

Editorial

Den meisten Leserinnen und Lesern der Zeitschrift für Evaluation sind die folgenden Fallskizzen oder die entsprechende Diskussion sicherlich vertraut:

- Die zwischen 2004 und 2012 in Deutschland erhobene Praxisgebühr von zehn Euro pro Quartal hatte zum Ziel, Praxisbesuche bei Bagatellfällen und die Anzahl der sogenannten ‚Selbstüberweisungen‘ zu fachärztlichen Behandlungen zu reduzieren, um damit die gesetzlichen Krankenversicherungen zu entlasten.
- Eines der Ziele der Bologna-Reformen im Hochschulsektor war, durch Einführung der gestuften Bachelor- und Masterstudiengänge sowie eines europaweit einheitlichen ECTS-Punktesystems die Mobilität von Studierenden zu steigern.
- Ganz aktuell wird im Kontext der andauernden Corona-Pandemie diskutiert, ob die Impfquote unter Erwachsenen eher durch mehr Aufklärung, einen möglichst niederschweligen Zugang zur Impfung, durch Anreize oder durch mehr oder weniger sanften Druck erhöht werden sollte.

Diesen vordergründig so unterschiedlichen Beispielen ist gemein, dass die jeweils diskutierten Maßnahmen auf theoretischen Annahmen darüber beruhen, wie und warum die Maßnahme zu den jeweiligen Zielen führt. Der Praxisgebühr lag eine (theoretische) Sanktions-Logik zugrunde, die den Praxisbesuch bei Bagatellfällen und ein zu häufiges ‚Hopping‘ zwischen fachärztlichen Praxen verhindern sollte. Bei den Bologna-Reformen war eine offensichtliche (theoretische) Annahme, dass national sehr unterschiedliche und nicht kompatible Formen der Studienorganisation und Leistungsanrechnung ein starkes Hindernis für die Mobilität von Studierenden darstellen. Den verschiedenen diskutierten Maßnahmen zur Steigerung der Impfquote liegen u.a. (theoretische) Annahmen zur Wirkungsweise der allgemeinen politischen Instrumente *Carrots, Sticks and Sermons* (Anreize, Sanktionierung und Überzeugung) zugrunde, wie sie im gleichnamigen Buch von Bemelmans-Videc, Rist und Vedung 2003 beschrieben wurden.

Zu allen drei Beispielen ließe sich ausführlich diskutieren, wie gut die genannten Interventionen tatsächlich im Sinne der jeweiligen Ziele oder intendierten Outcomes funktionieren bzw. funktionierten. Im Falle der Praxisgebühr etwa waren zwar Effekte auf die Frequenz von Praxisbesuchen, nicht aber eine dauerhafte Kostenreduktion für die gesetzlichen Kassen beobachtbar, sie wurde nach acht Jahren wieder abgeschafft; eine vermutete Nebenwirkung waren soziale Selbstselektions-effekte, die aufgrund unterlassener (notwendiger) Praxisbesuche eventuell langfristig sogar zu deutlich unerwünschten Effekten führten. Die Mobilität europäischer

Studierender dagegen hat seit den Bologna-Reformen tatsächlich deutlich zugenommen. Dafür sind andere Auswirkungen der Reformen bis heute Gegenstand kontroverser Auseinandersetzungen, so z.B. eine Tendenz zur ‚Verschulung‘ des Studiums durch die stärker reglementierten Studienpläne. Ähnlich ließe sich für die Instrumente *Carrots, Sticks and Sermons* berichten, dass die grundsätzlichen Wirkungsannahmen zwar zutreffen, die praktischen Effekte aber oft von Aspekten der konkreten Umsetzung, von Kontextfaktoren oder Merkmalen der Zielgruppen selbst abhängen. Ebenso gibt es oft nicht antizipierte gegenläufige Wirkungen, die die erzielten Effekte einer Maßnahme konterkarieren können.

Es zeigt sich also, dass es bei der Maßnahmengestaltung jedweder Art riskant sein kann, theoretische Überlegungen zur Wirkungsweise nur im Impliziten zu belassen. Denn wenn wir theoretische Annahmen der diskutierten Maßnahmen unterstellten, waren damit nicht notwendigerweise *explizite* theoretische Vorüberlegungen gemeint. Vielmehr werden die meisten Maßnahmen in der Praxis ja nicht ‚am Reißbrett‘ entworfen, ihre unterstellte Wirkungsweise wird also selten systematisch unter Zuhilfenahme der aktuellen wissenschaftlichen Expertise im jeweiligen Gegenstandsbereich antizipiert. Grund dafür ist weniger, dass das dafür erforderliche Wissen grundsätzlich fehlen würde. Denn auch wenn viele Maßnahmen innovativ sind, sind die ihnen zugrundeliegenden Wirkungsmechanismen etwa zur Veränderung von Verhaltensweisen ja selten gänzlich neu und unerforscht. Eher liegt es daran, dass soziale Maßnahmen, wie es Carol Weiss 1998 in ihrem Grundwerk *Evaluation* so schön beschrieben hat, komplexe Unterfangen sind, „ein Amalgam aus Träumen und Persönlichkeiten, Räumen und Theorien, Büroklammern und Organisationsstrukturen, Zielgruppen und Aktivitäten, Budgets und Fotokopien und großartigen Absichten“ (S. 48). Auch wenn dieses Zitat offensichtlich vor der Digitalisierung unserer Arbeitswelt entstanden ist, hat sich an der beschriebenen Tatsache wenig geändert, dass nämlich viele politisch und auch anders verantwortete Maßnahmen ein Kompromiss zwischen unterschiedlichen Interessen, Resultat wenig geplanter oder planbarer Entstehungsprozesse und nicht immer von klaren Ziel-Mittel-Überlegungen getrieben sind.

Daraus erklärt sich vielleicht auch, dass Wirkungsmodelle, der inhaltliche Schwerpunkt des Ihnen hier vorliegenden Themenhefts, bisher weniger im Kontext der Planung von Programmen und anderen Maßnahmen diskutiert wurden als in der Evaluation, die diese Gegenstände hinsichtlich ihrer Wirkungen untersucht und zu ihrer Verbesserung beitragen soll. Denn Wirkungsmodelle zeichnet aus, dass sie implizite Wirkungsüberlegungen explizit machen und damit ein Instrument sind, das sowohl im Sinne einer bewusster auf Wirkungen ausgerichteten Planung eingesetzt werden kann als auch zur Rekonstruktion und Untersuchung von Wirkungsmechanismen im Rahmen von wirkungsorientierten Evaluationen.

In der Entwicklung der Evaluation in den vergangenen Jahrzehnten stellt die zunehmende Verbreitung der Arbeit mit Wirkungsmodellen sicher eine der theoretisch und praktisch einflussreichsten methodologischen Weiterentwicklungen unseres Felds dar. Ursprünglich vor allem im Kontext theoriebasierter bzw. programmtheoriegesteuerter Evaluationsansätze diskutierte Überlegungen haben inzwischen breiten Eingang in die Evaluationspraxis gefunden und zu einer Vielfalt von Evaluati-

onsansätzen und Anwendungsvarianten beigetragen. Diese Breite der Verwendung sichtbar zu machen und daran anknüpfend den Mehrwert, aber auch mögliche Risiken und Kosten der Arbeit mit Wirkungsmodellen in der Evaluation zu diskutieren, war Ausgangsgedanke bei der Planung des Themenhefts. Die große Resonanz auf den im Juni 2020 publizierten Call for Abstracts dient als Beleg für die Ausgangsprämisse, dass Wirkungsmodelle inzwischen Eingang in den ‚Mainstream‘ der Evaluationspraxis gefunden haben. Entsprechend schwer fiel die Auswahl der Abstracts, deren Autorinnen und Autoren zur Ausarbeitung eines Vollmanuskripts eingeladen wurden. Letztendlich bilden fünf Beiträge den Themenschwerpunkt „Wirkungsmodelle und Programmtheorien in der Evaluationspraxis“ des aktuellen Themenhefts.

Der einführende Beitrag „Wie Wirkungsmodelle zur Wirkung kommen. Nutzungsvarianten, Mehrwert und Kosten ihrer Verwendung in der Evaluationspraxis“ von *Jan Hense* und *Sandy Taut* soll dem Themenheft als inhaltliche Rahmung dienen. Ausgehend von einem kurzen Rückblick auf die evaluationshistorische Entwicklung des Themenfelds und einiger begrifflicher Klärungen geben wir dort einen Überblick über die unterschiedlichen Einsatzvarianten von Wirkungsmodellen, die inzwischen weit über das ursprünglich die Diskussion dominierende ‚Öffnen der *black box*‘ hinaus gehen. Neben verschiedenen Bedingungen, die Wirkungsmodelle für die jeweiligen Einsatzvarianten erfüllen sollten, und einer Plausibilisierung, wie die jeweilige Einsatzvariante zu positiven Effekten auf die Evaluationsqualität oder auf der Gegenstandsebene führt, gehen wir darüber hinaus auf mögliche Kosten im Sinne negativer Effekte ein. Als Fazit plädieren wir für ein wirkungsorientiertes *Mindset* bei der Maßnahmenplanung in allen gesellschaftlichen Bereichen, welches von vornherein auch die Evaluation noch stärker ins Visier bringen würde.

Eine Entwurfsversion des Einführungsbeitrags wurde den Autorinnen und Autoren der weiteren Beiträge vorab zur Verfügung gestellt, verbunden mit der Einladung, in der Reflexion der eigenen Fallbeispiele darauf inhaltlich Bezug zu nehmen.

Thomas Fenzl, *Stella Lemke* und *Philipp Mayring* berichten im ersten dieser Beiträge mit dem Titel „Entwicklung und Erprobung eines Wirkmodells im Rahmen der Evaluation des baden-württembergischen Präventionskonzepts *stark.stärker.WIR*“ von einer theoriebasierten Evaluation im Bereich der Prävention und Gesundheitsförderung an Schulen. Die Evaluation verfolgte formative und summative Ziele. Das Wirkungsmodell wurde sowohl für die Überprüfung der Implementierungstreue als auch für die Überprüfung der Wirkungsannahmen der Präventionsmaßnahmen in Hinblick auf verschiedene Outcomes für Schülerinnen und Schüler genutzt. Das Bündel an Präventionsmaßnahmen, welche über die drei Säulen Suchtprävention, Gewaltprävention und Gesundheitsförderung auf unterschiedlichen Ebenen wirken und sowohl Verhalten wie auch Strukturen beeinflussen sollen, resultiert in einem hochkomplexen Wirkungsmodell. Letztendlich wurde eine Auswahl der Wirkungsannahmen über ein *Mixed-Methods-Design*, welches Prä-Post-Kontrollgruppenmessungen beinhaltet, überprüft. Die Autorengruppe präsentiert interessante Ergebnisse und leitet konkrete Optimierungsvorschläge für das Präventionskonzept und dessen Implementierung in den Schulen ab. Der Beitrag diskutiert abschließend die Kosten und den Nutzen der Verwendung von Wirkungsmodellen in dieser Evaluation. Insbesondere sieht die Autorengruppe die hohe Komplexität des

Modells im Nachhinein eher als Hindernis an. Eine rein quantitative Überprüfung der Annahmen wird für nicht sinnvoll erachtet. Die Überprüfung der Implementierungstreue auf der Basis des Wirkungsmodells wird als wichtiger Baustein der Evaluation diskutiert. Die starke Auseinandersetzung mit der Programmtheorie in Verbindung mit dem anspruchsvollen *Mixed-Methods*-Ansatz hat letztlich die Ableitung differenzierter Handlungsempfehlungen ermöglicht.

Der Beitrag „Wirkungsmodelle in der Eingliederungshilfe“ von *Sebastian Ottmann, Joachim König* und *Claudia Gander* berichtet aus einem noch andauernden Projekt aus dem Bereich der sozialen Arbeit. Insbesondere das Bundesteilhabegesetz, das seit 2017 bis 2023 schrittweise in Kraft tritt bzw. treten wird, hat hier die Notwendigkeit von Wirkungskontrollen forciert. Aufgrund der praktischen Hindernisse bei der Realisierung von (quasi-)experimentellen Untersuchungsdesigns schreiben die Autoren und die Autorin der Wirkungsplausibilisierung auf Basis von Wirkungsmodellen zukünftig ein besonderes Potenzial zu. Das konkret beschriebene Projekt bezieht sich einerseits auf den Berufsbildungsbereich, andererseits auf die teilstationäre Tagesbetreuung für erwachsene Menschen mit Behinderung nach dem Erwerbsleben. Für beide Bereiche wird ein mehrschrittiges Vorgehen beschrieben, bei dem unter Einbezug von Fachkräften und Zielgruppenmitgliedern jeweils ein Wirkungsmodell entwickelt wurde. Besonders interessant aus Sicht des Themenschwerpunkts ist die Diskussion zur Machbarkeit von ‚Ankerwirkungsmodellen‘, die ähnlich wie die *program archetypes* von Funnell und Rogers (2011) nicht mehr spezifisch für eine konkrete Intervention, sondern generisch für einen bestimmten Interventionstyp entworfen werden, um nicht für jede Evaluation die Entwicklung von neuem beginnen zu müssen.

„Partizipative Erstellung von Wirkungsmodellen“ lautet der Titel des Beitrags von *Susanne Giel, Susanne Mäder* und *Hannah Pütz*. Sie greifen die in der Diskussion häufig gestreifte, aber selten systematisch bearbeitete Frage auf, wer bei der Entwicklung eines Wirkungsmodells des Evaluationsgegenstands in welchem Maße beteiligt werden bzw. maßgeblichen Einfluss haben sollte. Unter Bezug auf ein der partizipativen Gesundheitsforschung entlehntes Stufenmodell plädieren sie für eine möglichst starke Einbeziehung von Beteiligten und Betroffenen, weil dadurch verschiedene positive Effekte für die Evaluation zu erwarten seien. Am Beispiel zweier Evaluationsprojekte aus dem Bereich der Gesundheitsförderung beschreiben die Autorinnen, welche Funktionen die Wirkungsmodelle und welche Rolle verschiedene Grade an Partizipation dabei jeweils hatten. Neben Chancen der Beteiligung werden anhand der Beispiele auch v.a. kontextbedingte Grenzen der Partizipation in der praktischen Umsetzung aufgezeigt und diskutiert.

Der Beitrag von *Melanie Ehren*, „Programme or Policy Theory: Lessons Learned from Two Decades of School Inspection Research“, kommt aus dem Bildungsbereich, und hier speziell aus dem Thema externe Schulevaluation oder Schulinspektion. Vor diesem Hintergrund beschreibt die Autorin den Gewinn der Verwendung von ‚policy theories‘ als Rekonstruktionen der Wirkungsweise von komplexen, diffusen Politikmaßnahmen mit starker ideologischer, normativer Komponente, wie dies beim Thema Bildungsqualität der Fall ist. Ehren greift auf eine reichhaltige internationale Erfahrung zurück und diskutiert verschiedene Herausforderungen bei

der theoriebasierten Arbeit: (a) die Notwendigkeit des Einbezugs sowohl normativer wie auch wirkungsorientierter Theorien nach Chen (2015), (b) die Schwierigkeit bei der Eingrenzung der Evaluation, wenn es sich um facettenreiche, nationale, oft auch unausgereifte Politikmaßnahmen handelt, (c) die Klärung von Verantwortlichkeiten und Rollen der Beteiligten bei der Erstellung von Theorien, und (d) die Wahl angemessener Methoden zur Überprüfung der Theorien je nach Phase der Evaluation. Sie illustriert diese Herausforderungen an vier Projekten, die sie insbesondere im Rahmen der niederländischen Schulinspektion, aber auch in einem EU-finanzierten Projekt in sechs europäischen Ländern/Regionen durchgeführt hat. Der Beitrag diskutiert abschließend, dass ein theoriebasiertes Vorgehen nur dann wirklich seine positive Wirkung im Sinne einer Verbesserung der Qualität von Politikmaßnahmen entfalten kann, wenn der politische Kontext ein konsensorientiertes Vorgehen unter Einbindung der Beteiligten ermöglicht, welches auf gegenseitigem Vertrauen und Respekt basiert.

Zwei weitere, ursprünglich fest für das Themenheft vorgesehene Beiträge konnten leider aufgrund externer, unvorhersehbarer Ursachen nicht wie geplant fertig gestellt werden. Besonders schmerzlich ist aus unserer Sicht, dass darunter auch der als kritischer Blick von außen eingeplante internationale Beitrag von Chris Coryn (Evaluation Center, Western Michigan University, USA) fiel. Wir haben versucht, diesen Ausfall insofern zu kompensieren, dass wir kritische Stimmen zu möglichen Dysfunktionalitäten beim Einsatz von Wirkungsmodellen in unseren einführenden Beitrag integriert haben, um auch der kritischen Fachperspektive angemessen Raum zu geben. Zusätzlich sei an dieser Stelle besonders der nicht mehr ganz aktuelle, aber in dieser Hinsicht dennoch sehr lesenswerte Überblicksartikel zur Praxis der theoriegesteuerten Evaluation von Coryn, Noakes, Westine und Schröter von 2011 empfohlen (A Systematic Review of Theory-Driven Evaluation Practice From 1990 to 2009, *American Journal of Evaluation*, 32 (2), S. 199-226).

Neben den dezidierten Beiträgen zum Themenheft runden wie immer weitere Beiträge die vorliegende Ausgabe ab.

In der Rubrik „Praxisberichte“ erwartet Sie der Beitrag „Monitoring und Evaluation wissenschaftlicher Weiterbildungsangebote an einer systemakkreditierten Universität“ von *Johanna Vogt* und *Ulrich Wacker*. Er berichtet aus einem Bereich, der neben Forschung und Lehre als drittes Tätigkeitsfeld von Hochschulen zumindest in Deutschland bisher noch eine relativ geringe Rolle spielt. Beschrieben wird ein zyklisches Monitoring- bzw. Evaluationsverfahren, das an der systemakkreditierten Universität Konstanz zur Qualitätssicherung der weiterbildenden Studiengänge entwickelt wurde.

Evelyn Funk bespricht in ihrer Rezension das Buch „Methoden als situierte Praxis. Die Gruppendiskussion in der Programmevaluation“ von Susanne Mäder.

In der Rubrik „DeGEval...Info“ berichten *Michael Bigos*, *Ann-Kristin Matthe* und *Hannah Pütz* von der Frühjahrstagung des DeGEval-Nachwuchsnetzwerks. *Jan Hense* stellt inhaltlich passend zum Themenschwerpunkt ein DeGEval-Wirkungsmodell vor, das im Vorstand der DeGEval vorrangig zur Unterstützung von Planungs- und Reflexionsprozessen entwickelt worden ist.

Danken möchten wir an dieser Stelle den Autorinnen und Autoren der Beiträge des Themenhefts sowie den anonymen Gutachterinnen und Gutachern dieser Beiträge. Dank gebührt auch den Kolleginnen und Kollegen, die Abstracts eingereicht hatten, welche in der Mehrzahl ebenfalls hochwertig waren, aber aus Platzgründen leider nicht im Themenheft berücksichtigt werden konnten. Ein besonderer Dank der Herausgebenden gilt Thomas Widmer (Institut für Politikwissenschaft, Universität Zürich) für die kritische Durchsicht und Begutachtung unseres einführenden Beitrags und dem Redakteur der Zeitschrift Hansjörg Gaus für die wie immer professionelle und geduldige Betreuung des Herausgabeprozesses.

Jan Hense und Sandy Taut