

## Rezension zu:

Bamberger, Michael J./Rugh, Jim/Mabry, Linda S.: *RealWorld Evaluation: Working Under Budget, Time, Data and Political Constraints*. 2. Auflage. Los Angeles/London/New Delhi/Singapore/Washington DC: Sage 2012. 712 Seiten plus Literaturangaben und Anhang, 40,99 £, ISBN 978-1-4129-7962-7

*Katharina Haas,<sup>1</sup> Sina Smid<sup>1</sup>*

Die zweite Auflage des Buches „RealWorld Evaluation: Working Under Budget, Time, Data and Political Constraints“ von Michael J. Bamberger, Jim Rugh und Linda S. Mabry thematisiert Programmevaluation vor dem Hintergrund der Restriktionen in den Bereichen Budget, Zeit, Daten und Politik, mit denen man in der Evaluationspraxis in der Regel konfrontiert wird. Es zeigt Wege und Möglichkeiten auf, um ein den entsprechenden Rahmenbedingungen angepasstes Evaluationskonzept zu entwickeln bzw. die passende Methodik auszuwählen und die Evaluation erfolgreich durchzuführen. Die Autoren und die Autorin präsentieren die 2012 erschienene 2. Auflage als Handbuch für eine breite Zielgruppe. Entgegen dem ersten Eindruck, den man anhand des Titelbildes gewinnt, beschränkt sich das 712-seitige Buch nicht auf den Bereich Entwicklungszusammenarbeit, aus dem sich der ‚RealWorld Evaluation‘-Ansatz entwickelte, sondern richtet sich an erfahrene und unerfahrenere Evaluationsanwender(innen), -organisator(inn)en und Projektverantwortliche in allen Kontexten der Evaluation. In dem stark methodenorientierten Buch legen die Autoren und die Autorin insbesondere Wert auf Impact-Evaluation, die Kombination quantitativer und qualitativer Methoden und die Einbettung der Evaluation in ein dem Programm zugrunde liegendes theoretisches Modell.

Die Autoren und die Autorin zeichnen sich durch langjährige Erfahrung im Evaluationsbereich aus. Bamberger und Rugh sind bereits jahrzehntelang als Evaluatoren, unabhängige Berater

und in der Lehre tätig. Im Auftrag der Weltbank und internationaler NGOs haben sie zahlreiche Programmevaluationen vor allem in Entwicklungsländern, z.B. in Lateinamerika, Afrika und Asien, durchgeführt. Mabry ist derzeit Professorin an der Washington State University. Ihre Forschungsschwerpunkte sind u.a. Programmevaluation und qualitative Sozialforschung. Sie alle hatten herausragende Positionen in mehreren Evaluationsassoziationen inne und sind Verfasser einer Reihe einschlägiger Publikationen.

Das Buch ist in drei Teile mit je einem thematischen Schwerpunkt und einem umfangreichen Anhang untergliedert. Im ersten Teil (Kapitel 1 bis 8) werden die sieben Schritte des ‚RealWorld Evaluation‘-Ansatzes vorgestellt, die dabei helfen sollen, Evaluationen unter den oben genannten Einschränkungen konzeptionell und methodisch angemessen durchführen zu können. Jedem der Schritte widmen die Autoren und die Autorin ein Kapitel. Im zweiten Teil gehen die Autoren und die Autorin vertiefend auf Evaluationsmethoden und -ansätze sowie Strategien im Umgang mit Restriktionen der ‚RealWorld‘ ein (Kapitel 9 bis 16). Organisation und Management von Evaluationen sind Gegenstandsbereich des dritten Teils (Kapitel 17 bis 19). Die Kapitel sind gut verständlich und in sich geschlossen geschrieben. Buchimmanente Querverweise und zahlreiche Literaturhinweise am Ende jedes Kapitels zum Vertiefen des jeweiligen Themas, ein Glossar, ebenso wie der 120 Seiten umfassende Anhang mit Arbeitsblättern und Checklisten für die Praxis, machen das Werk zu einem guten Nachschlagewerk und

---

1 Zentrum für Qualitätssicherung und -entwicklung der Johannes Gutenberg-Universität, Mainz

Arbeitsbuch. Hilfreich sind darüber hinaus zahlreiche Tabellen, Grafiken, Beispiel- und Definitionsbboxen.

Zunächst zu den sieben Schritten des ‚Real-World Evaluation‘-Ansatzes in Teil 1: Der erste Schritt beinhaltet die einleitende Planung der Evaluation. Es wird insbesondere auf die Bedeutung einer guten Abstimmung zwischen Auftraggeber(inne)n und Evaluators(inn)en bezüglich des Evaluationsziels und der Entwicklung eines Programm-Theorie-Modells hingewiesen. Der Umgang mit finanziellen und zeitlichen Einschränkungen, ebenso wie Wechselwirkungen zwischen diesen Restriktionen, werden in Schritt zwei und drei thematisiert. Lösungsansätze der Autoren und der Autorin hierfür sind beispielsweise das Ausweichen auf Sekundärdaten oder eine kleinere Stichprobe in Primärerhebungen. Thematisiert werden darüber hinaus die damit verbundenen Auswirkungen auf die Validität der Erhebung. Der vierte Schritt behandelt Strategien im Umgang mit Dateneinschränkungen im Hinblick auf den Anspruch, Wirksamkeitsevaluations durchzuführen. Insbesondere werden die Möglichkeiten aufgezeigt, im Nachhinein Ausgangsdaten zu rekonstruieren oder Vergleichsgruppen zu identifizieren, wenn Evaluators(inn)en nach Projektbeginn hinzugezogen werden. Im fünften Schritt wird die Gratwanderung zwischen dem Bewahren von Objektivität und der gleichzeitigen Restriktion der Evaluation durch teilweise auch unterschiedliche politische Interessen der Stakeholder diskutiert. Zentral ist hier die frühzeitige Identifikation der Anspruchsgruppen und möglicher Konflikte im Hinblick auf das Programm bzw. die Evaluation. Darauf folgend bezieht sich Schritt sechs der ‚RealWorld Evaluation‘ auf das Stärken des Evaluationsdesigns und der Validität der Schlussfolgerungen und Empfehlungen. Das Konzept der Validität wird ausführlich erläutert und im Kontext von qualitativer und quantitativer Methodenlehre sowie von Triangulation diskutiert. Anschließend wird ein Rahmen vorgestellt, welcher eine Beurteilung der Evaluationsdesigns, das Erkennen von Bedrohungen der Validitätskriterien der Evaluation und schließlich das Beseitigen von methodischen Schwächen erleichtern soll. Für diese Meta-Evaluation finden sich umfangreiche Checklisten im Anhang, die insgesamt über 60 Seiten umfassen. Appendix A bis E sind Muster-Arbeitsblätter, bestehend aus jeweils 3 Teilen (entsprechend

der Zielgruppen ‚senior management‘, ‚middle-level management‘ bzw. ‚evaluation manager‘, ‚evaluation specialist‘), mit Checklisten, mit deren Hilfe Bedrohungen der Validität zu unterschiedlichen Zeitpunkten der Evaluation (Design-Phase, Implementierungsphase, Berichtsphase(n)) identifiziert und bewertet werden können. Je nachdem, welche Methoden (quantitative, qualitative, mixed-methods) angewandt werden, stehen passende Arbeitsblätter zur Verfügung. In Schritt 7 geht es darum, die Zielgruppe für die Evaluation zu gewinnen und sie zu befähigen, Evaluationsergebnisse z.B. in Form eines selbstständig oder partizipativ entwickelten ‚Follow-up‘-Plans umzusetzen. Zunächst werden Gründe dafür aufgezeigt, warum Evaluationsergebnisse häufig nicht genutzt werden. Anschließend werden Kommunikationsstrategien vorgestellt, welche u.a. zu einem besseren Verständnis und einer Identifikation der Stakeholder mit der Evaluation und deren Ergebnissen beitragen können.

Im zweiten, methodischen Teil des Buchs besprechen die Autoren und die Autorin zunächst ethische Standards sowie ‚codes of conduct‘ für Evaluationen, welches eine Neuerung in der zweiten Auflage darstellt (Kapitel 9). Es stehen allerdings vor allem evaluatorsche Prinzipien und Richtlinien der USA im Vordergrund. Die Autoren und die Autorin gehen in Kapitel 10 vertiefend auf die Bedeutung eines der ‚Real-World Evaluation‘ zugrunde liegenden theoretischen Modells ein. Anhand dessen werden Hypothesen darüber aufgestellt, welche Interventionen, situative und individuelle Faktoren einen Einfluss auf den angestrebten Output, Outcome und letztendlich Impact des Programms haben. Darüber hinaus hilft das Modell dabei, kritische Punkte zu identifizieren, auf die der Fokus der Evaluation gelegt werden sollte. Je ein weiteres Kapitel widmen die Autoren und die Autorin der Auswahl eines angemessenen Evaluationsdesigns zur Wirksamkeitsanalyse (Kapitel 11), quantitativen und qualitativen Methoden (Kapitel 12 und 13) sowie deren Verknüpfung in ‚Mixed-Method‘-Designs (Kapitel 14). Neben Tipps im Hinblick auf die Datenerfassung und Ergebnisauswertung geben die Autoren und die Autorin dem Leser/der Leserin ‚best-practice‘-Beispiele an die Hand. Eine detailliertere Darstellung aller vorgestellten Evaluationsdesigns mit jeweils einem Fallbeispiel findet man darüber hinaus in Appendix F. Im Kontext der Aussagekraft der

Evaluationsergebnisse steht auch Kapitel 15 über Stichprobenziehung und -größe, eine der Stellschrauben bei zeitlichen und finanziellen Einschränkungen von RealWorld-Evaluationen. Anhand von zahlreichen Berechnungsbeispielen wird verdeutlicht, welche Auswirkungen eine Reduktion der Stichprobengröße auf die zu messende Effektstärke und Signifikanztests hat. Das in dieser Auflage neue Kapitel 16 widmet sich der Evaluation komplizierter und komplexer Entwicklungsinterventionen. Die Autoren und die Autorin erweitern nach einem einführenden, theoretischen Abschnitt das bis dahin behandelte Methodenspektrum durch im Evaluationskontext neuere Ansätze, z.B. das ‚concept mapping‘, die Netzwerk- oder Portfolioanalyse. Auch in diesem methodisch anspruchsvollen Feld wird auf Wirksamkeitsevaluation mit (quasi-)experimentellen Designs Wert gelegt, immer mit Blick darauf, wie mit der jeweiligen Methode die kontrafaktische Situation ermittelt wird. Die Methoden werden in diesem Kapitel leider ausschließlich anhand von Beispielen aus dem Feld der Entwicklungszusammenarbeit verdeutlicht. Zwar verweisen die Autoren und die Autorin auf die Anwendbarkeit von Definitionen und Methoden in anderen Themenfeldern, dennoch fällt die Einseitigkeit der Beispielauswahl auf. Das ‚concept mapping‘ als Evaluationsstrategie für komplexe, multidimensionale Programme wird in Appendix G ausführlich erläutert. Unklar bleibt, warum gerade dieser Ansatz noch einmal herausgegriffen wird. In Appendix H wird die Methodik zur Evaluation solcher komplexen Programme auf die Evaluation von Länderprogrammen angewandt.

Teil 3 enthält die der zweiten Auflage hinzugefügten Kapitel „Organizing and Managing Evaluations“ (Kapitel 17) und „Strengthening Evaluation Capacity at the Agency and National Levels“ (Kapitel 18). Diese zielen darauf ab – über finanzielle, zeitliche, Daten betreffende und politische Einschränkungen hinaus – organisatorische Schwierigkeiten zu thematisieren und die Evaluationskapazitäten auf verschiedenen Ebenen zu stärken. Für ein verbessertes Evaluationsmanagement werden in Kapitel 17 sechs Organisationsschritte aufgeführt, die in fast allen Evaluationen angewandt werden können. Sie bestehen aus Vorbereitungen und Personalbeschaffung für die Evaluation, Entwerfen des Evaluationsdesigns, Organisation der Durchführung, Bericht der Ergebnisse und zuletzt der Sicherstel-

lung ihrer Umsetzung. In Kapitel 18 wird zur Stärkung der Evaluationskapazitäten auf vier Schwerpunkte eingegangen: (1) Die Entwicklung eines Systems zur Qualitätssicherung der Evaluation. Dies beinhaltet beispielsweise die Verbesserung der Validität unter Hinzunahme der Arbeitsblätter des Appendix. (2) Die Bedeutung einer von Beginn an engen Zusammenarbeit zwischen Evaluators(inn)en und Management. (3) Die Stärkung der am Evaluationsprozess beteiligten Akteure (z.B. finanziell unterstützender Agenturen, der Evaluators(inn)en, der Nutzer(innen) der Evaluation, der profitierenden Zielgruppe und der öffentlichen Meinung). (4) Die Institutionalisierung des Evaluationssystems. Das abschließende Kapitel enthält einen zusammenfassenden Überblick über die Schwerpunkte des Buchs und künftigen Herausforderungen im Feld der Programmevaluation.

Liest man „RealWorld Evaluation“ im Ganzen, entsteht beim Leser der Eindruck von ständigen Wiederholungen, innerhalb und zwischen Kapiteln. Dies ist u.a. darauf zurückzuführen, dass Textinhalte in Grafiken und Tabellen mehrfach in unterschiedlicher Form aufbereitet werden. Positiv zu bewerten ist, dass dadurch auch einzelne Kapitel herausgegriffen werden können.

Für besonders geeignet halten wir das Werk für weniger Erfahrene und Evaluations-Einsteiger(innen) sowie zum Einsatz in der Lehre. Es bietet einen guten thematischen Einstieg durch grundlegende definitorische Abschnitte. Hervorzuheben sind hier Kapitel 7 zur Validität der Evaluation und Kapitel 16 zur Evaluation komplexer Interventionen. Die grundlegende Darstellung eines breiten Methodenspektrums enthält zwar vorwiegend die gängigen Methoden der empirischen Sozialforschung und wenig Innovatives, die Autoren und die Autorin behandeln diese allerdings mit Fokus auf den Umgang mit zeitlichen, finanziellen und politischen Beschränkungen sowie Einschränkungen in Bezug auf die Datenverfügbarkeit bei Programmevaluationen, also Hindernissen, die jedem Evaluator/jeder Evaluatorsin in der Praxis begegnen. In diesem engen Praxisbezug liegt auch der Mehrwert von „RealWorld Evaluation“ gegenüber anderen (Lehr-)Büchern zum Thema Programmevaluation. Sicherlich finden auch Praxiserfahrene die eine oder andere Anregung darin.

Die Erweiterungen der 2. Auflage sind inhaltlich sinnvolle Ergänzungen der 1. Auflage. Dies betrifft jeweils ein Kapitel zur Evaluation

komplexer Interventionen, zu ethischen Standards, zu Organisation und Management von Evaluationen sowie dem Stärken der Evaluationskapazitäten. Es wird Wert darauf gelegt, aktuelle methodische Diskussionen aufzugreifen. Hervorzuheben ist, dass die Autoren und die Autorin die Folgen der notwendigen Anpassungen und Beschränkungen des Evaluationsdesigns auf die Aussagekraft der Ergebnisse deutlich machen und darüber hinaus Wege aufzeigen, wie dennoch die größtmögliche Validität der Evaluationsergebnisse im Hinblick auf die Wirksamkeit von Programminterventionen erzielt werden kann.

Mit Blick auf das Themenspektrum im Kontext von Evaluationen würde unseres Erachtens ein Kapitel mit Praxistipps zur Angebotserstellung und Auftragseinwerbung hervorragend zu „RealWorld Evaluation“ passen.

In der Praxis sind zum Zeitpunkt der Ausschreibung genaue Zielsetzungen, ebenso wie Restriktionen der Evaluation noch nicht bekannt. Die Konzeptionierung und Kalkulation unterliegen somit diesen Beschränkungen.

Über das Buch hinaus erwähnenswert ist die dazugehörige Homepage [www.realworldevaluation.org](http://www.realworldevaluation.org). Hier findet sich neben Reviews und Zusatzmaterialien eine frei zugängliche, auf 37-Seiten komprimierte Zusammenfassung der 2. Auflage, die sich am Aufbau des Buchs orientiert und die jeweilige Zusammenfassung der Kapitel enthält – ein guter Einstieg, bevor man sich für das durchaus empfehlenswerte, aber sehr umfangreiche, nicht ganz günstige Buch für umgerechnet ca. 50 EUR (£ 40,99) entscheidet.

## Rezension zu:

Hennefeld, Vera/Stockmann, Reinhard (Hg.): *Evaluation in Kultur und Kulturpolitik. Eine Bestandsaufnahme*. Münster u.a.: Waxmann, 2013. 245 Seiten, 29,90 EUR, ISBN 973-3-8309-2819-5

*Anette Violet*<sup>1</sup>

Das Buch „Evaluation in Kultur und Kulturpolitik – eine Bestandsaufnahme“ richtet sich zum einen an Kulturschaffende und Beschäftigte in Kulturinstitutionen, die sich für die Themen Evaluation und Monitoring interessieren oder Berührungspunkte mit diesen Bereichen haben. Zum anderen eignet sich das Buch für Praktiker(innen), die selbst Evaluationen im Bereich Kultur und Kulturpolitik durchführen, sowie für Wissenschaftler(innen) aus diesem Politikfeld.

Sowohl in der deutschen Kultur und Kulturpolitik als auch in der Auswärtigen Kultur- und Bildungspolitik (AKBP) hat das Interesse an Evaluation in den letzten Jahren stetig zugenommen. „Evaluation in Kultur und Kulturpolitik – eine Bestandsaufnahme“ ist die erste Veröffentlichung im deutschen Sprachraum, die sich übergreifend mit dem Thema Evaluation in Kultur und Kulturpolitik beschäftigt. Ziel des von Vera Hennefeld und Reinhard Stockmann herausgegebenen Buches ist es, den Status quo des in der Evaluation im Kulturbereich, insbesondere in Deutschland, vorhandenen Wissens zu bündeln. Dies geschieht im Rahmen von neun Artikeln unterschiedlicher Autorinnen und Autoren einerseits auf der theoretischen und andererseits auf der praktischen Ebene. So soll dazu beigetragen werden, dass fruchtbare Diskussionen entstehen und dass sich Evaluation weiter in der Kultur und Kulturpolitik verbreitet und zunehmend Akzeptanz findet.

Das 245-seitige Buch ist in drei, jeweils wiederum drei Artikel umfassende Teile unterteilt. Im ersten Teil wird zunächst der Gegen-

standsbereich bestimmt. Im zweiten Teil werden ausgewählte Bereiche der Evaluationspraxis in Deutschland beleuchtet und im dritten Teil wird schließlich ein Einblick in die aktuelle Evaluationspraxis in der Kultur und Kulturpolitik gegeben, indem drei Best-Practice-Studien dargestellt werden. Alle der in „Evaluation in Kultur und Kulturpolitik“ veröffentlichten Autorinnen und Autoren haben langjährige theoretische und anwendungsbezogene Erfahrungen im Bereich Evaluation.

Im Vorwort des Buches stellen die Herausgeber fest, dass die enormen Herausforderungen unserer Zeit, wie Globalisierung, Digitalisierung oder Staatsschuldenkrise, knappere Ressourcen zur Folge haben, was sich auch im Politikfeld in einem Legitimationsdruck niederschlägt. Dass Evaluation im Kulturbereich eine zunehmende Rolle spielt, äußert sich unter anderem in der Gründung des Arbeitskreises Evaluation von Kultur und Kulturpolitik (2006) in der Gesellschaft für Evaluation (DeGEval) sowie darin, dass mehr und mehr Studien durchgeführt werden.

Der erste Teil des Buches, in dem zunächst der Gegenstandsbereich bestimmt wird, beginnt mit einem Artikel von Armin Klein, in dem die Rolle und Bedeutung von Evaluation in der deutschen Kulturpolitik beschrieben wird. Klein beschreibt zunächst den Aufbau des deutschen Kulturbetriebes und systematisiert unterschiedliche Arten von Kulturgütern. Er arbeitet heraus, dass die aktuellen Herausforderungen, wie „Globalisierung, demographische Verwerfung, Digi-

---

1 weltentwickeln – International Training + Evaluation, Berlin

„talisierung, Finanzkrise und Schuldenbremse“ (Klein 2013: 9) sich auch auf den öffentlich getragenen bzw. subventionierten Kulturbetrieb auswirken und fordert eine ordnungspolitische Debatte. Aspekte, wie die Fragen danach, welche kulturellen Leistungen staatlich gefördert und welche von anderen, zum Beispiel privatwirtschaftlich-kommerziellen Anbietern erbracht werden sollten, sowie was genau mit welchen konkreten Zielen für welchen Zeitraum gefördert werden soll, müssen laut Klein dringend geklärt werden. In diesem Zusammenhang stellt der Autor die Bedeutung von Evaluation – die Überprüfung, ob angestrebte Ziele im Kulturbereich tatsächlich erreicht wurden – heraus. Dabei betont Klein, dass bei Evaluationen im Kulturbereich ein partizipatives Vorgehen von besonders hoher Bedeutung ist, weil „im Sektor von Kunst und Kultur ganz spezifische Sensibilitäten zu berücksichtigen sind“ (ebd.: 30).

Es folgt ein Kapitel von Kurt-Jürgen Maaß, in dem die Rolle und Bedeutung von Evaluation in der deutschen AKBP analysiert wird. Maaß stellt zunächst die Entwicklung der deutschen AKBP sowie der Bedeutung von Evaluation innerhalb der deutschen AKBP von Ende des 2. Weltkrieges bis heute dar. Er stellt fest, dass seit einer Enquete Kommission in den 1970er Jahren im Auftrag des deutschen Bundestages Empfehlungen für eine bessere kulturelle Repräsentation der Bundesrepublik Deutschland im Ausland erarbeitet hat, keine auch nur annähernd vergleichbare Evaluation der AKBP stattgefunden hat (Maaß: 37). In den letzten zehn Jahren sei Evaluation jedoch wieder zu einem deutlich wichtigeren Thema in der AKBP geworden, wobei sich das Auswärtige Amt eher als „Evaluationsvermeider“ erweise (ebd.: 43). Anschließend führt Maaß einige praktische Beispiele an, um zu belegen, dass die Zurückhaltung des Auswärtigen Amtes unbegründet und AKBP evaluierbar ist. Er führt Beispiele für Instrumente an, die sich problemlos auf die AKBP übertragen lassen und schließt mit einem Plädoyer dafür, dass Thema Evaluation noch stärker und ohne Ressentiments in der deutschen AKBP zu etablieren.

Der erste Teil des von „Evaluation in Kultur und Kulturpolitik – eine Bestandsaufnahme“ schließt mit einem Artikel von Reinhard Stockmann, in dem die Methodik von Evaluationen im Politikfeld Kultur und Kulturpolitik dargestellt und diskutiert wird. Stockmann belegt zunächst, dass inzwischen sowohl aus Sicht der innerdeutschen Kulturförderung als auch aus der Perspek-

tive der AKBP Evaluation als notwendig erachtet wird. Anschließend zeigt er die notwendigen Voraussetzungen dafür auf, dass Evaluation in der Kultur und Kulturpolitik sinnvoll eingesetzt werden kann. Als zentrale Voraussetzung führt Stockmann die klare, operationale und wirkungsbezogene Formulierung von Programmzielen und Leistungsvereinbarungen mit Durchführungs- oder Mittlerorganisationen an (Stockmann 2013: 60). Schließlich geht der Autor auf Ziele, Aufgaben, Formen und Verfahren von Evaluation im Bereich der Kultur und Kulturpolitik ein. Unter anderem wird der von Stockmann entwickelte partizipative Evaluationsansatz des CEval dargestellt, der sich insbesondere für Evaluationen in Kultur und Kulturpolitik eignet, und die Evaluationsstandards der DeGEval – Nützlichkeit, Durchführbarkeit, Fairness und Genauigkeit – werden erläutert.

Der zweite Teil des Buches setzt sich aus ausgewählten Bereichen der Evaluationspraxis in Deutschland zusammen. Er beginnt mit einem Text von Gesa Birnkraut, in dem der Einfluss der föderalen Struktur der deutschen Kulturförderung auf die Bedeutung von Evaluationen und die Evaluationspraxis in diesem Bereich analysiert wird. Birnkraut gibt anfangs einen Überblick über die Strukturen der deutschen Kulturförderung und benennt die Herausforderungen für Evaluation in diesem Bereich. Es werden drei Ebenen der Evaluation im Bereich Kultur und Kulturpolitik aufgezeigt und diskutiert: Evaluation von Projektförderung, von institutioneller Förderung und Evaluation der übergreifenden Ebene, der Kulturförderung selbst. Anhand von zwei konkreten Beispielen aus dem Bereich der institutionellen Förderung zeigt Birnkraut anschließend, wo die Chancen und wo die Herausforderungen von ganzheitlichen Evaluationen in diesem Bereich liegen. Die Autorin kommt zu dem Schluss, dass das föderale System der deutschen Kulturförderung Evaluation einerseits erleichtert, weil viele verschiedene Ansätze von Evaluation in Eigenverantwortung der Länder umgesetzt werden können. Andererseits behindert das föderale System Evaluationen auch, da der Austausch und ein gemeinsames Verständnis für das Thema strukturell erschwert werden.

Es folgt eine Bestandsaufnahme der Wirkungsforschung zur kulturellen Bildung von Vanessa-Isabelle Reinwand. Ziel von Reinwands Artikel ist es, einen Überblick über die Wirkungsforschung in der kulturellen Bildung darzustellen. Sie beginnt damit, zentrale Begriffe wie „kulturelle Bildung“ zu bestimmen. An-

schließlich stellt sie unterschiedliche Ansätze und Studien dar, in denen die Wirkung von kultureller Bildung untersucht wird. In diesem Zusammenhang geht sie auf die Theorie der ästhetischen Erfahrung ein sowie auf kunstspartenspezifische Strukturanalysen, Biografieforschung, Transferforschung, Evaluationen, prozessorientierte Forschung und Kulturstatistiken. Die Autorin stellt eine der wenigen internationalen Vergleichsstudien dar, in der der Versuch einer weltweiten Bestandsaufnahme der Qualität künstlerischer Bildung unternommen wurde (vgl. Bamford 2006). Reinwands Artikel endet mit dem Fazit, dass zwar bereits zahlreiche Studien zur Wirkung kultureller Bildung existieren, diese jedoch zusammenhangslos über verschiedene Wissenschaftsdisziplinen verteilt sind. Sie weist darauf hin, dass disziplin- und länderübergreifende wissenschaftliche Gremien und Arbeitsgruppen fehlen, ebenso wie klare Begriffsbestimmungen und -abgrenzungen. Als weitere Herausforderung nennt die Autorin den Transfer wissenschaftlicher Ergebnisse in die Praxis.

Der zweite Teil wird durch einen Beitrag von Vera Hennefeld abgeschlossen. Ziel des Artikels ist es, die strategische Steuerung der Mittlerorganisationen im Bereich Kultur und Kulturpolitik durch mit dem Auswärtigen Amt geschlossene Zielvereinbarungen unter Einsatz von Evaluation darzustellen sowie den potenziellen Nutzen und Mehrwert zu diskutieren, der so in der Praxis erzielt werden kann. Hennefeld zeigt zunächst, dass das Auswärtige Amt zunehmend auf das Instrument der Zielvereinbarungen zurückgreift, um die Mittlerorganisationen strategisch zu steuern, und dass in diesem Zusammenhang auch Evaluation ein zentrales Thema ist. Anschließend gibt die Autorin eine Einführung in das aus der Betriebswirtschaftslehre stammende Konzept der Zielvereinbarungen und stellt dar, inwieweit dieses auf die öffentliche Verwaltung übertragen werden kann. Sie beleuchtet den Status quo des Einsatzes von Zielvereinbarungen in der AKBP und zeigt Optimierungspotenziale auf. Anschließend werden die Systematik und der Nutzen einer hierarchischen Zielpyramide dargestellt. Hennefeld macht deutlich, dass ein zentrales Kriterium für den Erfolg der strategischen Steuerung über Zielvereinbarungen die Akzeptanz des Systems und der Inhalte seitens der zentralen Akteure ist und gibt praktische Hinweise dafür, diese zu sichern. In ihrem Fazit stellt die Autorin u.a. fest, dass die Entwicklung einer qualitativ hochwertigen Zielvereinbarung zwar einerseits eine theore-

tisch, methodisch und vor allem auch politisch anspruchsvolle und zeitintensive Aufgabe ist. Andererseits wird der Abstimmungsaufwand im Nachgang reduziert, weil zum Beispiel nicht mehr jedes Einzelvorhaben gesondert geprüft und bewilligt werden muss, und die Mittlerorganisationen gewinnen höhere Freiheitsgrade bei der Konzeption von Maßnahmen.

Das Buch wird im dritten Teil durch einen Einblick in die Praxis abgerundet, indem drei Best-Practice-Studien beschrieben werden. Zunächst werden die Ergebnisse einer internationalen Metastudie von Volker Kirchberg und Robin Kuchar dargestellt. Die Autoren haben eine Metastudie von 16 Arbeiten aus 12 Ländern durchgeführt und sie hinsichtlich Inhalt, Methodik und Theoriegehalt auf Gemeinsamkeiten und Unterschiede hin analysiert. Ziel der Autoren ist dabei, die Vergleichbarkeit von Ergebnissen internationaler Studien zur Kulturnutzung zu evaluieren sowie Best-Practice-Beispiele als Grundlage für die inhaltlich, methodisch und theoretisch hochwertige Durchführung weiterer Studien zu identifizieren. Die Autoren kommen aufgrund ihrer Analyse u.a. zu dem Schluss, dass der Forschungsstand der repräsentativen Bevölkerungsforschung zu Kulturnutzung vor allem in Deutschland zu wünschen übrig lässt. Sie identifizieren insbesondere die flämische Studie des *Cultuurkijkers* zur Kulturnutzung einer repräsentativen Bevölkerungsstichprobe als wertvolle Benchmarkstudie. Kirchberg und Kuchar schließen mit Ausführungen zu den Bedingungen, die notwendig dafür sind, die Qualität und internationale Vergleichbarkeit entsprechender Studien zu steigern.

Weiterhin wird ein Ansatz zur Wirkungsevaluation von Veranstaltungen zur themenspezifischen Sensibilisierung vorgestellt (Hansjörg Gaus und Christoph Müller). Die Autoren stellen zunächst die methodischen Grundlagen der kausalen Wirkungsprüfung vor dem Hintergrund von Sensibilisierungsinterventionen dar. Sie ziehen anschließend die quasi-experimentelle Wirkungsevaluation einer Maßnahme zur themenspezifischen Sensibilisierung von Verbraucher(inne)n zum Thema Klimaschutz als Beispiel heran, um mit Wirkungsevaluationen verbundene Herausforderungen plastisch aufzuzeigen. Wenngleich es sich dabei nicht um ein Beispiel aus der Kultur und Kulturpolitik handelt, können die Ergebnisse und Implikationen ohne Weiteres auf dieses Feld übertragen werden. Gaus und Müller beschreiben ausführlich den Versuchsaufbau sowie die verwendeten Datenerhebungs-

und Auswertungsmethoden. Sie zeigen detailliert, wie es gelingen kann, trotz der Unmöglichkeit, Personen randomisiert einer Versuchs- und einer Vergleichsgruppe zuzuordnen und zwei Messzeitpunkte zu realisieren, Aussagen über die Wirkung einer Veranstaltung zu treffen.

Im letzten Beitrag wird ein kulturraumsensibles Monitoringinstrument sowie dessen Möglichkeiten und Grenzen beschrieben und diskutiert (Ute M. Metje und Peter Jablonka). Auftrag der Autorin/des Autors war, für das Goethe-Institut (GI) ein Monitoringinstrument zur Evaluierung des Besucherprogramms zu entwickeln. Im Rahmen des Besucherprogramms lädt das GI jährlich 400-500 ausländische Gäste aus unterschiedlichen Ländern nach Deutschland ein. Seit 2012 wird das Besucherprogramm durch das GI selbst evaluiert, die Entwicklung eines entsprechenden Monitoringinstruments wurde aber extern in Auftrag gegeben. Zentrale Anforderung an das Instrument war dabei, dass dieses kulturraumsensibel sein sollte: Menschen aus unterschiedlichen Kulturkreisen, mit unterschiedlichen Wert- und Normvorstellungen, sollten ein und dasselbe Instrument nutzen. Die Autorin/der Autor beschreiben ausführlich ihre Vorgehensweise bei der Entwicklung des Instruments.

In einem Exkurs gehen sie auf den Begriff der „Kulturraumsensibilität“ ein und erläutern,

warum sie bei der Entwicklung des Monitoringinstruments das Messmodell des ereignisorientierten Ansatzes wählten.

Sie gehen auf Aspekte der technischen Umsetzung ein und stellen erste Befragungsergebnisse dar. Die Bewertung der Kulturraumsensibilität des Instruments kann allerdings erst nach weiteren, umfangreichen Datenerhebungen erfolgen.

Fazit: Der von den Herausgebern formulierte Anspruch, eine aktuelle Bestandsaufnahme der Evaluation im Kulturbereich zusammenzustellen, ist voll und ganz geglückt. Besonderer Verdienst des Buches ist die gelungene Balance zwischen theoretischen Ausführungen zu Gegenstandsreich und Methodik und den Best-Practice-Beispielen. Da grundlegende Begriffe geklärt werden, eignet sich das Buch sowohl für Leserinnen und Leser, die bislang wenig Erfahrungen im Bereich Evaluation in der Kultur und Kulturpolitik gesammelt haben, als auch für erfahrene Theoretiker(innen) und Praktiker(innen), die insbesondere im Rahmen des dritten Teils durch die dargestellten Best-Practice-Studien wertvolle Hinweise und Gedankenanstöße für die Planung und Durchführung von Evaluationen erhalten. Alles in allem bietet das Buch einen sehr guten Einblick in die Evaluation in Kultur und Kulturpolitik und ist ein wertvoller Beitrag zur aktuellen Diskussion.



## Rezension zu:

Widmer, Thomas/De Rocchi, Thomas: Evaluation. Grundlagen, Ansätze und Anwendungen. Zürich: Rüegger Verlag, 2012. 188 Seiten, 16,50 EUR, Reihe Kompaktwissen CH, Bd. 16, ISBN: 978-3-7253-0980-1

*Anita Pachner*<sup>1</sup>

Der in der Sachbuchreihe „Kompaktwissen CH“ erschienene Band verfolgt die Zielsetzung, einem breiten Publikum „in einer übersichtlichen Form einen fundierten Einblick in das facettenreiche Instrumentarium der Evaluation mit seinen Möglichkeiten und Grenzen“ (S. 7f.) zu bieten. Bei der Einführung handelt es sich folglich nicht um ein Lehrbuch, sie will vielmehr „der Information und Aufklärung“ (S. 8) dienen. Dies geschieht vor dem Hintergrund, dass sich „Evaluation als soziale Praxis“ (S. 7) zunehmender Bedeutung erfreut, was sich nicht nur in einer wachsenden Anzahl von Fachpublikationen niederschlägt, sondern auch in der Gründung nationaler und internationaler Fachgesellschaften sowie einer Vielzahl von Aus- und Weiterbildungsangeboten. Evaluation ist damit inzwischen als Profession und/oder Disziplin fest in den meisten Bereichen staatlichen Handelns in Europa und insbesondere in der Schweiz verankert. Dabei wird Evaluation in der vorliegenden Publikation als „Instrument zur systematischen und transparenten Bewertung eines Gegenstandes im öffentlichen Raum“ (S. 7) aufgefasst und somit abgegrenzt von wissenschaftlicher (Grundlagen-) Forschung.

Das Buch umfasst viele im Rahmen einer Einführung wichtigen Aspekte von Evaluation zuzüglich einer Schwerpunktsetzung auf ihre politischen Rahmenbedingungen und – entsprechend der Reihe, in der das Werk erschienen ist – die Evaluationspraxis in der Schweiz.

Im Anschluss an eine auf das Thema hinführende **Einleitung** gibt Kapitel 2 einen Über-

blick über **Grundbegriffe der Evaluation**, nennt wichtige Definitionen und nimmt Abgrenzungen zu verwandten Konstrukten vor. Dabei wird der Facettenreichtum des Evaluationsansatzes deutlich. Als leitende Arbeitsdefinition für den Band wird formuliert: „Unter Evaluation wird eine wissenschaftliche Dienstleistung verstanden, die sich mit der systematischen und transparenten Bewertung eines Gegenstandes befasst“ (S. 11). Wissenschaftlichkeit, der Dienstleistungsgedanke und die Zielsetzung der Bewertung werden im Kontext dieser Einführung folglich als zentrale Elemente von Evaluation aufgefasst. Im Weiteren werden Bewertungskriterien expliziert, Evaluation als Prozess und als Produkt dargestellt und der Begriff „Evaluationsforschung“ kritisch beleuchtet. Im zweiten Teil des Kapitels wird Evaluation unterschieden von alltäglicher Bewertung, von wissenschaftlicher Forschung und von verschiedenen Instrumenten der (öffentlichen) Verwaltungspraxis. Der dritte Abschnitt widmet sich dem Verhältnis zu verwandten Ansätzen bzw. speziellen, meist prospektiven Evaluationsverfahren von der Gesetzesfolgenabschätzung bis hin zur Finanzbewertung. Der letzte Teil des zweiten Kapitels geht auf die wichtigsten Dimensionen ein, entlang derer Evaluationen ausgestaltet werden können. Diese sind Gegenstand, Funktion, Wirkungsdimension, Verortung und das Verhältnis Evaluandum – Evaluation.

In Kapitel 3 werden in einem historischen Zugriff Stufen der **Entwicklung der Evaluation** nachgezeichnet. Dies erfolgt unter enger Bezug-

1 Fachgebiet Pädagogik, Technische Universität Kaiserslautern

nahme auf die sich wandelnden politischen Rahmenbedingungen. Zunächst wird kurz die thematische Entwicklung des Evaluationsfeldes ausgehend von den 1960er Jahren in den USA bis in die Gegenwart geschildert. Es wird konstatiert, dass heute Evaluation in fast allen Gesellschaftsbereichen und Politikfeldern zu finden ist und damit eine große thematische Breite erlangt hat. Im zweiten Abschnitt des Kapitels wird die konzeptionelle Entwicklung des Evaluationsfeldes anhand des „Generationenmodells“ von Guba und Lincoln (1989) nachvollzogen. Sie unterscheiden die vier Entwicklungsschritte Messen, Beschreiben, Bewerten und Verhandeln. Der dritte und ausführlichste Teil dieses Kapitels widmet sich politischen Rahmenbedingungen und ihrem Einfluss auf die Evaluationspraxis. Entlang der folgenden fünf Entwicklungslinien werden die sogenannten Evaluationskonjunkturen (S. 41) näher beschrieben: sozialstaatliche Expansion, neoliberaler Sparstaat, New Public Management, post-moderne Multi-level-Governance und Evidence-based Policy Making. Abschließend wird der Frage nachgegangen, welchen Beitrag umgekehrt Evaluation für die liberal-demokratische Entwicklung eines Staates leisten kann.

Kapitel 4 expliziert grundlegende **Theorien, Konzepte und Modelle** der Evaluation, wobei insbesondere Effektivitäts-Modelle, ökonomische Modelle und akteursorientierte Modelle unterschieden werden. Die Effektivitäts-Modelle werden weiter untergliedert in Modelle, die einen zielorientierten (wie z.B. die Objective-based Evaluation), einen resultatorientierten (wie z.B. die Goal-free Evaluation) oder einen erklärungsorientierten Ansatz verfolgen. Letzterer wird in attributionsorientierte Ansätze (wie z.B. die Randomized Controlled Trials) und theoriebasierte Ansätze (wie z.B. das CIPP-Modell) unterteilt. Bei den ökonomischen Modellen werden Produktivitäts-Modelle (wie z.B. das Cost-productivity Model) und Effizienz-Modelle (wie z.B. die Cost-effectiveness Analysis) unterschieden. Die akteursorientierten Modelle schließlich werden ausdifferenziert in klientenorientierte (wie z.B. die Utilization-focused Evaluation) und stakeholderorientierte (wie z.B. die Responsive Evaluation) Ansätze.

Kapitel 5 ist **Methoden** der Evaluation und ihren Besonderheiten gewidmet und informiert in knapper Form über spezifische Instrumente der Evaluation. Insbesondere in diesem Kapitel wird sichtbar, dass es sich bei dem einführenden Werk nicht um ein Lehrbuch handelt. In einem ersten

Abschnitt wird kurz auf sozialwissenschaftliche Methodik im Allgemeinen eingegangen. Dazu werden die gängigsten Untersuchungsdesigns und Untersuchungstechniken zur Datenerhebung und -auswertung angerissen. Der zweite Teil des Kapitels beschreibt als Besonderheiten der Evaluation die Kombination von Methoden und Verfahren, die Zeitdimension, die geringe Bedeutung der Generalisierung, die Vergleichsdimensionen, die Bewertungsdimension und den wirtschaftlichen Einsatz von Design und Verfahren. Der dritte Abschnitt nennt als spezifische Instrumente und Verfahren der Evaluation verschiedene Bewertungsverfahren sowie Leistungs- und Wirkungsindikatorensysteme. Insbesondere in diesem Kapitel wird wiederholt auf den Unterschied von Evaluationen und sozialwissenschaftlicher Forschungstätigkeit abgehoben.

In Kapitel 6 wird der **Evaluationsprozess – Schritt für Schritt**, also das idealtypische Vorgehen bei der Durchführung einer Evaluation, erläutert. Es hat damit fast den Charakter eines Leitfadens bzw. einer Anleitung zur Lancierung von Evaluationsprojekten. So werden zentrale Fragen für die Planung eines Evaluationsprojektes aufgeführt. Anschließend wird die Ausarbeitung eines Pflichtenheftes und die Auftragsvergabe vom Einholen von Offerten über deren Beurteilung bis hin zum Vertragsabschluss beschrieben. Die folgenden beiden Abschnitte beschäftigen sich schließlich mit der Durchführung und Begleitung der Evaluationsaktivitäten sowie der Berichterstattung und Verbreitung der Resultate.

Kapitel 7 – mit Kapitel 4 das umfangreichste Kapitel des Bandes – stellt unter der Überschrift **Evaluation im Kontext** die Evaluationspraxis in der Schweiz vor. Zunächst wird dazu auf die institutionelle Einbettung der Evaluationsfunktion entlang ihrer historischen Entwicklung in der Schweiz eingegangen. Besonders bemerkenswert ist dabei die Verankerung der Überprüfung der Wirksamkeit von Maßnahmen des Bundes in der Verfassung (Art. 170 BV). Darüber hinaus werden die Formulierung von Evaluationsklauseln und die organisatorische Einbettung von Evaluation in der Praxis dargestellt. Der zweite Teil des Kapitels beleuchtet die Evaluationspraxis in der Schweiz auf der Ebene des Bundes sowie der Kantone und Gemeinden. Der dritte Teil setzt sich mit der Evaluationsnutzung unter den Stichworten Nutzungsforschung und Nutzungstypen, „Theory of Influence“ sowie „Evidence-based Policy Making“ auseinander.

Der abschließende vierte Abschnitt wendet sich der Qualität von Evaluation zu und beschäftigt sich neben der Evaluationsqualität mit der Entwicklung von Standards für die Evaluation, wozu die „SEVAL-Standards“ aufgelistet werden. Außerdem wird die Anwendung von Standards allgemein sowie in Meta-Evaluationen, wo sie als Bewertungskriterien fungieren können, behandelt. Zum Schluss erfolgt eine Würdigung der Evaluationsstandards als Instrument zur Bewertung der Qualität von Evaluationen.

Die **Diskussion** in Kapitel 8 schließlich fasst den aktuellen Stand der Evaluation zusammen, indem sie ihn kritisch würdigt, und kommt zu dem Schluss, dass erstens „die Evaluation im schweizerischen politischen System durchaus intakte Zukunftsaussichten hat. Zudem ließen sich diese zweitens sogar noch erheblich verbessern, wenn in Zukunft noch selektiver, zielorientierter und differenzierter evaluiert würde“ (S. 167).

Wie aus diesem Überblick über die Inhalte des Buches bereits deutlich wird, besteht die Besonderheit dieser Einführung in die Evaluation in der sehr detaillierten Auseinandersetzung mit politischen Rahmenbedingungen von Evaluation einerseits und der Fokussierung auf die Evaluationspraxis in der Schweiz andererseits.

Insbesondere die in dem Band vorgenommene Unterscheidung von Evaluation und wissenschaftlicher Forschung, womit v.a. sozialwissenschaftliche Grundlagenforschung gemeint ist und die sich wie ein roter Faden durch die Publikation zieht, ist nicht immer nachvollziehbar.

So gibt es auch in der sozialwissenschaftlichen Forschungspraxis anwendungsorientierte Forschungsbemühungen, die nicht eindeutig der Grundlagenforschung zuzurechnen sind, von staatlicher Seite in Auftrag gegeben werden (in Deutschland z.B. vom Bundesministerium für Bildung und Forschung, kurz BMBF) und damit ähnlichen Gesetzmäßigkeiten unterliegen, wie sie die Autoren hier ausschließlich für „die Evaluation“ im Kontext von Politik und Verwaltung verbuchen. Ein ergänzender Untertitel etwa hätte deutlich machen können, dass Evaluation hier als in politischen Anwendungskontexten stattfindende soziale Praxis betrachtet wird.

Lässt sich der Leser jedoch auf das dargelegte Verständnis von Evaluation ein, dann handelt es sich bei der Einführung um eine leicht lesbare, weitgehend verständliche und gut strukturierte Annäherung an das komplexe Feld von Evaluation im Kontext staatlichen Handelns. Dazu tragen insbesondere die praxisnahen Beispiele, die zahlreichen Veranschaulichungen komplexer Sachverhalte sowie die mittels Überleitungen und Zusammenfassungen gut gelungene Leserführung bei. Der Band eignet sich damit gut als Informationsquelle und Orientierungshilfe für ein breites, an verwaltungs- oder politiknahen Anwendungsfeldern von Evaluation interessiertes Publikum, umso mehr, wenn es von der schweizerischen Perspektive zu gewinnen vermag.



Hans Anand Pant, Petra Stanat,  
Ulrich Schroeders, Alexander Roppelt, Thilo Siegle,  
Claudia Pöhlmann (Hrsg.)

## **IQB-Ländervergleich 2012**

**Mathematische und naturwissenschaftliche  
Kompetenzen am Ende der Sekundarstufe I**

2013, 420 Seiten, br., 39,90 €  
ISBN 978-3-8309-2990-1

In diesem Band wird über den dritten Ländervergleich des Instituts zur Qualitätsentwicklung im Bildungswesen (IQB) berichtet. Im Fach Mathematik und in den naturwissenschaftlichen Fächern Biologie, Chemie und Physik werden die Kompetenzstände verglichen, die in den Ländern der Bundesrepublik Deutschland von Schülerinnen und Schülern der 9. Jahrgangsstufe im Jahr 2012 erreicht wurden. Die IQB-Ländervergleichsstudien überprüfen, inwieweit in den Ländern die mit den länderübergreifenden Bildungsstandards der Kultusministerkonferenz gesetzten Ziele erreicht werden. Sie ersetzen im Bereich der Sekundarstufe I die nationalen PISA-Ergänzungsstudien, in denen zuletzt 2006 mathematische und naturwissenschaftliche Kompetenzen der Schülerinnen und Schüler in den Ländern Deutschlands verglichen wurden.



**WAXMANN**  
Münster · New York · München · Berlin

## 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 Univision 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.