

## Rezension zu:

Strobl, Rainer/Lobermeier, Olaf/Heitmeyer, Wilhelm (Hg.): Evaluation von Programmen und Projekten für eine demokratische Kultur. Wiesbaden: Springer VS 2012. 234 Seiten, 39,95 EUR, ISBN 978-3-531-18306-0

*Gerd-Michael Hellstern*<sup>1</sup>

Wie Evaluationen zur Identifizierung, Bewertung und Weiterentwicklung wirksamer Strategien für die Entwicklung einer demokratischen Kultur beitragen können, ist eine hochaktuelle Frage in der Tradition der aufklärerischen Funktion der Evaluation. Die Veröffentlichung der Beiträge der von *Rainer Strobl* und *Olaf Lobermeier* von der proVal-Gesellschaft für sozialwissenschaftliche Forschung in Hildesheim und *Wilhelm Heitmeyer* vom Institut für interdisziplinäre Konflikt- und Gewaltforschung der Universität Bielefeld 2011 am Zentrum für interdisziplinäre Forschung der Universität Bielefeld durchgeführten Tagung zielt auf eine Bestandsaufnahme, wie Evaluation zur Förderung einer pluralistischen und demokratischen Kultur beitragen kann.

Die Beiträge der internationalen Autoren aus Deutschland, Großbritannien, der Schweiz und den USA ordnen die Herausgeber drei übergreifenden Fragestellungen zu:

- der Frage nach adäquaten Evaluationsansätzen und Methoden,
- den Erfahrungen mit nutzungsorientierten Evaluationsstrategien und
- der gesellschaftlichen Bedeutung von Evaluation.

In der programmatischen Einleitung verweist *Wilhelm Heitmeyer* auf die aktuelle Bedeutung und die Schwachstellen bisheriger Forschungen zum Thema rechtsextremer, antisemitischer und fremdenfeindlicher Straf- und Gewalttaten, die als Reaktion auf spektakuläre Ereignisse eher kurzfristig angelegt, methodisch unzureichend

unter den politischen Rahmenbedingungen kaum nützliche Informationen erarbeiten können. Brennpunkthaft spiegeln sich daher in den Forschungen zu diesen Themen die methodischen, praktischen und grundsätzlichen Anwendungsprobleme einer Evaluation und besitzen daher über den hier thematisierten Forschungsbereich hinaus Bedeutung für die Praxis der Evaluation.

Der einflussreiche Ansatz einer theoriegeleiteten Evaluation, der maßgeblich von *Huey Chen* unter der Anleitung von Peter Rossi entwickelt wurde, wird von diesem in seinem (englischsprachigen) Beitrag („Theory-driven evaluation: Conceptual framework, application and advancement“) vorgestellt. Ausführlich wird die Entwicklung des Ansatzes diskutiert und der theoriegeleitete Ansatz wird von der Black-Box-orientierten Impact-Forschung sowie methodenorientierten Ansätzen einer Evaluation durch die explizite Formulierung eines an den Stakeholdern einer Evaluation entwickelten Handlungsmodells abgegrenzt, das die Kontextfaktoren und Implementationsbedingungen modelliert. Es ist jedoch weniger die fehlende theoretische Grundlage alternativer Modelle als der Versuch einer expliziten Programmtheorie, die Chens Handlungsmodell auszeichnet. Am Beispiel eines Anti-Drogenprogramms illustriert er den Einsatz der einzelnen Bausteine und an mehreren Beispielen – eines Alkohol- und Drogenvermeidungsprogramms, eines Ernährungsprogramms und eines Bildungsprogramms – die unterschiedlichen Typen von Programmtheorien (lineare Modelle,

1 Institut für Betriebswirtschaftslehre, Universität Kassel

multi-determinierte und dynamische Modelle). So richtig und wünschenswert diese umfassende Modellierung mit den Schritten ‚Einbezug der Stakeholder‘ und die ‚Entwicklung eines Handlungsmodells‘ erscheinen, nicht immer wird eine so umfassende Evaluierung bei begrenzten Mitteln möglich sein. Von besonderem Interesse ist der Versuch, die traditionellen Kriterien der externen und internen Validität im Sinne einer Orientierung an den Stakeholdern zu erweitern. Nach Chen orientierten sich diese Kriterien bisher an traditionellen Methoden und Techniken der Sozialwissenschaften, als eine anwendungsorientierte Wissenschaft benötigt eine Evaluation jedoch eigene Methoden und Kriterien. Hierzu entwickelt er ein „integrative validity Model“, das Wirksamkeits-, Machbarkeits- und Übertragbarkeitskriterien beinhaltet, die allerdings, wie auch deutlich wird, nicht immer gleichermaßen erreicht werden können und daher einem gegenseitigen Abwägungsprozess unterliegen. In einem Bottom-up-Ansatz, der zuerst Machbarkeit überprüft, dem die Wirksamkeitsprüfung folgt und der schließlich die Generalisierbarkeit überprüft, sieht er eine erfolgreiche Umsetzung der Kriterien am ehesten verwirklicht.

*Thomas Widmers* Beitrag zur Wirkungsevaluation von Maßnahmen der Demokratieförderung verdient Beachtung, da er die häufig sehr vielfältigen Begrifflichkeiten zur Beschreibung von Wirkungen (Input, Prozess, Output, Outcome und Impact) und deren Umsetzungsprobleme illustriert. Dabei sieht er einen Wandel in der Steuerung der Programme von einer hierarchischen Steuerung zu einer kooperativen Steuerung, einer ‚governance‘. Für die Evaluation bedeutet dies einen Übergang von einer input- und verhaltensorientierten Prozess-Steuerung hin zu einer Steuerung über Leistungs- und Wirkungsvorgaben. Diese führt zu einer Veränderung der Legitimationserwartungen. Die ursprüngliche demokratische Input-Legitimation und (regelkonformorientierte) Prozess-Legitimation in einer Evaluation werden zunehmend ersetzt durch den verstärkten Bedarf an Output-Legitimationen. Am Beispiel des Phänomens Rechtsextremismus illustriert Widmer, wie ein Gegenstand zu untersuchen ist, der manifeste und latente Merkmale umfasst, der unerwünschte Handlungen und verdecktes Handeln beinhaltet, unterschiedliche Zielgruppen umfasst, der emotional und moralisch aufgeladen ist und für den häufig eine Symptombekämpfung charakteristisch ist. Eine Evaluation, die seinen Kriterien

nach wissenschaftlich, nutzenbringend und bewertend ist, bedarf der Klarheit über die untersuchten Wirkungen, sie hat sich angesichts der Komplexität auf spezifische Fragestellungen zu konzentrieren, ihre Grenzen zu respektieren und diese offen darzulegen, um glaubwürdig zu bleiben.

*Manuel Eisner, Tina Malti, Denis Ribeaud und Barbara Müller* geben einen Überblick über den Stand kriminologischer Forschungen und berichten über ihre Erfahrungen aus einem Projekt zur sozialen Entwicklung von Kindern in Zürich – einem Feldversuch, kombiniert mit einer Längsschnittstudie –, das aggressivem Problemverhalten von Kindern begegnen sollte. Dabei zeigen sie auf, dass sich die Generalisierbarkeit von klinischen Wirksamkeitsstudien mit kleinen Teilnehmerzahlen und strikter Kontrolle der Intervention von Studien mit großer Fallzahl und begrenzter Umsetzungsqualität deutlich unterscheidet, insbesondere wenn die Wirkungen über längere Zeiträume betrachtet werden. Daher fordern sie kontrollierte Wirkungsstudien und Feldversuche, die langfristige Effekte und reale Umsetzungsbedingungen zur Grundlage haben. Dabei versuchen sie aufzuzeigen, wie die Frage nach der Implementation, der internen und externen Validität sowie der Qualität der Messinstrumente zu beantworten ist. Dies erfordert, die Bedürfnisse der Beteiligten zu verstehen und einzubeziehen, d.h. (1) die Umsetzung eines möglichst überprüften, theoretischen Modells organisatorisch vorzubereiten, (2) die bestehenden Strukturen und Ressourcen zu berücksichtigen (um Anschlussfähigkeit sicherzustellen) sowie (3) eine hohe Bereitschaft für die Teilnahme zu erzeugen. Innerhalb des Forschungsdesigns kommt (1) der Auswahl, (2) der Zuordnung zu Kontroll- und Experimentalgruppen sowie (3) der Wahl der Zeitpunkte große Bedeutung zu. Bei der Messung mit mehreren Instrumenten (multi-method) ist vor allem auch der Erfassung der Umsetzungstreue der Programme große Beachtung zu schenken. Der Bericht liefert durch seine zusammenfassenden Handlungsanweisungen und Informationen über Forschungseinrichtungen einen hilfreichen Einstieg in die experimentelle Feldforschung und diskutiert auch die dabei auftretenden ethischen Fragen. Sie zeigen, wie anspruchsvolle experimentelle Feldstudien sinnvoll durchzuführen sind. Kritisch bleibt jedoch zu vermerken, dass diese an Voraussetzungen gebunden sind, die in der Praxis nicht immer gegeben sind, und nicht

immer haben sich großangelegte Feldexperimente in den USA angesichts der komplexen Wirkungsverläufe als erfolgreich erwiesen.

Neben der ausgezeichneten Diskussion der theoretischen und methodischen Grundlagen einer Evaluation von Programmen widmet der Band den Fragen nach der Nützlichkeit von Evaluationen breiten Raum. Wie kann die Nutzung von Evaluationen zur Förderung einer demokratischen Kultur erreicht werden? *Wolfgang Bewyl* und *Susanne Giel* von der Univation in Köln sprechen von einer „Nutzenkrise der Evaluation“ und sehen in den u.a. durch *Stufflebeam* und *Patton* aufgeworfenen Fragen nach einer stakeholder-orientierten, nutzungsfokussierten Evaluation einen Perspektivenwechsel in der Evaluation. Das ursprünglich akademisch dominierte Forschungsparadigma wird durch einen nutzungsfokussierten Ansatz abgelöst, in dem die Steuerungsfunktion von der Nutzungsperspektive der Beteiligten und Betroffenen ausgeht. Dabei unterscheiden sie die Nutzung (utilization) von dem Nutzen (use), dem Vorteilsgewinn und der Nützlichkeit (utility), wie er sich häufig im Prozessnutzen bei der Durchführung einer Evaluation im Unterschied zum Ergebnissenutzen durch die Nutzung der Evaluationsergebnisse ergibt. Die vorausschauende partizipative Nutzenklärung wird damit zur Voraussetzung für eine erfolgreiche Evaluation. Am Beispiel der multizentrischen Evaluation von 22 Modellprojekten zur Prävention von rechtsextremem rassistischem Verhalten des Programms „Vielfalt tut gut“ werden die Arbeitsschritte vorgestellt: Identifizierung der Nutzungsabsichten in einer Strukturlandkarte, Strukturierung auf der Grundlage eines logischen (Wirkungs-)Modells, dem Programmbaum, und Analyse der Wirkungszusammenhänge durch Fallstudien, die die Wirkungsannahmen rekonstruieren, überprüfen, validieren und vergleichen. *Wolfgang Bewyls* und *Susanne Giel's* Studie verdeutlicht dabei die Schwierigkeiten einer partizipativen Strategie der Vielfalt, den unterschiedlichen Erwartungen gerecht zu werden.

*Kerstin Sischka* von der Internationalen Akademie für innovative Pädagogik, Psychologie und Ökonomie (INA) an der FU Berlin berichtet in ihrem Beitrag „Zum Nutzen eines konfliktsensiblen Ansatzes in der wissenschaftlichen Projektbegleitung“ über die Vorgehensweise des wissenschaftlichen Begleiteams zum Bundesprogramm „Vielfalt tut gut. Jugend für Vielfalt, Toleranz und Demokratie“. Die Evalua-

tion erfolgte vor allem über sukzessive, die 18 Modellprojekte reflektierenden Gesamtworshops, die als „Erfahrungsräume“ betrachtet werden für Prozesse der Interaktion und Diskussion sowie als entwicklungsfördernder Rahmen für eine sich als formativ verstehende Evaluation. Ergänzt wurden diese durch Projektbesuche vor Ort und eine abschließende Wirkungseinschätzung durch eine Outcome-Mapping-Matrix, die versuchte, die Veränderungen durch die Modellprojekte in den relevanten Wirkungsfeldern (kulturell, strukturell, personal und relational) zu beschreiben. *Kerstin Sischka* versteht den Ansatz als eine emergente Entwicklungsevaluation (development evaluation), die versucht, über Interaktion, Feedbackzyklen und Bewertungen der Themencluster im Diskurs Forschungsergebnisse über Rechtsextremismus, Demokratie, Jugend und Familie mit systematischer Praxisreflexion zu verbinden. Dieser psychosoziale Ansatz zielt, wie in dem Beitrag „Aufbau und Evaluation von Netzwerken für eine demokratische Kultur“ von *Rainer Strobl* und *Olaf Lobermeier* von der proVal-Gesellschaft für sozialwissenschaftliche Analyse, Beratung und Evaluation, auf eine entwicklungsfördernde Vernetzung der beteiligten Akteure. Ausgehend von der Annahme, dass eine gut vernetzte Gemeinschaft soziales Kapital (Bourdieu) mobilisiert, sehen *Rainer Strobl* und *Olaf Lobermeier* in der Einbindung eines breiten Spektrums von demokratischen Kräften und Akteuren die Möglichkeit, demokratische Kultur zu stärken. Sie unterscheiden zum einen natürliche (primäre und sekundäre) sowie künstliche (tertiäre) Netzwerke, die der Erreichung bestimmter Ziele dienen. Für deren Planung sind die Interessen der verschiedenen Beteiligten und Betroffenen durch eine Stakeholderanalyse zu berücksichtigen. Beratung und Begleitung beim Ausbau und Management eines Netzwerkes wird damit zur Aufgabe einer formativen Evaluation. Zum anderen unterscheiden sie in Ihrer Analyse komplementäre, homogene Netzwerke, die vorrangig zur Durchsetzung von Interessen dienen, von Innovationsnetzwerken, die eher neue Lösungen anstreben. Inzwischen erlauben zahlreiche Softwareprogramme, Netzwerkanalysen mit entsprechenden Kennzahlen (Dichte, Zentralität, Ausgrenzung) durchzuführen. Zu Recht verweisen die Autoren darauf, dass dabei zwischen Qualität der Vernetzung und Qualität der Vernetzungsergebnisse zu unterscheiden ist. Zur Förderung einer demokratischen Kultur sind die

Orientierung an den Wirkungen der Vernetzung und die Qualität der Vernetzungsergebnisse entscheidend.

Orientiert sich der Beitrag von *Strobl* und *Lobermeier* vor allem an dem Management von Netzwerken, so resümiert der lesenswerte Beitrag von *Gesa Schirmacher* aus dem Niedersächsischen Ministerium für Soziales, Frauen, Familie, Gesundheit und Integration zum Thema „Was sich Auftraggebende von Evaluationen und Evaluationspraxis wünschen“ vor allem ihre Erfahrungen mit der Nutzung von Evaluationen. Kritisch spricht sie die Tendenz zu einer inflationären Verwendung nichtprofessioneller Evaluationen, aber auch methodisch-technische Probleme an: Valide Evaluationen erfordern langfristige Untersuchungen, die angesichts drängender sozialer Fragen häufig nicht möglich sind, auch sind Wirkungen kaum monokausal, daher lässt sich der Erfolg nur schwer messen, schließlich stellt sich die ethische Frage nach Verzicht von Maßnahmen bei Bildung von Kontrollgruppen für valide Untersuchungen. Auf der Suche nach praktikablen Verfahren jenseits von Do-it-yourself-Verfahren und undifferenzierten Datensammlungen und Berichtspflichten sieht sie drei unterschiedliche Nutzungsziele: (1) die Verbesserung von Programmen vor allem durch eine formative Evaluation, (2) die Ermittlung kostengünstiger Alternativen sowie (3) die Unterstützung (Legitimation) wirksamer Programme. Wichtige Voraussetzung sieht sie in einer Vertrauensbasis durch offene Auftragsverhandlungen zwischen Auftraggebenden und Evaluator(inn)en, die vor allem durch eine wirksame Kommunikation der Grundlagen der Evaluationstechnik mit einer der jeweiligen Zielgruppe entsprechenden überzeugenden Darstellung in den Berichten zu erzielen ist. Der kritischen Sicht gegenüber der Auftragsforschung als Pseudo-Evaluationen setzt sie das Modell einer ausdifferenzierten Evaluationslandschaft gegenüber, in dem die Auftragsforschung durch externe Evaluationen ergänzt wird. Dem Vorwurf der Pseudo-Evaluationen begegnet sie mit der Feststellung, dass Auftraggeber eher ein Interesse an differenzierten Berichten zur effektiven Steuerung von Programmen haben und rechtfertigt das Steuerungsinteresse an politisch gesetzten Zielen mit den Grundprinzipien einer demokratisch verfassten Gesellschaft. Angesichts der begrenzten finanziellen Spielräume empfiehlt sie zur Vermeidung der Nichtnutzung von Ergebnissen einer Evaluation, praktikable Zwischenlösungen aufzuzeigen und

ein „Handlungsraaster von mehreren Schritten“ großen Lösungen und Empfehlungen vorzuziehen.

Im dritten Teil des Bandes widmen sich *Reinhard Stockmann* und *Wilhelm Heitmeyer* aus unterschiedlichen Perspektiven sehr grundsätzlichen Fragen der gesellschaftlichen Bedeutung und Wirkungen der Evaluation für die Entwicklung einer demokratischen Kultur. Der kompetenten Bestandsaufnahme und zukunftsweisenden Perspektive einer professionellen Evaluation von *Reinhard Stockmann* setzt *Wilhelm Heitmeyer* eine eher pessimistische Sicht zukünftiger Entwicklungen in der Evaluation gegenüber. Die Perspektiven sind jedoch weniger divergierend, insofern, als beide positive Beiträge und kritische Entwicklungen aufzeigen. Als Lösung fordert *Reinhard Stockmann* in seinem Beitrag zur „Evaluation und Gesellschaft“, nach einer informativen Übersicht zur weltweiten Entwicklung der Evaluationsforschung, in kritischer Sicht der populären Verwendungen nichtwissenschaftlicher „Alltagsevaluationen“ striktere Qualitätskontrollen, die die Einhaltung wissenschaftlicher Kriterien und fachlicher Standards (Nützlichkeit, Durchführbarkeit, Fairness, Genauigkeit) sicherstellen. Qualifizierung und Professionalisierung von Evaluator(inn)en sowie der Aufbau entsprechender ausdifferenzierter Evaluationskapazitäten sind hierfür entscheidende Voraussetzungen. In seinem ausdifferenzierten Modell der gesellschaftlichen Funktion von Evaluation unterscheidet er drei gesellschaftliche Zweckbestimmungen einer Evaluation: (1) die Aufdeckung von Lernpotenzialen mit dem Ziel der Programm- und Projektsteuerung, (2) die Sicherstellung von Transparenz zur Kontrolle von demokratisch legitimierten Politikzielen sowie (3) die Überprüfung der Werte, Normen und Dialogformen in einer der gesellschaftlichen Aufklärung dienenden, nachhaltigen, dialogorientierten Evaluation. Über die Kritik der bisherigen Formen von Evaluationen weist er auf die unterschiedlichen Nutzungen von Evaluationen hin: (1) den direkten, instrumentellen Nutzen, (2) den konzeptionellen Nutzen, (3) den „Überzeugungs“-Nutzen sowie (4) den Prozessnutzen. Während die instrumentelle Bedeutung der Evaluation verbreitet aufzufinden ist, fehlt es bisher an einer Evaluationskultur, die neben der politikgetriebenen Evaluation auch eine weitergehende, unabhängige wissenschaftliche Evaluation sicherstellt, um zur gesellschaftlichen Aufklärung beizutragen. *Werner Heitmeyers* Resümee „Evaluations-

forschung. Parameter und Auswirkungen für die Gesellschaft“ betont dagegen die Grenzen einer Steigerung rationaler Entscheidungen durch die Evaluation. Er sieht fünf Problembereiche: (1) betriebswirtschaftliche Kalküle dominieren weitgehend die Entscheidungskriterien in Evaluationen, (2) Evaluationen behindern die Innovations- und Risikobereitschaft für eine freiheitliche, plurale und demokratische Gesellschaftsentwicklung, (3) Evaluationen dienen häufig als Teil einer ideologischen Instrumentalisierung und Inszenierung, (4) sie wirken als Kontrolle politisch unliebsamer Entwicklungen und führen (5) zu einer Standardisierung, die die soziale Vielfalt nicht zur Kenntnis nimmt. Die Annahme einer Kultur gesellschaftlichen Lernens durch Evaluation ist daher kritisch zu hinterfragen und Öffentlichkeit, eine strikte Praxis der Veröffentlichung von Evaluationsergebnissen, ist dafür eine zentrale Voraussetzung.

Für alle, die sich mit aktuellen Entwicklungen und unterschiedlichen Verfahren der Evaluation vertraut machen, bietet die Veröffentlichung der Tagungsbeiträge eine nützliche Quelle, die die internationale Diskussion aufgreift und durch die interdisziplinäre Ausrichtung einen gelungenen Überblick über die vielfältigen Ansätze einer nutzenorientierten Evaluation ermöglicht. Während die Beiträge nur begrenzt konkrete, inhaltliche Ergebnisse der Forschungen zur Förderung einer demokratischen und pluralistischen Kultur enthalten, ermöglicht der Band einen hervorragenden Überblick über die Entwicklung und den Stand der methodischen Diskussion und mit den abschließenden Beiträgen von *Reinhard Stockmann* und *Wilhelm Heitmeyer* einen kompetenten Ausblick auf die Herausforderungen, die sich für eine gesellschaftliche Nutzung der Evaluation stellen.