verständnis und einer entsprechend differenzierten Beurteilung von Interventionen zu gelangen (Mühlhauser, Lenz, & Meyer, 2012; DeGEval, 2013)¹³². Dabei wird die Vorrangstellung einzelner Methoden mit Verweis auf unterschiedliche Potenziale und Grenzen abgelehnt:

Because each method, or even multiple methods, involves its own strengths and weaknesses, there realistically is no one best method that can universally apply to every evaluation situation. (Chen, 1990, 27)

Die grosse Herausforderung liegt allerdings in der Art und Weise der Integration der verschiedenen Methoden und Ergebnisse, damit tatsächlich ein reichhaltigeres Bild entsteht und nicht einfach ein Nebeneinander von sich zum Teil ergänzenden, zum Teil sich aber auch potenziell widersprechenden Informationen. Letztlich ist es diese Verknüpfung der Ergebnisse von quantitativen und qualitativen Methoden, welche eine mixed-methods Studie ausmachen und dazu führen, dass das Gesamtergebnis mehr ist als die Summe der Ergebnisse aus den einzelnen Teilstudien (Bryman, 2007)¹³³.

Die wichtigsten Forschungsinstrumente sind dabei nicht, wie oft angenommen wird, die Beherrschung des Kanons von mechanistisch anwendbaren Methoden zur Datenerhebung und –auswertung, sondern die umfassende Erfahrung mit der konkreten Problemstellung, die Kenntnis der ungeklärten Fragestellungen, die Fähigkeit, möglichst viele Indizien für unterschiedliche Theoreme bzw. Theorien zu finden, die Kompetenz, diese zu einem konsistenten Gesamtbild zusammen zu stellen, und die Bereitschaft, die Resultate laufend in Frage zu stellen bzw. der Kritik durch Dritte auszusetzen. (Uhl, 2009, 35)

In den untersuchten Evaluationen wird die Kombination von quantitativen und qualitativen methodischen Zugängen mehrheitlich dazu genutzt, um ergänzend zu den Wirkungsmessungen Hinweise zur Wirkungsweise zu erhalten (E1, E3, E9). Mit den qualitativen Zugängen werden einerseits die Messergebnisse interpretiert und plausibilisiert, andererseits Verbesserungspotentiale herausgearbeitet. In zwei anderen Evaluationen (E4, E7) ist die Wirkungsbeurteilung klar qualitativ ausgerichtet, mit punktueller Nutzung von Zahlen aus der Nutzerstatistik. In der Evaluation E2 leistet der qualitative Teil der Evaluation neben der quantitativen Wirkungsanalyse einen bedeutenden Teil der Wirkungsbeurteilung. Mittels qualitativer Methodik werden primär Wirkungen auf der Organisationsebene beurteilt, während die quantitativen Zugänge dazu genutzt werden, um die Wirksamkeit der Interventionen auf der Ebene der individuellen Gesundheit zu messen. In den anderen Evaluationen werden entweder nur quantitative (E5, E8) oder nur qualitative methodische Zugänge (E6) zur Wirkungsbeurteilung eingesetzt.

Impact Evaluationen sind in der Gesundheitsförderung und Prävention nur schwer durchführbar, und zwar aus folgenden Gründen:

- Bei settingorientierten Interventionen scheitert die Durchführung oft daran, dass nicht ausreichend viele Interventions- und Kontrollsettings zur Verfügung stehen oder eine entsprechende Studie vom Aufwand her nicht realisierbar ist.
- Ausserhalb des Labors, in realen Interventionssituationen wird die Qualität der Studien durch vielfältige Kompromisse und Verzerrungen beeinträchtigt

¹³² Mühlhauser et al. (2012) haben fünf neuere massgebliche Handbücher und Leitfäden zu Evaluationsstudien und systematischen Reviews untersucht und zeigen, dass alle die Berücksichtigung sowohl quantitativer wie auch qualitativer Methoden für die Entwicklung und Evaluation komplexer Methoden vorsehen.

¹³³ Tashakkori und Creswell (2007) sprechen von „quasi-mixed studies“, wenn die Ergebnisse unterschiedlicher Methoden nicht wirklich miteinander verbunden werden.

- Die Dynamik und Unvorhersehbarkeit komplexer sozialer Systeme führt in Kombination mit der Langfristigkeit der Wirkungshorizonte dazu, dass die Studienbedingungen in der Regel nicht in allen Settings und über die gesamte Untersuchungszeit konstant gehalten werden können.
- Oft ist eine vollständige Randomisierung nicht möglich (vgl. E1) und eine Blindung macht in der Regel keinen Sinn.

Auch wenn Evaluationen als experimentelle Studien ausgewiesen werden, offenbaren sich bei genauem Hinsehen häufig die oben erwähnten Schwächen, die die eigentlichen Vorteile des Designs, nämlich die eindeutige Wirkungszuschreibung, gleich wieder zunichtemachen. So zum Beispiel in E1, wo die versprochene Randomisierung an Ansprüchen im Feld scheitert (andere eindrucksvolle Beispiele in Eisner & Ribeaud, 2007):

Neun Berufsschulen nahmen an der Studie teil, 3 Schulen jedoch nur unter der Voraussetzung, dass alle an der Studie teilnehmenden Klassen auch von der Intervention profitieren könnten. Zwei Schulen nahmen nur unter der Voraussetzung teil, dass der Klassenverband nur durch die Befragungszeit mit den Fragebögen belastet werden würde. Die Klassen dieser fünf Schulen wurden gesamthaft entweder zur Interventions- oder Kontrollgruppe zugeordnet. (E1, 16)

Neben den Unzulänglichkeiten aufgrund praktischer Hindernisse werden (quasi-) experimentelle Studien auch wegen ihrer begrenzten Aussagekraft kritisiert. Dies sind einige Gründe dafür:

- Komplexe Interventionen unterscheiden sich erfahrungsgemäss auch bei identischem Konzept in der Umsetzung deutlich. Zudem wird in experimentellen Designs der Einfluss von Kontextfaktoren durch die Randomisierung weitgehend neutralisiert, obwohl die spezifischen Kontextfaktoren höchst wirkungsrelevant sind. So wird letztlich die Wirkung einer durchschnittlichen Intervention in einem durchschnittlichen Kontext gemessen, was bei der vorliegenden Komplexität dem Einzelfall kaum gerecht wird. Es ist deshalb nicht erstaunlich, dass in vielen Bereichen kaum Unterschiede zwischen Interventions- und Kontrollgruppen gefunden werden (vgl. zum Beispiel Eisner & Ribeaud, 2007).
- Das isolierte Herausgreifen einzelner unabhängiger und abhängiger Variablen aus komplexen Ursache-Wirkungsgefügen mit vielfältigen Wechselwirkungen und Rückkoppelungen stellt eine starke Komplexitätsreduktion dar. Eine darauf aufbauende Wirkungsevaluation kann nur über den gewählten Ausschnitt Aussagen machen.
- In den Berechnungsmodellen kann nur kontrolliert werden, was an Einflüssen bekannt ist. Unbekannte, aber potenziell hoch wirksame Einflüsse werden dadurch vernachlässigt. Das komplexe Zusammenspiel von Einflussfaktoren (systemintern und übergreifend) lässt sich in den stark vereinfachenden Berechnungsmodellen nicht abbilden, berücksichtigte und tatsächliche Wirkungsfaktoren können stark auseinanderklaffen.
- Oft lässt sich nicht eindeutig trennen, was Ursache und was Wirkung ist (Bewegungsmangel kann zu Übergewicht führen, Übergewicht kann aber auch den Bewegungsmangel fördern). Den vielfältigen Rückkoppelungen in komplexen Systemen wird in der Regel nicht Rechnung getragen.
- Bei Untergruppenvergleichen in RCT's, zum Beispiel wenn innerhalb der Interventionsgruppe intensive Umsetzungen mit weniger intensiven verglichen werden, ist die Randomisierung ausgeschaltet und die Ergebnisse sind wegen wahrscheinlicher Verzerrungen mit Vorsicht zu genießen.
- Oftmals haben die untersuchten Interventionen gar nicht das Potenzial, eine messbare Wirkung zu erzeugen, da sie im Vergleich zu den vielen anderen Einflussfaktoren nur einen sehr bescheidenen Beitrag an eine Veränderung leisten können.
- Die Untersuchung einzelner Teilprojekte mittels randomisierter kontrollierter Studien ist zwar möglich, vermag aber häufig keine Wirkungen in den langfristigen Zielindikatoren nachzuweisen, da diese nur über das Zusammenspiel der verschiedenen Massnahmen und zusammen mit bedingenden Einflussfaktoren zustande kommen.

Uhl (2012) bringt diese Kritikpunkte mit folgendem Zitat auf den Punkt:

Aber selbst wenn ein ideales Experiment durchführbar ist, ist der direkte Vergleich zwischen Experimental- und Kontrollgruppe meist wenig aufschlussreich, weil sich die relevanten Prozesse in komplexen Zusammenhängen mit vielen Beteiligten und nicht berücksichtigten Variablen manifestieren. (Uhl, 2012, 67)

In Bezug auf die begrenzten Möglichkeiten, komplexe Phänomene zu verstehen und Wirkungen nachzuweisen, fordert er zu Bescheidenheit und Ambiguitätstoleranz auf:

Es muss uns möglich sein zuzugeben, dass wir Phänomene ganz und gar nicht verstehen und dass wir in manchen Situationen große Zweifel haben, ohne unsere Kompetenz als Wissenschaftler und Evaluatoren generell in Frage zu stellen. (Uhl, 2012, 76)

Bei der Entscheidung, in welchen Fällen Impact Evaluationen sinnvoll sind, kann die Unterscheidung zwischen einfachen, komplizierten und komplexen Interventionen (Funnell und Rogers 2011, 90) eine wichtige Orientierungshilfe sein, die sich auf Glouberman und Zimmerman (Glouberman & Zimmerman, 2002) abstützt (vgl. Tab. 2 dieser Arbeit). Wenn Interventionen der Gesundheitsförderung und Prävention z.B. sog. ‚einfache‘ Interventionselemente enthalten (wie z.B. Kurse zur Bewegungsförderung in Gesundheitsförderungsprogrammen für ältere Menschen), können randomisierte kontrollierte Studien dazu genutzt werden, diese Elemente auf ihre Wirksamkeit hin zu überprüfen. Dies vor allem, wenn es sich um verhaltensorientierte Massnahmen handelt, die sich direkt an Individuen richten. Da allerdings auch scheinbar ‚einfache‘ Interventionen durch ihre soziale Natur gegeben komplex sind, sind die Ergebnisse grundsätzlich mit Vorsicht zu geniessen und die Wirkungsnachweise sollten über theoretische Fundierung und über die Rekonstruktion der Wirkungsmechanismen plausibilisiert werden. Ursache-Wirkungszusammenhänge, die mittels (quasi-) experimentellen Studien untersucht werden, müssen auf Programmtheorien basieren, die theoretisch begründet sind und von den Akteuren vor Ort als adäquat eingeschätzt werden. Wenn angenommen werden muss, dass eine Programmtheorie für den konkreten untersuchten Kontext nicht relevant ist, macht eine darauf aufbauende randomisierte kontrollierte Studie keinen Sinn – und wird vermutlich auch keine Wirkungen nachweisen können. Theoretische Fundierung, Kontextanalyse und die Abstimmung dieser beiden bilden eine gute Basis für die Durchführung einer Impact Evaluation.

In den untersuchten Evaluationen bauen die Wirkungsbeurteilungen und entsprechende klare Wirkungsaussagen auf unterschiedlichen Arten von Vergleichen auf und beschränken sich bei weitem nicht auf quantitative Kontrollgruppenvergleiche. Es haben sich Alternativen zu experimentellen und quasi-experimentellen Studien gezeigt, die ebenfalls Wirkungsnachweise proklamieren. Kontrafaktische Vergleiche auf Basis von Fallstudien, detaillierte Rekonstruktionen von Wirkungsmechanismen, das Abstützen auf Wirkungsurteile von Beteiligten und vielfältige Kombinationen dieser Strategien. Die Zuschreibung von beobachteten Veränderungen zu Interventionen wird dabei weitgehend als grosse Herausforderung anerkannt und entsprechend fallen viele Wirkungsaussagen zurückhaltend aus. Trotzdem werden in den meisten Evaluationen auch ganz klare Wirkungsaussagen gemacht. Weil es sich nicht erkennbar auf methodische oder andere Unterschiede zurückführen lässt, wann Wirkungsaussagen als gesichert dargestellt werden und wann nicht, muss vermutet werden, dass dies stark von den Evaluatorinnen und Evaluatoren selber abhängt; die einen sind freizügiger mit Wirkungszuschreibungen, andere sind vorsichtiger.

Um Unsicherheit in den Wirkungszuschreibungen zu reduzieren, werden Wirkungsaussagen auf unterschiedliche Weise plausibilisiert. Unterschiedliche methodische Verfahren können grundsätzlich mehr oder weniger plausible Hinweise auf Wirkungszusammenhänge und Wirkungen liefern. Es kann zum grossen Teil nicht beurteilt werden, inwiefern die angeführten Plausibilisierungen gezielt (aus)gesucht wurden, um eine Position zu bestätigen.

Plausibilisierung von Wirkungsaussagen

Die Schwierigkeit, Wirkungen in komplexen Zusammenhängen nachzuweisen, bewahrt nicht davor, Entscheidungen treffen und handeln zu müssen:

Wir können nicht umhin, Entscheidungen zu treffen und diese fallen besser aus, wenn wir die Auswirkungen unseres Handelns möglichst präzise antizipieren können; also wenn wir auf gut fundierte Ursache-Wirkungs-Hypothesen zurückgreifen können. Aber selbst wenn unsere Prognosen bloß auf schwachen Beinen stehen, können wir die meisten Entscheidungen weder bis zur Klärung aufschieben noch gänzlich unterlassen. (Uhl 2012, 66)

Wie deutlich aufgezeigt werden konnte, ist der Spielraum der Evaluation, wie sie zu Wirkungsaussagen kommt und in welcher Deutlichkeit sie formuliert werden, gross. Es stellt sich die Frage, wie plausibel und wie gut nachvollziehbar derartige Wirkungsaussagen sind. An vielen Beispielen konnte gezeigt werden, wie sehr Wirkungskonstruktionen von Verkürzungen, Verzerrungen, Verfälschungen, Unsicherheiten etc. begleitet sind. Auch die kritischen Beiträge in der Literatur dazu stützen diese prinzipielle Unzulänglichkeit von ‚eindeutigen‘ Wirkungsaussagen (z.B. Uhl 2009; Uhl 2012; Eisner und Ribeaud 2007).

Wirkungsevaluationen kommt in komplexen Zusammenhängen also nicht die Aufgabe zu, Wirkungen zu ‚beweisen‘, sondern möglichst plausible Erkenntnisse über Wirkungspotenziale, Wirkungszusammenhänge und eingetretene Wirkungen zu liefern, um damit Entscheide hinsichtlich der Ausgestaltung und Durchführung von Interventionen zu erleichtern. Wirkungsevaluation soll helfen, komplexe Wirkungsmechanismen besser zu verstehen, dadurch adäquater zu intervenieren und Fehler, die aus der Komplexität der untersuchten Systeme herrühren, möglichst zu vermeiden, resp. adäquat auf komplexe Dynamiken zu reagieren.

Kausalitäten können nicht beobachtet werden, es sind immer soziale Konstruktionen und diese sind – wenn auch statistisch erdrückend - immer probabilistisch. Auch in Impact Evaluationen werden verschiedene Strategien angewendet, um die Validität gefundener Kausalitäten zu überprüfen und die Evidenzbasis von Wirksamkeitsevaluationen zu stärken (Chen 1990, 146 f). Umso bedeutender ist eine derartige Plausibilisierung in komplexen Zusammenhängen. Die untersuchten Evaluationen haben eine grosse Vielzahl unterschiedlicher Plausibilisierungen zur Absicherung der Erkenntnisse gezeigt (vgl. Kapitel 7.4), und sie liefern interessante Ansatzpunkte für eine Plausibilisierungsstrategie. Die wichtigsten sollen hier noch einmal zusammengefasst dargestellt werden. Es sind Plausibilisierungen durch:

- sozialwissenschaftliche Theorien,
- Ergebnisse anderer Studien,
- Einsatz verschiedener Erhebungs- und Auswertungsmethoden,
- Berücksichtigung unterschiedlicher Perspektiven,
- nachträgliche zusätzliche Erhebungen,
- Bewertung der Qualität der Programmkonzeption,
- Expertenmeinungen,
- und Überprüfung an der Praxis.

Die drei Strategien der Wirkungszuschreibung, die in der empirischen Analyse identifiziert werden konnten, bieten einen weiteren Ansatzpunkt für eine Plausibilisierungsstrategie. In den untersuchten Evaluationen wurden verschiedene Kombinationen dieser Strategien angewendet:

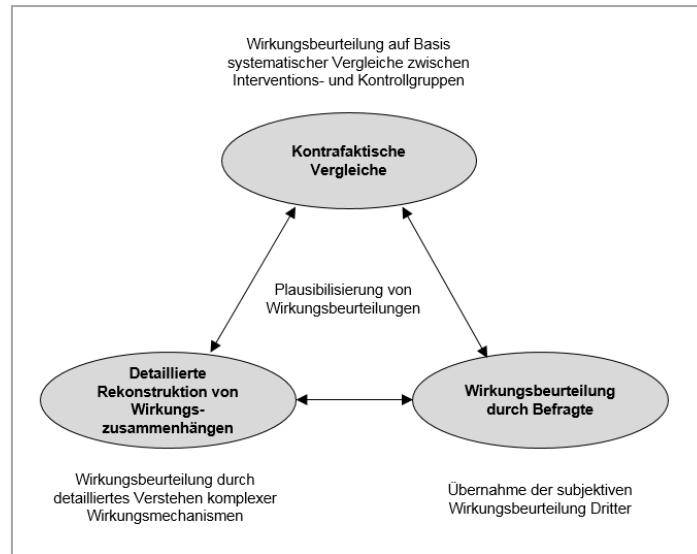


Abb. 29 Nutzung der drei Strategien der Wirkungszuschreibung für Plausibilisierungen

Allerdings werden Hinweise zur Plausibilisierung von Erkenntnissen nur punktuell angegeben, und es wird nicht ersichtlich, wie systematisch Plausibilisierungen durchgeführt werden. Auch in der Evaluationsliteratur ist das Thema der Plausibilisierung wenig systematisch aufgearbeitet; es scheint aber in einzelnen Anwendungen immer wieder auf, so z.B. in der Literatur rund um ‚mixed methods‘ und Triangulationen oder auch in der ‚Contribution Analysis‘. Bei der ‚Contribution Analysis‘ handelt es sich um einen innovativen Ansatz zur Evaluation komplexer Programme (Mayne, 2001; Mayne, 2011). Dieser liefert ein prozessorientiertes Evaluationsmodell, welches dazu anleitet, entlang von Wirkungsvermutungen Schritt für Schritt bestehende Evidenzen zu beurteilen, durch Evaluationsmassnahmen neue Erkenntnisse zu generieren, alles zu einer fundierten Gesamtbeurteilung zu verdichten und so den Beitrag einer Intervention zur beobachteten Veränderung zu schätzen. Methodisch macht die ‚Contribution Analysis‘ keine Vorgaben. Lücken in der Argumentationskette der ‚contribution story‘ sollen mit angemessenen Forschungsmethoden gefüllt werden; dabei kann es sich sowohl um hypothesenprüfende quantitative als auch um explorative, qualitative Zugänge handeln. Um den Beitrag des eigenen Programms abschätzen zu können, müssen konsequenterweise auch externe Einflussfaktoren identifiziert und deren Beiträge abgeschätzt werden. Auch hierzu wird auf bestehende Evidenzen, auf Daten Dritter, auf Experteneinschätzungen und gegebenenfalls auf eigene Erhebungen und Analysen zurückgegriffen. Dieser Evaluationsrahmen verbindet zwei Strategien, die für Wirkungsevaluationen in komplexen Zusammenhängen bedeutsam sind: die Verwendung unterschiedlicher Methoden zur Untersuchung unterschiedlicher Wirkungsaspekte und die Idee der Plausibilisierung.

Generalisierung von Erkenntnissen in komplexen Zusammenhängen

Die Generalisierung von Wirkungsnachweisen und die Übertragung auf andere Kontexte stellt die Evaluation der Gesundheitsförderung und Prävention vor zusätzliche Herausforderungen. Wenn davon ausgegangen wird, dass sich Wirkungen in komplexen sozialen Systemen in selbstorganisierten Prozessen entfalten und dass jedes soziale System und seine System-Umweltbeziehungen einzigartig sind, haben Wirkungsaussagen zunächst nur für jenen Kontext Gültigkeit, in welchem sie entstanden sind. Wenn sich ein Gesundheitsförderungsprojekt in einer Schule bewährt hat, dann kann nicht automatisch darauf geschlossen werden, dass sich dasselbe Projekt auch in der Nachbarschule bewährt. Und selbst wenn sich das Projekt in zehn Schulen bewährt hat, kann es in der elften scheitern. In den untersuchten Evaluationen wird wiederum sehr unterschiedlich mit Verallgemeinerungen um-

gegangen, und es haben sich unterschiedliche Zugänge und Spannungsfelder gezeigt. Verallgemeinernde Aussagen werden einerseits durch möglichst grosse, repräsentative Stichproben begründet, andererseits durch eine breite Abdeckung und tiefgehende Analyse unterschiedlicher Fälle. Die einen Evaluationen suchen differenziertes kontextspezifisches Wissen, um Aussagen über Wirkungen und Wirkungsbedingungen absichern zu können und um praxistaugliche Erkenntnisse zu generieren. Wirkungsnachweise, die unter kontrollierten Bedingungen zustande kommen, seien für die Praxis mit ihren unkontrollierbaren Bedingungen kaum aussagekräftig. Andere wiederum beanspruchen für ihre in Kontrollstudien nachgewiesenen Wirkungen breite Gültigkeit – wenn auch die Formulierung hin und wieder vorsichtig bleibt:

Diese hohe Beteiligung hat zur Folge, dass unsere Daten belastbar und generalisierbar auf andere Kindergärten und Regionen Baden-Württembergs scheinen. (E9, 97)

Die Wirkungsentfaltung ist einerseits system- und kontextspezifisch, andererseits werden Ergebnisse gewünscht, die kontextübergreifend gültig und verwendbar sind. Für die Wirkungsmessung soll die Intervention möglichst standardisiert werden, obwohl dies einer höheren Wirksamkeit entgegenstehen kann:

Die moderne Interventionsforschung hat wiederholt gezeigt, dass gerade die Flexibilität und lokale Anpassbarkeit von Interventionen zu mehr Erfolg, Motivation und Selbsterhaltung von Interventionen und Projekten führt. Da die Fachkräfte in der Standard-Intervention jedoch strenge Vorgaben zur Durchführung des Programms haben, kann eine Anpassung des Programms „Komm mit in das gesunde Boot“ an das lokale Umfeld nur in sehr begrenztem Maße stattfinden (E9, 90)

An diesem Beispiel zeigt sich das Spannungsfeld zwischen kontextspezifischer Wirkungsentfaltung und den Ansprüchen an das Forschungsdesign und Generalisierung.

Die ‚Realist Evaluation‘ (Pawson & Tilley, 1997) stellt System-Kontextbeziehungen konsequent ins Zentrum der Betrachtung. Dieser in der Komplexitätsdiskussion vielbeachtete Ansatz geht davon aus, dass sich Wirkung nicht unabhängig vom Kontext entfalten kann, sondern dass Kontextbedingungen entscheidender Teil jeder Wirkungsentstehung sind:

Whether mechanisms ‘fire’ (operate at a particular moment) depends on the context. If I am standing on land when I release the tennis ball, gravity will draw it to the ground. If I happen to be underwater, a different mechanism (buoyancy) will cause the ball to float. In fact both mechanisms operate in both contexts, but gravity is the stronger force in air and buoyancy the stronger force under water. The two mechanisms compete, and the context determines which mechanism ‘wins’. (Westhorp, 2014, 6)

Entsprechend untersucht die ‚Realist Evaluation‘, wie sich Interventionen in unterschiedlichen Kontexten unterschiedlich entwickeln. Sie fragt nicht danach, ob eine Intervention wirkt oder nicht, sondern wie, bei wem und unter welchen Umständen (Pawson & Tilley, 1997). Die Realist evaluation bleibt allerdings nicht bei Aussagen über Einzelfälle stehen, sondern reduziert Komplexität wieder, indem sie typische Muster herausarbeitet, sogenannte Kontext-Mechanismus-Outcome-Konfigurationen. Auf diese Weise lassen sich Sachverhalte zunächst kontextspezifisch untersuchen und dann im Sinne von „unter diesen Umständen führen diese Mechanismen zu diesen Wirkungen“ generalisieren.

Wirkungsevaluation bewegt sich im Spannungsfeld zwischen generalisierten und kontextspezifischen Ursache-Wirkungszusammenhängen. Am Anfang steht meist generalisiertes Wissen in Form von wissenschaftlicher Evidenz. Während hypothesenprüfende Verfahren mehrheitlich auf der Ebene generalisierter Aussagen bleiben, tauchen explorative Verfahren in Kontextspezifisches ab, um die Erkenntnisse aus diesen Analysen in einem zweiten Schritt mehr oder weniger stark wieder zu generalisierten Aussagen zu verdichten. Der Zwischenschritt der kontextspezifischen Analyse führt einerseits zu breiter abgestützten Ergebnissen, andererseits zu wertvollen Einsichten für die Steuerung und

Optimierung der untersuchten Interventionen. Auch in Impact Evaluationen wird der Mehrwert durch ergänzende kontextspezifische Analysen zunehmend thematisiert:

The underlying problem here is the complexity of the context. By this I mean that the influence of X on Y is confounded by factors Z that are impossible fully to enumerate, of uncertain or highly variable value, difficult to separate, and/or impossible fully to control. Additional complexity arises if the nature and value of X and/or Y is also uncertain. (...) The point is not to secure universal agreement, but to be as clear and precise as possible about what can reasonably be expected in a given context. (Copestake, 2014, 417)

In der Generalisierung wird Komplexität reduziert, indem auf ausgewählte kontextübergreifende Elemente fokussiert wird und kontextspezifische Faktoren ausgeblendet werden. Dies trifft auch für argumentative Verallgemeinerungen zu.

8.7 Synthese

Unterschiedlicher Umgang mit Komplexität

Die Analyse hat gezeigt, dass die Evaluationen sehr unterschiedlich mit der Komplexität der untersuchten Systeme, konkret mit Interaktionsvielfalt, komplexen Wirkungsmechanismen, Selbstorganisation und Unvorhersehbarkeit umgehen. Es zeigen sich für diese Aspekte unterschiedliche Herausforderungen, Spannungsfelder und Strategien. Die Strategien lassen sich danach unterscheiden, ob die Evaluation tendenziell Komplexität ausblendet (Black Box Evaluationen) oder ob sie Komplexität bewusst berücksichtigt (komplexitätssensible Evaluation).

	Interaktionsvielfalt	Komplexe Wirkungsmechanismen	Selbstorganisation	Unvorhersehbarkeit	Wirkungsbeurteilung
Zentrale Herausforderungen	Identifikation relevanter Wirkungsfaktoren	Verstehen der Wirkungsmechanismen	Berücksichtigen von Prozessen der Selbstorganisation	Erkennen nicht-intendierter Entwicklungen	Attribution/ Kontribution
Spannungsfelder	Ausweitung - Reduktion der Faktoren	Einfache – komplexe Mechanismen	Determinismus - Selbstorganisation	Planbarkeit – Unvorhersehbarkeit	Potenzial und Grenzen von Wirkungsevaluation
Strategien von Black Box Evaluationen	Deduktive Grenz- ziehung	Einfache Modellie- rungen	Externe Wirkungs- konstruktion	Hypothesen- prüfung	Kontrafaktischer Vergleich (quantitativ) ⇒ Messen
Strategien komplexitätssensibler Evaluationen	Induktive Grenz- ziehung	Komplexe Model- lierungen	Eigenlogik der untersuchten Systeme	Exploration	Plausibilisierung durch detaillierte Rekonstruktion ⇒ Verstehen

Tab. 20 Zentrale Aspekte der Wirkungsevaluation im Umgang mit Komplexität

Innerhalb dieses Rahmens lassen sich in den untersuchten Evaluationen unterschiedliche Ansätze der Reduktion und Ausweitung von Komplexität erkennen. Die folgende Zusammenstellung basiert auf dem Handlungsspielraum zwischen Rekonstruktionen geringer Komplexität und Rekonstruktionen hoher Komplexität, der aus dem empirischen Material herausgearbeitet wurde (vgl. Tab. 19, Seite 158):

	Ansätze zur Reduktion von Komplexität	Ansätze zur Ausweitung von Komplexität
Grenzziehung	<ul style="list-style-type: none"> - Deduktive Grenzziehung und starre Handhabung der Systemgrenzen. - Ausblenden von neu auftauchenden Faktoren. 	<ul style="list-style-type: none"> - Induktive Grenzziehung und flexible Handhabung der Systemgrenzen. - Integration neu auftauchender Faktoren.
Beurteilungsreferenz	<ul style="list-style-type: none"> - Einschränkung auf vordefinierte Wirkungsziele als Referenz für die Wirkungsbeurteilung. - Überprüfen externer Wirkungskonstruktion. - Abstützen auf stark generalisierte Wirkungsfaktoren. 	<ul style="list-style-type: none"> - Kritische Reflexion der vorgegebenen Wirkungsziele. Berücksichtigung von nicht-intendierten Nebenwirkungen. - Herausarbeiten der Eigenlogik und der Selbstorganisationsprozesse der untersuchten Systeme. - Suche nach kontext- und zielgruppenspezifischen Wirkungsfaktoren.
Evaluationsdesign	<ul style="list-style-type: none"> - Hypothesenprüfender Zugang. - Starres Evaluationsdesign. Beschränkung auf eingangs festgelegte Fragestellungen, Perspektiven und Methoden. - Distanzierte Beobachtung des Untersuchungsgegenstandes. - Fokussierung auf wenige Datenquellen und Perspektiven. - Vergleich weniger Momentaufnahmen. - Fokussierung auf eine oder wenige Erhebungsmethoden. - Beschränkung auf die vordefinierten Variablen. - Vermeidung kritischer Reflexion der eingesetzten Methoden. 	<ul style="list-style-type: none"> - Explorativer Zugang. - Emergentes Evaluationsdesign. Anpassung der Fragestellungen, Perspektiven und Methoden an die Entwicklung von Intervention und Evaluation. - Herstellen von Nähe zum Untersuchungsgegenstand. - Berücksichtigung einer Vielfalt von Datenquellen und Perspektiven. - Fokussierung auf Prozesse. - Nutzen und Kombination verschiedener Erhebungsmethoden. - Einfangen von kontext- und zielgruppenspezifischen Erklärungen, von Hintergründen und Zwischentönen. - Kritische Reflexion des Potenzials und der Grenzen der verwendeten Methoden.
Auswertung	<ul style="list-style-type: none"> - Zusammenfassende, verallgemeinernde Auswertungen. - Vergleiche weniger Variablen und Zeitpunkte. - Orientierung an einfachen, stark verallgemeinernden Wirkungsmodellen mit linearer Anordnung der Wirkungszusammenhänge. 	<ul style="list-style-type: none"> - Differenzierte Auswertungen, kontext- und zielgruppenspezifische Analysen - Differenzierte Vergleiche und Analyse dynamischer Prozesse. - Induktive Entwicklung von differenzierten, kontextspezifischen Wirkungsmodellen. Detaillierte Rekonstruktion von komplexen Wirkungszusammenhängen.
Wirkungsbeurteilung	<ul style="list-style-type: none"> - Abstützen der Wirkungsurteile auf quantitativ angelegte kontrafaktische Vergleiche. - Unkritische Proklamierung eindeutiger Wirkungsnachweise. - Unkritische Verallgemeinerungen der Wirkungsaussagen. - Starke Verdichtung der Erkenntnisse. 	<ul style="list-style-type: none"> - Abstützen der Wirkungsurteile auf explorativ ausgerichtete, detaillierte Rekonstruktionen von Wirkungszusammenhängen. - Selbstkritische Reflexion der Wirkungszuschreibungen und Nutzung verschiedener Formen der Plausibilisierung. - Kontextspezifische Wirkungsbeurteilungen und vorsichtige Generalisierungen. - Belassen von Vielfalt.

Tab. 21 Strategien der Reduktion und Ausweitung von Komplexität

Die untersuchten Evaluationen nutzen und kombinieren die verschiedenen Strategien der Reduktion und Ausweitung von Komplexität vielfältig. Auf einer sehr groben Ebene lassen sich die verschiedenen Formen zu einer Typologie des Umgangs mit Komplexität verdichten.

Typologie zum Umgang der Evaluationen mit Komplexität

In der vorliegenden Studie haben sich u.a. mit Blick auf den Prozess der Wirkungskonstruktion (Abb. 25) zusammenfassend vier Typen des Umgangs mit Komplexität herauskristallisiert. Diese Typologie basiert stark vereinfachend darauf, wie viel Komplexität die Evaluationen zu welchem Zeitpunkt berücksichtigen: entsprechend können ein linienförmiger, ein linsenförmiger, ein zitronenförmiger und ein kegelförmiger Typ unterschieden werden:

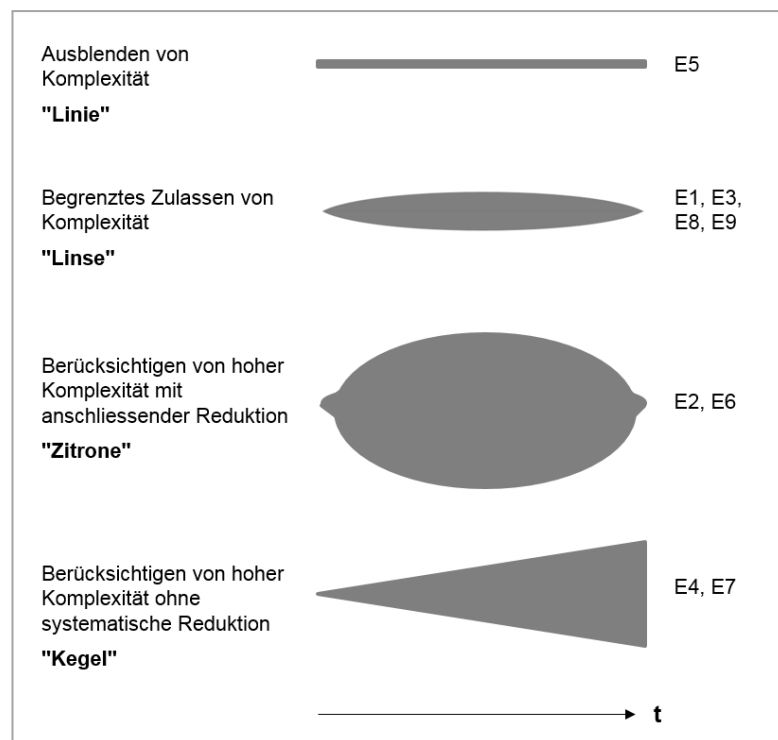


Abb. 30 Typen des Umgangs mit Komplexität

Die vier Typen zeichnen sich durch unterschiedliche Merkmale aus und es lassen sich je spezifische Vor- und Nachteile benennen. Nachfolgend werden die vier Typen erläutert:

Ausblenden von Komplexität (Linie)

Evaluationen dieses Typs beschränken sich darauf, Zusammenhänge zwischen ganz wenigen isolierten Faktoren zu untersuchen. Im Extremfall gehen lediglich die Intervention als Ganzes als unabhängige Variable und ein einziger Wirkungsparameter als abhängige Variable in das Forschungsdesign ein. Es wird von einem linearen Wirkungszusammenhang ausgegangen, der mittels hypothesenprüfenden Verfahren untersucht wird. In der Untersuchung werden keine explorativen Verfahren angewendet, um zusätzliche Einflussfaktoren zu suchen oder komplexere Wirkungszusammenhänge zu analysieren. Die untersuchten Interventionen werden nicht als Systeme rekonstruiert. Wirkungszusammenhänge interessieren in diesem Evaluationstyp nicht, die ‚Black Box‘ wird nicht geöffnet.

Vorteile dieses Evaluationstyps sind die Übersichtlichkeit des Designs aufgrund einer stark begrenzten Anzahl berücksichtigter Faktoren, klare Wirkungsaussagen sowie einfach verstehbare Ergebnisse. Ein

wesentlicher Nachteil ist, dass kein Wissen verfügbar ist über Wirkungsweisen, über Wirkungen und Wirkungspotenziale einzelner Massnahmen, über kontextspezifische Wirkungspotenziale und über Nebenwirkungen. Wenn in einer Evaluation keine Wirkung nachgewiesen werden kann, bleibt unklar, *weshalb* sich keine Wirkung hat entfalten können. Zudem fehlen bei Evaluationen dieses Typs viele Formen der Plausibilisierung, die von anderen Evaluationen genutzt werden (Rekonstruktion von Wirkungszusammenhängen, Perspektiven- und Methodentriangulation etc.). Dies erhöht das Risiko für Fehlinterpretationen.

Begrenztes Zulassen von Komplexität (Linse)

Dieser Typ zeichnet sich im Kern ebenfalls durch ein einfaches Ursache-Wirkungsverständnis aus, und im Zentrum der Evaluation steht auch eine hypothesenprüfende Untersuchung weniger Wirkungszusammenhänge. Im Unterschied zum linienförmigen Typ sind aber entweder die den Berechnungen zugrundeliegenden Modelle komplexer oder aber die quantitativen Untersuchungen werden durch explorative Elemente ergänzt, um auf diese Weise Hinweise auf Ursache-Wirkungszusammenhänge zu sammeln. Die explorativen Elemente begründen in diesem Typ allerdings keine eigenständigen Wirkungsanalysen, sondern dienen lediglich dazu, die gemessenen Wirkungen zu interpretieren und Hinweise auf Verbesserungspotenziale zu geben. Der Deckel der ‚Black Box‘ wird nur leicht angehoben, um etwas Komplexität zuzulassen.

Vorteile dieses Typs sind ebenfalls klare Wirkungsaussagen und einfach verständliche Ergebnisse. Im Vergleich zum linienförmigen Typ gibt es hier einen Mehrwert durch die Öffnung des Blicks: einzelne Hinweise auf Wirkungsmechanismen, auf förderliche und hinderliche Kontextfaktoren oder auf Nebenwirkungen können einen zusätzlichen Erkenntnisgewinn liefern und bei der Interpretation und Plausibilisierung der Ergebnisse helfen. Im Vergleich zu den folgenden beiden Typen sind diese Erkenntnisse aber sehr begrenzt.

Berücksichtigen von hoher Komplexität mit anschliessender Reduktion (Zitrone)

Evaluationen, die diesem Typ zugeordnet werden, öffnen den Blick rasch für komplexe Zusammenhänge, und sie berücksichtigen u.a. mithilfe von explorativen Methoden eine Vielzahl von Einflussfaktoren auf die Wirkungsindikatoren. Die untersuchten Interventionen werden als komplexe soziale Systeme mit den entsprechenden Merkmalen (vgl. Kapitel 4.2) rekonstruiert. Während im linsenförmigen Typ die explorativen Elemente primär der Interpretation der Messergebnisse dienen, erhalten sie bei diesem Typ eine eigenständige Rolle bei der Wirkungsbeurteilung: Wesentliche Wirkungsaussagen basieren auf Erkenntnissen, die mittels explorativer Methoden gewonnen wurden. Evaluationen dieses Typs enthalten wesentliche explorative Teile, können aber auch mit hypothesenprüfenden kombiniert sein. Die ‚Black Box‘ wird weit geöffnet, um komplexe Wirkungsmechanismen zu erkennen und zu untersuchen. Am Schluss wird die ‚Black Box‘ wieder geschlossen. Und hier liegt die zentrale Herausforderung dieses Evaluationstyps: die Verdichtung der Komplexität zu klaren Wirkungsaussagen.





Diese Evaluationen erlauben tiefgehende Einblicke in (kontextspezifische) Wirkungsmechanismen und liefern detaillierte und kontextspezifische Erkenntnisse über Wirkungszusammenhänge und Entwicklungen. Sie ermöglichen es, die Wirkungsweise der untersuchten Systeme weitgehend zu verstehen. Die Erkenntnisse sind z.B. durch die Nutzung unterschiedlicher Methoden oder Perspektiven breit abgestützt. Ein Nachteil dieses Typs ist die schwer einschätzbare Güte von subjektiven Wirkungseinschätzungen, die u.a. mithilfe von explorativen Befragungen eingeholt werden. Zudem sind zitierte Einzelaussagen in ihrer Bedeutung für eine Wirkungsbeurteilung nur schwer einschätzbar. Ein anderer Nachteil liegt in der schwer vermittelbaren Komplexität.

Berücksichtigen von hoher Komplexität ohne systematische Reduktion (Kegel)

Dieser Typ setzt bei der Wirkungsbeurteilung vollumfänglich auf explorative Methoden und weitet den Blick kontinuierlich aus, um mehr und mehr Faktoren in die Betrachtung aufzunehmen. Die untersuchten Systeme werden gleich wie im vorangehenden Typ mit hoher Komplexität rekonstruiert, unter Berücksichtigung der Merkmale komplexer sozialer Systeme. In Evaluationen diesen Typs wird die ‚Black Box‘ ebenfalls weit geöffnet, allerdings gelingt es hier nicht, diese auch wieder zu schliessen. Es bleibt eine grosse Komplexität, die nicht mehr systematisch reduziert wird - die Verdichtung zu klaren Wirkungsaussagen bleibt die Evaluation schuldig.





Wie beim zitronenförmigen Typ wird auch hier viel wertvolles, kontextspezifisches Detailwissen über Wirkungszusammenhänge geboten. Der grosse Nachteil dieses Typs liegt aber in der Gefahr, dass die Nutzenden der Evaluation von der Komplexität überfordert sind, die Ergebnisse nur schwer begreifen und deren Bedeutung einschätzen können und infolgedessen selber eine Komplexitätsreduktion vornehmen, die dem Gegenstand nicht angemessen ist. Die Ergebnisse sind in ihrer Komplexität nur schwer vermittelbar und kommunikativ verwertbar.

Zusammenfassend ergibt sich folgende Übersicht:

Typ	Merkmale	Vorteile	Nachteile
 Ausblenden von Komplexität <i>E5</i>	<ul style="list-style-type: none"> - Auf sehr wenige, isolierte Faktoren begrenzt - Keine Ausweitung während der Durchführung der Evaluation - Lineares Ursache-Wirkungsverständnis, lineare Modelle - Kein systemischer Zugang - Hypothesenprüfendes Design 	<ul style="list-style-type: none"> - Übersichtliches Design - Klare Wirkungsaussagen - Einfach verständliche Ergebnisse 	<ul style="list-style-type: none"> - Keine Erkenntnisse über: <ul style="list-style-type: none"> - Wirkungszusammenhänge (weshalb Interventionen wirken oder nicht wirken) - Wirkungen unterschiedlicher Umsetzungen - Bedeutung von Kontextfaktoren - Nebenwirkungen - Fehlende Plausibilisierungen -> Risiko für Fehlinterpretationen
 Begrenztes Zulassen von Komplexität <i>E1, E3, E8, E9</i>	<ul style="list-style-type: none"> - Auf wenige Faktoren fokussiert - Zum Teil mit explorativen Methoden ergänzt - Einfaches Ursache-Wirkungsverständnis, tendenziell lineare Modelle - Kein systemischer Zugang - Primär hypothesenprüfendes Design 	<ul style="list-style-type: none"> - Klare Wirkungsaussagen - Einfach verständliche Ergebnisse - Mehrwert durch erweiterte Perspektive: Hinweise auf Wirkungsweisen, Umsetzungsunterschiede, Kontextfaktoren, Nebenwirkungen -> Verbesserte Interpretation und Plausibilisierung 	<ul style="list-style-type: none"> - Limitierte Erkenntnisse über: <ul style="list-style-type: none"> - Wirkungszusammenhänge (weshalb Interventionen wirken oder nicht wirken) - Wirkungen unterschiedlicher Umsetzungen - Bedeutung von Kontextfaktoren - Nebenwirkungen
 Hohe Komplexität mit anschließender Reduktion <i>E2, E6</i>	<ul style="list-style-type: none"> - Detaillierte Analyse von Wirkungsmechanismen - Komplexe Modelle, systemischer Zugang (oder Ansätze dazu) - Verdichtung der Ergebnisse - Exploratives Design, z.T. mit hypothesenprüfendem kombiniert 	<ul style="list-style-type: none"> - Klare Wirkungsaussagen - Detaillierte, kontextspezifische Erkenntnisse über Wirkungsweisen, Umsetzungsunterschiede, Kontextfaktoren, Nebenwirkungen - Breit abgestützte Ergebnisse 	<ul style="list-style-type: none"> - Schwer einschätzbare Güte subjektiver Wirkungseinschätzungen - Schwer einschätzbare Bedeutung von zitierten Einzelaussagen - Verdichtung z.T. schwer nachvollziehbar
 Hohe Komplexität ohne systematische Reduktion <i>E4, E7</i>	<ul style="list-style-type: none"> - Detaillierte Analyse von Wirkungsmechanismen - Komplexe Modelle, systemischer Zugang (oder Ansätze dazu) - Keine systematische Verdichtung - Differenzierte Ergebnisse ohne klare Wirkungsaussagen - Exploratives Design 	<ul style="list-style-type: none"> - Detaillierte, kontextspezifische Erkenntnisse über Wirkungsweisen, Umsetzungsunterschiede, Kontextfaktoren, Nebenwirkungen - Breit abgestützte Ergebnisse 	<ul style="list-style-type: none"> - Schwer einschätzbare Güte subjektiver Wirkungseinschätzungen - Schwer einschätzbare Bedeutung von zitierten Einzelaussagen - Komplexität schwer vermittelbar - Keine klaren Wirkungsaussagen - Gefahr willkürlicher Komplexitätsreduktion durch Nutzende

Tab. 22 Vor- und Nachteile der vier Typen im Umgang mit Komplexität

Wenn die vier Typen mit den Merkmalen komplexer sozialer Systeme kombiniert werden, ergeben sich die folgenden Unterscheidungsmerkmale (Tab. 23):

	Black Box	Interaktionsvielfalt		Komplexe Wirkungsmechanismen		Selbstorganisation		Unvorhersehbarkeit		Wirkungsbeurteilung	
	Umgang mit der Black Box	deduktive Grenz-ziehung	induktive Grenz-ziehung	Ver-dichtung	Erkenntnisse über Wirkungsmechanismen	Externe Wirkungs-konstruktion	Eigenlogik Zielsysteme	hypothesen-prüfendes Design	exploratives Design	Kontra-faktischer Vergleich	Detaillierte Rekon-struktion
 E5	Black Box bleibt geschlossen	x				x		x		x	
 E1, E3, E8, E9	Der Deckel wird leicht angehoben	x	(x)	(x)	(x)	x	(x)	x	(x)	x	
 E2, E6	Black Box wird geöffnet und wieder geschlossen	(x)	x	x	x	(x)	x	(x)	x	(x)	x
 E4, E7	Black Box wird geöffnet, aber nicht mehr geschlossen	(x)	x		x	(x)	x		x		x

(x): möglich, aber nicht Schwerpunkt

Tab. 23 Umgang mit den Merkmalen komplexer sozialer Systeme

9 Zusammenfassung, Schlussfolgerungen und Ausblick

Wirkungsbeurteilungen stellen die Evaluation von Gesundheitsförderung und Prävention vor grosse Herausforderungen. Das Dilemma zwischen Evidenzansprüchen, die von der medizinischen Forschung herrühren, und der Komplexität der zu untersuchenden Interventionssysteme verunsichert; nicht nur die Praxis, sondern auch die Wissenschaft. In der Gesundheitsförderung und Prävention entzündet sich die Diskussion um die Angemessenheit von Evaluationsverfahren vor allem an den randomisierten kontrollierten Studien, welche unter idealen Bedingungen als Goldstandard für Wirkungsbeurteilungen gelten. Deren Nutzen wird zwar nicht grundsätzlich in Abrede gestellt, aber von vielen für die Evaluation von komplexen Interventionen als nicht adäquat abgelehnt. Objektivistische Forschungsverständnisse und entsprechende Forschungsmethoden werden für komplexe Interventionen der Gesundheitsförderung und Prävention weitgehend als unangemessen zurückgewiesen und eine flexible Anwendung und Kombination unterschiedlicher Methoden und Perspektiven für die Beurteilung von Interventionen gefordert (WHO, 2001; Robert Koch Institut, 2012).

9.1 Zusammenfassende Beantwortung der Forschungsfrage

Die vorliegende Untersuchung ist vor diesem Hintergrund der Frage nachgegangen, wie Evaluation der Gesundheitsförderung und Prävention bei der Beurteilung von Wirkungen mit der Komplexität der untersuchten Interventionen umgeht. Sie lässt sich zusammenfassend wie folgt beantworten:

- Evaluationen der Gesundheitsförderung und Prävention gehen sehr unterschiedlich mit Komplexität um. Sie verwenden eine grosse Vielfalt von methodischen Zugängen, die in unterschiedlichen Forschungsdesigns gebündelt werden, um Wirksamkeit und/oder Wirkungsweise der untersuchten Interventionen zu beurteilen. Dabei zeigen sich Wirkungen als soziale Konstruktionen im Spannungsfeld zwischen Wünschbarem und Machbarem, als Ergebnis von Interaktionen zwischen den Evaluierenden, den Auftrag Gebenden und weiteren Stakeholdern, dem Wissenschaftssystem und der Praxis.
- Bei der Wirkungskonstruktion hat die Evaluation einen grossen Handlungsspielraum. Bei der Eingrenzung, Erfassung und Verarbeitung von Komplexität sowie bei der Wirkungsbeurteilung fällt sie laufend Entscheide über den Umfang und die Art berücksichtigter Komplexität (vgl. Tab. 19, Seite 158) und nutzt entsprechende Strategien der Komplexitätsreduktion und -ausweitung (vgl. Tab. 21, Seite 220). Es ist nicht die Komplexität der Intervention, die zu mehr oder weniger Komplexität in der Evaluation führt, sondern die Wahrnehmung von Komplexität und entsprechende methodische Entscheide durch die Evaluation. Dabei baut die Evaluation - als selbstreferentielles System - auf ihrem eigenen Evaluationsverständnis und fachlichem Know-how auf.
- Die unterschiedlichen Wirkungskonstruktionen bewegen sich im Spektrum zwischen einfach modellierten hypothesenprüfenden Wirksamkeitsevaluationen (Black Box Evaluationen) und explorativ ausgerichteten Evaluationen, die den Untersuchungsgegenstand als komplexe soziale Systeme rekonstruieren (komplexitätssensible Evaluationen). Evaluation bewegt sich dabei immer im Spannungsfeld zwischen Einfachheit, die Wichtiges vernachlässigt, und Komplexität, die nicht mehr zu bewältigen ist. Es hat sich gezeigt, dass weder weniger Komplexität noch mehr Komplexität zu klareren Wirkungsaussagen führt.
- Black Box Evaluationen gründen auf einfachen Modellen, die Wirkungskonstruktion erfolgt weitgehend bereits vor der Datenerhebung. Sie basiert dabei primär auf den Wirkungsvermutungen der Auftrag Gebenden und auf wissenschaftlichem Wissen. Die Systemgrenzen werden deduktiv gezogen, von aussen an die untersuchten Systeme herangetragen und bleiben starr. Black Box Evaluationen beantworten die Frage nach der Wirksamkeit unter Nutzung kontrafaktischer Ver-

gleiche, sie führen aber nicht zu Erkenntnissen über Wirkungsmechanismen. Diese Begrenzung wird zum Teil durch die Kombination mit explorativen Methoden wettgemacht.

- Im Gegensatz dazu entwickeln komplexitätssensible Evaluationen unter Nutzung explorativer Verfahren komplexe Modelle. Die Systemgrenzen sind nicht starr, sondern werden im Laufe der Evaluation erweitert. Die Wirkungsbeurteilung erfolgt über detaillierte Rekonstruktionen von Wirkungszusammenhängen. Explorative Verfahren erweisen sich für eine komplexitätssensible Evaluation als besonders wertvoll, weil sie zu zielgruppen- und kontextspezifischen Erkenntnissen über Wirkungsweisen führen. Zudem öffnen sie den Blick für nicht-beabsichtigte (positive oder negative) Nebenwirkungen. Die Attribution von Wirkungen zu Massnahmen erfolgt hier nicht über kontrafaktische Vergleiche, sondern über detaillierte Rekonstruktionen von Wirkungszusammenhängen.
- Die beobachteten Wirkungskonstruktionen werden aufgrund der zu verarbeitenden Komplexität von vielen Limitierungen, Schwierigkeiten, Verzerrungen und Ungenauigkeiten begleitet, welche die Wirkungsbeurteilung beeinflussen und die Aussagekraft der Ergebnisse einschränken.
- Zentrale Merkmale komplexer sozialer Systeme wie Dynamik, Multidimensionalität, Selbstorganisation oder Unvorhersehbarkeit werden nur selten thematisiert und im Studiendesign explizit berücksichtigt. So werden beispielsweise oft nur Messergebnisse zu zwei Zeitpunkten miteinander verglichen, was keine Aufschlüsse über Dynamiken gibt; oder es werden zwar Veränderungen auf mehreren Wirkungsebenen untersucht, ohne aber Zusammenhänge zwischen diesen zu berücksichtigen; oder die Evaluationen versteifen sich auf die Messung von Zielen und die Beurteilung von Wirkungshypothesen und verpassen dabei nicht-intendierte Nebenwirkungen.
- Evaluationsansätze, welche auf system- oder komplexitätstheoretischen Grundlagen basieren, werden kaum verwendet. Wenn, dann wird Komplexität primär über traditionelle explorative Forschungsmethoden berücksichtigt (Beobachtungen und Interviews). Wenn hypothesenprüfende und explorative Verfahren in derselben Evaluation genutzt werden, dann geschieht dies mehrheitlich getrennt. Einsichten komplexer Art fließen nicht in die Berechnungen ein, sondern werden höchstens ergänzend genutzt, um Messergebnisse zu interpretieren oder zu plausibilisieren.
- Wirkungen werden weitgehend als messbare Veränderungen in meist vorgängig definierten Wirkungsindikatoren verstanden. Prozessorientierte Wirkungsverständnisse, welche partizipative Empowermentprozesse als eigentliche Wirkungen ausweisen, werden nur am Rande einiger explorativ angelegter Evaluationen sichtbar.
- Obwohl fast jede der untersuchten Evaluationen anders mit Komplexität umgeht, lassen sich vier Typen des Umgangs mit Komplexität identifizieren: 1) Vollständiges Ausblenden von Komplexität, 2) begrenztes Zulassen von Komplexität, 3) Berücksichtigen von hoher Komplexität mit anschließender Reduktion und 4) Berücksichtigen von hoher Komplexität ohne systematische Reduktion. In den meisten Evaluationen werden verschiedene Erhebungsmethoden kombiniert; allerdings wird das Potenzial von Mixed method-Designs bei weitem nicht ausgeschöpft (Greene 2007), da die verwendeten Methoden kaum systematisch aufeinander bezogen werden.
- Operationalisierte Wirkungsziele und die darin sichtbaren Wirkungsvermutungen bieten der Evaluation eine wertvolle Referenz für die Beurteilung von Wirkungen. Sie verleiten aber dazu, den Evaluationsblick auf die antizipierten Veränderungen hin zu verengen. Je offener die angestrebten Wirkungen formuliert sind, desto eher wird ein breites Suchen nach Wirkungen begünstigt.
- Die Evaluationen nutzen vielfältige Formen des Strukturierens von Komplexität. Dabei erweisen sich insbesondere Programmtheorien und Wirkungsmodelle als wertvolle Unterstützungshilfen im Umgang mit Komplexität. Sie machen vermutete Wirkungszusammenhänge, aber auch den Grad der berücksichtigten Komplexität explizit und dienen bei der Wirkungsbeurteilung als wichtige Orientierung. Sie stellen aber letztlich immer eine starke Komplexitätsreduktion dar und können sozi-

ale Systeme in ihrer Komplexität nie wirklich abbilden. Auch eine noch so starke Ausdifferenzierung löst diese Grundproblematik nicht, sie verschiebt sie nur auf tiefere Ebenen.

- Wirkungsnachweise werden nicht allein über randomisierte kontrollierte Studien erbracht. Neben kontrafaktischen Vergleichen werden Wirkungsaussagen vor allem mittels detaillierter Rekonstruktionen von Wirkungszusammenhängen begründet. Zudem stützen sich Evaluationen auf Wirkungsbeurteilungen von Befragten ab. Diese drei Strategien werden zudem vielfältig kombiniert.
- Die Evaluierenden sind sich der Schwierigkeit von Wirkungsnachweisen in sozialen Systemen mehrheitlich bewusst und nutzen unterschiedliche Formen der Plausibilisierung, um ihre Wirkungsaussagen abzustützen.

In der empirischen Untersuchung konnte gezeigt werden, wie Evaluationen der Gesundheitsförderung und Prävention Wirkungen systematisch konstruieren. Sie haben beim Aufbau und dem Erzählen ihrer Wirkungsgeschichten einen grossen Handlungsspielraum, den sie weitgehend arbiträr nutzen. Es werden keine klaren Kriterien sichtbar, die eine Bewertung der Güte der verschiedenen Formen der Wirkungsnachweise in komplexen sozialen Systemen erlauben würden und die Frage schlüssig beantworten könnten, wie viel Komplexität wann angemessen ist. In einem nächsten Kapitel sollen auf der Basis der theoretischen und empirischen Erkenntnisse dieser Arbeit einige Reflexionen zur Wahl des Komplexitätsgrades angestellt werden.

9.2 Reflexionen zur Angemessenheit des Komplexitätsgrades in Evaluationen

Derselbe Evaluationsgegenstand kann sehr unterschiedlich komplex evaluiert werden. Ein Zielsystem wie ein Betrieb oder eine Schule kann als relativ homogene Einheit in eine Wirkungsevaluation einfließen oder aber sehr detailliert untersucht werden. Ebenso kann eine Intervention als *eine* Variable in eine Wirkungsmessung einfließen oder aber sie wird in hoher Komplexität rekonstruiert. Die folgende Darstellung mag die Spannungsfelder, die Herausforderungen und das Potenzial verdeutlichen, welche mit der Komplexität sozialer Systeme verbunden sind. Eine Black Box Evaluation würde sich idealtypisch nur für die Auswirkungen eines Gesundheitsförderungsprojekts (A) direkt auf das Körpergewicht der teilnehmenden Kinder interessieren (B) und nicht für die dazwischengeschalteten intermediären Wirkungsebenen und Wirkungsmechanismen. Komplexitätssensible Evaluation würde die Wirkungsmechanismen innerhalb der Black Box detailliert untersuchen, um die Wirkungsweise verstehen zu lernen, Messergebnisse besser interpretieren zu können und um Erkenntnisse darüber zu gewinnen, wie die Wirkung des Projekts verbessert werden kann:

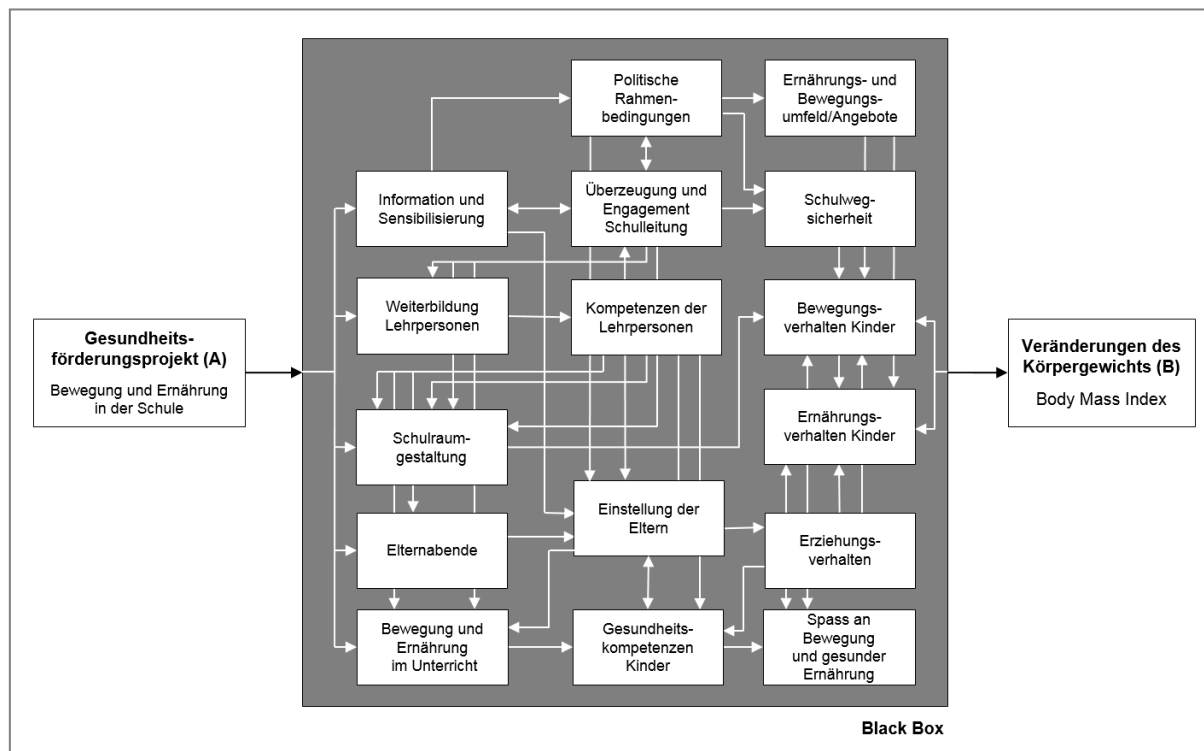


Abb. 31 Ein Blick in die Black Box

Black Box Evaluationen untersuchen im Wesentlichen, ob es einen statistischen Zusammenhang zwischen einer Intervention (Gesundheitsförderungsprojekt) und einer beobachteten Veränderung (Körpergewicht) gibt. Falls sich ein solcher Zusammenhang belegen lässt, ist dies je nach Design und Güte der Untersuchung ein mehr oder weniger plausibler Hinweis auf die Wirksamkeit der Intervention. Die Komplexität der Wirkungsentstehung (Dynamik, Selbstorganisation, Emergenz etc.) ist bei dieser Art der Analyse nicht relevant, von Bedeutung ist das Gesamtergebnis dieser Prozesse. Black Box Evaluationen sind deshalb nicht dazu geeignet, Wirkungszusammenhänge zu verstehen oder Wirkungen zu erklären. Aber auch komplexitätssensible Evaluationen, welche die Wirkungszusammenhänge in der Black Box untersuchen, sind grundsätzlich mit derselben erkenntnistheoretischen Limitierung konfrontiert. Die Untersuchung jedes einzelnen Wirkungszusammenhangs (z.B. zwischen den Kompetenzen der Lehrpersonen und der Einstellung der Eltern) und auch des Zusammenspiels zwischen diesen stellt letztlich eine grobe Simplifizierung von Wirklichkeit dar und die vermutete Komplexität von sozialen, aber auch psychischen und physischen Systemen macht ein gänzlich Verstehen unmöglich. Die Kästchen in der Black Box sind also wieder Black Boxen und diese bestehen wiederum aus Black Boxen etc. Komplexität liesse sich fast beliebig auffächern. Es ist entsprechend nicht die Intervention, die per se mehr oder weniger komplex ist, sondern die Evaluation entscheidet über den Grad der Komplexität, mit welcher die Intervention beobachtet wird; sie entscheidet, wie tief sie in die Komplexität eintauchen will. Je nach Komplexitätsgrad sind andere Erkenntnisse möglich. Die Frage nach der Angemessenheit des Komplexitätsgrades einer Wirkungsevaluation lässt sich infolgedessen nicht generell beantworten, sondern sie ist davon abhängig, wie wirkungsrelevant ein Beobachter Komplexität einschätzt und welche Komplexität der Ergebnisse er wünscht. Ein Betrieb kann entsprechend als komplexes, dynamisches System wahrgenommen werden, dessen Komplexität als hoch wirkungsrelevant eingeschätzt wird und die Evaluation kann sich für die kontextspezifischen Prozesse der Wirkungsentfaltung interessieren. Dann wird sie Interventionen in Betrieben eher mittels komplexitätssensibler Evaluation beurteilen. Oder sie kann die Komplexität eines Betriebes und die Unterschiede zwischen Betrieben als nicht relevant für eine Wirkungsbeurteilung einschätzen, eher eine Black Box Evaluation wählen und Erkenntnisse erhalten, die zwar eine

Beurteilung auf übergeordneter Ebene erlauben, aber keine Erkenntnisse über Interventions- und Kontextunterschiede oder über Wirkungsmechanismen enthalten. Aus der vorliegenden Arbeit lassen sich einige Anhaltspunkte ableiten, die für die Bestimmung eines geeigneten Komplexitätsgrades Orientierung geben:

	Black Box Evaluation	Komplexitätssensible Wirkungsevaluation
Komplexität der Zielsysteme ¹³⁴ <i>im Hinblick auf die Förderung von Gesundheit</i>	Gering	Hoch
Komplexität der Intervention ¹³⁴	Gering	Hoch
Standardisierungsgrad und Stabilität der Intervention	Hoch	Gering
Unvorhersehbare Entwicklungen	Eher unwahrscheinlich oder nicht wirkungsrelevant	Wahrscheinlich und wirkungsrelevant
Erkenntnisse über Wirkungsmechanismen	Nicht erforderlich	Erwünscht
Zielgruppen- und kontextspezifische Erkenntnisse	Nicht erforderlich	Erwünscht
Ähnlichkeit der Zielsysteme bei mehreren Umsetzungen	Sehr ähnlich	Sehr unterschiedlich
Ähnlichkeit der Kontexte bei mehreren Umsetzungen	Sehr ähnlich	Sehr unterschiedlich
Zeithorizont der Interventionswirkung	Eher kurzer Wirkungshorizont	Eher langer Wirkungshorizont
Plausibilität der Interventionslogik	Hohe Plausibilität	Hohe Unsicherheit
Alternative Wirkungserklärungen	Keine oder vergleichsweise unbedeutende andere Einflüsse	Bedeutende alternative Wirkungserklärungen wahrscheinlich
Art der Interventionen	Bereits gut untersuchte, bewährte Interventionen ¹³⁵	Innovative Interventionen

Tab. 24 Anhaltspunkte zur Bestimmung des Komplexitätsgrades einer Evaluation

Diese Anhaltspunkte mögen auf den ersten Blick bestechend sein. Erst auf den zweiten Blick offenbaren sie ihre Relativität. So können zum Beispiel zwei benachbarte Schulen als sehr ähnlich wahrgenommen werden (kleine Schulen, gleiche Organisationsstruktur etc.), und auch ihr Kontext erscheint vergleichbar (überschaubare Gemeinden, landwirtschaftlich geprägt etc.). Dieselben Schulen und Kontexte unterscheiden sich aber enorm, wenn auf der Ebene einzelner Subsysteme und Interaktionen verglichen wird. Und diese Unterschiede können hoch gesundheitsrelevant sein. Auch die Komplexität und der Standardisierungsgrad einer Intervention können sehr unterschiedlich wahrgenommen werden. So kann eine Rückenschulung in einem Betrieb als vergleichsweise wenig komplex und hoch standardisiert wahrgenommen werden. Wenn aber die einzelnen Interaktionen zwischen der Kursleitung und den Teilnehmenden, die einflussenden individuellen Bedürfnisse und Besonderheiten sowie die Feinheiten in der Gruppendynamik berücksichtigt werden, zeigt sich eine hohe Komplexität und geringe Standardisierung. Es ist wichtig, dass sich Evaluationen der Relativität der Komplexitätseinschätzungen bewusst sind und entsprechende Einstufungen begründen können. Folgende Aspekte scheinen bei der Einschätzung, wie komplex Evaluationen einen Untersuchungsgegenstand wahrnehmen und welche forschungsmethodischen Entscheide sie davon ableiten, eine Rolle zu spielen:

¹³⁴ Vgl. die Merkmale komplexer sozialer Systeme (Kap. 4.2).

¹³⁵ Wesentliche Wirkungsmechanismen und potenzielle Nebenwirkungen sind für vergleichbare Kontexte bekannt.

Handlungsrelevanz und Anschlussfähigkeit an Entscheidungs- und Handlungssysteme: Die Evaluationsergebnisse müssen im politischen System, in der Verwaltung oder bei den Zielgruppen anschlussfähig sein, sie müssen sich bei der Bewältigung der gesundheitlichen Herausforderungen und bei der Lösung der sozialen Probleme bewähren. Sie müssen Lern- und Entwicklungsprozesse in den Interventions- und Zielsystemen sowie in den Entscheidungsgremien anregen und zu politischen und fachlichen Entscheidungen führen, die zu einer Verbesserung der Ausgangssituation führen. So ist gut verständlich, wenn sich Gesundheitsförderungsfachleute, welche im Alltag mit der Komplexität ihres Handlungsfeldes konfrontiert sind, eher dazu neigen, die Interventionen, Zielsysteme und Kontexte als hoch komplex wahrzunehmen. Interventionen der Gesundheitsförderung und Prävention, denen eine hohe Wirksamkeit nachgewiesen wurde, sollten zu positiven Entwicklungen in den Kontexten führen, für welche die Wirksamkeit proklamiert wurde. Da die Wirklichkeit dem Menschen nicht unmittelbar zugänglich ist, können der Wahrheitsgehalt oder die Plausibilität von Erkenntnissen und damit von Wirkungsaussagen nicht direkt an der Realität überprüft werden. Wenn die Erkenntnisse aus Wirkungsevaluationen aber umgesetzt werden, können sie sich – wiederum über Beobachtungen vermittelt - mehr oder weniger bewähren.

Wissenschaftliche Konventionen: Zum einen haben sich in wissenschaftlichen Subsystemen Konventionen darüber herausgebildet, wie viel Komplexität in welcher Situation als angemessen eingeschätzt wird. In Wirksamkeitsstudien von Medikamenten ist der Mensch als Untersuchungseinheit eine akzeptierte Grösse und es gilt auch als allgemein akzeptiert, dass Kontextfaktoren durch Randomisierung als wenig wirkungsrelevant behandelt werden. In Bezug auf Wirksamkeitsstudien in der Gesundheitsförderung und Prävention scheiden sich die Geister und es werden intensive Diskurse über das geeignete Mass an Komplexität geführt. Während die einen Exponent/-innen in clusterrandomisierten Forschungsdesigns ganze Kindergärten oder Schulen als homogene Einheiten in Wirksamkeitsanalysen einfließen lassen (z.B. die hier untersuchte Evaluation De Bock & Fischer, 2011) und interne Komplexität auf dieser Beurteilungsstufe als nicht ergebnisrelevant erachten, ist dies für andere nicht akzeptabel (vgl. das Kapitel zur Evidenzdebatte in der Gesundheitsförderung).

Methodisches Knowhow: Man wird für eine Evaluation nicht einen Komplexitätsgrad wählen, den man mit den zur Verfügung stehenden Methoden und Kompetenzen nicht bewältigen kann oder der nicht in das Portfolio eines Evaluationsinstituts passt. So dürften Evaluator/-innen, die sich auf explorative Forschungsmethoden spezialisiert haben, Zielsysteme und Interventionen mit einer vergleichsweise hohen Komplexität rekonstruieren wollen, also beispielsweise eher Differenzen in Interventionen, Zielsystemen und Kontexten suchen als von Beginn an Gleichheit zu unterstellen. Umgekehrt dürfte es RCT-Spezialist/-innen eher dazu drängen, ein Gesundheitsförderungsprojekt als Variable ‚Intervention‘ in ein Forschungsdesign aufzunehmen, als in die Komplexität der vielfältigen Interaktionen einzutauchen.

Passung mit der Intervention: Dieser Aspekt bezieht sich vor allem auf die Vergleichbarkeit von Zielsystemen und Kontexten. In experimentellen Studien wird darauf geachtet, dass die untersuchten Systeme möglichst vergleichbar sind. Es werden beispielsweise nicht Menschen und Hunde als Probanden in dieselbe Medikamentenstudie aufgenommen, sondern nur Menschen und eventuell nur Menschen eines bestimmten Geschlechts oder Alters. Doch auf welcher Ebene sind diese Menschen wirklich vergleichbar? Sobald man tiefer in die Komplexität eintaucht, unterscheiden sich dieselben Menschen plötzlich stark. Sie haben sehr unterschiedliche Lebenserfahrungen, sind genetisch und gesundheitlich unterschiedlich vorbelastet, sie reagieren unterschiedlich sensibel auf Umwelteinflüsse etc. Die Bestimmung der Vergleichbarkeit der Zielsysteme und Kontexte kann bis zu einem gewissen Grad gelingen, indem diese in Bezug zur Intervention gesetzt werden. Menschen und Hunde dürften sehr unterschiedlich sein, wenn es um die Evaluation eines Sprachförderungsprogramms geht. In Bezug auf eine blutstillende Massnahme sind sie unter Umständen eher vergleichbar. Wenn Programme evaluiert werden, die zum Ziel haben, dass Kinder mehr zu Fuss zur Schule gehen, dürften sich ländliche und städtische Schulen viel stärker unterscheiden als im Hinblick auf Präventionsmassnahmen im Klassenzimmer. In experimentellen Medikamentenstudien, in welchen primär die Frage

der Wirksamkeit (ja/nein) beantwortet werden soll, geht der einzelne Mensch als relativ homogene Einheit in die Untersuchung ein. Soll aber die Wirkungsweise eines Medikamentes detailliert beleuchtet werden, muss der Mensch als hoch komplexes bio-chemisches System betrachtet werden.

Anschlussfähigkeit an Wissenschaftssysteme: Zusätzlich müssen Wirkungsnachweise auch im wissenschaftlichen Kontext auf Akzeptanz stossen, sowohl bezüglich ihrer Methodik als auch bezüglich der Ergebnisse. Dabei dürfen sie durchaus innovativ sein, sie müssen aber an die aktuellen wissenschaftlichen Diskurse anschliessen und sich in diesen Diskursen bewähren. Aufgrund der allgemeinen methodischen Verunsicherung, die angesichts der Komplexität sozialer Systeme herrscht, gibt es viele unterschiedliche Diskurse, Strömungen und Anschlusspunkte. Dabei kann nicht eine klare Entwicklungstendenz festgestellt werden, sondern der alte Methodenstreit zwischen Verfechtern hypothesenprüfender Verfahren und Verfechtern explorativer Verfahren scheint sich auch in dieser neuen Thematik fortzusetzen. Vorsichtige Einigkeit zeichnet sich höchstens an den Rändern der (wahrgenommenen) Komplexität ab. Bei scheinbar ‚einfachen‘, individuumsbezogenen sozialen Interventionen billigen die meisten Exponentinnen und Exponenten experimentellen Studien eine hohe Nachweiskraft zu. Auf der anderen Seite des Spektrums, in hoch komplexen Zusammenhängen mit sich ständig verändernden Gegebenheiten, scheinen die Grenzen experimenteller und quasi-experimenteller Verfahren offensichtlich zu sein. Dazwischen scheint man sich uneinig und sucht befriedigende Lösungen zunehmend in der Kombination der verschiedenen Ansätze. So werden beispielsweise explorative Verfahren in experimentellen Designs ergänzend genutzt, um Hinweise auf nicht-intendierte Nebenwirkungen oder förderliche und hinderliche Faktoren zu erhalten.

Die häufig gestellte Frage, ob hoch komplexe Interventionen genauso komplex evaluiert werden müssen, muss grundsätzlich verneint werden. Evaluationen können, auch wenn sie noch so vielfältig sind, nie die Komplexität der untersuchten sozialen Systeme abbilden. Aus logischen Gründen ist der Komplexitätsgrad der Rekonstruktion des Untersuchungsgegenstandes immer geringer als der Komplexitätsgrad des Untersuchungsgegenstandes selber, Evaluation reduziert immer Komplexität. Die Frage, wie viel Komplexität in der Evaluation einer Intervention angemessen ist, lässt sich zusammenfassend und vereinfacht wie folgt beantworten: Wer komplexe Antworten will, muss komplex evaluieren. Wer Informationen über dynamische Entwicklungs- oder Selbstorganisationsprozesse benötigt, um beteiligte Akteure zielgerichtet unterstützen zu können, sollte ein entsprechendes Mass an Komplexität berücksichtigen und eine angemessene komplexitätssensible Evaluation wählen. Eine Grenze ist der komplexitätssensiblen Evaluation vor allem durch die Verstehbarkeit der Ergebnisse gegeben. Sind Evaluationsergebnisse so komplex, dass sie von den Adressaten nicht mehr verstanden werden, verlieren sie ihre Handlungsrelevanz. Wer hingegen auf einem sehr geringen Komplexitätsniveau evaluiert, läuft ebenfalls Gefahr, dass die Ergebnisse nicht mehr handlungsrelevant sind, da wichtige wirkungsrelevante Faktoren vernachlässigt wurden:

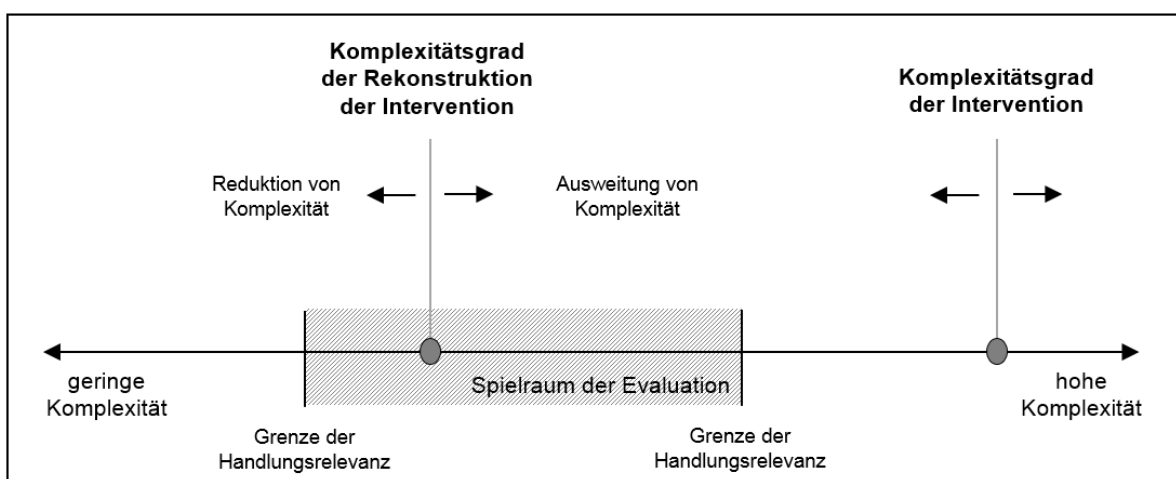


Abb. 32 Spielraum der Evaluation zwischen Ausweitung und Reduktion von Komplexität

Es empfiehlt sich, in der Evaluation immer dann eine Stufe mehr Komplexität zuzulassen, wenn es bezüglich eines Aspekts Anzeichen dafür gibt, dass ein Mehr an Komplexität wichtige Entscheidungs- und handlungsrelevante Erkenntnisse verspricht. Durch die zusätzliche Analysetiefe können Evaluationsergebnisse plausibilisiert und gefestigt werden. Baut eine Wirkungsevaluation zum Beispiel darauf auf, dass eine Intervention in verschiedenen Schulen standardisiert erfolgt oder dass sich die Kontexte dieser Schulen nicht wirkungsrelevant unterscheiden, dann lohnt es sich, diese Annahmen durch genaues Hinsehen zu überprüfen, wenn diesbezüglich Zweifel aufkommen. Die Ergänzung von Black Box Evaluationen mit explorativen Elementen ist ein gutes Beispiel dafür, wie durch eine Ausweitung von Komplexität Hinweise auf Wirkungsmechanismen, auf die Bedeutung von Kontextfaktoren und auf Nebenwirkungen generiert werden können, die die Erkenntnisse vertiefen und festigen.

Der Umgang mit der Komplexität sozialer Systeme ist in der Evaluation noch wenig systematisiert. Die im nächsten Kapitel vorgestellte Kriterienliste soll dazu dienen, diesen Umgang zu reflektieren und zu helfen, Wirkungskonstruktionen der Evaluation transparenter zu machen, bewusster und angemessener mit Komplexität umzugehen und Möglichkeiten der Ausweitung von Komplexität aufzuzeigen. Mit der Kriterienliste, welche aus den theoretischen und empirischen Erkenntnissen dieser Studie entwickelt wurde und auf den referenzierten Autorinnen und Autoren aufbaut, soll ein Diskussionsbeitrag in Richtung einer Bestimmung komplexitätssensibler Evaluation geleistet werden.

9.3 Umrisse einer komplexitätssensiblen Evaluation

Während experimentelle und quasiexperimentelle Studien in der Methodenliteratur ausführlich beschrieben sind (z.B. in Khandker et al., 2010; Gertler et al., 2011), fehlen vergleichbar klare Umschreibungen von ‚komplexitätssensiblen‘ Evaluationen noch weitgehend¹³⁶. Im diesem Kapitel sollen auf der Grundlage der Erkenntnisse dieser Arbeit die Umrisse einer Evaluation skizziert werden, welche die Komplexität sozialer Interventionen zum Ausgangspunkt der Betrachtung nimmt und Wirkungsmechanismen und Selbstorganisationsprozesse in ihrer Analyse mit berücksichtigt, um breit abgestützte und vielfältig verwendbare Erkenntnisse zu erhalten.

1) Systemische Perspektive

Interventionen der Gesundheitsförderung und Prävention realisieren sich in komplexen sozialen Systemen, selbst dann, wenn die Massnahmen nicht Struktur- sondern lediglich Verhaltensveränderungen beabsichtigen. Die einzelnen Menschen sind in diese Systeme eingebunden, werden in wechselseitigen Interaktionen von ihnen beeinflusst und gestalten sie ihrerseits mit. Die Entfaltung von Wirkung in der Gesundheitsförderung und Prävention ist in diese komplexen Systemmechanismen eingebunden, die Einflussfaktoren auf Gesundheit sind beinahe unbegrenzt. Um Wirkungen beurteilen zu können, muss die Evaluation die zentralen Wirkungsfaktoren identifizieren und die Wirkungsmechanismen in ihren Grundzügen verstehen lernen. Ohne diese Wirkungsmechanismen zu kennen, läuft die Evaluation Gefahr, festgestellte Veränderungen vorschnell als Wirkungen fehl zu interpretieren.

Komplexitätssensible Wirkungsevaluation...

- betrachtet den Untersuchungsgegenstand aus einer systemischen Perspektive. Sie versucht, komplexe Wirkungsmechanismen zu rekonstruieren und zu verstehen und verlässt sich nicht auf die isolierte Betrachtung einzelner Kennzahlen oder Wirkungsketten.
- beginnt mit einer Analyse der zu untersuchenden Systeme, um deren Funktionsweise zu verstehen.

¹³⁶ Ansätze dazu liefern beispielsweise Patton (2011); Robert Koch Institut (2012); Pawson and Tilley (1997); DeGEval (2014); Forss, et al. (2011).

- arbeitet potenziell wirkungsrelevante Perspektiven und Interessen von Stakeholdern in den untersuchten Systemen heraus.
- untersucht strukturelle Kopplungen zwischen verschiedenen relevanten sozialen Systemen und zwischen sozialen, psychischen und physischen Systemen.
- hat gleichzeitig die verschiedenen Handlungsebenen der Gesundheitsförderung (Individuum, Gruppe, Organisation, Lebenswelt, Politik/Gesellschaft) im Blick und berücksichtigt Interaktionen zwischen diesen Handlungsebenen.
- reflektiert Interventionen der Gesundheitsförderung und Prävention im Hinblick auf eine systemische Perspektive kritisch.

2) Induktive und transparente Grenzziehung

Die Ursache-Wirkungszusammenhänge in der Gesundheitsförderung und Prävention sind potenziell unendlich vielfältig. Die Evaluation muss, wie auch die Intervention selber, diese Komplexität reduzieren und sich auf die Untersuchung ausgewählter, besonders wirkungsrelevant erscheinender Wirkungszusammenhänge beschränken. Sie muss eine Balance finden zwischen einerseits Offenheit, um nicht Wichtiges und neu Auftauchendes aus dem Blick zu verlieren, und andererseits Einschränkung, um sich nicht in der Unendlichkeit potenzieller Wirkungseinflüsse zu verlieren.

Komplexitätssensible Wirkungsevaluation...

- unterscheidet sich von anderen Evaluationen dadurch, dass sie sich nicht auf die Berücksichtigung von wenigen zu Beginn festgelegten Einflussfaktoren beschränkt, sondern proaktiv und kontinuierlich nach wirkungsrelevanten Faktoren sucht und diese in die Wirkungsbeurteilung aufnimmt.
- macht transparent, welche Einflüsse auf ein Interventionssystem als besonders wirkungsrelevant eingeschätzt werden, auf welcher Grundlage dies geschieht und wie die Evaluation diese berücksichtigt.
- bezieht bei der Selektion und Grenzziehung einerseits das aktuelle wissenschaftliche Wissen und andererseits kontextspezifisches Praxis- und Expertenwissen, insbesondere auch aus den Zielsystemen, mit ein. Die Übertragung wissenschaftlicher Erkenntnisse über Ursache-Wirkungszusammenhänge auf den jeweiligen Kontext wird genauso kritisch reflektiert wie die Wirkungsvermutungen der Stakeholder vor Ort.
- öffnet in einem ersten Schritt das Spektrum an Komplexität, um in einem zweiten Schritt fundiert eingrenzen zu können. Entsprechende Selektionen und Grenzziehungen werden begründet.
- überprüft und erweitert bei Bedarf die Grenzziehung mittels explorativer Methoden.

3) Modellierung von Wirkungszusammenhängen

In der Gesundheitsförderung und Prävention entfaltet sich Wirkung in dynamischen Entwicklungsprozessen. Gemessene Veränderungen können leicht als Wirkungen missinterpretiert werden, wenn diese Entwicklungsprozesse nicht bekannt sind. Programmtheorien und Wirkungsmodelle machen vermutete Wirkungszusammenhänge in der Regel auf einer sehr generellen Ebene (Oberflächenstruktur) explizit. Sie sind als stark vereinfachende Rahmenmodelle zu verstehen. Komplexitätssensible Evaluation kann derartige Modelle als Ausgangspunkt nehmen, muss in ihrer Analyse aber detaillierter werden und versuchen, in die Tiefenstruktur einzudringen. Eine ausgearbeitete ‚theory of action‘ kann dabei helfen, explorativ gewonnene Erkenntnisse über Feinheiten in Wirkungszusammenhängen systematisch zu strukturieren.

Komplexitätssensible Wirkungsevaluation...

- legt vermutete Wirkungszusammenhänge zum Beispiel im Rahmen von Programmtheorien oder Wirkungsmodellen offen.

- nutzt wissenschaftliche Theorien zur Reflexion und Ergänzung der Programmtheorien und Wirkungsmodelle.
- untersucht ausgehend von diesen allgemeinen Konstruktionen Wirkungszusammenhänge detailliert und kontextspezifisch, stützt sich dabei auf Untersuchungen im Feld und auf Erklärungsmuster der Zielgruppen und ihrer Umgebung (lokale Theorie) und passt die entworfenen Modelle neuen Erkenntnissen an.
- rekonstruiert die untersuchten Interventionen detailliert, um wirkungsrelevante kontext- und zielgruppenspezifische Besonderheiten zu erkennen.
- begnügt sich nicht mit einfachen Vorher-Nachher-Vergleichen sondern untersucht dynamische Entwicklungsprozesse und deren Einfluss auf die Wirkungsentfaltung.

4) Kontext- und zielgruppenspezifische Analysen

Interventionen der Gesundheitsförderung und Prävention entfalten ihre Wirkung in der konkreten Interaktion mit den Zielsystemen vor Ort. Da diese aufgrund ihrer Komplexität prinzipiell einzigartig sind, kommt dem entsprechenden Kontext eine grosse Bedeutung bei der Wirkungsentfaltung zu.

Komplexitätssensible Wirkungsevaluation...

- nimmt Abstand von der Vorstellung, dass Wirksamkeit eine Eigenschaft eines Projekts oder Programms ist, welche unabhängig vom jeweiligen Interventionskontext existiert.
- versteht und untersucht Wirkung kontext- und zielgruppenspezifisch und sucht dazu die Nähe zu den untersuchten Interventions- und Zielsystemen.
- interessiert sich für Unterschiede in den Umsetzungen der Intervention und deren Bedeutung für die Wirkungsentfaltung.
- macht wirkungsrelevante System/Umfeld-Bezüge sichtbar und begründet die (Nicht-) Berücksichtigung dieser Faktoren.
- überprüft die Übertragung von Erkenntnissen aus anderen Studien auf den eigenen Kontext und die eigenen Zielgruppen kritisch.

5) Bewusster Umgang mit Ungewissheit

Aufgrund der komplexen Wirkungsmechanismen ist davon auszugehen, dass Interventionen in der Realität vom Plan abweichen, unvorhersehbare Entwicklungen und nicht-beabsichtigte Wirkungen entstehen. Exploration erübrigt sich auch dann nicht, wenn sich eine Intervention in Pilotumsetzungen andernorts bewährt hat. Interaktionen in sozialen Systemen können zwar potenziell unendlich vielfältig sein, sie entstehen aber nicht zufällig, sondern sind Gesetzmässigkeiten unterworfen und bilden entsprechende emergente Ordnungsstrukturen heraus.

Komplexitätssensible Wirkungsevaluation...

- versucht, die ein System und seine Umweltbeziehungen bestimmenden Regeln, Mechanismen und Druckpunkte aufzudecken.
- interessiert sich für komplexe Systemdynamiken und für ungeplante Entwicklungen; Elemente, welche andere Evaluationen oftmals lieber ausblenden.
- beschränkt sich nicht auf die Untersuchung von intendierten Wirkungszielen, sondern interessiert sich ebenso für Wirkungen jenseits der Zielsetzungen und Wirkungsvermutungen der Projektträger und Auftraggeber. Sie fragt darüber hinaus, um auch ungeplante Mechanismen, Übersteuerungen und unbeabsichtigte (Neben-) Wirkungen in den Blick zu bekommen.
- verfolgt kein starres Evaluationsdesign, sondern passt sein Evaluationsdesign und seine Methodik bei Bedarf den Entwicklungen im Interventionssystem und neu auftauchenden Fragestellungen an (emergentes Evaluationsdesign).

- braucht nicht zwingend operationalisierte Wirkungsziele. Offenheit in Bezug auf die Wirkungen trägt im Gegenteil dazu bei, unvoreingenommen nach Wirkungen und Nebenwirkungen zu suchen.

6) Analyse von Selbstorganisationsprozessen

Soziale Systeme sind operativ geschlossen und entscheiden auf Basis eigener Reflexionen selber, wie sie auf Interventionen, die von aussen an sie herangetragen werden, reagieren. Dieser Umstand charakterisiert alle Interventions- und Zielsysteme der Gesundheitsförderung und Prävention und ist für deren Wirkungsweise hoch relevant. Es ist nicht die externe Intervention, welche von aussen ‚etwas im Zielsystem bewirkt‘, sondern die Wirkung entsteht durch die Art und Weise der Rezeption der externen Anstösse durch die Zielsysteme selber. Nachhaltige Gesundheitsförderung und Prävention unterstützt die Zielsysteme in ihrer Selbstorganisation.

Komplexitätssensible Wirkungsevaluation...

- geht davon aus, dass es sich bei Wirkungen um das Ergebnis von gemeinsamen Entwicklungsprozessen zwischen den Zielsystemen und ihrer Umgebung handelt (inkl. Interventions- und Evaluationssystem).
- interessiert sich für die strukturellen Kopplungen zwischen Interventions- und Zielsystemen und dafür, wie die Zielsysteme die Angebote der Interventionssysteme aufnehmen und verarbeiten.
- interessiert sich entsprechend für Selbstorganisationsprozesse in Zielsystemen und untersucht, inwiefern intervenierende Systeme die Selbstorganisation der Zielsysteme unterstützen und inwiefern die Zielsysteme durch die Intervention befähigt und ermächtigt werden, selber angemessen und besser mit den entsprechenden gesundheitlichen Herausforderungen umzugehen.
- stellt dabei das Wissen und die Mitwirkung der Stakeholder der Interventions- und Zielsysteme in den Mittelpunkt.
- unterstützt selber die Selbstorganisationsprozesse der untersuchten Interventionssysteme durch die Mitgestaltung gemeinsamer Lernprozesse, sucht die Nähe zu den untersuchten Systemen und bringt sich in diese ein.
- reflektiert die eigene Rolle und den Einfluss auf die Wirkungsentfaltung in den Interventions- und Zielsystemen kritisch.

7) Kritisch reflektierte Wirkungsnachweise

Sogenannte ‚echte‘ Wirkungsevaluationen, welche eindeutige Wirkungsnachweise liefern, sind in der Gesundheitsförderung und Prävention eine Illusion. Wirkungsevaluationen in komplexen Feldern sind immer mit grossen Unsicherheiten behaftet. Unterschiedliche methodische Verfahren können grundsätzlich mehr oder weniger plausible Hinweise auf Wirkungszusammenhänge und Wirkungen liefern.

Komplexitätssensible Wirkungsevaluation...

- trachtet nicht danach, Wirkungen zu beweisen, sondern ist sich der Potenziale und Grenzen von Wirkungszuschreibungen in komplexen sozialen Systemen bewusst und macht diese transparent. Sie lässt in Bezug auf Wirkungsaussagen Bescheidenheit erkennen.
- sucht nicht „absolute“ oder „allgemeine“ Wirkungsnachweise, sondern nachvollziehbare Aussagen über das Wirkungspotenzial, die Wirkungsweise und die Wirksamkeit von einzelnen Interventionen.
- fragt nicht primär danach, ob eine Intervention wirkt oder nicht, sondern was die Intervention unter welchen Umständen in den untersuchten Systemen und bei den Zielgruppen auszulösen vermag.

8) Plausibilisierungen

Da Wirkungsaussagen in Evaluationen der Gesundheitsförderung und Prävention immer mit Limitierungen und Unsicherheiten verbunden sind, werden systematische Verfahren der Absicherung einzelner Aussagen benötigt. Die Evaluation kann dazu unterschiedliche Plausibilisierungsformen nutzen.

Komplexitätssensible Evaluation...

- entwickelt Plausibilisierungsstrategien, um Wirkungsaussagen zu festigen. Sie kombiniert dabei eigene empirische Erkenntnisse mit Erkenntnissen aus anderen Studien sowie mit theoretischem Wissen.
- nutzt das Potenzial von Daten-, Methoden- und Perspektivtriangulation, um Wirkungsfragen aus unterschiedlichen Perspektiven zu beleuchten und Wirkungsbeurteilungen zu plausibilisieren.
- kombiniert bei der Wirkungsbeurteilung wenn möglich kontrafaktische Vergleiche (z.B. aus experimentellen oder quasiexperimentellen Studien), detaillierte Rekonstruktionen von Wirkungszusammenhängen und Wirkungsbeurteilungen von Akteuren vor Ort.
- kann als iterativer Prozess konzipiert werden, in welchem Ursache-Wirkungszusammenhänge auf Basis von wissenschaftlicher Evidenz und lokalen Datenerhebungen kontinuierlich verfeinert und verfestigt werden.

Komplexitätssensible Evaluationen haben in der Gesundheitsförderung und Prävention ein grosses Potenzial, da sie sich bei der Beantwortung von Wirkungsfragen für die Dynamiken und partizipativen Prozesse interessieren, welche die Interventionen der Gesundheitsförderung ausmachen. Erkenntnisse zu kontextspezifischen Wirkungsweisen sind geeignet, wirkungsvolle Praxis der Gesundheitsförderung und Prävention weiter zu entwickeln. Sie bilden eine Brücke zwischen hypothesenprüfenden Wirksamkeitsevaluationen und Prozessevaluationen, indem sie nicht nur die Frage beantworten, ob Interventionen wirken, sondern wie sie unter welchen Bedingungen wirken. Sie liefern kontextspezifisches Detailwissen, Hintergründe und Zwischentöne, die für zwischenmenschliche Interaktionen bedeutend sind.

Durch die Berücksichtigung einer grossen Vielfalt von Wirkungsfaktoren und von komplexen Wirkungszusammenhängen laufen komplexitätssensible Evaluationen allerdings Gefahr, vom Interventionssystem nicht verstanden und infolgedessen nicht genutzt zu werden. Komplexitätssensible Evaluation muss der transparenten Reduktion von Komplexität besondere Beachtung schenken und sich um die Anschlussfähigkeit an die Interventions- und Zielsysteme bemühen.

9.4 Schlussfolgerungen für die Evaluation von Gesundheitsförderung und Prävention

Gesundheit entsteht in hoch komplexen Interaktionsprozessen, in denen Interventionen der Gesundheitsförderung und Prävention nur einen kleinen Ausschnitt darstellen. Es ist nicht möglich, die Komplexität dieser Prozesse vollumfänglich zu erfassen oder Interventionen isoliert zu erbringen. Entsprechend können die Wirkungsmechanismen der Gesundheitsförderung und Prävention auch nicht vollständig abgebildet oder verstanden werden. Obwohl Interventionen der Gesundheitsförderung und Prävention nicht trivial sind¹³⁷, muss ihnen immer in gewisser Weise Trivialität unterstellt werden, um sie planen, steuern und evaluieren zu können. Auch wenn in Wirkungsmodellen Komplexität aufgefächert wird, um Wirkungsmechanismen zu untersuchen, so sind doch die einzelnen Bezüge innerhalb dieser Modelle noch immer starke Vereinfachungen sozialer Wirklichkeit. Wenn Evaluationen die Wirkungen von Gesundheitsförderung und Prävention untersuchen, sind sie immer mit limitierter Erkenntnismöglichkeit durch starke Komplexitätsreduktion, Verzerrungen und Ungenauigkeiten konfrontiert. Entsprechend kann es der Evaluation nicht darum gehen, absolute Gewissheit zu erlangen und ein-

¹³⁷ Im Sinne trivialer Maschinen (vgl. von Foerster (2009)).

deutige Wirkungsnachweise im Sinne von ‚Beweisen‘ zu erbringen, sondern Ungewissheit in Bezug auf Wirksamkeit und Wirkungsweisen zu reduzieren und Planungssicherheit zu erhöhen.

Wie viel Komplexität ist angemessen?

Die Untersuchung hat ein breites Spektrum an möglichen Formen der Wirkungsbeurteilung gezeigt. Es hat sich dabei kein ‚Goldstandard‘ herauskristallisiert. Vielmehr bringen unterschiedliche Formen des Umgangs mit Komplexität unterschiedliche Vor- und Nachteile mit sich. Qualitativ hochwertige experimentelle Studien können auf einer allgemeinen Ebene mit einer hohen Plausibilität die Frage beantworten, ob eine beobachtete Veränderung einer bestimmten Intervention zugeschrieben werden kann oder nicht. Auch in der Gesundheitsförderung und Prävention sind derartige Studien grundsätzlich möglich (wenn auch selten in ausreichender Qualität durchführbar). Dabei werden aber weder Unterschiede in den Umsetzungen noch Kontexteinflüsse berücksichtigt, es werden keine Wirkungsmechanismen oder Nebenwirkungen sichtbar, und Selbstorganisationsprozesse oder subjektive Elemente werden nicht beachtet. Die Aussagekraft der Erkenntnisse ist entsprechend beschränkt, ermöglicht aber eine grobe Einschätzung des Wirkungspotenzials von Interventionen der untersuchten Art. Obwohl klare Kriterien dafür fehlen, wann welche Art von Komplexität angemessen ist, lassen sich einige Anhaltspunkte geben (vgl. Tab. 24, Seite 229). Black Box Evaluationen (z.B. experimentelle Studien) erscheinen vor allem dann sinnvoll,

- wenn die Vergleichbarkeit der Zielsysteme, die Vergleichbarkeit der Kontexte, der Standardisierungsgrad der Intervention und die Stabilität während der Umsetzungen als hoch oder als nicht wirkungsrelevant eingeschätzt werden,
- wenn die Komplexität der Intervention und der Zielsysteme als gering oder als nicht wirkungsrelevant eingeschätzt wird,
- wenn unvorhersehbare Entwicklungen als eher unwahrscheinlich oder als nicht wirkungsrelevant beurteilt werden,
- wenn der Zeithorizont der Wirkung eher kurz ist oder als nicht relevant eingeschätzt wird,
- wenn Erkenntnisse über Wirkungsmechanismen nicht erforderlich sind,
- wenn die Programmlogik hoch plausibel ist und alternative Wirkungserklärungen unwahrscheinlich erscheinen,
- und wenn die Interventionen bereits gut evaluiert und wesentliche Wirkungsmechanismen weitgehend bekannt sind (vgl. Tab. 24, Seite 229).

Da es sich bei diesen Anhaltspunkten jedoch zu einem grossen Teil um Aspekte handelt, die je nach Situation oder Beobachter/-in sehr unterschiedlich eingeschätzt werden können, bleibt die Bestimmung des Komplexitätsgrades immer bis einem bestimmten Grad relativ und bedarf der Aushandlung. Die Anschlussfähigkeit an Entscheidungs- und Handlungssysteme sowie an relevante Wissenschaftssysteme bieten dabei wichtige Bezugspunkte. Die Ergebnisse von Wirkungsevaluationen in der Gesundheitsförderung und Prävention müssen einerseits von den Adressaten verstanden werden können und sie müssen Lern- und Entscheidungsprozesse massgeblich unterstützen.

Wirkungsevaluation hat ein weit breiteres Aufgabenspektrum als den Nachweis, ob eine Intervention wirkt oder nicht. In der Gesundheitsförderung und Prävention, in der es darum geht, Menschen zu ermächtigen, ihren Handlungsspielraum für die Gestaltung gesundheitsförderlicher Lebensbedingungen zu nutzen und zu erweitern, scheint die Aufgabe lohnenswerter, Lernprozesse in Richtung Wirkungsentfaltung und Selbstorganisation zu unterstützen, als sich auf das Erbringen stark verallgemeinernder Wirkungsbelege zu beschränken, die entscheidende Fragen offen lassen, und sich nur bedingt auf andere Kontexte übertragen lassen.

Bewusster Umgang mit Komplexität

Aufgrund der hohen Komplexität der Gesundheitsförderung und Prävention müssen Wirkungsevaluationen transparent machen, wie sie mit dieser Komplexität umgehen. Dies bedeutet in Bezug auf die Merkmale komplexer sozialer Systeme:

- Interaktionsvielfalt (*Konnektivität und Interdependenz, Offenheit*): Die Evaluation muss darlegen, welche Elemente und Interaktionen als wesentlich für die Wirkungsentstehung betrachtet werden und wie diese in die Evaluation einfließen. Sie muss weiter darlegen, welche externen Einflüsse auf das Interventionssystem einwirken, welche dabei als besonders wirkungsrelevant eingeschätzt werden und wie die Evaluation diese berücksichtigt.
- Komplexe Wirkungsmechanismen (*Feedback, Multidimensionalität/Nicht-Linearität/Zirkularität, Dynamik*): Sie muss die vermuteten Wirkungsmechanismen explizit machen und begründen, worauf sie bei der Wirkungsbeurteilung fokussiert. Sie muss Potenziale und Grenzen in Bezug auf den gewählten Komplexitätsgrad nachvollziehbar machen.
- Selbstorganisation (*Adaption/Co-Evolution, Reflexivität, Selbstorganisation/operative Geschlossenheit*): Sie muss darlegen, wie sie die Perspektiven und Selbstorganisationsprozesse der Interventions- und Zielsysteme in der Evaluation berücksichtigt. Sie muss ihre eigene Rolle im Zusammenspiel mit diesen Systemen klären und transparent machen.
- Unvorhersehbarkeit (*Emergenz, Unvorhersehbarkeit/Ungewissheit*): Die Evaluation muss aufzeigen, wie sie mit unvorhersehbaren Entwicklungen im Interventionssystem umgeht und wie sie nicht-intendierte Nebenwirkungen berücksichtigt.

Ausweitung von Komplexität

Wenn Gesundheit als Ergebnis komplexer Interaktionsprozesse verstanden wird und Interventionen der Gesundheitsförderung als dynamische partizipative Prozesse, die in verschiedenen Kontexten unterschiedliche Wirkungsdynamiken entfalten, und das Wissen um diese Wirkungsdynamiken als steuerungsrelevant eingeschätzt wird, dann bieten sich komplexitätssensible Evaluationen für deren Beurteilung an. Komplexitätssensible Evaluation lässt sich in ihrer Analyse von den Merkmalen komplexer Systeme leiten und interessiert sich entsprechend für Systemdynamiken, vielfältige Interaktionen zwischen Systemebenen, ungeplante Entwicklungen, wechselseitige Rückkoppelungen und ähnliche Elemente, welche andere Evaluationen gerne ausblenden. Sie untersucht, inwiefern durch die Interventionen Bedingungen geschaffen werden, die es den Zielsystemen ermöglichen, selbständig und angemessen mit komplexen gesundheitlichen Herausforderungen umzugehen. Wenn Wirkungsevaluationen Komplexität zu stark reduzieren, ist es nicht möglich, diese Prozesse und damit die Wirkungsweisen von Interventionen zu verstehen. Nachgewiesene Wirksamkeit wird so zu einer Zuschreibung, die von sehr begrenztem Wert für künftige Interventionsplanungen ist. Je mehr sich die Kontexte und die darin realisierten Interventionen unterscheiden, umso weniger aussagekräftig sind mittels experimenteller Studien gewonnene Erkenntnisse. Es ist nicht möglich, Umsetzungsempfehlungen abzugeben, wenn nicht klar ist, welche Art von Intervention in welchem Kontext wirksam war. Zudem fehlt wertvolles Wissen über Wirkungsmechanismen, das genutzt werden könnte, um künftige Interventionen möglichst optimal an den jeweiligen Kontext anzupassen und dadurch das Wirkungspotenzial zu erhöhen.

Komplexitätsreduktion

Auf der einen Seite entwickeln sich komplexe (soziale) Systeme prinzipiell unvorhersehbar. Gleichzeitig etablieren sich in sozialen Systemen immer Strukturen (Interaktionsmuster, Regeln etc.), auf denen die Systementwicklung und damit auch die Wirkungsentfaltung (selbstreferentiell) aufbauen. Die Kenntnis dieser Strukturen und damit das Verstehen der Wirkungslogik der untersuchten Systeme sind für ein gezieltes Intervenieren wesentlich. Komplexitätssensible Evaluation eignet sich dazu, die-

se Systemstrukturen und Wirkungsmechanismen aufzudecken und ist daher für die Gesundheitsförderung und Prävention besonders wertvoll. Durch iterative Prozesse der Ausweitung und Reduktion von Komplexität, durch den ständigen Wechsel von Hinein- und wieder Herauszoomen, lassen sich einerseits wirkungsrelevante Strukturmerkmale und Prozesse erkennen (Induktion) und andererseits wieder an der Komplexität im Feld testen und plausibilisieren (Deduktion). Die qualitative Sozialforschung hat ein breites Methodenrepertoire entwickelt, welches sich der Verdichtung und (qualitativen) Generalisierung von komplexen Sachverhalten annimmt und auf welches auch Wirkungsevaluation zurückgreifen kann.

Da die Ergebnisse von Wirkungsevaluationen an die relevanten Systeme der Politik, Verwaltung, Wissenschaft und Praxis anschlussfähig sein müssen, muss ihr Komplexitätsgrad auf diese Systeme abgestimmt werden, damit sie von diesen als relevante Umwelt akzeptiert werden. Nur so haben Evaluationsergebnisse das Potenzial, Lernprozesse in diesen Systemen anzuregen und gesundheitsförderliche Veränderungen zu begünstigen. Dabei kann es wichtig sein, Evaluationsergebnisse für die verschiedenen Systeme in unterschiedlichem Mass zu reduzieren und aufzubereiten. Das politische System wird in der Regel weniger Komplexität fordern und verarbeiten können als beispielsweise die umsetzende Praxis oder die Wissenschaft. Komplexitätssensible Evaluation ist gefordert, die durch Öffnung erlangte Komplexität bis zu einem bestimmten Grad wieder zu reduzieren.

Attribution und Kontribution

Die Zuschreibung von beobachteten Veränderungen zu umgesetzten Interventionen ist die zentrale Herausforderung der Wirkungsevaluation. Es ist nicht ‚die Intervention‘ an sich, die wirkt, sondern ein komplexes Zusammenspiel zwischen Interventionssystem, Zielsystemen und Umfeld. Die Intervention kann in den Zielsystemen Entwicklungen anstossen, die vielfältige Selbstorganisationsprozesse in Gang setzen. So entwickelt sich in jedem Kontext ein unterschiedliches komplexes Wirkungsgefüge, in welchem es schwierig ist, Einflüsse der Intervention von anderen Einflüssen zu unterscheiden und beobachtete Veränderungen eindeutig einzelnen Faktoren zuzuordnen. Intervention und Kontext lassen sich nicht trennen. In randomisierten kontrollierten Studien sind plausible Wirkungsnachweise möglich, die sich auf das Potenzial der Intervention im Sinne eines wirkungsvollen Anstosses von Selbstorganisationsprozessen beziehen. Da sie aber die unterschiedlichen ausgelösten Wirkungsmechanismen nicht beurteilen, sondern Komplexität mehrheitlich ausblenden, ist ihre Aussagekraft und Nützlichkeit begrenzt. Zudem gefährdet die Praxisrealität, zum Beispiel durch begrenzte Verfügbarkeit oder durch besondere Ansprüche von Akteuren, die Qualität eines experimentellen Studiendesigns. Die detaillierte und transparente Rekonstruktion von Ursache-Wirkungszusammenhängen hingegen kann zu nachvollziehbaren und hoch plausiblen Wirkungszuschreibungen führen. Die Kombination kontrafaktischer Vergleiche mit detaillierter Rekonstruktion von Wirkungszusammenhängen bietet ein grosses Potenzial zur gegenseitigen Absicherung. Wirkungsbeurteilungen von Betroffenen können die Erkenntnisse ergänzen und zusätzlich plausibilisieren.

Die systematische Beurteilung der Kontribution von Interventionen zu beobachteten Veränderungen ist bisher ein unterbelichtetes Thema in der Evaluation. Gerade in der Gesundheitsförderung und Prävention, wo sich einzelne Interventionen als Mosaiksteine in grössere Wirkungszusammenhänge einbetten, ist die Frage nach dem Wirkungsbeitrag der Intervention bedeutsam. Oft scheitern experimentelle Studien daran, dass sie keine signifikanten Wirkungen auf Ebene der Gesundheitsziele zeigen können, weil der Umfang der Intervention im Verhältnis zu anderen Einflussfaktoren zu unbedeutend ist, um statistisch nachweisbar zu werden. Wirkungsanalysen auf intermediären Zielebenen können dabei helfen, den Beitrag an eine Gesamtwirkung besser abzuschätzen, wie Ackermann et al. (2015) am Beispiel der Evaluation einer nationalen Strategie für ein gesundes Körpergewicht aufzeigen. Methodische Ansätze für eine systematische Analyse von Wirkungsbeiträgen gibt es erst wenige; die Contribution Analysis (Mayne, 2001) zeigt eine vielversprechende Möglichkeit.

Eine Antwort auf das Evidenzdilemma

Angesichts der starken Involvierung der Auftrag Gebenden und weiterer Stakeholder und Zielgruppen in die Wirkungskonstruktion lässt sich die Idee einer objektiven, neutralen Wirkungsbeurteilung durch die Evaluation nicht aufrechterhalten. Auf der anderen Seite darf sich die Evaluation auch nicht auf die Wiedergabe von Wirkungsbeurteilungen von Befragten beschränken, wenn sie ihrer Funktion als Beobachterin zweiter Ordnung gerecht werden will. Evaluation erhält ihren Wert vielmehr durch ein Oszillieren zwischen ihrem eigenen Standpunkt auf Basis wissenschaftlicher Methodik, Theorie und Evidenz und der Perspektive von Akteuren im untersuchten Feld mit ihrer Praxisexpertise (lokale Evidenz). Die ‚Wahrheit‘ liegt weder in der Expertensicht noch in den vielfältigen subjektiven Sichtweisen der Akteure und Betroffenen im Feld allein. Relevante Erkenntnis über Wirkungen entsteht in der Interaktion zwischen diesen beiden Welten, in kontinuierlichen Prozessen der Ausweitung und Reduktion von Komplexität. Evidenzbasierte Gesundheitsförderung und Prävention kann entsprechend nicht darauf reduziert sein zu beantworten, welche Interventionen wirken oder nicht wirken, sondern sie muss Erkenntnisse darüber liefern, welche Art von Interventionen in welchen Kontexten und bei welchen Zielgruppen welches Wirkungspotenzial haben. Der Ansatz der ‚Realist Review‘ (Pawson, Greenhalgh, Harvey & Walshe, 2005) eine auf der ‚Realist Evaluation‘ aufbauende innovative Methodik des systematischen Reviews und einer entsprechenden Evidenzbegründung liefert wertvolle Grundlagen für eine Ausweitung des Evidenzbegriffs in Richtung mehr Komplexität.

Evaluation der Gesundheitsförderung und Prävention liegt im Spannungsfeld zwischen Ansprüchen der Evidence Based Medicine einerseits und der Komplexität sozialer Systeme andererseits. Randomisierte kontrollierte Studien können einen Beitrag zur Beurteilung von Interventionswirkungen liefern. Sie können aber keinen Goldstandard begründen. Die Lösung im Streit um die ‚richtige‘ Form der Evaluierung, wie er derzeit in der Evidenzdebatte der Gesundheitsförderung und Prävention geführt wird (vgl. Kap. 2), liegt nicht in einer einseitigen Stellungnahme und Ablehnung der Gegenseite, sondern vielmehr in einer fundierten und ernsthaften Diskussion des Potenzials und der Grenzen einzelner methodischer Ansätze und insbesondere ihrer Kombination, wie sie im Rahmen der ‚Impact Evaluation‘ teilweise begonnen wurde (Copestake, 2014). In der Evaluationspraxis der Gesundheitsförderung und Prävention ist die Kombination unterschiedlicher Methoden bereits Realität, wenn auch systemische Ansätze noch kaum Verwendung finden. Wenn die vorliegende Studie einen Beitrag zu einem besseren gegenseitigen Verständnis leisten kann, dann hat sich die Mühe gelohnt.

9.5 Synthese

In diesem Kapitel werden die Erkenntnisse der vorliegenden Arbeit noch einmal verdichtet und dargelegt, welchen Beitrag sie zur aktuellen Diskussion rund um Evaluation und Komplexität in der Gesundheitsförderung und Prävention leisten. Der Wert der Arbeit wird darin gesehen, dass sie erstens eine detaillierte Beschreibung der Wirkungskonstruktion in Evaluationen der Gesundheitsförderung und Prävention leistet, zweitens mit der Skizzierung einer komplexitätssensiblen Evaluation eine Alternative zu Black Box Evaluationen für Wirkungsbeurteilungen beschreibt und drittens darauf aufbauend eine kritische Reflexion des Umgangs von Evaluationen mit Komplexität bei der Wirkungsbeurteilung von Interventionen der Gesundheitsförderung und Prävention bietet.

Detaillierte Beschreibung der Wirkungskonstruktion in Evaluationen

Anhand von neun methodisch sehr unterschiedlich angelegten Evaluationen aus dem Feld der Gesundheitsförderung und Prävention wurde vergleichend beschrieben, wie verschiedenen Evaluationen ihren Untersuchungsgegenstand rekonstruieren. Sie haben einen grossen Handlungsspielraum bei der Eingrenzung, Erfassung und Verarbeitung von Komplexität sowie bei der Beurteilung von Wirkungen. Sie rekonstruieren die untersuchten Systeme teilweise mit sehr hoher, teilweise aber nur mit sehr geringer Komplexität. Das aus der Analyse herauskristallisierte Handlungsspektrum ist auf Seite 158 in Tabelle Tab. 19 zusammenfassend dargestellt. Davon haben sich verschiedene Strategien der Komplexitätsreduktion und -ausweitung ableiten lassen, die in Evaluationsdesigns unterschiedlich kombiniert vorkommen (Tab. 21, Seite 220).

Bei den untersuchten Wirkungsnachweisen handelt es sich entsprechend um soziale Konstruktionen, die über einen mehrstufigen systematischen Prozess verlaufen. Mit der Eingrenzung von Komplexität und der Wahl des Evaluationsdesigns werden wichtige Weichen für die Wirkungskonstruktion gestellt – mit grossem Einfluss auf die Art und den Umfang der Erkenntnisse und der ausgewiesenen Wirkungen. Dabei hat sich gezeigt, dass sich Evaluationen bei ihren Wirkungsnachweisen keineswegs nur auf experimentelle oder quasiexperimentelle Evaluationsdesigns abstützen. Vielmehr haben sich idealtypisch zwei grundsätzlich unterschiedliche Zugänge zu Wirkungsbeurteilungen und Wirkungsnachweisen herauskristallisiert: Black Box Evaluationen, die über kontrafaktische Vergleiche und einen messenden Zugang zu Wirkungsnachweisen gelangen und komplexitätssensible Evaluationen, die über detaillierte Rekonstruktionen von Wirkungszusammenhängen und damit über einen verstehenden Zugang zu Wirkungsnachweisen gelangen. Diese beiden Zugänge gehen sehr unterschiedlich mit den Merkmalen komplexer sozialer Systeme und den entsprechenden Herausforderungen und Spannungsfeldern um (vgl. Tab. 20, Seite 219):

Interaktionsvielfalt: Bei der Eingrenzung der untersuchten Systeme und der Identifikation relevanter Wirkungsfaktoren gehen Black Box Evaluationen primär deduktiv vor, indem sie die als relevant betrachteten Faktoren hauptsächlich aus den Interventionszielen und der Literatur ableiten und ihre Messungen darauf abstützen. Sind die Systemgrenzen einmal gezogen, bleiben sie weitgehend starr und neu auftauchende Faktoren werden ausgeblendet, um vergleichende Vorher-Nachher-Messungen nicht zu gefährden. Komplexitätssensible Evaluationen verfolgen einen primär induktiven Zugang, indem Interventionsziele kritisch hinterfragt werden, aktiv nach neuen Hinweisen auf wirkungsrelevante Faktoren gesucht wird und die Systemgrenzen laufend erweitert und angepasst werden.

Komplexität der Wirkungsmechanismen: Black Box Evaluationen reduzieren die Komplexität der Wirkungsmechanismen in Interventionen der Gesundheitsförderung und Prävention auf wenige stark generalisierte Ursache-Wirkungszusammenhänge, die sich in einfachen Wirkungsmodellen abbilden lassen. Komplexitätssensible Evaluationen hingegen fächern eine Vielfalt von Wirkungszusammen-

hängen auf, untersuchen deren Dynamiken und suchen nach relevanten kontext- und zielgruppenspezifischen Unterschieden bei der Umsetzung der Interventionen.

Selbstorganisation: Black Box Evaluationen überprüfen Wirkungskonstruktionen, die von aussen an die Interventionssysteme herangetragen werden, aus einer distanzierten Beobachterperspektive. Komplexitätssensible Evaluationen suchen die Nähe zu den untersuchten Systemen, versuchen systemeigene Sichtweisen und Verständnisse der Interventionslogik herauszuarbeiten und analysieren Prozesse der Selbstorganisation in den untersuchten Systemen.

Unvorhersehbarkeit: Black Box Evaluationen können nur sehr begrenzt auf unvorhergesehene Entwicklungen eingehen, da diese nicht nachträglich in ein (quasi-)experimentelles Design aufgenommen werden können. Vielmehr beschränken sie sich auf die Überprüfung weniger, zu Beginn festgelegter Wirkungszusammenhänge. Komplexitätssensible Evaluationen hingegen nutzen explorative methodische Zugänge, um offen nach Wirkungen zu fragen und um unvorhergesehene Entwicklungen und nichtintendierte (Neben-) Wirkungen erkennen zu können.

Wirkungsbeurteilung: Black Box Evaluationen bauen ihre Wirkungsnachweise auf quantitativen kontrafaktischen Vergleichen auf. Von Wirkung wird dann gesprochen, wenn sich eine Interventionsgruppe im Vergleich zu einer möglichst randomisiert gebildeten Kontrollgruppe signifikant positiver im Sinne der angestrebten Wirkungsziele entwickelt hat. Wirkungen werden als Messergebnisse auf einer stark generalisierten Ebene ausgewiesen. Komplexitätssensible Evaluationen begründen ihre Wirkungsnachweise auf einer ganz anderen Logik. Wirkungen gelten dann als belegt, wenn die detaillierte Rekonstruktion entsprechender Wirkungszusammenhänge aus unterschiedlichen Perspektiven diese als logisch nachvollziehbar und plausibel erscheinen lässt. Wirkungsbeurteilungen werden kontext- und zielgruppenspezifisch vorgenommen, Wirkungszuschreibungen selbstkritisch reflektiert und Wirkungsnachweise sehr vorsichtig verallgemeinert.

Skizzierung einer komplexitätssensiblen Evaluation

Experimentelle und quasi-experimentelle Forschungsdesigns sind in der Literatur umfassend beschrieben. Obwohl es eine Vielzahl von Evaluationsansätzen gibt, die sich unter Berücksichtigung von mehr Komplexität mit der Beurteilung von Wirkungen befasst, fehlen übergreifende Darstellungen, was unter einer komplexitätssensiblen Evaluation verstanden werden könnte. Diese Arbeit will dazu einen Beitrag leisten, indem sie die Erkenntnisse der Untersuchung zu Umrissen einer komplexitätssensiblen Evaluation verdichtet.

Die in Kapitel 9.3 skizzierten Kriterien komplexitätssensibler Evaluation können Evaluationen darin unterstützen, ihre Wirkungskonstruktionen fundiert und transparent zu gestalten und dabei den Merkmalen komplexer sozialer Systeme so weit wie möglich Rechnung zu tragen. Programmtheorien und Wirkungsmodelle können dabei als wichtige Strukturierungshilfe dienlich sein. Sie helfen, trotz komplexer Wirkungsmechanismen den Überblick zu wahren und die vielfältigen Daten theoriebasiert zu bündeln. Nicht nur wissenschaftliches Wissen, sondern auch kontextspezifisches Praxiswissen (lokale Theorie) sind wichtige Grundlagen dafür. Die Untersuchungsergebnisse legen nahe, dass das Bewusstsein für die Komplexität der untersuchten sozialen Systeme zwar teilweise gegeben ist, dass Evaluationen in der Gesundheitsförderung und Prävention die untersuchten Interventionen und Zielsysteme aber kaum als dynamische, adaptive und selbstorganisierende Systeme konzipieren und kaum auf entsprechende systemische Methoden und Instrumente zu deren Analyse zurückgreifen.

Für Wirkungsevaluationen in der Gesundheitsförderung und Prävention bieten sich komplexitätssensible Evaluationen besonders an, um über die Herausarbeitung von Wirkungsmechanismen der hochkomplexen Interventionen kontext- und zielgruppenspezifische Aussagen über die Wirksamkeit der Interventionen und differenziertes steuerungsrelevantes Wissen zu generieren. Je eher Interventionen den Prinzipien der Gesundheitsförderung entsprechen, also partizipative Empowermentprozesse auf verschiedenen Handlungsebenen anregen, desto eher scheinen komplexitätssensible Evaluationen

für eine Wirkungsbeurteilung angemessen zu sein; die Konzepte der Gesundheitsförderung und die Theorien komplexer sozialer Systeme haben eine hohe konzeptionelle Entsprechung. Die Komplexitäts- und die Systemtheorie bilden einen geeigneten Rahmen für die Untersuchung und das Verständnis komplexer sozialer Systeme und bieten sich als theoretische Basis für die Evaluation von Gesundheitsförderung und Prävention an. Indem sie den Blick auf die Interaktionsvielfalt, auf die komplexen Wirkungsmechanismen, auf Selbstorganisation und Unvorhersehbarkeit sozialer Entwicklungen lenken, ermöglichen sie Einblicke in das Wirken von Gesundheitsförderung und Prävention, die weit über die Erkenntnisse experimenteller Forschung hinausreichen.

Vor dem Hintergrund der System- und Komplexitätstheorie sollten Wirkungsevaluationen in der Gesundheitsförderung und Prävention stärker auf Strukturen und Prozesse der Selbstorganisation fokussieren und untersuchen, inwiefern eine Intervention die Selbstorganisation in den Zielsystemen zur Bewältigung gesundheitlicher Herausforderungen tatsächlich unterstützt und so zu einer nachhaltigen Wirkungsentfaltung beiträgt. Dazu bedarf es eines Sinneswandels in der Evaluation: Interventionswirkungen sollen nicht nur an den anvisierten Veränderungen auf Ebene Gesundheitsindikatoren gemessen werden, sondern vermehrt an der Stärkung der Selbstorganisation der Zielsysteme und an entsprechenden Kompetenzen und Ressourcen.

Kritische Reflexion des Umgangs der Evaluation mit Komplexität

Die vorliegende Arbeit liefert eine vertiefte Auseinandersetzung mit dem spannungsgeladenen Thema „Evaluation und Komplexität in der Gesundheitsförderung und Prävention“. Die zentralen Erkenntnisse und herausgearbeiteten Positionen lassen sich wie folgt zusammenfassen:

Jede soziale Intervention ist komplex und jede Evaluation ist grundsätzlich unterkomplex. Auch die Berücksichtigung von mehr Komplexität in einer Evaluation führt nicht aus diesem Dilemma heraus – die Grenzen der Erkenntnis zeigen sich auf jeder Komplexitätsstufe von neuem. Der in der Evidenzdebatte oft gehörte Vorwurf, dass (quasi-) experimentelle Evaluationen die Komplexität sozialer Realität nicht erfassen würden, trifft zu; aber nicht nur für diese Art von Evaluationen, sondern – in unterschiedlichem Ausmass - für alle Evaluationen. Dass experimentelle Studien für die Wirkungsevaluation von Interventionen der Gesundheitsförderung und Prävention *grundsätzlich* abgelehnt werden, ist angesichts der Komplexität dieser Interventionen auf den ersten Blick zwar nachvollziehbar, aber theoretisch nicht haltbar. Jede Evaluation muss Komplexität stark reduzieren und diese Komplexitätsreduktion mit der Aussagekraft der Erkenntnisse legitimieren können. Gut gemachte (quasi-) experimentelle Studien können auf einer übergeordneten Ebene wertvolle Hinweise auf das Wirkungspotenzial von Interventionen geben. Da sie allerdings kaum Rückschlüsse auf die Wirkungsrelevanz zielgruppen-, kontext- oder interventionsspezifischer Besonderheiten ermöglichen und keine Erkenntnisse über differenzierte Wirkungsmechanismen und über nicht vorhersehbare Nebenwirkungen liefern, sind sie für fundiertes Entscheiden und Handeln in der Gesundheitsförderung und Prävention nur begrenzt tauglich. Ein grosses Erkenntnispotenzial liegt in der Kombination von (quasi-) experimentellen Studien mit explorativen Ansätzen.

Das Komplexitätsniveau, auf welchem evaluiert wird, muss für jede Wirkungsevaluation situationsspezifisch festgelegt werden, und zwar mit Blick auf die geforderte Differenziertheit der Erkenntnisse, auf die von den Adressaten noch verstehbare und bewältigbare Komplexität, auf die Grenzen, die sich in der Forschungs- und Handlungspraxis ergeben sowie auf die Anschlussfähigkeit der Untersuchungen und Erkenntnisse an die relevanten Wissenschaftssysteme. Bei der Bestimmung eines angemessenen Komplexitätsgrades können die in Kapitel 8.7 herausgearbeiteten Anhaltspunkte Orientierung geben. Für einen höheren Komplexitätsgrad der Evaluation spricht zum Beispiel, wenn die Ziel- und Interventionssysteme als hoch komplex eingeschätzt werden, die Intervention wenig standardisiert ist und als wenig stabil eingeschätzt wird, wenn Erkenntnisse über Wirkungsmechanismen, zielgruppen- und kontextspezifische Besonderheiten gewünscht werden, wenn bei mehreren parallelen Umsetzungen die Kontexte und Zielsysteme sehr unterschiedlich sind und wenn die Plausibilität der Wirkungs-

logik eher gering ist (vgl. insbesondere Tab. 24, Seite 229). Diese Kriterien dürfen aber nicht darüber hinwegtäuschen, dass entsprechende Einschätzungen letztendlich immer relativ bleiben – eine Erschwernis, die von der prinzipiell unendlichen Komplexität sozialer Systeme herrührt und nicht vermieden werden kann.

Auf Basis der theoretischen und empirischen Erkenntnisse dieser Arbeit wird hier für einen Mittelweg plädiert zwischen Positionen, die experimentelle Studien grundsätzlich als ‚Goldstandard‘ für Wirkungsevaluationen proklamieren und Positionen, die experimentelle Studien als unangemessen für die Gesundheitsförderung und Prävention zurückweisen. Iterative Prozesse der Ausweitung und Reduktion von Komplexität erscheinen vielversprechend, um wirkungsrelevante Struktur- und Prozessmerkmale zu erkennen, intermediäre Wirkungsbeiträge zu erfassen, den Beitrag einer Intervention zu einer beobachteten Gesamtveränderung abzuschätzen und so zu einer plausiblen übergeordneten Wirkungsbeurteilung im Feld der Gesundheitsförderung und Prävention zu gelangen. Die Soziologie kann mit ihrem theoretischen Fundament und ihren forschungsmethodischen Ansätzen wesentlich zum Verständnis von Wirkungszusammenhängen in diesem Handlungsfeld beitragen, welches sich durch ein komplexes Zusammenspiel von sozialen Interaktionen auf der Mikro-, Meso- und Makroebene des gesellschaftlichen Zusammenlebens auszeichnet. In diesem Sinne erhält der in letzter Zeit vermehrt gehörte Ruf nach soziologischem und systemtheoretischem Denken seine Berechtigung.

9.6 Ausblick

Erkenntnisse der Komplexitätstheorie werden in der Evaluation erst seit wenigen Jahren systematisch aufgenommen und bearbeitet. Vorläufer, die sich in der Evaluation mit der Komplexität sozialer Systeme auseinandergesetzt und auf dieser Basis einen spezifischen Evaluationsansatz begründet haben, sind beispielsweise Pawson und Tilley (1997) mit ihrem Ansatz der Realist Evaluation, John Mayne (2001) mit der Contribution Analysis oder Earl und Smutylo (2001) mit dem Outcome Mapping. In den letzten Jahren sind einige bedeutende Publikationen erschienen, die die Themen Evaluation und Komplexität explizit verbinden (u.a. Patton, 2011; Robert Koch Institut, 2012; Forss et al., 2011).

Aufgrund der Untersuchung und eigener Erfahrungen in diesem Feld darf davon ausgegangen werden, dass in Zusammenhang mit komplexen sozialen Systemen eine grosse Lücke zwischen den in der Literatur diskutierten Evaluationsthemen und -ansätzen und der aktuellen Evaluationspraxis klafft. Die Komplexität sozialer Systeme ist in der Evaluationspraxis kaum Thema, und es werden entsprechend auch kaum Wege zu einem systematischen Umgang mit dieser Komplexität sichtbar. Innovative Methoden, die sich auf die Beurteilung von Wirkungen in komplexen sozialen Systemen konzentrieren, finden in der Evaluationspraxis noch wenig Verwendung. Selbstorganisationsprozesse, die für die Wirkungsentfaltung in der Gesundheitsförderung und Prävention entscheidend sind, werden nur in Einzelfällen vertieft untersucht. Vorherrschend sind klassische Evaluationsansätze, wie sie seit vielen Jahren praktiziert werden. Immerhin zeigt sich in der oft vorgefundenen Kombination von hypothesenprüfenden und explorativen Evaluationsmethoden, dass das Bewusstsein für eine gegenseitige Befruchtung dieser beiden Forschungsverständnisse gegeben ist. Wenn die Komplexität sozialer Systeme die Fachdiskussionen in der Evaluation und in der Gesundheitsförderung in den nächsten Jahren weiterhin herausfordert und anregt, stehen interessante Entwicklungen an. Folgende Schritte könnten sich dabei als fruchtbar erweisen:

- Annäherung zwischen Vertreterinnen und Vertretern hypothesenprüfender und explorativer Verfahren, um das Potenzial ihrer Methoden und deren Kombination zur Wirkungsbeurteilung komplexer Interventionen gemeinsam und ohne Berührungängste weiter zu entwickeln.

- Weiterentwicklung und Systematisierung komplexitätssensibler Evaluationsformen, um Potenzial und Grenzen von Evaluationen, die auf der Theorie komplexer sozialer Systeme gründen, herauszuarbeiten, zu diskutieren und zu schärfen.
- Fundierte Auseinandersetzung mit Attribution und Kontribution auf der Grundlage der Komplexitätstheorie. Die ‚Contribution Analysis‘ ist hier als vielversprechender Ansatz zu nennen.
- Ermutigung der Evaluationspraxis, systemische und komplexitätssensible Evaluationsansätze zu nutzen und für Wirkungsbeurteilungen nutzbar zu machen.
- Kritische Reflexion des Potenzials partizipativer Ansätze für Wirkungsevaluationen. Diskussion der Chancen und Risiken einer verstärkten Einbindung von Akteuren aus den Interventions- und Zielsystemen in die Planung und Durchführung der Evaluation sowie in die Interpretation der Ergebnisse.
- Verstärkte Fokussierung von Selbstorganisationspotenzialen in Wirkungsevaluationen der Gesundheitsförderung und Prävention und Entwicklung entsprechender Methoden und Indikatoren.
- Untersuchung des Kosten-Nutzen Verhältnisses von verschiedenen Formen der Wirkungsevaluation im Spektrum zwischen Black Box Evaluationen und komplexitätssensiblen Evaluationen.
- Entwicklung von Konzepten zur systematischen Plausibilisierung von Wirkungsnachweisen.
- Weiterentwicklung von Methoden zur Reduktion von Komplexität in Evaluationen, u.a. zur Verdichtung von komplexen, kontextspezifischen Evaluationsergebnissen zu fundierten Wirkungsaussagen.

Ob die Konzepte der Komplexität zu einem *Paradigmenwechsel* im sozialwissenschaftlichen Forschungsverständnis führen werden, wie dies teilweise gefordert wird, bleibt abzuwarten. Wenn sich neben traditionellen Evaluationsformen komplexitätssensible Evaluationsansätze für Wirkungsbeurteilungen entwickeln und etablieren können, welche Zielsysteme und Interventionen als dynamische, selbstorganisierte, adaptive Systeme verstehen und untersuchen, und wenn das Potenzial fundiert reflektiert wird, welches die Kombination von (quasi-) experimentellen und komplexitätssensiblen Ansätzen bietet, ist bereits viel gewonnen. Eine *Revolution*, welche hypothesenprüfende Verfahren gänzlich verwirft (und das Kind gewissermaßen mit dem Bade ausschüttet), erscheint auf Basis des momentanen Erkenntnisstandes nicht erstrebenswert.

10 Anhang

10.1 Begriffe

Es gehört auch zur Komplexität sozialer Systeme, dass Begriffe unterschiedlich genutzt werden. Das folgende Glossar legt das Begriffsverständnis dar, welches dieser Arbeit zugrunde liegt.

Anspruchsgruppen einer Intervention: Anspruchsgruppen (Stakeholder) einer Intervention sind Gruppen, deren Mitglieder aufgrund ihrer Rolle bestimmte Erwartungen haben bzw. entsprechende Ansprüche an die Intervention geltend machen. Zu den Anspruchsgruppen zählen Personen, die an der Intervention beteiligt sind, die die Intervention beeinflussen oder die selbst von der Intervention beeinflusst werden (Zielgruppen). (Gesundheitsförderung Schweiz, www.quint-essenz.ch/de/concepts).

Attribution: Attribution im Kontext der Wirkungsevaluation meint die Zuschreibung von geplanten und tatsächlich festgestellten Veränderungen zu den umgesetzten Interventionen. Da diese Zuschreibung alles andere als einfach ist, spricht man in der Wirkungsevaluation vom Attributionsproblem (Forss et al., 2011).

Emergenz: Herausbildung von Strukturen und Eigenschaften eines Systems durch die Interaktion der einzelnen Systemkomponenten. Die sich herausbildenden Strukturen und Eigenschaften sind nicht in den einzelnen Systemkomponenten bereits angelegt und deshalb nicht durch diese erklärbar (vgl. Willke, 2006).

Evidenzbasierung: Unter Evidenzbasierung ist die wissenschaftliche Absicherung gemeint, die sich aus der Synthese von Ergebnissen empirischer Studien ergibt. "Evidenzbasierung schliesslich setzt die Wirkungsevaluation mehrerer gleichartiger Interventionsstudien voraus, denn aus deren Zusammenfassung und Beurteilung wird die Evidenzbasis gebildet." (Bödeker, 2012, 33)

Evidence Based Medicine (EBM): „EBM ist der gewissenhafte, eindeutige und kenntnisreiche Gebrauch der gegenwärtig am besten belegten Verfahren zur Entscheidung der Behandlung des individuellen Patienten. Die Umsetzung von Nachweis gestützter Medizin bedeutet die individuelle klinische Erfahrung mit dem besten objektiven Nachweis aus der systematischen Forschung zu verknüpfen.“ (Rossboth & Gay, 2007, 35)

Evidenzbasierte Gesundheitsförderung: Evidenzbasierte Gesundheitsförderung meint die wissenschaftliche Abstützung von Entscheiden und Handeln der Gesundheitsförderung durch die Erkenntnisse aus empirischen Studien. Im Gegensatz zur Evidence Based Medicine orientiert sich die evidenzbasierte Gesundheitsförderung aufgrund der Komplexität ihrer Interventionen nicht an der Evidenzhierarchie der EBM. (vgl. Elkeles & Broesskamp-Stone, 2012)

Evaluation: "Sie ist eine methodisch kontrollierte, verwertungs- und bewertungsorientierte Form des Sammelns und Bewertens von Informationen. Empirisch-wissenschaftliche Forschung wird zur Evaluation, wenn sie einer intersubjektiv geltenden (normativen) Bewertung eines Sachverhaltes dient, die in einem objektivierten Verfahren und anhand explizit gemachter Kriterien und Maßstäbe vorgenommen wird." (Elkeles, 2012, 79). Oder mit Bezug zu Komplexität ausgedrückt: "Evaluation is the systematic investigation of the merit or worth of an object (program) for the purpose of reducing uncertainty in decision making" (Mertens) zit. in (Lee, 2006, 139). Evaluation bezeichnet einerseits eine Methodik, andererseits wird der Begriff in dieser Arbeit aber auch verwendet, um das Evaluationssystem mit seinen Repräsentanten, den Evaluator/-innen, zu bezeichnen.

Experimentelle Studie: Eine experimentelle Studie ist eine Interventionsstudie, in welcher die Ergebnisse eines quantitativen Vorher- Nachher-Vergleichs in einer Interventionsgruppe mit den Ergebnissen in einer Kontrollgruppe verglichen werden, wobei die Gruppenbildung per Zufallsprinzip erfolgt sein muss (auch ‚RCT – Randomized Controlled Trial‘ genannt). In quasiexperimentellen Studien er-

folgt die Zuteilung zu Interventions- und Kontrollgruppe nicht zufällig, sondern nach bestimmten Matchingverfahren.

Gesundheitsförderung und Prävention: Mit dem Begriffspaar „Gesundheitsförderung und Prävention“ sind Public Health-Interventionen gemeint, die das Ziel verfolgen, auf Ebene der Gesamtbevölkerung oder einzelner Bevölkerungsgruppen die Gesundheit zu fördern und Krankheiten zu verhindern. Während sich Krankheitsprävention von ihrer Tradition her primär an zu verhindernden Risikofaktoren orientiert, nimmt die Gesundheitsförderung nicht die Krankheiten, sondern die Determinanten für Gesundheit und Wohlbefinden in den Blick (Altgeld & Kolip, 2010) und zielt darauf ab, gesundheitserhaltende und gesundheitsfördernde Schutzfaktoren zu stärken. In Programmen und Projekten der Gesundheitsförderung und (Primär-)Prävention, um deren Evaluation es in der vorliegenden Studie geht, kommen häufig beide Interventionsarten kombiniert, aber mit unterschiedlicher Ausprägung zur Anwendung. Da primärpräventive Ansätze sich heute weitgehend an den Grundlagen und Prinzipien der Gesundheitsförderung orientieren und es immer schwieriger wird, die Ansätze zu trennen, wird in der vorliegenden Arbeit weitgehend von „Gesundheitsförderung und Prävention“ als einem Handlungsfeld gesprochen.

Impact Evaluation: Oberbegriff für quantitative Wirksamkeitsevaluationen. Die Wirkungsbeurteilung erfolgt mittels kontrafaktischem Vergleich (in der Regel zwischen einer Interventionsgruppe und einer Kontrollgruppe ohne die entsprechende Intervention). Im Vordergrund stehen experimentelle und quasiexperimentelle Designs (eine umfassende Übersicht bieten Khandker et al., 2010).

Intervention: Unter einer Intervention wird ganz allgemein ein Eingriff in ein sonst anders ablaufendes Geschehen verstanden (Trojan, 2012). Als Intervention kann ein Programm oder ein Projekt der Gesundheitsförderung und Prävention gelten, aber auch lediglich eine einzelne Massnahme.

Komplexität: Unter Komplexität wird der Grad der Vielschichtigkeit, Vernetzung und Folgelastigkeit eines Entscheidungsfeldes in Bezug auf ein bestimmtes Problem, welches für ein bestimmtes System in einer bestimmten Situation Entscheidungen erfordert, verstanden. (Willke, 2006) Eigenschaften komplexer Systeme sind etwa: Hohe Konnektivität und Interdependenz, Offenheit, Feedback, Nichtlinearität, Dynamik, Adaption und Co-Evolution, Reflexivität, Selbstorganisation/operative Geschlossenheit, Emergenz und prinzipielle Unvorhersehbarkeit (vgl. Kapitel 4.2 dieser Arbeit).

Kontrafaktische Situation: Die Situation, die (hypothetisch) eingetreten wäre, wenn es keine Maßnahme bzw. kein Projekt für die betreffenden Zielgruppen gegeben hätte. Deren „Herstellung“ (mittels fabrizierter Daten) wird in der experimentaldesigngesteuerten Evaluation und der quasiexperimentaldesigngesteuerten Evaluation auch als „fundamentales Evaluationsproblem“ bezeichnet. (Univation: http://www.univation.org/index.php?class=Calimero_Webpage&id=9025, 18.4.2015).

Kontribution: Im Wissen um die vielfältigen Einflussfaktoren auf gesellschaftliche Veränderungen geht die Wirkungsevaluation davon aus, dass eine bestimmte evaluierte Intervention lediglich einen Beitrag an eine gewünschte Veränderung leistet. Kontribution bezeichnet diesen Beitrag und das Kontributionsproblem verweist auf die Schwierigkeit, den Beitrag einer Intervention an eine beobachtete Veränderung zu bestimmen (vgl. Mayne, 2011).

Mentale Modelle: Mentale Modelle sind modellhafte Repräsentationen der Wirklichkeit in den Köpfen der Menschen (Repräsentationssysteme). (Johnson-Laird, 1983)

Messen: Messen wird definiert als Zuordnung von Zahlen zu Objekten oder Ereignissen, sofern diese Zuordnung eine homomorphe (eindeutige) Abbildung eines empirischen Relativs in ein numerisches Relativ ist. In den Sozialwissenschaften wird der Begriff aber auch breiter verwendet, beispielsweise für die Messung von Einstellungen, Kompetenzen oder auch Wirkungen, wobei hier eine Präzision und Eindeutigkeit suggeriert wird, die streng genommen nicht gegeben ist.

Monitoring: Kontinuierliche oder periodische systematische Datenerfassung zur Überwachung von Prozessen und Ergebnissen (Monitoring bewertet nicht).

Operative Geschlossenheit: Autopoietische Systeme bauen in ihren Operationen immer und ausschliesslich auf ihren aktuellen internen Strukturen und Prozessen auf, sie sind also selbstbezogen und innengesteuert (Simon, 2008), in ihrer Tiefenstruktur geschlossen, damit gänzlich unabhängig und unbeeinflussbar von ihrer Umwelt und dadurch nur um den Preis ihrer eigenen Zerstörung von aussen steuerbar (Willke, 2006; Simon, 2008).

Programm/Interventionsprogramm: Ein Programm ist ein aus einer übergeordneten Strategie abgeleitetes, zielgerichtetes längerfristiges aber zeitlich befristetes Vorhaben, das in der Regel aus mehreren untereinander koordinierten Projekten besteht. (Gesundheitsförderung Schweiz, www.quint-essenz.ch/de/concepts, 18.4.2015)

Projekt/Interventionsprojekt: Ein Projekt bezeichnet ein einmaliges strukturiertes Vorhaben, um bis zu einem bestimmten Zeitpunkt festgelegte Ziele zu erreichen. (Gesundheitsförderung Schweiz, www.quint-essenz.ch/de/concepts, 18.4.2015)

Randomized Controlled Trials (RCT): s. Definition „Experimentelle Studie“.

Selbstevaluation: Unter Selbstevaluation werden systematische, datenbasierte Verfahren der Beschreibung und Bewertung verstanden, bei denen die praxisgestaltenden Akteure identisch sind mit den evaluierenden Akteuren. (socialnet und Univation, www.selbstevaluation.de)

Selbstorganisation: Das Prinzip der Selbstorganisation geht davon aus, dass es weder eine externe Steuerung noch eine interne zentralisierte Kontrollinstanz gibt, welche die Entwicklung eines Systems steuern. Vielmehr entstehen und verändern sich Systemstrukturen über komplexe Interaktionen zwischen dem System und seiner Umwelt.

Wirkungen: Beobachtete Veränderungen, die ‚eindeutig‘ einem Projekt oder Programm zugeschrieben werden können. Wirkungen können positiv oder negativ sein, sie können beabsichtigt wie auch unbeabsichtigt sein. Als wirksam wird eine Intervention dann bezeichnet, wenn vorgängig definierte Wirkungen wissenschaftlich nachgewiesen sind. Im englischen Sprachgebrauch wird zwischen ‚effectiveness‘ und ‚efficacy‘ unterschieden, wobei unter efficacy die potenzielle Wirksamkeit unter idealen, kontrollierten Bedingungen und unter effectiveness die tatsächliche Wirksamkeit unter realen Bedingungen verstanden wird.

Wirkungsevaluation: „Mit einer Wirkungsevaluation wird der Frage nachgegangen, ob Ereignisse Folge einer Intervention sind. Wirkungsevaluation zielt auf das Aufdecken der Ursachen des Ereignisses und betrifft damit eine Kernfrage der Erkenntnistheorie.“ (Bödeker, 2012, 33) Weil Wirkung resp. Wirksamkeit die Frage der Attribution von Wirkungen zu Massnahmen stellt, sind für viele Evaluationstheoretiker Wirkungsevaluationen nur mittels ‚rigoroser‘ Evaluationsdesigns umsetzbar, welche systematische Kontrollgruppenvergleiche beinhalten (vgl. Caspari, 2009; Khandker et al., 2010). Der Begriff Wirkungsevaluation wird in dieser Studie etwas breiter für all jene Evaluationen genutzt, welche sich mit der systematischen Beurteilung der Wirkungen von Interventionen beschäftigen.

10.2 Merkmale komplexer sozialer Systeme

Eigenschaften komplexer sozialer Systeme

Konnektivität und Interdependenz: Hohe Konnektivität und Interdependenz in Sozialsystemen führt über vielfältige Rückkoppelungen und Wechselwirkungen zu hoher Dynamik und Komplexität. Ein sehr hoher Grad an strukturierenden Regeln, welcher die Freiheitsgrade für Entscheidungen reduziert, vermindert allerdings die Kontingenz und damit auch die Komplexität von Systemen wieder.

Offenheit: Soziale Systeme definieren sich in Abgrenzung zu ihrer Umwelt. Ihre Identität erhalten sie nicht nur als Einheit, sondern als System-Umwelt-Differenz. In der Umwelt gibt es beinahe unendlich viele Anschlussmöglichkeiten für soziale Systeme, was die Komplexität grundsätzlich unfassbar macht.

Feedback: Vielfältige negative und positive Rückkoppelungen auf unterschiedlichsten Ebenen und in unterschiedlichen Dimensionen führen zu einem unüberschaubaren Netz an wechselseitigen Beeinflussungen. Während negative Rückkoppelungen stabilisierend wirken, verstärken positive Rückkoppelungen destabilisierende Entwicklungen. Selbst kleinste positive Feedbacks können unverhältnismässig grosse Dynamiken auslösen und Auslöser sein für den Übergang in chaotische Zustände.

Multidimensionalität und Nicht-Linearität: Ursache-Wirkungszusammenhänge in komplexen Systemen sind in der Regel nicht linear, sondern vielschichtig und wechselseitig. Kleine Aktivitäten können unbeabsichtigt große Wirkungen entfalten, und vielfältige Nebenwirkungen sind eher die Regel als Ausnahmen. Zirkuläre Prozesse führen dazu, dass Ursachen und Wirkungen nicht eindeutig unterschieden werden können, weil durch die Rückkoppelungen Ursachen zu Wirkungen werden und umgekehrt.

Dynamik: Vielfältige Interaktionen und Wechselwirkungen innerhalb von einzelnen Systemen aber auch zwischen verschiedenen Systemen, die sich über die Zeit laufend verändern, führen zu einer hohen Dynamik und Komplexität.

Adaption und Co-Evolution: Soziale Systeme sind fähig, sich an veränderte Umweltbedingungen anzupassen - sie entwickeln sich laufend mit den sie umgebenden Systemen weiter.

Reflexivität: Soziale Systeme zeichnen sich dadurch aus, dass sie sich selber thematisieren und sich als Umwelt anderer Systeme begreifen können. Dies ermöglicht ihnen, externe Erwartungen oder Entwicklungen zu erahnen und sich antizipierend zu adaptieren.

Selbstorganisation/operative Geschlossenheit: Soziale Systeme sind trotz Offenheit gegenüber ihrer Umwelt in ihrer Tiefenstruktur autonom, resp. operativ geschlossen. Sie nehmen Umwelteinflüsse als Irritationen oder Perturbationen wahr und organisieren ihre Reaktion darauf selbstorganisierend.

Emergenz: Vielfältige Interaktionen und Wechselwirkungen zwischen Elementen komplexer Systeme führen dazu, dass immer wieder Neues entsteht, das nicht erklärbar oder berechenbar ist auf Basis der Eigenschaften der isolierten Elemente. Hier gilt der Ausdruck: „das Ganze ist mehr als die Summe seiner Einzelteile“.

Unvorhersehbarkeit/Ungewissheit: Die Vielschichtigkeit der Beziehungen zwischen Systemkomponenten untereinander und zu Systemen ihrer Umwelt, die hohe Dynamik und die Emergenz von Strukturen auf der Makroebene führen zu komplexen, nicht-linearen Ursache-Wirkungszusammenhängen, die multidimensional, nicht proportional und dadurch weitgehend unvorhersehbar und ungewiss sind.

10.3 Codesystem

Konzepte/Kategorien 1.-3. Ebene	Kategorien 4. Ebene	Anzahl Codes Ebene 1 und 2
Komplexe Systeme		450
Interaktionsvielfalt		115
Konnektivität und Interdependenz, Offenheit	Unterschiedliche Kontexte; Unterschiedliche Interventionen	
Komplexe Wirkungsmechanismen		194
Feedback		
Rückkoppelungen und Wechselwirkungen beachten		
Nicht-Linearität/Zirkularität		
Dynamik		
Adaption/Co-Evolution		
Komplexitätsreduktion		
Selbstorganisation		96
Reflexivität	Selbstreflexion der Evaluation; Interventionsart beeinflussen; Hinweise auf Nebenwirkungen der Evaluation	
Selbstorganisation/ operative Geschlossenheit		
Emergenz		33
Emergenz		
Unvorhersehbarkeit		
Komplexität eingrenzen		738
Nutzerperspektive		94
Direkte Umsetzbarkeit		
Am Anfang stehen Wirkungshypothesen		
Übergeordnete Programme		
Orientierung an den Interventionszielen	Unterschiedliche Zielformulierungen; Fehlende Referenz	
Orientierung an Praxisnutzen		
Ansprüche der Träger		
Orientierung ökonomisch		
Orientierung an Behörden Politik		
Verständlichkeit und Nachvollziehbarkeit der Ergebnisse		
Methodische Einengung		
Wissenschaftliches Wissen		223
Expertenwissen		
Publikation der Ergebnisse		
Ungenügende/unklare Datenlage		
Bedingungen		
Orientierung an Gesundheit und Ges.-Problemen		
Orientierung an Literatur/Wissenschaft	In Wirkungsaussagen anderer Studien einbetten	
Wirkungsmodelle	Implizite Modelle; Modelle aus der Literatur; Modellierungen; Programmlogik	
Orientierung an wissenschaftlichen Standards		
Wissenslücken		
Orientierung an anderen Interventionsprojekten		
Interdisziplinär vorgehen		
Orientierung an Grundlagen zum Gegenstand		
Orientierung an Voruntersuchungen		
Orientierung an bewährter Forschung		
Methodische Erfordernisse		
Hohe interne Validität		
Praxisrealität		72
Begrenzte Datenverfügbarkeit		
begrenzte Ressourcen/zu hoher Aufwand		
Aussicht auf bessere Ergebnisse		
Beschränkte Verfügbarkeit		
Begrenzte Projekt oder Beobachtungsdauer		
Sich verändernder Kontext	Zeitliche Verzögerungen; Saisonale Schwankungen	
Diverses		
Grenzziehung		344
Deduktive Grenzziehung		
Induktive Grenzziehung		

Mehrere Wirkungsebenen berücksichtigen		
Vielfalt von Einflussfaktoren berücksichtigen		
Verschiedene Perspektiven einfangen	Aus verschiedenen Perspektiven beurteilen; Breites Spektrum an ZG berücksichtigen	
grosse Abdeckung anstreben	Eine hohe Anzahl Personen berücksichtigen; Umfangreiche Daten sammeln	
Identifizieren von Faktoren	Ausschliessen von Faktoren; Auf leicht verfügbare Faktoren abstützen	
Sampling	Verweigerungen; Ausschlüsse; Dropouts; Selektion nicht begründen	
Nebenwirkungen beachten		
Komplexität erfassen		504
Kontexte beschreiben		
Fehlende Daten		
Fehlende Perspektiven		
Mehr Erhebungsaufwand		
Anpassung der Methodik	Grössere Stichprobe erforderlich	
Bestehende Daten nutzen		31
Monitoringdaten nutzen		
Standardisiert erheben		110
Komplexität über Zahlen fassbar machen	Messen; Operationalisieren; Indexbildung; Mit Proxys arbeiten; quantifizieren; nivellieren; Faktoren durch andere Faktoren ausdrücken; Kompromisse machen	
offen (qualitativ) herangehen		41
Nicht voll standardisierte Gespräche		
Offene Fragen in schriftlicher Befragung		
offen erheben		
Erkenntnis durch Nähe erreichen		
beobachten		
breites Spektrum offen einfangen		
Methodenmix		53
Auf Andere abstützen		114
Mit Akteuren im Feld partizipieren	Erhebungen koordinieren	
Auf subjektive Einschätzungen bauen		
Priorisierung den Untersuchten überlassen		
Personen in Gruppen befragen		
Zielgruppen befragen		
Schlüsselpersonen befragen		
stellvertretend befragen		
Zufällig zu Befragende auswählen		
Partizipation an Evaluation kontrollieren		
Retrospektiv fragen		
Prospektiv fragen		
Hypothetisch fragen - Potenziale erfragen		
persönlich befragen (face to face)		
nach Bewertungen fragen		
Nach Ursache-Wirkungszusammenhängen fragen		48
Delegation der Wirkungskonstruktion		
schätzen		28
konservative Schätzung		
Exemplarisch untersuchen (u.a. auch Pilot)		6
Nachfassen		
Grenzen der Erfassung		50
Schwierigkeiten bei der Erfassung managen		
Zu viel des Guten		
Nicht alles untersuchen		
Diverses		12
Komplexität verarbeiten		1266
Spezielle Auswertungen		
Ordnen		163
Begriffe vorgeben	Kategorien im Voraus festlegen; Begriffsverständnis voraussetzen	
klassifizieren	Homogenisieren	
fokussieren		
priorisieren		

zuordnen	
werten	
betonen hervorheben	
kombinieren	
modellieren	
Auf Literatur abstützen	
systematisieren	
visualisieren	
Ausschliessen	
sortieren/in Reihenfolge bringen	
verdichten, zusammenfassen	
Filtern	
Vergleiche anstellen	436
Soll-Ist-Vergleiche quant	
Soll-Ist-Vergleiche qual	
Vorher-Nachher-Vergleich qual und quant	
Entwicklungen über die Zeit	Grössere Zeiträume beobachten; Entwicklungen aufzeigen; Kleine Zeitintervalle beachten; Veränderungen visuell darstellen
Gruppen vergleichen	Randomisieren; Verschiedene Gruppen vergleichen; Gruppen bewusst weglassen; Interventionsgruppe mit Kontrollgruppe vergleichen; Untergruppen vergleichen; Festlegung Gruppengrösse; Nicht Unterschiedliches vergleichen; Ein-Gruppen-Design
Vergleichende Fallstudien	Gemeinsamkeiten
Vergleich mit hypothetischer Situation	
Vergleich mit theoretischer Referenz	
Vergleiche mit normativer Referenz	
Faktoren einander gegenüberstellen	
Erschwerte Vergleichbarkeit	
Differenziert auswerten	Ergebnisse ausdifferenzieren; Differenzierungen sekundär behandeln
Berechnen	131
Potenziale untersuchen - quant.	
zusammenfassen - nivellieren	
statistisch 'kontrollieren'	
Korrelationen untersuchen	
auf statistische Wahrscheinlichkeiten bauen	
Hochrechnen	
statistisch auswerten	
gewichten (statistisch)	
hypothetisch berechnen	
korrigieren	
bereinigen	
Abdeckung berechnen	
Verschiedene Masse vergleichen	
Auf- oder abrunden	
Varianten prüfen	
isolieren	
Berechnungsart wählen und begründen	
Ungenauigkeiten	
Beschreiben	469
Evaluationsfragen beantworten	
Methodik erläutern	
Ergebnisse begründen	
Kontexte beschreiben	
Interventionen beschreiben	
Interventionsprozesse nachzeichnen	
Potenziale/Bedingungen untersuchen	
Wirkungsmechanismen analysieren	
Berechnungsergebnisse erläutern	
Veränderungen präsentieren	
Ergebnisse diskutieren	
auf Nebenwirkungen hinweisen	
Intervention beurteilen	

Ergebnisse mit anderen Studien vergleichen	
Empfehlungen abgeben	
Veränderungen über die Zeit	
Exemplarisch illustrieren/Beispiele anführen	
Passung beurteilen	
Erklären	
Soll-Ist Vergleiche beschreiben	
In die Tiefe gehen	
Unterschiedliche Meinungen nebeneinander stehen lassen	
Deskriptive Statistik	
Kontextsensibel untersuchen	65
Mitbedingende Faktoren berücksichtigen	
Ergebnisse in den Kontext stellen	
Wirkungen beurteilen	1023
Klare Wirkungsaussagen	174
Wirkungsnachweise	Veränderungen in die gewünschte Richtung proklamieren; Nur kurzfristige Wirkungen
Hinweise auf Wirkungen	
Teilwirkungen	
Ungenügende/geringe Wirkungen	
Nichtwirkung	Umgekehrte Wirkung als beabsichtigt
Wirkungskonstruktionen der Befragten wiedergeben	
Wirkungen nicht nachweisen können	18
Unsichere undeutliche Wirkungsaussagen	105
Spannweite für Ergebnisse angeben	
Auf Schätzungen basieren	
Nur Korrelationen	
Wirkungsfrage offen	
Vorsichtige Formulierung von Veränderungen und Wirkungen	
Wahrscheinliche Wirkungen proklamieren	Auf real grössere Wirkungen hinweisen
auf Vermutungen basieren	Vermutungen überprüfen; Auf mögliche weitere Wirkungen verweisen; Annahmen treffen
Ambivalente Wirkungsaussagen machen	
Hinweise auf Ursache-Wirkungszusammenhänge	274
Nutzen von Wirkungsmodellen	
Wirkungen auf Massnahmen zurückführen - Attribution	
Beitrag an Wirkungen proklamieren - Kontribution	
Veränderungen in Wirkungsindikatoren konstatieren	
Wirkungspotenziale	
Nachweise für veränderte Bedingungen für Wirkungen	
Nachhaltigkeit untersuchen	
Lesarten wählen	
Auf Gründe für Wirkung/Nichtwirkung verweisen	
Intervention und Wirkung in Zusammenhang bringen	
auf hinderliche Faktoren hinweisen	
Ergebnisse verallgemeinern/generalisieren	23
Eingeschränkte Generalisierbarkeit	
Repräsentativität anstreben	
recht ungenaue Angaben machen	9
Weitere Konsequenzen	
Zusammenfassend beurteilen	94
Ergebnisse zusammenfassen	Verdichten
Durchschnitte berechnen und vergleichen	
Wirkungsaussagen plausibilisieren	92
Vorab-Plausibilisierung	
Folgestudien verlangen	
Plausible Erklärungen suchen	
Potenzial und Grenzen der Wirkungskonstruktion	135
Grenzen der Intervention	

Wirkungen zur Diskussion stellen	Varianten diskutieren; Szenarien durchspielen; Auf Lücken verweisen; Intuitiv nachvollziehen; Interpretieren; Intervention zur Diskussion stellen; Werturteil abgeben (normativ); Wirkungen in die Zukunft projizieren; Wirkungen aus verschiedenen Perspektiven beurteilen; über die Güte entscheiden
Nutzen kombinierter Methoden	
Hohe Relevanz	
Ausblickend fragen	
Langfristigkeit betonen	
Diskussion RCT	
Bescheidenheit	Ergebnisse vorsichtig formulieren
Grenzen der Aussagekraft	98
Fehlende Informationen	
Schwierigkeiten der Wirkungsprüfung eingestehen	
Grenzen der Beeinflussung durch Intervention anerkennen	
Hinweise auf zweifelhafte/eingeschränkte Daten - Bias	
Grenzen aufgrund zu kleiner Fallzahlen	
Diverses	84
Valorisieren	73
Empfehlungen abgeben	Empfehlungen der Befragten wiedergeben; Verbesserungen anregen
Lernen	
Design	11
begleiten	
Evaluation hat selber Interventionscharakter	
Anonymisieren	
Total Codes	4065

10.4 Auszug Arbeitstabelle axiales Kodieren

Bedingungen Kontext <i>Unter welchen Bedingungen kommen die versch. Strategien zum Einsatz?</i>	Strategien <i>und Konkretisierungen</i>	Konsequenzen <i>bzgl. Wirkungsbeurteilung in GF/P</i>	Bemerkungen <i>Aufgefallenes, offene Fragen etc.</i>
<ul style="list-style-type: none"> - Unterschiedliche Wirkungsebenen, Evaluationsfragen, Indikatoren, die nach unterschiedlicher Methodik rufen. - Komplexität der Wirkungsentstehung, die nicht mit einer Methodik befriedigend beurteilt werden kann. - Unterschiedliche Ansprüche an die Evaluation (Nachweis der Wirksamkeit und Lernen bzgl. Wirkungsentstehung) 	Verschiedene Methoden einsetzen Quantitativ orientierte Methodik eher für Feststellung von Veränderungen und für standardisierte Erfassung von Faktoren von Wirkungsmechanismen. Qualitativ orientierte Methodik für die diff. Beurteilung von Wirkungsmechanismen, Rahmenbedingungen, Prozessen, Zwischentönen,... Kombination von Methoden für Plausibilisierungen, Methoden z.T. Teil aufeinander aufbauend eingesetzt	<ul style="list-style-type: none"> - Breiteres und differenzierteres Bild, macht aber Zusammenführung erforderlich. - Gegenseitige Befruchtung. Ermöglicht komplexeres Verstehen der Wirkungsentstehung. - Plausibilisierung durch Methodentriangulation - Kombination von Wirkungsnachweis und Wirkungsmechanismen 	<ul style="list-style-type: none"> - Vergleiche versus ‚dichte Beschreibung‘ (resp. Kombination) - Jede Methode reduziert Komplexität - z.T. mehr auf Forscherseite (z.B. Entwicklung Erhebungsinstrumente,...), z.T. mehr auf Seite der Untersuchten
<ul style="list-style-type: none"> - Ohne Standardisierung geht es nie (vgl. Begriffe, Leitfäden,...), aber der Grad der Standardisierung variiert - wenn quant. Vergleiche oder Kalkulationen anstehen, dann ist eine hohe Standardisierung erforderlich - Anschluss an die Literatur (Begriffe, Modelle,...) bringt Standardisierung mit sich 	Standardisiert erheben standardisiert fragen, Vorlagen nutzen, Begriffe vorgeben, Monitoringdaten nutzen etc.	<ul style="list-style-type: none"> - Grenzziehung (Grenze der Standardisierung) immer erforderlich - mit steigender Standardisierung weniger Offenheit - Vorteile: Vergleichbarkeit, - Nachteile: Nicht standardisiert Erhobenes kann verborgen bzw. nicht berücksichtigt bleiben -> geringere 	<ul style="list-style-type: none"> - Standardisierung versus Freiheitsgrade bei Modellen (wie bindend sind Modelle aus der Literatur für versch. Kontexte?) - Wie vertragen sich Standardisierung und untersch. Kontexte generell? - Die Standardisierung darf nicht so eng sein, dass Unvorhergese-

Bedingungen Kontext <i>Unter welchen Bedingungen kommen die versch. Strategien zum Einsatz?</i>	Strategien <i>und Konkretisierungen</i>	Konsequenzen <i>bzgl. Wirkungsbeurteilung in GF/P</i>	Bemerkungen <i>Aufgefallenes, offene Fragen etc.</i>
		<i>Tiefe, Detailliertheit -> weniger Verstehen und Lernen bzgl. Wirkungsweisen</i>	<i>henes verborgen bleibt</i>
<ul style="list-style-type: none"> - tendenziell wenn Überprüfung von Wirksamkeit und nicht detaillierte Wirkungsweisen interessieren - Grundlage für Berechnungen - Wenn mengenmässige Aussagen interessieren (Anteil Personen, die...) 	<p>Komplexität über Zahlen erfassen Operationalisieren, Kategorien festlegen, Zählen, messen, Indexe bilden ... Eine besondere Art der Ordnung, welche Reihenfolgen, quant. Vergleiche, Berechnungen,... erlaubt</p>	<ul style="list-style-type: none"> - einfache Vergleichbarkeit über Zahlen, aber Gefahr ungleicher Vergleiche (Äpfel mit Birnen) 	<ul style="list-style-type: none"> - läuft bzgl. Wirkungen auf quantitative Vergleiche hinaus - erfordert eine hohen Grad an Standardisierung - Operationalisierung ist immer auch Selektion (worauf schaue ich?) - Verweis auf Orientierung an... - Impact Evaluation
<ul style="list-style-type: none"> - wenn Wirkungsmechanismen, Hintergründe und Gründe, Details,... interessieren, dann eher offen bei Unkenntnis der interessierenden Indikatoren/Daten - Offenheit beidseitig: offene Methoden brauchen auch Offenheit von der Gegenseite 	<p>Offen herangehen Mit unterschiedlicher Methodik: nicht voll standardisierte Gespräche (u.a. Gruppendiskussionen), offene Fragen in schriftlicher Befragung, beobachten, Erkenntnis durch Nähe erreichen, Kombination versch. offener Methoden</p>	<ul style="list-style-type: none"> - Stärke: Unvorhergesehenes erfassen - Diff. Wirkungsweisen besser erfassen - Dichtes Bild - Bei Offenheit kann die Erfassung an die Situation angepasst werden (-> emergentes Design) - Gefahr des Sich-Verlierens 	<ul style="list-style-type: none"> - Dichte Beschreibung ermöglichen - Offenheit als Schlüssel für das Verständnis von komplexen sozialen Systemen - Bei Offenheit geschieht die Komplexitätsreduktion z.T. stark bei den Untersuchten (Photo elicitation, Interview)
<ul style="list-style-type: none"> - Wirkungen sind immer Konstruktionen - wer macht die Wirkungskonstruktion? Sowohl in schriftlichen Befragungen als auch in Interviews werden Befragte direkt nach Wirkungszusammenhängen (Programmwirkungen, Nebenwirkungen,...) gefragt. - Aber auch über Kennzahlen sind solche Kausalitäten heraus lesbar (z.B. Absenzen aufgrund von Krankheit) 	<p>Nach Kausalitäten fragen Die Konstruktion von Wirkungszusammenhängen wird von der Evaluation teilweise an die Befragten delegiert.</p>	<ul style="list-style-type: none"> - Es handelt sich zunächst lediglich um subjektive Wirkungseinschätzungen - Gefahr, dass "Wirkung" und "Intervention" ganz unterschiedlich gefasst werden - Subjektiv empfundene Wirkungen werden als Wirkungen deklariert, zum Teil aber nicht transparent als 'subjektive' Sichtweisen offen gelegt - Wirkungsaussagen werden z.T. sehr exemplarisch angeführt 	<ul style="list-style-type: none"> - Was ist der Stellenwert subjektiver Wirkungseinschätzungen - Offenes Fragen gibt mehr Raum für nicht Vorhergesehenes, geschlossenes Fragen gibt die Möglichkeit, zuvor Definiertes (erwünschte Wirkungen, Befürchtetes,...) abzufragen. - Link zu Wirkungsmechanismen! - Solche Aussagen sollen transparent gemacht und wenn möglich über andere Quellen plausibilisiert werden
<ul style="list-style-type: none"> - Gründe für die Auswahl berücksichtigter konkreter (potentieller) Nebeneffekte (pos. oder neg): Nebeneffekte bereits im Projektkonzept enthalten, von der Evaluation eingebracht und abgefragt, emergent - in offenen Erhebungen erfahren (nicht unbedingt explizit erhoben) 	<p>Nebenwirkungen erfassen Unbeabsichtigte Nebenwirkungen werden nur sehr punktuell thematisiert und systematisch gesucht</p>	<ul style="list-style-type: none"> - Im besten Fall Überarbeitung während des Prozesses, ansonsten evtl. für eine spätere Umsetzung - Einfließen in Empfehlungen 	<ul style="list-style-type: none"> - Nebenwirkungen wie auch Wirkungen sind zunächst kontextspezifisch zu verstehen - Nebenwirkungen sind auch Konstruktionen, die unterschiedlich und von unterschiedlichen Akteuren konstruiert werden - Wie bringe ich Wirkungen und Nebenwirkungen in ein Verhältnis? Wie wäge ich ab? vgl. EBM: klinische Praxis des Arztes - Übertragung liegt beim Arzt)
<ul style="list-style-type: none"> - Beschränkte Verfügbarkeit von Datenquellen und Akteuren - Unverhältnismässig hoher Aufwand - Nicht-Kenntnis von Einflussfaktoren 	<p>Grenzen der Erfassung anerkennen Erfassung als Grundproblem, so sind z.B. die meisten Erfassungen nur Annäherungen (self-reported, aber auch z.T. Beobachtungen,...) Oft sind Monitoringdaten nicht</p>	<ul style="list-style-type: none"> - Über- oder Unterschätzung von Effekten (z.B. Wirkungen, Kosten-Nutzen Verhältnisse,...) - Ausweichen auf Einschätzungen - Reduzierte Aussage- 	<ul style="list-style-type: none"> -

Bedingungen Kontext <i>Unter welchen Bedingungen kommen die versch. Strategien zum Einsatz?</i>	Strategien <i>und Konkretisierungen</i>	Konsequenzen <i>bzgl. Wirkungsbeurteilung in GF/P</i>	Bemerkungen <i>Aufgefallenes, offene Fragen etc.</i>
<ul style="list-style-type: none"> - Schwierige Operationalisierung und fehlende Instrumente - Unterschiedliche Kontextbedingungen verhindern strenge Standardisierung von Intervention und Evaluation. - Nutzen/Wirkung schwer isolierbar und messbar (Teile eines Ganzen) - emergente Entwicklungen - Sehr viele kleine Geschehnisse können aus Ressourcen- und methodischen Gründen nicht alle erfasst werden - Zeithorizont nicht passend (Wirkungsentfaltung braucht länger als Projektdauer) 	<p>differenziert genug, um sie nutzen zu können</p>	<p>kraft</p> <ul style="list-style-type: none"> - Beschränkte Verallgemeinerbarkeit (weil Kontextbedingungen nicht im Griff) 	

10.5 Literaturverzeichnis

- Abels, H. (2007). *Einführung in die Soziologie. Band 2: Die Individuen in ihrer Gesellschaft* (3rd ed.). *Hagener Studententexte zur Soziologie*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Ackermann, G., Kirschner, M., Guggenbühl, L., Schulte-Abel, B., Klohn, A., & Mattig, T. (2015). Measuring success in obesity prevention: A synthesis of Health Promotion Switzerland's long-term monitoring and evaluation strategy. *Obesity Facts*, 8(1), 17–29.
- Ackermann, G., Bergman, M. M., Heinzmann, C., & Läubli-Loud, M. (2009). Komplexitätsreduktion durch Klassifikationsmodelle in der Gesundheitsförderung und Prävention. In W. Kirch & R. S. T. 2.-2. Middeke Martin und Rychlik (Eds.), *Aspekte der Prävention* (pp. 20–29). Stuttgart: Thieme.
- Ackermann, G., & Pfyl, R. (2010). Vom Umgang der Evaluation mit Komplexität. *LeGes - Gesetzgebung & Evaluation*, (1), 73–83.
- Ackermann, G., & Conrad Zschaber, C. (2011). Wirkungsorientierte Gesundheitsförderung in der schule. In W. Dür (Ed.), *Lehrbuch schulische Gesundheitsförderung* (1st ed., pp. 277–290). Bern: Huber.
- Altgeld, T., & Kolip, P. (2010). Konzepte und Strategien der Gesundheitsförderung. In K. Hurrelmann, T. Klotz, & J. Haisch (Eds.), *Lehrbuch Prävention und Gesundheitsförderung* (3rd ed., pp. 45–56). Bern: Hans Huber.
- American Evaluation Association AEA. (2004). *Guiding Principles for evaluators: Revisions reflected herein ratified by the AEA membership*. Retrieved from <http://www.eval.org/p/cm/ld/fid=51>
- Antonacopoulou, E., & Chiva, R. (2007). The social complexity of organizational learning. *Management learning*, 38, 277–295.
- Bakewell, O. & Garbutt, A. (2005). *The use and abuse of the logical framework approach*. Stockholm.
- Baraldi, C., Corsi, G., & Esposito, E. (2008). *GLU: Glossar zu Niklas Luhmanns Theorie sozialer Systeme* (1. Aufl., [Nachdr.]). Frankfurt am Main: Suhrkamp. Retrieved from <http://www.worldcat.org/oclc/552123292>
- Barić, L., & Conrad, G. (1999). *Gesundheitsförderung in Settings: Konzept, Methodik und Rechenschaftspflicht zur praktischen Anwendung des Settingansatzes der Gesundheitsförderung*. Gamburg: Verlag für Gesundheitsförderung.
- Bateson, G. (1981). Kybernetische Erklärung. In G. Bateson (Ed.), *Ökologie des Geistes* (pp. 515–529). Frankfurt a. M.: Suhrkamp.
- Beck, U. (1986). *Risikogesellschaft. Auf dem Weg in eine andere Moderne*. Frankfurt a. M.: Suhrkamp.
- Becker, H. S. (1990). Generalizing from case studies. In E. W. Eisner & A. Peshkin (Eds.), *Qualitative inquiry in evaluation* (pp. 233–242). New York: Teachers College Press.
- Berger, P. L., & Luckmann, T. (2009). *Die gesellschaftliche Konstruktion der Wirklichkeit* (22nd ed.). Frankfurt a.M.: Fischer.
- Bergman, M. M. (2010). On Concepts and Paradigms in Mixed Methods Research. *Journal of Mixed Methods Research*, 4, 171–175.
- Bertalanffy, L. von. (1969). *General system theory. Foundation Development Applications* (2nd ed.). New York: Braziller.
- Beywl, W. (2006). Demokratie braucht wirkungsorientierte Evaluation - Entwicklungspfade im Kontext der Kinder- und Jugendhilfe. In Projekt exe (Ed.), *Wirkungsevaluation in der Kinder- und Jugendhilfe - Einblicke in die Evaluationspraxis* (pp. 25–46). München: DJI.
- Blank, S., & Löwenbein, O. (2009). Inwiefern kann Evaluation mittels systemtheoretischem Ansatz und Netzwerkansatz die Steuerung von Projekten durch Lernprozesse befördern? *Zeitschrift für Evaluation*, 8(1), 75–85.
- Blättner, B., & Niebuhr, D. (2015). Methodik weiterentwickeln: Evidenzinformierte Entscheidungsfindung in der Gesundheitsförderung. *Prävention Extra*, 38(1).
- Bödeker, W., & Kreis, J. (Eds.). (2006). *Evidenzbasierung in Gesundheitsförderung und Prävention*. Bremerhaven: Verlag für neue Wissenschaft.

- Bödeker, W. (2006). Evidenzbasierung in Gesundheitsförderung und Prävention - Der Wunsch nach Legitimation und das Problem der Nachweisstrenge. In W. Bödeker & J. Kreis (Eds.), *Evidenzbasierung in Gesundheitsförderung und Prävention* (pp. 1–12). Bremerhaven: Verlag für neue Wissenschaft.
- Bödeker, W. (2012). Wirkungen und Wirkungsnachweis bei komplexen Interventionen. In Robert Koch Institut (Ed.), *Evaluation komplexer Interventionsprogramme in der Prävention: Lernende Systeme, lehrreiche Systeme? Beiträge zur Gesundheitsberichterstattung des Bundes*. (pp. 33–42). Berlin: RKI.
- Bodstein, R. (2007). The complexity of the discussion on effectiveness and evidence in health promotion practices. *Promotion & Education*, 14, 16–20.
- Bohnsack, R. (2003). *Rekonstruktive Sozialforschung. Einführung in qualitative Methoden* (5th ed.). Opladen: Leske und Budrich.
- Bortz, J., & Dörig, N. (2003). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. Berlin. Heidelberg. New York: Springer.
- Bowker, G. C., & Leigh Star, S. (1999). *Sorting things out. Classification and its consequences*. Cambridge. London: MIT.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, 3, 77–101.
- Broesskamp-Stone, U. (2008). Best Practice in der Gesundheitsförderung und Prävention - Konzept und Leitlinien für Entscheidungsfindung und fachliches Handeln. In I. Spicker & G. Sprengseis (Eds.), *Gesundheitsförderung stärken. Kritische Aspekte und Lösungsansätze*. Wien: Facultas.
- Broesskamp-Stone, U. (2009). Gute, vielversprechende, beste Praxis? Der Best-Practice-Rahmen für Gesundheitsförderung und Prävention. In P. Kolip & V. Müller (Eds.), *Qualität von Gesundheitsförderung und Prävention* (pp. 115–136). Bern: Huber.
- Bryman, A. (2007). Barriers to integrating quantitative and qualitative research. *Journal of Mixed Methods Research*, 1(1), 8–22.
- Byford, S., & Sefton, T. A. J. (2003). Economic evaluation of complex health and social care interventions. *National Institute Economic Review*, (186), 98–108.
- Byrne, D. (1998). *Complexity theory and the social sciences: An introduction*. London. New York: Routledge.
- Campbell, M., Fitzpatrick, R., Andrew Haines, A., Kinmonth, A. L., Sandercock, P., Spiegelhalter, D., & Tyrer, P. (2000). Framework for design and evaluation of complex interventions to improve health. *BMJ*, 321(7262).
- Capra, F. (2005). Complexity and life. *Theory Culture and Society*, 22(5), 33–44.
- Caspari, A. (2009). 'Rigore' Wirkungsevaluation - methodische und konzeptionelle Ansätze der Wirkungsmessung in der Entwicklungszusammenarbeit. *Zeitschrift für Evaluation*, 8(2), 183–2013.
- Caspari, A. (2012). Chancen der Wirkungsorientierung für die entwicklungspolitische Bildungsarbeit. *Zeitschrift für internationale Bildungsforschung und Entwicklungspädagogik*, 35(2), 11–17.
- Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis*. London: Sage.
- Chen, H.-F. (1990). *Theory-driven evaluations*. London: Sage.
- Cilliers, P. (1998). *Complexity & postmodernism: Understanding complex systems*. New York: Routledge.
- Cilliers, P. (2002). Why we cannot know complex things completely. *Emergence*, 4(1/2), 87–84.
- Cilliers, P. (2005). Complexity, deconstruction and relativism. *Theory, Culture and Society*, 22, 255–267.
- Cochrane, A. L. (1972). *Effectiveness and efficiency: Random reflections on health services*. London: Burgess & Son.
- Coombes, Y. (2004). Evaluating according to purpose and resources: Strengthening the evidence base incrementally. In M. Thorogood & Y. Coombes (Eds.), *Evaluating health promotion: practice and methods* (2nd ed., pp. 27–39). Oxford: Oxford Medical Publications.
- Coombes, Y., & Thorogood, M. (2004). Introduction. In M. Thorogood & Y. Coombes (Eds.), *Evaluating health promotion: practice and methods* (2nd ed., pp. 3–9). Oxford: Oxford Medical Publications.

- Copestake, J. (2014). Credible impact evaluation in complex contexts: Confirmatory and exploratory approaches. *Evaluation*, 20(4), 412–427.
- Corbin, J., & Strauss, A. (1990). Grounded Theory Research: Procedures, Canons, and Evaluative Criteria. *Zeitschrift für Soziologie*, 19(6), 418–427.
- Craig, P., Dieppe, P., Michie, S., Nazareth, I., & Petticrew, M. (2008). Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ*, 337, 694–696. Retrieved from <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2769032/>
- CSDH. (2008). *Closing the gap in a generation: health equity through action on the social determinants of health: Final Report of the Commission on Social Determinants of Health*. Geneva.
- Dahlgren, G. & Whitehead, M. (1991). *Policies and strategies to promote social equity in health*. Stockholm.
- De Bock, F. & Fischer, J. E. (2011). *Gesundheitsförderung im Kindergarten. Evaluation des Programms „Komm mit in das gesunde Boot“ der Baden-Württemberg Stiftung in Kindergärten in Baden-Württemberg*.
- DeGEval. (2008). *Standards für Evaluation*. Mainz. Retrieved from <http://www.degeval.de/degeval-standards/begleitmaterial>
- DeGEval. (2013). *Systemische Ansätze in der Evaluation*. Diskussionspapier der Arbeitsgruppe "Systemische Ansätze" des Arbeitskreises Evaluation von Entwicklungszusammenarbeit in der DeGEval.
- DeGEval. (2014). *Komplexität und Evaluation: Positionspapier des Vorstandes der DeGEval Gesellschaft für Evaluation*.
- Denzin, N. K., & Lincoln, Y. S. (2005). *The Sage handbook of qualitative research*. London: Sage.
- Dörner, D. (2008). *Die Logik des Misslingens. Strategisches Denken in komplexen Situationen* (7th ed.). Reinbek bei Hamburg: Rowolts.
- Dür, W., Pelikan, J. M., & Waldherr, K. (2010). *Theory of health and health promotion in organizations: Part 1 of 6*. Wien.
- Earl, S., Carden, F., & Smutylo, T. (2001). *Outcome mapping. Building learning and reflection into development programs*: International development research centre.
- Egger, G., Spark, R., & Donovan, R. (2005). *Health promotion strategies and methods* (2nd ed.): McGraw-Hill Australia Pty Ltd.
- Ehmig, S., Dunkel, A., & Lenares, J. (2009). *Evaluation des Modulprojekts Kidz-Box im Kanton Zug: Ergebnisse aus Befragungen und Beobachtungen*. Lugano.
- Eisner, M. & Ribeaud, D. (2007). *Zur Evaluation von Gewaltpräventionsmassnahmen. Drei Analysen zur Wirksamkeit von Interventionen* (Forschungsbericht aus der Reihe z-proso No. 06). Zürich.
- Elder-Vass, D. (2007). Luhmann and emergentism: competing paradigms for social systems theory? *Philosophy of the social sciences*, 37(4), 408–432.
- Elkeles, T. (2006). Evaluation von Gesundheitsförderung und Evidenzbasierung? *Zeitschrift für Evaluation*, (1), 39–70.
- Elkeles, T. (2012). Mögliche und machbare Evaluationsdesigns – Gedanken zur Evaluation oder: von Kanonenkugeln und Köchen. In Robert Koch Institut (Ed.), *Evaluation komplexer Interventionsprogramme in der Prävention: Lernende Systeme, lehrreiche Systeme? Beiträge zur Gesundheitsberichterstattung des Bundes*. (pp. 79–88). Berlin: RKI.
- Elkeles, T. & Broesskamp-Stone, U. (2012). *Evidenzbasierte Gesundheitsförderung*. Retrieved from www.bzga.de/leitbegriffe/
- Ezemenari, K., Rudqvist, A., & Subbarao, K. (1999). *Impact evaluation: a note on concepts and methods*. Washington D.C.
- Fetterman, D. M., & Wandersman, A. (2005). *Empowerment Evaluation. Principles in Practice*. New York. London: Guilford.
- Feyerabend, P. (1980). *Erkenntnis für freie Menschen*. Veränderte Ausgabe. Frankfurt a. Main: Suhrkamp.
- Fleck, L. (1980). *Entstehung und Entwicklung einer wissenschaftlichen Tatsache: Einführung in die Lehre vom Denkstil und Denkkollektiv*. Mit einer Einleitung herausgegeben von Lothar Schäfer und Thomas Schnelle. Frankfurt a. M: Suhrkamp.
- Foerster, H. von. (2009). Entdecken oder erfinden: Wie lässt sich Verstehen verstehen? In H. Gumin & H. Meier (Eds.), *Einführung in den Konstruktivismus* (11th ed., pp. 41–88). München Zürich: Piper.

- Forss, K., Marra, M., & Schwartz, R. (Eds.). (2011). *Comparative policy evaluation: Vol. 18. Evaluating the complex: Attribution, contribution, and beyond*. New Brunswick, New Jersey: Transaction Publishers.
- Frank, M. W., Weihofen, W., Duetz Schmucki, M., Nocera, S., & Paccaud, F. (2013). *Public Health Workforce in Switzerland: A National Census*. Zürich.
- Funnell, S. C., & Rogers, P. J. (2011). *Purposeful program theory: Effective use of theories of change and logic models*. San Francisco: Jossey Bass.
- Furrer, C. & Balthasar, A. (2011). *Evaluation der bfu-Präventionskampagne im Schneesport* (bfu-Report No. 66). Bern.
- Galea, S., Hall, C., & Kaplan, G. A. (2009). Social epidemiology and complex system dynamic modeling as applied to health. *International Journal of Drug Policy*, 20, 209–216.
- Gay, J., & Rossboth, D. W. (2007). Wissenschaftstheorie. In D. Rossboth, J. Gay, & V. Lin (Eds.), *Einführung in Evidence Based Medicine. Wissenschaftstheorie, Evidence Based Medicine und Public Health* (pp. 5–31). Wien: Verlags- und Buchhandels AG.
- Gerhardus, A. (2010). Evidence-based Public Health: Ein Gebiet in Entwicklung. In A. Gerhardus, J. Breckenkamp, O. Razum, N. Schmacke, & H. Wenzel (Eds.), *Evidence-based Public Health* (pp. 17–29). Bern: Huber.
- Gertler, P. J., Martinez, S., Premand, P., Rawlings, L. B., & Vermeersch, C. M. (2011). *Impact evaluation in practice*. Washington D.C.: The World Bank. Retrieved from http://siteresources.worldbank.org/EXTHDOFFICE/Resources/5485726-1295455628620/Impact_Evaluation_in_Practice.pdf
- Gladwell, M. (2002). *The tipping point: How little things can make a big difference*. New York: Back Bay Books.
- Glaser, B. G., & Strauss, A. L. (2010). *Grounded Theory. Strategien qualitativer Forschung* (3rd ed.). Bern: Huber.
- Gloor, D. & Meier, H. (2004). *Evaluationsbericht zum Projekt "MuM - mutige Mädchen - Integration von gewaltbereiten Mädchen im unteren Kleinbasel: Im Auftrag des Treffs für Mädchen und junge Frauen, Basler Freizeitaktion BFA, und der mobilen Jugendarbeit Basel*. Zürich.
- Glouberman, S. & Zimmerman, B. (2002). *Complicated and complex systems. What would successful reform of Medicare look like?* Discussion paper. Retrieved from http://www.plexusinstitute.org/resource/collection/6528ED29-9907-4BC7-8D00-8DC907679FED/ComplicatedAndComplexSystems-ZimmermanReport_Medicare_reform.pdf
- Gmel, G. & Wicki, M. (2010). *Effekt der Einschränkung der Erhältlichkeit von Alkohol auf Alkohol-Intoxikationen im Kanton Genf: Abschlussbericht*. Forschungsbericht Nr. 54-A. Lausanne.
- Gmel, G., Labhart, F., Venzin, V., & Marmet, K. (2010). *Kurzinterventionen bei risikoreich Alkohol konsumierenden Berufs- und Mittelschülerinnen und -schülern im Kanton Zürich: Abschlussbericht* (Forschungsbericht aus der Reihe z-proso No. 55). Lausanne.
- Gomez, P., & Probst, G. (1999). *Die Praxis des ganzheitlichen Problemlösens: Vernetzt denken. Unternehmerisch handeln. Persönlich überzeugen* (3rd ed.). Bern. Stuttgart. Wien: Haupt.
- Green, L. W. (2006). Public health asks of systems science: to advance our evidence based-practice, can you help us get more practice-based evidence? *American Journal of Public Health*, 96(3).
- Greene, J. C. (2007). *Mixed methods in social inquiry*. San Francisco: Jossey Bass.
- Grossman, J. (2008). A couple of the nasties lurking in evidence-based medicine. *Social epistemology*, 22(4), 333–352.
- Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation*. London: Sage.
- Gutzwiller, F., & Paccaud, F. (2007). Public Health, Sozial- und Präventivmedizin, Partnerdisziplinen. In F. Gutzwiller & F. Paccaud (Eds.), *Sozial- und Präventivmedizin. Public Health* (3rd ed., pp. 13–21). Bern: Huber.
- Hafen, M. (2007 a). *Grundlagen der systemischen Prävention. Ein Theoriebuch für Lehre und Praxis*. Heidelberg: Carl-Auer-Systeme.
- Hafen, M. (2007 b). *Mythologie der Gesundheit*. Heidelberg: Carl-Auer Systeme.
- Halkow, A. (2004). Mut zum Dialog! Partizipatorische Evaluationsstrategien und ihre Potenziale für die Gesundheitsförderung. In E. Luber & R. Geene (Eds.), *Qualitätssicherung und Evidenzbasierung in der Gesundheitsförderung* (pp. 177–196). Frankfurt am Main: Marbus.
- Hart, D. (2000). Evidenz-basierte Medizin (EBM) und Gesundheitsrecht: Überlegungen zu rechtlichen Konsequenzen der Verwissenschaftlichung der Medizin. *Medizinrecht*, 18(1), 1–5.

- Hawe, P., Shiell, A., & Riley, T. (2004). Complex interventions: How "out of control" can a randomised controlled trial be?: Complex interventions are more than the sum of their parts, and interventions need to be better theorised to reflect this. *British Medical Journal*, 328(7455), 1561–1563.
- Hills, M., & Carroll, S. (2004). Health promotion evaluation, realist synthesis and participation. *Ciência & Saúde Coletiva*, 9(3), 536–539.
- Högnabba, S., Julkunen, I., Kainulainen, S., Korteniemi, P., Lindqvist, T., & Peitola, P. (2005). Steps into realistic evaluation in social work in Finland. In P. Sommerfeld (Ed.), *Evidence-based social work - towards a new professionalism?* (pp. 109–126). Bern: Peter Lang.
- Hoverstadt, P. (2010). The Viable System Model. In M. Reynolds & S. Holwell (Eds.), *Systems approaches to managing change: a practical guide* (pp. 87–134). London: Springer Press.
- Hsieh, H.-F. & S. S. E. (2005). Three approaches to Qualitative Content Analysis. *Qualitative health research*, 15(9), 1277–1288.
- Huber, M., Knottnerus J. André, Green, L., van der Horst, H., Jadad, A. R., Kromhout, D., ... (2011). How should we define health? *British Medical Journal*, (343:d4163).
- Israel, B., Schulz, A., Parker, E., & Becker, A. (1998). Review of community-based research: Assessing partnership approaches to improve public health. *Annual Review of Public Health*, 19, 173–202.
- Jenny, G., Inauen, A., Brauchli, R., Füllemann, D., Müller, F., Bauer, G., ... (2011). *Projekt SWING - Schlussbericht der Evaluation*. Zürich.
- Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. London: Cambridge University Press.
- Jones, H. (2011a). *Taking responsibility for complexity: When is a policy problem complex, why does it matter and how can it be tackled?* Briefing paper 68. London.
- Jones, H. (2011b). *Taking responsibility for complexity. How implementation can achieve results in the face of complex problems*. Working Paper 330. London.
- Khandker, S. R., Koolwal, G. B., & Samad, H. A. (2010). *Handbook on Impact evaluation. Quantitative methods and practices*. Washington D.C: The World Bank.
- King, N., & Horrocks, C. (2010). *Interviews in qualitative research*. Thousand Oaks: Sage.
- Klir, G. J. (Ed.). (2001). *IFSR International series on systems science and engineering: Vol. 15. Facets of systems science* (2nd ed.). New York: Kluwer Academic/Plenum Publishers.
- Kneer, G., & Nassehi, A. (2000). *Niklas Luhmanns Theorie sozialer Systeme* (4th ed.). München: Wilhelm Fink.
- Kolberg, L. C. (n.d.). *Case studies and generalization: Paper submitted to doctoral course DRE 1002 Case studies methodology*. Retrieved from http://www.google.ch/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CCUQFjAA&url=http%3A%2F%2Fwww.bi.no%2Fupload%2FHjemmesider%2Ffgl00076%2FPaper%2520DRE%25201002.pdf&ei=PljCU_nnNuii0QXY_4CgAw&usg=AFQjCNE3DI5m62Q0_YiCzoZ3L6BZ1y2EZg&bvm=bv.70810081,d.d2k&cad=rja
- Kolip, P., Ackermann, G., Ruckstuhl, B., & Studer, H. (2012). *Gesundheitsförderung mit System: quintessenz - Qualitätsentwicklung in Projekten*. Bern: Huber.
- Kuhn, J., Lampert, T., & Ziese, T. (2012). Einführung ins Thema: Komplexe Interventionen - komplexe Evaluationen? In Robert Koch Institut (Ed.), *Evaluation komplexer Interventionsprogramme in der Prävention: Lernende Systeme, lehrreiche Systeme? Beiträge zur Gesundheitsberichterstattung des Bundes*. (pp. 9–14). Berlin: RKI.
- Kuhn, T. S. (1976). *Die Struktur wissenschaftlicher Revolutionen* (2nd ed.). Frankfurt a. M: Suhrkamp.
- Laat, B. de. (2013). Evaluator, evaluand, evaluation commissioner. A tricky triangle. In M. Läubli Loud & J. Mayne (Eds.), *Enhancing evaluation use. Insights from internal evaluation units* (pp. 15–36). London: Sage.
- Lee, B. (2006). Theories of evaluation. In R. Stockmann (Ed.), *Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder* (pp. 137–176). Münster: Waxmann.
- Lessard, C., & Birch, S. (2010). Complex problems or simple solutions? Enhancing evidence-based economics to reflect reality. In I. Shemilt, M. Mugford, L. Vale, K. Marsh, & C. Donaldson (Eds.), *Evidence-based Decisions and Economics: Health Care, Social Welfare, Education and Criminal Justice* (pp. 162–172). Oxford: Blackwell Publishing.
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. London: Sage.

- Luber, E. (2004). Wessen Qualität ist gemeint? In E. Luber & R. Geene (Eds.), *Qualitätssicherung und Evidenzbasierung in der Gesundheitsförderung* (pp. 37–58). Frankfurt am Main: Marbus.
- Luber, E., & Geene, R. (Eds.). (2004). *Qualitätssicherung und Evidenzbasierung in der Gesundheitsförderung*. Frankfurt am Main: Marbus.
- Luhmann, N. (1970). *Soziologische Aufklärung: Aufsätze zur Theorie sozialer Systeme* (Vol. 1). Opladen: Westdeutscher Verlag.
- Luhmann, N. (1987). *Soziale Systeme. Grundriss einer allgemeinen Theorie*. Frankfurt a. Main: Suhrkamp.
- Luhmann, N. (1992). *Die Wissenschaft der Gesellschaft*. Suhrkamp: Suhrkamp.
- Luhmann, N. (2008). *Einführung in die Systemtheorie* (4th ed.). Heidelberg: Carl Auer.
- Lüssi, P. (1995). *Systemische Sozialarbeit. Praktisches Lehrbuch der Sozialberatung*. Bern: Haupt.
- Mader, L. (2009). Die institutionelle Einbettung der Evaluationsfunktion in der Schweiz. In T. Widmer, W. Beywl, & C. Fabian (Eds.), *Evaluation. Ein systematisches Handbuch* (pp. 52–63). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Mainzer, K. (n.d.). *Was sind komplexe Systeme? Komplexitätsforschung als integrative Wissenschaft*. Retrieved from http://www.integrative-wissenschaft.de/Archiv/dokumente/Mainzer-14_10_04.pdf
- Malik, F. (1993). *Systemisches Management, Evolution, Selbstorganisation*. Bern Stuttgart Wien: Haupt.
- Malik, F. (2008). *Strategie des Managements komplexer Systeme*. Bern Stuttgart Wien.
- Mathews, M. K., White, M. C., & Long, R. G. (1999). Why study the complexity sciences in the social sciences? *Human relations*, 52(4), 439–462.
- Mayne, J. (2001). Addressing attribution through contribution analysis: Using performance measures sensibly. *The Canadian Journal of Program Evaluation*, 16(1), 452–92.
- Mayne, J. (2011). Contribution analysis: Addressing cause and effect. In K. Forss, M. Marra, & R. Schwartz (Eds.), *Comparative policy evaluation: Vol. 18. Evaluating the complex. Attribution, contribution, and beyond* (pp. 53–95). New Brunswick, New Jersey: Transaction Publishers.
- Mayring, P. (2007). Generalisierung in qualitativer Forschung. *Forum: Qualitative Sozialforschung*, 8(Art 26). Retrieved from <http://www.qualitative-research.net/index.php/fqs/article/view/291>
- Mayring, P. G.-Z. M. (Ed.). (2008). *Pädagogik. Die Praxis der qualitativen Inhaltsanalyse* (2nd ed.). Weinheim und Basel: Beltz.
- Mc Queen, D. (2000). Perspectives on health promotion: theory, evidence, practice and the emergence of complexity. *Health Promotion International*, 15(2), 95–97.
- Mc Queen, D. V. (2001). Strengthening the evidence base for health promotion. *Health Promotion International*, 16(3), 261–268.
- Mc Queen, D. V. (2007). Critical issues in theory for health promotion. In D. Mc Queen, I. Kickbusch, L. Potvin, J. M. Pelikan, L. Balbo, & T. Abel (Eds.), *Health and modernity. The role of theory in health promotion* (pp. 21–42). Berlin. Heidelberg. New York: Springer.
- Mc Queen, D. V. (2013). The nature of causality: Beyond traditional evidence. In D. V. Mc Queen (Ed.), *Global Handbook on Noncommunicable Diseases and Health Promotion* (pp. 37–49). New York: Springer.
- Mc Queen, D. V., & Anderson, L. M. (2001). What counts as evidence: issues and debates. In WHO (Ed.), *European Series. Evaluation in health promotion. Principles and perspectives* (pp. 63–81). WHO Regional Publications.
- Mc Queen, D., Kickbusch, I., Potvin, L., Pelikan, J. M., Balbo, L., & Abel, T. (Eds.). (2007). *Health and modernity. The role of theory in health promotion*. Berlin. Heidelberg. New York: Springer.
- Means, R., & Smith, R. (1988). Implementing a pluralistic approach to evaluation in health evaluation. *Policy and Politics*, 16(1), 17–28.
- Merton, R. K. (1968). *Social theory and social structure*: Free press.
- Montuori, A. (2003). The complexity of improvisation and the improvisation of complexity: *Human relations*, 56(2), 237–255.
- Morecroft, J. (2010). System dynamics. In M. Reynolds & S. Holwell (Eds.), *Systems approaches to managing change: a practical guide* (pp. 25–85). London: Springer Press.
- Mowles, C. (2014). Complex, but not quite complex enough: The turn to the complexity sciences in evaluation scholarship. *Evaluation*, 20(2), 160–175.

- Mühlhauser, I., Lenz, M., & Meyer, G. (2012). Entwicklung, Bewertung und Synthese von komplexen Interventionen - eine methodische Herausforderung. In Robert Koch Institut (Ed.), *Evaluation komplexer Interventionsprogramme in der Prävention: Lernende Systeme, lehrreiche Systeme? Beiträge zur Gesundheitsberichterstattung des Bundes*. (pp. 43–55). Berlin: RKI.
- Müller-Benedict, V. (2011). *Grundkurs Statistik in den Sozialwissenschaften*. Lehrbuch (5th ed.). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Naaldenberg, J., Vaandrager, L., Koelen, M., Wagemakers, A.-M., Saan, H., & Hoog, K. de. (2009). Elaborating on systems thinking in health promotion practice. *Global Health Promotion*, 16(1), 39–47.
- Ninacs, W. A. (2008). *Empowerment et intervention*. Laval: Les Presses de l'Université.
- Nitsch, M., Waldherr, K., Denk, E., Griebler, U., Marent, B., & Rorster, R. (2013). Participation by different stakeholders in participatory evaluation of health promotion: A literature review. *Evaluation and program planning*, 40, 42–54.
- Noack, H. (2002, November). *Evidenzbasierte Gesundheitsförderung als Strategie einer QM-gestützten kommunalen Gesundheitspolitik*. Referat auf dem Forum der Akademie für öffentliches Gesundheitswesen, Düsseldorf.
- Norad (Ed.). (1999). *The Logical Framework Approach (LFA): Handbook for objectives-oriented planning* (4th ed.). Oslo: Norad. Norwegian Agency for Development Cooperation.
- Nowotny, H. (2005). The increase of complexity and its reduction: emergent interfaces between the natural sciences, humanities and social sciences. *Theory, Culture and Society*, 22, 15–31.
- Nutbeam, D. (1998). Evaluating health promotion - progress, problems and solutions. *Health Promotion International*, 13(1), 27–44.
- Nutbeam, D. (2000). Health literacy as a public health goal: a challenge for contemporary health education and communication strategies into the 21st century. *Health Promotion International*, 15(3), 259–267.
- Oliver, S., Harden, A., Rees, R., Shepherd, J., Brunton, G., Garcia Jo, & Oakley, A. (2005). An emerging framework for including different types of evidence in systematic reviews for public policy. *Evaluation*, 11, 428–446.
- Oxford Centre for Evidence Based medicine. (2009). *Levels of evidence*. Retrieved from www.cebm.net
- Page, S. E. (2008). Uncertainty, difficulty and complexity. *Journal of theoretical politics*, 20, 115–149.
- Parker, R. M., Schaller, J., & Hansmann, S. (2003). Catastrophe, chaos and complexity models and psychosocial adjustment to disability. *Rehabilitation counselling bulletin*, 46, 234–241.
- Patton, M. Q. (2008). Advocacy impact evaluation. *Journal of Multidisciplinary Evaluation*, 5(9), 1–10.
- Patton, M. Q. (2011). *Developmental evaluation. Applying Complexity concepts to enhance innovation and use*. New York: Guilford.
- Pawson, R., & Tilley, N. (1997). *Realistic evaluation*. London: Sage.
- Pawson, R., Greenhalgh, T., Harvey, G., & Walshe, K. (2005). Realist review – a new method of systematic review designed for complex policy interventions. *Journal of Health Services Research & Policy*, 10(1), 21–34.
- Pelikan, J. M. (2007a). Understanding differentiation of health in late modernity by use of sociological systems theory. In D. Mc Queen, I. Kickbusch, L. Potvin, J. M. Pelikan, L. Balbo, & T. Abel (Eds.), *Health and modernity. The role of theory in health promotion* (pp. 74–102). Berlin. Heidelberg. New York: Springer.
- Pelikan, J. M. (2007b). Zur Rekonstruktion und Rehabilitation eines absonderlichen Funktionssystems - Medizin und Krankenbehandlung bei Niklas Luhmann und in der Folgerezeption. *Soziale Systeme*, 13(1 + 2), 290–303.
- Peters, M. & Wapf, B. (2006). *Evaluation der Wirkungen lokaler Alkoholpolitiken – eine vergleichende Fallstudienanalyse: Schlussbericht*. Zürich.
- Petticrew, M. (2011). When are complex interventions 'complex'? When are complex interventions 'simple'? *European Journal of Public Health*, 21(4), 397–399.
- Plottu, B., & Plottu, E. (2009). Contraintes et vertus de l'évaluation participative. *Revue française de gestion*, 40(2), 31–56.
- Potvin, L. (2007). Managing uncertainty through participation. In D. Mc Queen, I. Kickbusch, L. Potvin, J. M. Pelikan, L. Balbo, & T. Abel (Eds.), *Health and modernity. The role of theory in health promotion* (pp. 103–128). Berlin. Heidelberg. New York: Springer.

- Potvin, L., & Richard, L. (2001). Evaluating community health promotion programmes. In WHO (Ed.), *European Series. Evaluation in health promotion. Principles and perspectives* (pp. 213–240). WHO Regional Publications.
- Prigogine, I., & Stengers, I. (1984). *Order out of chaos: man's new dialogue with nature*. New York: Bantam.
- Prigogine, I. (1997). *The end of certainty: Time, chaos, and the new laws of nature*. New York: Free press.
- Ramalingam, B., & Jones, H. (2008). *Exploring the science of complexity: Ideas and implications for development and humanitarian efforts*: Overseas development Institute.
- Reynolds, M. (2014). Equity-focused developmental evaluation using critical systems thinking. *Evaluation*, 20(1), 75–95.
- Robert Koch Institut (Ed.). (2012). *Evaluation komplexer Interventionsprogramme in der Prävention: Lernende Systeme, lehrreiche Systeme?* Beiträge zur Gesundheitsberichterstattung des Bundes. Berlin: RKI.
- Robertson, A. (1998). Shifting discourses on health in Canada: from health promotion to population health. *Health Promotion International*, 13, 155–166.
- Rogers, E. S. (1968). Public health asks of sociology...: Can the health sciences resolve society's problems in the absence of a science of human values and goals? *Science*, 159, 506–508. Retrieved from <https://vpn1.mobile.unibas.ch/+CSCO+dh756767633A2F2F6A6A6A2E667076726170727A6E742E626574++/content/159/3814/506.full.pdf>
- Rogers, P. J. (2008). Using programme theory to evaluate complicated and complex aspects of interventions. *Evaluation. Sage publications*, 14(1), 19–48.
- Rosenbrock, R. (2004). Qualitätssicherung und Evidenzbasierung - Herausforderungen und Chancen für die Gesundheitsförderung. In E. Luber & R. Geene (Eds.), *Qualitätssicherung und Evidenzbasierung in der Gesundheitsförderung*. Frankfurt am Main: Marbus.
- Rossboth, D. W., & Gay, J. (2007). Nachweis gestützte Medizin (EBM - Evidence Based Medicine). In D. Rossboth, J. Gay, & V. Lin (Eds.), *Einführung in Evidence Based Medicine. Wissenschaftstheorie, Evidence Based Medicine und Public Health* (pp. 32–153). Wien: Verlags- und Buchhandels AG.
- Rossboth, D., Gay, J., & Lin, V. (Eds.). (2007). *Einführung in Evidence Based Medicine. Wissenschaftstheorie, Evidence Based Medicine und Public Health*. Wien: Verlags- und Buchhandels AG.
- Rothwell, P. M. (2005). External validity of randomised controlled trials: "To whom do the results of this trial apply?". *Lancet*, 365(1), 82–93.
- Ruckstuhl, B. (2003). Evidenzbasierte Gesundheitsförderung. In BZgA Bundeszentrale für Gesundheitliche Aufklärung (Ed.), *Blickpunkt Gesundheit: Vol. 6. Leitbegriffe der Gesundheitsförderung* (4th ed., pp. 34–36). Peter Sabo.
- Ruckstuhl, B. (2011). *Gesundheitsförderung. Entwicklungsgeschichte einer neuen Public Health-Perspektive. Grundlagentexte Gesundheitswissenschaften*. Weinheim und München: Juventa.
- Rutter, H. (2012). The single most important intervention to tackle obesity...: Editorial. *International Journal of Public Health*, (57), 657658.
- Saan, H. (2005). The road to evidence: the European path. *Promotion & Education*, 12(1), 6–7.
- Sackett, D. L., Rosenberg, W. M., Gray, J. A. M., Haynes, R. B., & Richardson, W. S. (1996). Evidence based medicine: what it is and what it isn't. *British Medical Journal*, 312(7023), 71–72. Retrieved from <http://www.bmj.com/content/312/7023/71.full>
- Sawyer, K. R. (2005). *Social emergence. Societies as complex systems*: Cambridge University Press.
- Scriven, M. (1991a). *Evaluation Theasaurus* (4th ed.). Newbury Park. London. New Dehli: Sage.
- Scriven, M. (1991b). Prose and cons about goal-free evaluation. *Evaluation Practice*, 12(1), 55–76.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Belmont: Wadsworth.
- Shiell, A., Hawe, P., & Gold, L. (2008). Complex interventions or complex systems? Implications for health economic evaluation. *British Medical Journal*, 336(7656), 1281–1283.
- Simon, F. B. (2008). *Einführung in Systemtheorie und Konstruktivismus* (3rd ed.). Heidelberg: Carl Auer.

- Sommerfeld, P. (Ed.). (2005). *Evidence-based social work - towards a new professionalism?* Bern: Peter Lang.
- Sommerfeld, P., Baumgartner, E., Baur, R., Klemenz, R., Müller, S., & Piller, E. M. (2006). *Evaluation "Midnight Basketball" Projekte. Bericht im Auftrag der Gesundheitsförderung Schweiz.*
- Speller, V., Learmouth, A., & Harrison, D. (1997). The search for evidence of effective health promotion. *British Medical Journal*, 315, 361–362.
- Spencer, B., Broesskamp, U., Ruckstuhl, B., Ackermann, G., Spoerri, A., & Cloetta, B. (2008). Modelling the results of health promotion activities in Switzerland: development of the Swiss Model for Outcome Classification in Health Promotion and Prevention. *Health Promotion International*, 23(1), 86–97.
- Staub-Bernasconi, S. (2007). *Soziale Arbeit als Handlungswissenschaft. Systemische Grundlagen und professionelle Praxis – ein Lehrbuch.* Stuttgart: UTB.
- Steinke, I. (1999). *Kriterien qualitativer Forschung.* Weinheim: Juventa.
- Steinke, I. (2007). Gütekriterien qualitativer Forschung. In U. Flick, E. von Kardorff, & I. Steinke (Eds.), *Qualitative Forschung. Ein Handbuch* (5th ed., pp. 319–331). Rohwolts.
- Stern, E., Stame, N., Mayne, J., Forss, K., Davies, R., & Befani, b. (2012). *Broadening the range of designs and methods for impact evaluations: Report of a study commissioned by the Department for International Development.* Working paper 38.
- Stevens, W. (2004). Economic evaluation of health promotion interventions. In M. Thorogood & Y. Coombes (Eds.), *Evaluating health promotion: practice and methods* (2nd ed., pp. 41–55). Oxford: Oxford Medical Publications.
- Stockmann, R. (2010). *10 Thesen zur Evaluation der Wirksamkeit der Entwicklungszusammenarbeit* (CEval Arbeitspapiere No. 18). Saarbrücken.
- Strauss, A., & Corbin, J. (1996). *Grounded Theory: Grundlagen Qualitativer Sozialforschung.* Weinheim: Psychologische Verlagsunion.
- Strübing, J. (2008). *Grounded theory.: Zur sozialtheoretischen und epistemologischen Fundierung des Verfahrens der empirisch begründeten Theoriebildung.* (2nd ed.). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Stufflebeam, D. L. (2003). *The CIPP Model for evaluation - an update.* Annual conference of the Oregon Program Evaluators Network. Portland.
- Taleb, N. N. (2007). *The black swan: The impact of the highly improbable.* London: Penguin.
- Tang, C. C. (2007). Struktur/Ereignis: Eine unterentwickelte, aber vielversprechende Unterscheidung in der Systemtheorie Niklas Luhmanns. *Soziale Systeme*, 13(1+2), 86–98.
- Tang, K. C., Ehsani, J., & Mc Queen, D. V. (2003). Evidence-based health promotion: Recollections, reflections and reconsiderations. *Journal of Epidemiology and Community Health*, 57, 841–843.
- Thorogood, M., & Coombes, Y. (Eds.). (2004). *Evaluating health promotion: practice and methods* (2nd ed.). Oxford: Oxford Medical Publications.
- Thrift, N. (1999). The Place of complexity. *Theory, Culture and Society*, 16(3), 31–69.
- Tones, K. (1997). Beyond the randomized controlled trial: a case for "judicial review". *Health Education Research*, 12(2), i–iii.
- Tones, K. (2000). Evaluating health promotion: a tale of three errors. *Patient Education and Counseling*, 39, 227–236. Retrieved from http://cys.bvsalud.org/lildbi/docsonline/2/8/182-88-Evaluating_health_promotion.pdf
- Tremblay, M.-C., & Richard, L. (2011). Complexity: a potential paradigm for a health promotion discipline. *Health Promotion International*, 29(2), 378–388. Retrieved from <http://heapro.oxfordjournals.org/content/early/2011/09/07/heapro.dar054.full>
- Trojan, A. (2006). Zu Chancen und Grenzen der Evidenzbasierung komplexer sozialer Intervention. In W. Bödeker & J. Kreis (Eds.), *Evidenzbasierung in Gesundheitsförderung und Prävention.* Bremerhaven: Verlag für neue Wissenschaft.
- Trojan, A. (2012). Vor dem Messen und Rechnen: Die Landschaft beschreiben.: Überlegungen für eine Klassifizierung und einheitliche Terminologie von Gesundheitsförderungsinterventionen für Evaluation und Evidenzbasierung. In Robert Koch Institut (Ed.), *Evaluation komplexer Interventionsprogramme in der Prävention: Lernende Systeme, lehrreiche Systeme? Beiträge zur Gesundheitsberichterstattung des Bundes.* (pp. 21–32). Berlin: RKI.
- Uhl, A. (2009). Absurditäten in der Suchtforschung. *Wiener Zeitschrift für Suchtforschung*, 32(3/4), 19–39.

- Uhl, A. (2012). Methodenprobleme bei der Evaluation komplexerer Sachverhalte: Das Beispiel Suchtprävention. In Robert Koch Institut (Ed.), *Evaluation komplexer Interventionsprogramme in der Prävention: Lernende Systeme, lehrreiche Systeme? Beiträge zur Gesundheitsberichterstattung des Bundes*. (pp. 57–78). Berlin: RKI.
- Ulrich, W., & Reynolds, M. (2010). Critical Systems Heuristics. In M. Reynolds & S. Holwell (Eds.), *Systems approaches to managing change: a practical guide* (pp. 243–292). London: Springer Press.
- Unicef (Ed.). (2008). *Bridging the gap: The role of monitoring and evaluation in evidence-based policy-making*.
- Urry, J. (2005). The complexities of the global. *Theory, Culture and Society*, 22, 235–254.
- Urry, J. (2006). Complexity. *Theory Culture and Society*, 23, 111–115.
- van Ongevalle, J., Huyse, H., & van Petegem, P. (2014). Dealing with complexity through actor-focused planning, monitoring and evaluation (PME). *Evaluation*, 20(4), 447–466.
- Vester, F. (2007). *The art of interconnected thinking. Tools and concepts for a new approach to tackling complexity*. München: MCB Verlag GmbH.
- Walach, H. (2006). Zirkulär statt hierarchisch - methodische Überlegungen zur Evaluation komplexer therapeutischer Massnahmen. In W. Bödeker & J. Kreis (Eds.), *Evidenzbasierung in Gesundheitsförderung und Prävention*. Bremerhaven: Verlag für neue Wissenschaft.
- Walby, S. (2007). Complexity theory, systems theory, and multiple intersecting social inequalities. *Philosophy of the social sciences*, 37(4), 449–470.
- Waldrop, M. M. (1992). *Complexity. The emerging science at the edge of order and chaos*. New York. Toronto. London. Sydney: Simon & Schuster.
- Watts, D. J. (2004). *Six degrees. The science of a connected age*. London: Vintage.
- Westhorp, G. (2014). *Realist Impact Evaluation: An introduction*. London.
- Westley, F., Zimmerman, B., & Patton, M. Q. (2007). *Getting to maybe. How the world is changed*. Toronto: Vintage.
- WHO. (1986). *Ottawa Charta zur Gesundheitsförderung*. WHO-autorisierte Übersetzung.
- WHO. (1998a). *Health promotion evaluation: recommendations for policy-makers*. Retrieved from <http://www.infosihat.gov.my/artikelHP/bahanrujukan/HEandPenilaian/HP%20Evaluation.pdf>
- WHO. (1998b). *Health Promotion Glossary*. Geneva.
- WHO. (1998c). *Resolution of the executive board of the WHO on health promotion*. Adopted at the 101st session. Agenda Item 9, 24.
- WHO (Ed.). (2001). *European Series. Evaluation in health promotion: Principles and perspectives: WHO Regional Publications*.
- WHO. (2003). *Social determinants of health: the solid facts*. Copenhagen.
- WHO. (2006). *Constitution of the World Health Organization (Basic Documents)*.
- WHO. (2008). *Milestones in Health Promotion. Statements from global conferences*. Geneva.
- WHO. (2013). *Global action plan for the prevention and control of noncommunicable diseases 2013-2020*.
- Widmer, T., & Beywl, W. (2009). Divergente Entwicklungspfade der Evaluation: Beobachtungen, Thesen und Perspektiven einer feldabhängigen und ungleichzeitigen Professionalisierung in Deutschland, Österreich und der Schweiz. In T. Widmer, W. Beywl, & C. Fabian (Eds.), *Evaluation. Ein systematisches Handbuch* (pp. 509–527). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Widmer, T. (2006). *Meta-Evaluation. Kriterien zur Bewertung von Evaluationen*. Bern Stuttgart Wien: Haupt.
- Widmer, T., Landert, C., & Bachmann, N. (2000). *Evaluations-Standards der Schweizerischen Evaluationsgesellschaft (Seval-Standards)*. Retrieved from <http://www.seval.ch/de/standards/>
- Wieser, S., Kauer, L., Schmidhauser, S., Pletscher, M., Brügger, U., Jeanrenaud, C., ... (2010). *Synthesebericht – Ökonomische Evaluation von Präventionsmassnahmen in der Schweiz: Bericht im Auftrag des Schweizerischen Bundesamtes für Gesundheit*. Winterthur.
- Williams, B., & Hummelbrunner, R. (2009). *Systems concepts in action: A practitioner's toolkit*. Stanford: Stanford University Press.
- Willke, H. (2001). *Systemtheorie III: Steuerungstheorie* (3rd ed.). Stuttgart: Lucius & Lucius.
- Willke, H. (2005). *Systemtheorie II: Interventionstheorie* (4th ed.). Stuttgart: Lucius & Lucius.

- Willke, H. (2006). *Systemtheorie I: Grundlagen* (7th ed.). Stuttgart: Lucius & Lucius.
- Windeler, J., Antes, G., Behrens, J., Donner-Banzhoff, N., & Lelgemann, M. (2008). Randomisierte kontrollierte Studien. Kritische Evaluation ist ein Wesensmerkmal ärztlichen Handelns: Die gegen randomisierte kontrollierte Studien vorgebrachten Einwände sind nicht überzeugend und zeigen oft Informationsdefizite. *Deutsches Ärzteblatt*, 105(11), 565–570.
- World Health Organization. (1986). Ottawa Charter for Health Promotion. *Health Promotion*, 1, iii–v.
- Wright, M. T. (2006). Auf dem Weg zu einer theoriegeleiteten, evidenzbasierten, qualitätsgesicherten Primärprävention in Settings. In: *Jahrbuch für Kritische Medizin*, 43, 55–73.
- Wright, M. T. (Ed.). (2010). *Partizipative Qualitätsentwicklung in der Gesundheitsförderung und Prävention*. Bern: Hans Huber.
- Wright, M. T., Roche, B., Unger, H. von, Block, M., & Gardner, B. (2010). Zur Notwendigkeit einer wissenschaftlichen Fundierung der Partizipativen Gesundheitsforschung. In M. T. Wright (Ed.), *Partizipative Qualitätsentwicklung in der Gesundheitsförderung und Prävention*. (pp. 221–232). Bern: Hans Huber.
- Zeeb, H., & Donath, S. (2010). Epidemiologische Methoden in Evidence-based Public Health. In A. Gerhardus, J. Breckenkamp, O. Razum, N. Schmacke, & H. Wenzel (Eds.), *Evidence-based Public Health* (pp. 69–82). Bern: Huber.