

Rezension zu:

Birnkraut, Gesa: *Evaluation im Kulturbetrieb*. Kunst- und Kulturmanagement. Wiesbaden: VS Verlag 2011. 128 Seiten, 19,95 EUR, ISBN 978-3-531-16731-2

Ute Marie Metje,¹ *Vera Hennefeld*²

Die Autorin des 2011 erschienenen Buches „Evaluation im Kulturbetrieb“ richtet sich mit ihrer Publikation insbesondere an Kulturinstitutionen und Kulturschaffende mit dem Ziel, die Vorbehalte gegenüber Evaluationen zu entkräften und die Potenziale dieses Instruments zur Initiierung langfristiger Lernprozesse herauszustellen. Im deutschsprachigen Raum handelt es sich erst um die zweite Buchveröffentlichung, die sich grundlegend mit *Evaluationen in Kulturbetrieben* befasst.³ Einleitend verweist die Autorin auf die zunehmende Bedeutung von Evaluation im Kultursektor und macht damit gleichzeitig die Absicht ihrer Ausführungen deutlich: Die Kulturförderung in Deutschland folgt der klassischen föderalen Struktur und der Kultur kommt in den Bundesländern und Stadtstaaten eine unterschiedliche Bedeutung zu. Gleichwohl wird Kultur zu einem erheblichen Anteil mit öffentlichen Geldern finanziert und Rechenschaftslegung in der Kultur immer wichtiger. Für Kulturschaffende und Kulturbetriebe stellt Evaluation jedoch immer noch oftmals ein Schreckgespenst dar, mittels dessen versucht wird, zu kontrollieren, was nicht kontrolliert werden kann. Ziel der Publikation ist es daher, die Potenziale von Evaluationen für Kulturbetriebe aufzuzeigen und zu vermitteln. Die Autorin ist geschäftsführende Gesellschafterin von BIRNKRAUT/PARTNER, einer Kulturmanagementberatung, und Vorsitzende des Instituts für Kulturkonzepte Hamburg e.V. Seit 2011 hat sie eine Professur für strategisches Management in Non-Profit-Organisationen an der Hochschule Osnabrück in der Fakultät für Wirtschafts- und Sozialwissenschaften inne.

Bei dem vorliegenden Buch handelt es sich um ein 128 Seiten umfassendes Werk, das in vier kurzen Abschnitten an die aktuelle Diskussion und Überlegungen der Evaluationscommunity anknüpft und sieben praxisorientierte Kapitel bietet. Die Zielgruppe sind Kulturschaffende und Beschäftigte in Kulturinstitutionen, die bisher wenig Erfahrung mit Evaluation haben oder dieser kritisch gegenüberstehen. Birnkraut begründet zunächst die Notwendigkeit einer solchen Publikation und formuliert ihre Überzeugung, dass Evaluationen im Kultursektor einen entscheidenden und erheblichen Mehrwert für Einrichtungen und Projekte erbringen können. Ergänzend zur anfänglichen Einführung in Begrifflichkeiten liefert die Autorin am Beispiel dreier europäischer Länder (Großbritannien, Niederlande, Schweiz) einen Überblick über deren jeweilige Kulturförderung und Kulturpolitik sowie die daraus resultierende Handhabung von Evaluation. Im Anschluss daran wird der Zusammenhang von Kulturpolitik und Kulturförderung und Evaluation in Deutschland skizziert. Danach folgt der praxisorientierte Teil mit jeweils kurzen Abschnitten zu Formen und Ablauf von Evaluation, Kennzahlen und Indikatoren sowie Instrumenten. Ergänzt wird dies durch ein Kapitel, in dem *Kleine Helfer* für die Durchführung von Evaluationen vorgestellt werden. Abschließend widmet sich die Verfasserin den Wirkungen und Veränderungen durch Evaluation und spricht Empfehlungen an Kulturinstitutionen und die Kulturpolitik aus.

Dem Verständnis der Verfasserin folgend trägt *Evaluation* dazu bei, aufzuklären, „ob eine Ein-

1 Evaluation & wissenschaftliche Beratung, Hamburg

2 Centrum für Evaluation (CEval), Saarbrücken

richtung systematisch und kontinuierlich an der Sicherstellung der eigenen Wirksamkeit arbeitet. Dabei geht es um die Qualität und Nachhaltigkeit dieser Sicherstellung. (...) Die Ergebnisse einer Evaluation sind Diskussionsgrundlage und führen zu gemeinsam getragenen Veränderungsprozessen, die kontinuierlich durch Evaluation begleitet werden“ (S. 18). Die Autorin konzentriert sich in ihrer Arbeit auf die Erkenntnisfunktion von Evaluation, die Entwicklungs- und Lernprozesse auslösen (kann) und plädiert an die Kulturpolitik, Evaluation nicht nur als Kontrollinstrument einzusetzen. Diesen Teil abschließend setzt Birnkraut Evaluation in den Kontext zu Instrumenten wie Controlling, Monitoring und Qualitätsmanagement.

In den zwei folgenden Kapiteln erhält der Leser einen Überblick darüber, wie *Evaluationen* in *Großbritannien*, den *Niederlanden*, der *Schweiz* und *Deutschland* gehandhabt werden, und welche Bedeutung sie im Rahmen der jeweiligen Kulturpolitik und Kulturförderung einnehmen. Hierzu gibt Birnkraut einen kurzen Einblick in die unterschiedliche Tradition und Praxis von Evaluation in diesen Ländern und liefert Hinweise auf erfolgreich eingesetzte Evaluationen in Kulturinstitutionen. So verfügt beispielsweise Großbritannien über eine 50 Jahre währende Tradition und Erfahrung mit Evaluation im kulturellen Sektor. Unterschiede im Umgang mit Evaluation in den drei Ländern werden insbesondere im Vergleich zu Deutschland deutlich. Die Besonderheit des deutschen Kultursektors liegt in dessen föderaler Strukturierung, da dieser Bereich in jedem Bundesland von anderen Ressorts verwaltet und finanziert wird. Daraus resultieren ein unterschiedlicher Stellenwert und eine andere Gewichtung von Kultur in den einzelnen Bundesländern; bundeseinheitliche Regelungen existieren nicht. Aufgrund dieser großen Unterschiede plädiert die Autorin dafür, Evaluation in der Kultur zu differenzieren und auf nationaler und kommunaler bzw. auf Ebene der Städte zu verorten. Ihrer Erfahrung nach sind die Aufgaben von Evaluation auf diesen Ebenen sehr verschieden. Während es auf nationaler Ebene eher um die Überprüfung von ganzen Subventionssystemen und Handlungsempfehlungen zu Strukturveränderungen an die Politik geht, werden Evaluationen auf kommunaler Ebene häufiger

Kontroll- und Legitimationsfunktionen zugeschrieben und die Ergebnisse betreffen direkt Kulturinstitutionen bzw. Kulturbetriebe. Inzwischen wird in Deutschland Kultur aber auch als Wirtschafts- und Imagefaktor erkannt und Evaluationen machen einen „return on cultural investment“ (S. 35) sichtbar. Dieser Aspekt ist, so Birnkraut, in den angloamerikanischen Ländern schon lange Teil der kulturpolitischen Diskussion.

Anschließend an diese theoretischen Ausführungen wendet sich die Autorin in den folgenden Abschnitten ausführlich und sehr praxisorientiert Evaluationen zu, beginnend mit zwei Abschnitten über *Formen* und *Ablauf von Evaluation*. Hier fügt Birnkraut zur Verdeutlichung viele praktische Beispiele an, die ihre Ausführungen und Überlegungen sehr gut veranschaulichen und auch Laien ein klares Bild von den Anforderungen an Evaluation bieten. Für erfahrene Experten von Evaluation bietet die Autorin jedoch nichts Neues, vielmehr orientiert sie sich im Aufbau dieser Abschnitte am zeitlichen Ablauf von Evaluationen und stellt die bekannten aber wichtigen Fragen: Warum werden Evaluationen in Auftrag gegeben? Wer soll sie durchführen, also intern oder extern und zu welchem Zeitpunkt? Birnkraut definiert überdies die Begriffe *Input*, *Output*, *Outcome* und *Impact*, erläutert den Nutzen von Evaluationen von Projekten oder in Institutionen und benennt die Voraussetzungen für eine funktionierende Evaluation. Abschließend werden der Evaluationszyklus dargestellt und Formen der Berichterstattung aufgeführt.

Im nächsten Kapitel widmet sich die Verfasserin der Entwicklung von *Kennzahlen* und *Indikatoren* zur Analyse und Bewertung von Zielen und Maßnahmen.⁴ An dieser Stelle erfolgt der Hinweis auf die Diskussion, dass qualitative Indikatoren im Kulturbereich oft schwer zu bedienen sind. Eine Lösung sieht Birnkraut darin, sich vom künstlerischen Kernprozess zu entfernen und mehr auf die Managementprozesse einzugehen. Die Autorin gibt Beispiele für quantitative und qualitative Kennzahlen und Indikatoren und liefert Hinweise auf Interpretationsmöglichkeiten bzw. deren Aussagekraft. Dieser Abschnitt endet mit der – aufgrund ihrer starken inhaltlichen Fokussierung auf nur einen Aspekt diskussionswürdigen – Feststellung, dass es „al-

3 Die zahlreichen Veröffentlichungen zum Thema Besucherforschung und Evaluation in Museen bleiben hier ausgespart.

4 Birnkraut verwendet die Begriffe *Kennzahlen* und *Indikatoren* synonym.

les in allem bei den Kennzahlen und Indikatoren darum geht, den Return on Investment oder auch den Return on Cultural Investment zu definieren“ (S. 71) und zu beziffern.

Im Anschluss daran stellt Birnkraut zahlreiche *Erhebungsinstrumente* vor, die sich ihrer Erfahrung nach bei Evaluationen in Kulturbetrieben bewährt haben. Dazu gehören klassische Instrumente wie Fragebögen, Interviews und Beobachtungen sowie auch anspruchsvolle Instrumente wie beispielsweise der Einsatz des Balanced Score Card Systems. Die Balanced Score Card dient im Wesentlichen dazu, die Strategie einer Institution in der Steuerung umzusetzen und betrachtet üblicherweise die Perspektive der Kunden, der Finanzen sowie interne Arbeitsprozesse und den Bereich Lernen und Wachsen, wozu u.a. auch die Personalentwicklung gehört (vgl. S. 86). Um den Einsatz dieses Instruments in der Praxis zu verdeutlichen, beschreibt die Verfasserin eine selbst durchgeführte Evaluation und diskutiert Chancen und Grenzen dieser Methode.

In den letzten drei kurzen Abschnitten werden zunächst *Kleine Helfer* vorgestellt: Eine Auflistung wenig aufwendiger aber wirkungsvoller Methoden, die zu Evaluationsaktivitäten ermuntern und zugleich aufzeigen sollen, dass hilfreiche Befunde zu spezifischen Aspekten auch außerhalb von groß angelegten Evaluationen im kleinen Rahmen gewonnen werden können. In den folgenden Überlegungen unter dem Titel *Wirkungen und Veränderungen durch Evaluation* geht die Verfasserin noch einmal auf die klassischen Bedenken gegenüber Evaluation in Kulturinstitutionen ein und gibt positive Beispiele für die Wirksamkeit des Instruments, wenn es denn sinnvoll eingesetzt wird und Veränderungen herbeigeführt werden sollen. Abschließend plädiert Birnkraut in dem Kapitel *Empfehlungen an Kulturinstitutionen und Kulturpolitik* für die Nutzung von Evaluation und verweist auf die Chancen, die sich durch den Einsatz dieses Instruments in der Kultur eröffnen können. Die Autorin versteht Evaluation im kulturellen Sektor immer auch als politischen Prozess und zugleich als Möglichkeit zur Verbesserung und Professionalisierung der Kulturarbeit.

Fazit: Mit dem vorgestellten Buch ist eine Publikation gelungen, die insbesondere für Praktiker(innen) aus Kulturbetrieben und -institutionen einen leicht verständlichen Einstieg und guten Überblick über Evaluation und ihre Möglichkeiten und Grenzen im kulturellen Sektor bietet. Hier müssen zum einen vor allem die vie-

len Beispiele hervorgehoben werden, an denen sich die Leser(innen) orientieren können und die die inhaltlichen und theoretischen Überlegungen gut verständlich machen; zum anderen ist der Überblick über Handbücher zum Themenfeld zu betonen, die seitens der Autorin kommentiert werden. Dankenswert und wichtig ist auch der Ansatz, der immer noch starken Skepsis gegenüber Evaluationen in diesem Bereich entgegenzuwirken, indem viele Beispiele für eine fruchtbare und nachhaltige Nutzung von Evaluation und damit für deren Wirksamkeit und Veränderungskraft aufgezeigt werden. Der Text ist sprachlich klar formuliert und Fachbegriffe werden mit Verweis auf die Quellen definiert und erläutert. Den Leser(inne)n erschließt sich durch die Lektüre die Perspektive der Autorin auf Evaluationen und deren Zweck und Nutzen im kulturellen Sektor. Dies ist Stärke und Schwäche zugleich: Das Buch ist als Einstiegslektüre für diejenigen nützlich, die sich zum ersten Mal mit Evaluation im Kulturbereich auseinandersetzen (müssen), weil es einen allgemeinen Überblick bietet und – dies ist ein besonderes Verdienst der Publikation – weil auch die Kulturpolitik und Kulturförderung Deutschlands zu drei anderen europäischen Ländern in Bezug gesetzt wird. Gleichzeitig ist dies ein hoher Anspruch, der unserer Ansicht nach zu kurz geraten ist und deshalb fragmentarisch wirkt und unvollständig bleiben muss. Gerade dieser Abschnitt hätte mit Blick auf die Professionalisierungsbemühungen vieler Mitglieder der Evaluationscommunity noch stärker an die aktuelle Diskussion von Evaluation in der Kultur anknüpfen und sich intensiver auf Best-Practice-Beispiele oder gute Erfahrungen in anderen europäischen Ländern beziehen sollen. Dies wäre insbesondere auch hinsichtlich des seit 2005 existierenden UNESCO-Übereinkommens zum Schutz und zur Förderung der Vielfalt kultureller Ausdrucksformen wünschenswert gewesen, da auch Deutschland Vertragspartei ist und durch dieses Übereinkommen Veränderungen in der Kulturpolitik und Kulturförderung Deutschlands notwendig werden. Trotz dieser Kritik kommt der Publikation eine wichtige Funktion zu, da sie die Diskussion über Evaluation in Kulturinstitutionen (hoffentlich) weiter vorantreibt und einen Einblick in gelingende aber auch herausfordernde Aspekte liefert.

Rezension zu:

Sandermann, Mary: Die Bedeutung von Soft Skills für Evaluationsnutzungen. Eine komparative Analyse. Wiesbaden: VS Verlag für Sozialwissenschaften 2011. 250 Seiten, 39,95 EUR, ISBN 978-3-531-17887-5

*Andreas Balthasar*¹

Die Nutzungsorientierung gehört zu den Wesensmerkmalen von Evaluationen. Ob und unter welchen Bedingungen Evaluationen Nutzen stiften, ist Gegenstand der evaluationswissenschaftlichen Verwendungsforschung, welche in den 1960er Jahren mit den Arbeiten von Carol Weiss ihren Anfang nahm. Die Dissertation von Mary Sandermann knüpft an diese Forschungstradition an. Bisherige Untersuchungen haben insbesondere den Einfluss von Prozessmerkmalen einer Evaluation, von Umfeldfaktoren oder von unterschiedlichen institutionellen Arrangements der Durchführung einer Evaluation auf die Nutzung thematisiert. Die vorliegende Dissertation stellt demgegenüber den Einfluss sozialer Kompetenzen auf die Nutzung von Evaluationsergebnissen ins Zentrum. Zu den sozialen Kompetenzen zählt die Autorin die Gestaltung des Evaluationsprozesses, die Art und Weise der Interaktion zwischen den Evaluator(inn)en und der Gruppe der Entscheidungsträger(innen) sowie die Persönlichkeit der Evaluator(inn)en. Die Hauptfragestellung lautet: Welche Bedeutung haben die Interaktionsgestaltung während einer Evaluation und die sozialen Kompetenzen der Hauptinteraktionspartner(innen) für die Evaluationsnutzungen?

Die Dissertation umfasst einen ausführlichen theoretischen und einen empirischen Teil. Im theoretischen Teil werden die Entwicklungsphasen der Interaktionsperspektive in der evaluationsbezogenen Nutzungsforschung zwischen 1971 und 2009 beleuchtet. Im historischen Abriss zeigt sich, dass die Wichtigkeit von interper-

sonellen und sozialen Kompetenzen stets zugezogen hat. Allerdings beruht die Bewertung dieser Entwicklung nicht auf einem stabilen theoretischen Fundament. Die Autorin geht diese Lücke dadurch an, dass die Konstrukte „soziale Kompetenzen“, „Interaktionsgestaltung“ und „Umfeld- und Prozessmerkmale“ für die Untersuchung empirisch fassbar gemacht werden. Dazu werden verschiedene Modelle aus unterschiedlichen Forschungsdisziplinen analysiert und geprüft, ob sie für die drei genannten Konstrukte einen Beitrag zur Untersuchung von weichen Faktoren bei der Evaluationsnutzung leisten können.

Die sozialen Kompetenzen werden aus psychologischer Sicht, aus der Sicht der Evaluationsnutzungsforschung und aus den von der DeGEval herausgegebenen Empfehlungen der Kompetenzfelder zur Aus- und Weiterbildung von Evaluator(inn)en beleuchtet. Als Schnittmenge der drei Zugänge werden für die Nutzung wichtige Eigenschaften von Evaluatorinnen und Evaluatoren identifiziert. Zu diesen Eigenschaften werden Durchsetzungsfähigkeit, Flexibilität, Führungsverhalten, Handlungsorientierung, Kontaktfähigkeit, Empathie, Reflexion, Soziabilität und Teamkompetenz gezählt.

Die soziale Interaktionsgestaltung wurde für die Untersuchung durch die aus der Interaktionstheorie bekannten Aspekte „Gestaltung der Kommunikation“, „Verfügbarkeit bei Bedenken von Teilnehmenden“ und „wertschätzende Grundhaltung“ aufgebaut.

1 Institut für Politikwissenschaft, Universität Luzern

Umfeld- und Prozessmerkmale der Evaluation flossen als kontrollierende Größen in die Untersuchung ein. Einbezogen wurden die Integration von Nutzungsprozessen in den Ablauf einer Evaluation, die Orientierung der Evaluation an den Informationsbedürfnissen der potenziell Nutzenden, die Einhaltung formaler Qualitätskriterien und das politische Klima, in welchem die Evaluation durchgeführt wurde.

Das auf diesen drei Faktorbündeln aufgebaute Untersuchungsmodell von Sandermann geht davon aus, dass die Interaktionsgestaltung und die sozialen Kompetenzen der Hauptinteraktionspartner(innen) ausschlaggebend sind für die Evaluationsnutzung.

Im empirischen Teil wird zuerst eine Einschränkung des Untersuchungsgegenstands vorgenommen. Die Dissertation beschränkt sich auf die Analyse der Nutzung von Evaluationen, welche universitäre Mentoringprogramme betreffen. Solche Programme zielen darauf ab, die ungleiche Geschlechterverteilung an Universitäten durch den Austausch zwischen einem erfahrenen Mentor/einer erfahrenen Mentorin und einer Mentee zu reduzieren. Die Autorin hat 59 Mentoringprogramme an deutschen Universitäten identifiziert und sieben für die vertiefende Untersuchung ausgewählt. Ein wichtiges Auswahlkriterium war, dass ein Programm extern evaluiert worden war. Mit je zwei Beteiligten der ausgewählten Mentoringprogramme wurden halbstandardisierte Interviews geführt. Dabei wurde unter anderem auf das *Bochumer Inventar zur berufsbezogenen Persönlichkeitsbeschreibung* zurückgegriffen, um die sozialen Kompetenzen der Beteiligten zu erheben. Befragt wurden sowohl die Evaluatorinnen und die Evaluatoren der Programme als auch die Hauptinteraktionspartner auf Programmseite.

Die Auswertung der Daten erfolgte unter Heranziehung der *Fuzzy Set Qualitative Comparative Analysis* (fsQCA). Es handelt sich dabei um eine wissenschaftliche Methodologie für den systematischen Fallvergleich, welche als Mittelweg zwischen statistischen Standardverfahren und qualitativen Fallstudien konzipiert wurde. Jeder Fall wird als Konfiguration von Bedingungen verstanden, die zu einem bestimmten Outcome führen, wobei notwendige von hinreichenden Bedingungen unterschieden werden. Um eine fsQCA durchführen zu können, mussten die Informationen aus den Interviews und Fragebogen kalibriert werden.

Die deskriptive Auswertung zeigte, dass in sechs der sieben untersuchten Mentoringpro-

grammen Evaluationsnutzung stattgefunden hat. In all diesen Fällen wurde die Interaktionsgestaltung positiv bewertet. Neun der 14 befragten Evaluationsakteur(inn)e(n) wurden durchschnittliche oder gar überdurchschnittliche soziale Kompetenzen zugesprochen. Die Autorin differenzierte in der Auswertung zwischen direkter und indirekter Evaluationsnutzung. Unter indirekter Evaluationsnutzung subsumierte sie die Verwendung von Evaluationsergebnissen, die nicht direkt im Zusammenhang mit den Fragestellungen der Evaluation standen. Diesbezüglich zeigt sich, dass weder die Interaktionsgestaltung noch die Umfeld- und Prozessmerkmale oder die beruflichen sozialen Kompetenzen der Beteiligten die indirekte Evaluationsnutzung beeinflussen. Relevant scheinen einzig die Interessen der beteiligten Personen zu sein.

Die Auswertung der Daten, die direkte Evaluationsnutzung mit der fsQCA betreffend, identifizierte günstige Umfeld- und Prozessmerkmale als eine notwendige Bedingung für direkte Nutzung. Auch die positive Interaktionsgestaltung zwischen einem Evaluator/einer Evaluatorin und dem Hauptinteraktionspartner/der Hauptinteraktionspartnerin wirkt sich positiv auf die direkte Evaluationsnutzung aus. Die Auswertung zeigte sogar, dass eine positive Interaktionsgestaltung Defizite bei der Sozialkompetenz aufzuheben vermag. Weiter ließ sich erkennen, dass eine teamorientierte Ansprechperson aufseiten des Evaluationsteams eine positive Auswirkung auf die spätere Evaluationsnutzung hat. Insgesamt weist die Dissertation darauf hin, dass günstige Umfeld- und Prozessmerkmale eine notwendige Bedingung für die direkte Nutzung von Evaluationen darstellen. Zudem lässt sich erkennen, dass die bislang wenig berücksichtigte Interaktionsgestaltung in den untersuchten Fällen eine hinreichende Bedingung für die direkte Nutzung von Evaluationen darstellt.

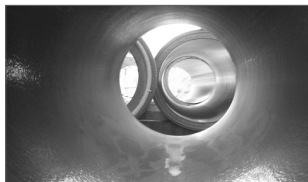
Die kurze Zusammenfassung zeigt, dass die Dissertation von Mary Sandermann einen wertvollen Beitrag im Hinblick auf eine differenzierte Theorie der Verwendung von Evaluationen leistet. Allerdings beruhen die Ergebnisse auf einer sehr schmalen Datenbasis. Indem nur sieben Evaluationsprogramme für die Untersuchung berücksichtigt wurden, wird auch die minimale Anwendungsbedingung von fsQCA nicht erreicht. Gemeinhin wird nämlich davon ausgegangen, dass in fsQCA zwischen zehn und 50 Fälle einbezogen werden sollten. Ein weiteres Problem besteht darin, dass nicht wirklich klar wird, wen die Autorin zu den Hauptinteraktions-

partner(inne)n einer Evaluation zählt. Einmal sind die Evaluatorinnen und Evaluatoren sowie die Evaluierten bzw. Interviewpartner(innen) gemeint, ein anderes Mal sind es die Evaluierten oder Adressat(inn)en der Evaluation. Hier wäre eine konsistente und eindeutige Begriffsverwendung wünschenswert gewesen. Schließlich erschwert die Komplexität des Untersuchungsmodells die Nachvollziehbarkeit der Ergebnisse.

Man würde sich ein einfacheres und eleganteres Modell wünschen, um der zentralen Frage der Arbeit nachzugehen.

Dennoch verdient die Dissertation von Mary Sandermann Beachtung, da sie die Dimension der sozialen Kompetenzen und der Interaktionsgestaltung theoretisch fundiert und empirisch gestützt in die evaluationswissenschaftliche Verwendungsforschung einbringt.

Klassische Konzepte zur Durchführung von Programmevaluationen stoßen regelmäßig an Grenzen der Praxis oder können wissenschaftlichen Ansprüchen nicht genügen. Als einen alternativen Zugang präsentiert das Buch theoriebasierte Evaluationskonzepte. Diese ermöglichen Evaluationen, die für Beteiligte und Betroffene einen Nutzen erzeugen, realistisch umsetzbar sind, zu genauen Ergebnissen führen und den Beteiligten gegenüber fair sind. Die Publikation verfolgt ausdrücklich eine anwendungsorientierte Absicht und richtet sich an die Praxis sowie die Aus- und Fortbildung von Evaluatorinnen und Evaluatoren. Abgeleitet aus den jeweiligen Nutzungspotentialen grundlegender methodischer Zugänge entwickelt die Autorin ein Konzept integrativer Methodennutzung zur Realisierung theoriebasierter Evaluationen. Die einzelnen Analyseschritte werden anhand der Beispielevaluation einer internetbasierten Lernumgebung illustriert.



Susanne Giel

Theoriebasierte Evaluation

Konzepte und methodische Umsetzungen

2013, 306 Seiten, br., 34,90 €
ISBN 978-3-8309-2855-3



WAXMANN
Münster · New York · München · Berlin

Rezension zu:

Patton, Michael Q.: *Essentials of Utilization-focused Evaluation*. Los Angeles: Sage 2012. 488 Seiten, 58,00 US \$, ISBN 978-1-4129-7741-8

*Frans L. Leeuw*¹

In the preface to this book, Patton informs the reader that one of the reasons to write *Essentials of Utilization-focused Evaluation* (UF-E) was that the original book on this topic doubled its number of pages from 303 in 1978 to 667 in 2008, which may have caused unintended side effects (reducing the use and impact of it). One was that the 667 pages of the most recent edition were used as a ‘substitute pillow’ during a trip in an overnight train in India rather than as a study book, while unexpected costs to have the book ‘lifted’ in and out of a house of a reader was (jokingly) another unintended and negative side effect of the almost 700 pages thick volume. Needless to say that these experiences for Patton, a man who is – and rightly so – keen on making a difference, i.e. realizing an impact with his scholarly work, triggered (single loop) learning. The result of this learning has been this new book *Essentials of UF-E* (2012), published by Sage and 488 pages thin (or long).

The structure of the *Essentials*-book is rather straightforward. Apart from sections covering the introduction, conclusions and summary, the book consists of seventeen steps described in the same number of chapters. The landscape of the book is articulated nicely and is designed to function as the central theme; it can be found in many places throughout the book. Patton calls this ‘landscape’ “the complex dynamic and adaptive systems graphic”, picturing interactions among all 17 steps, but he also uses a simple ‘stairway metaphor’, which gets higher and higher as the book progresses covering more steps. I assume that the underlying assumption of

this design is to make sure that no reader is left behind, gets lost or stops reading.

The 17 steps start with the inevitable assessment of the program and organizational readiness for UF-E and the assessment of the evaluator’s readiness and competence to do the job. The book then moves on to the identification and engagement of primary users of evaluations, and to identifying primary intended uses. A little higher up in the stairway the question is addressed if fundamental areas for evaluations are being adequately addressed (like implementation, outcome and attribution questions). Step 9 moves to determining what the underlying intervention model or theory is that is evaluated, while subsequent steps discuss the process of negotiating which methods to generate findings are appropriate. From step 12 onwards the focus is fully dedicated towards questions like how to stimulate the use of evaluations, how to gather data with an ongoing attention to use, how to present data and findings to relevant audiences and how not to forget the “follow up with primary intended users to facilitate and enhance use”. The final step is the meta-evaluation of the use: “be accountable, learn and improve” (p. 388 ff.).

It is a rich book, describing in detail how to handle an evaluation trying to stimulate utilization, but that also invites the reader to join short digressions. The first one is a section on ‘evaluation anxiety’, aka as the “pre-evaluation stress syndrome” (p. 23), which is immediately linked to a ‘what to do-activity’, i.e. to create a positive atmosphere and vision of the evaluation, together with the expected future user(s) of the evalu-

1 University of Maastricht and National Justice Research Center WODC, Den Haag, The Netherlands

ation. Another digression (p. 83) has to do with the several temptations evaluators can be lured into. Patton lists 10 of them. One of them is that “evaluators make themselves the primary decision-makers and therefore the primary users” of ‘their’ evaluation. Another one is assuming that the funder of the evaluation is automatically the primary stakeholder. An interesting but disputable one is temptation no. 7, “taking a stance of staying above the fray of people and politics”. Patton suggests that trying to do this leads to “irrelevance ... evaluation is action-oriented and action is in the fray”. This statement sheds light on the epistemological frame that characterizes this book (and that I’ll come to a bit later). The interesting section on six alternative evaluation goals (p. 115) is also a must-read, although I almost totally disagree with what he is writing about ‘developmental evaluations’. Sometimes, Patton is a true believer. No references can be found to unintended side effects of the M(onitoring) & E(valuation) industry (à la lowering policy ambitions for example) or to the many other side effects, when auditing marches in (teaching to the test, the performance paradox, protocollization etc.), although a little later Patton indeed shortly discusses the performance measurement mantra (p. 152).

A very informative chapter is the one of formulating evaluation questions. A typology of questions is produced of which the IT question (what is IT?, does IT work? etc.) I like best (p. 179). IT has nothing to do with ICT, but IT is ‘The Intervention’ or ‘The Program’. Here, Patton holds a plea to open up the black box of what in many evaluations often is kept silent, as IT usually is only the label of the intervention, program, law or other type of policy measure that has to be evaluated. Unfortunately, Patton is not addressing methodological literature on how to distinguish types of (evaluation) problems, and the methodological quality of the way in which problems are formulated. What makes research problems ill or even erroneously formulated? These are problems that are formulated against a background that consists of at least one incorrect statement. The background “is constituted by the antecedent knowledge and, in particular, by the presuppositions of the problem. The presuppositions of the problem are the statements that are somehow involved but not questioned in the statement of the problem and in the inquiry prompted by it” (Bunge 1997: 194).

The chapter on the intervention theory/logic is, as are many other parts of the book, basic and again loaded with charts and graphs. Unfortunately, not much is said about the methodology how to find (‘detect’, ‘reconstruct’) underlying assumptions, i.e. the intervention or program theory. Over the last decade a number of methodological suggestions have been formulated, including making use of argument mapping software, that can increase the level of professionalism when doing this kind of work. Now the reader may either wonder how to ‘find’ this intervention theory or, which is more dangerous, just starts to ‘write the theory/assumptions up’. Tilley (1992), more than a decade ago, warned already for the danger of sloppy reconstructions of theories.

The most well-known chapters in the book are the ones on how to ‘sell’ evaluation findings to your audiences. They focus on how to write the evaluation report in a way that it is almost unputdownable, how to get stakeholders engaged in dialogues focusing on learning from the evaluations, to study the way in which ‘information is beautiful’-websites help you to transfer the sometimes voluminous and heavy sets of data to the audiences involved and even to watch Hans Rosling (Swedish statistician) movies/you tubes, who animates statistical findings and trends like they are the European Soccer Finale.

Numerous exhibits, menus and boxes are part of the book. A shortcut to characterize the book would therefore be to label it the ABC3EFM-toolbox of UF-E. ABC3EFM stands for *Arrows, Boxes, Checklists, Charts, Cartoons, Exhibits, Figures & Menus*. Many examples from a wide number of fields and studies are also part of the book showing what in the respective practice has been going on and what can be learned from them.

Although exhibits, boxes and checklists certainly give guidance to readers and in particular to newcomers in the field, they also leave an impression of UF-E as something mechanical, and even ‘techno-ish’: evaluators should possess x, y and z, should practice Q, have participants share or do this and that, engender commitment to ... etc. The actions to be taken during the 17 steps include the descriptions of situations where and when to carry them out, often with an impression of what the consequences will be. This, together with the large number of (normative) suggestions and guiding principles, makes me to bring up this

critique. Moreover, and in line with a recent presentation by Gary Henry,² the question should be asked how much systematic evidence there is for all the guidelines and normative statements. Of course Patton's personal decades of experiences doing utilization-focused evaluations is an important source as is the literature list, but I doubt that systematic and robust evidence on what works best, when and how, and for whom when transferring and using results from evaluations by practitioners, policy makers, politicians, managers, teachers, journalists and the public is available to underpin all of Patton's claims, suggestions and recommendations. Recently, I searched for this type of evidence in the field of knowledge transfer, translation, utilization in repositories of social, behavioral and technological studies, but found amazingly little systematic evidence.³ Case studies, narrative material and surveys seem to dominate (Brandon/Singh 2009).

Moreover, there are two other serious and epistemological points of critique on the book. The first concerns its clearly articulated one-sidedness. One of the central arguments in the book is that almost everything that is in the hands of the evaluation commissioner, the evaluators and the users of the evaluation has to be done to make sure that evaluations are designed and carried out in such a way that they indeed are used. Non-utilization looks like a total waste of resources. How perfectly clear this statement is, it misses some points. The first is that there are good and bad evaluations. Bad evaluations should not be carried out, but for sure not be utilized. They wrongly inform society, policy makers and others; the error costs can be high (Leeuw 2011). The second point is that as evaluation has turned into a booming business with its own evaluation machines (Dahler-Larsen 2000), this increases the likelihood that other incentives than 'speaking truth to power' and 'learning' are on the agenda when doing and using evaluation results (like shopping for future clients and research money, bureau politics, teaching to the test). Dahler-Larsen's (2000: 70) finding that almost never evaluations are used to terminate policies of Danish municipalities is an exciting though somewhat worrying one. "Out of 600

evaluations, less than 1% led to termination of the evaluated activities, whereas 78% of the evaluations led to adjustments of program activities". There is no guarantee that these 'adjustments' de facto and letztlich will lead to more effective policies; the opposite phenomenon that 'interventions breed interventions' is also possible, but often neglected.

The second problematic aspect of Patton's book is the attention paid to collaboration between evaluators and primary users. His definition of UF-E is that it is "a process for making decisions about (these) issues in collaboration with an identified group of primary users focusing on their intended use of evaluations" (p. 6). This process is "highly personal and situational" (p. 4).

Although Patton refers to standards for evaluation like accuracy and accountability, in my opinion he does not pay enough attention to the cognitive-behavioral consequences of the way in which 'his' UF-E should be carried out. Take this example: "if evaluations are ignored or misused, we have to look at where our own practices and processes may have been inadequate" (p. 5). The possibility that the very act of collaboration between evaluators and evaluands may have contributed or even have caused evaluation findings not to be used, because officials are no longer certain about the independence and credibility of evaluators as well as the methodological quality of their designs and data, is not part of Patton's cognitive frame of reference. In my own world, which crosses over several decades of evaluations in Europe and in fields ranging from crime and justice, education and development aid to environment and energy and public administration, I have experienced many times that it is exactly the 'beloved' collaboration between evaluators and evaluands, including stakeholders, funders and others, which causes non-utilization or misuse.

Finally, a few other points:

- I miss the role internet and, more broadly, 'living in the web' plays in and for UF-evaluations. A couple of times Patton refers to it, but a systematic analysis of the strengths and challenges that developments like 'Big Data' (logs, mobile trans-

2 European Evaluation Society Congress, Helsinki, Session: Theories in Evaluation, October 2012.

3 The keynote was given at the EIPPEE Conference 2012: Advancing the use of research in education across Europe, Den Haag: "Transferring research (results) to policy makers and practitioners: Experiences from the fields of security and justice and development aid".

actions of many kinds, online user-generated content, such as blog posts and Tweets, online searches, satellite images, GPS-travel data etc.; see UN Global Pulse, www.globalpulse.org), digital policies and interventions and new ways of data collection (involving clients) by using apps, sensors or the ethnocoder, remain untouched. The same is true for the competition to come between the wisdom (sometimes: the stupidity) of the crowd and the wisdom of a few collaborating evaluators. As a profession, we seriously have to think about that.

- Comparing evaluations with DNA (see the double helix picture on the front page) and, further on in the book, linking evaluation to Gordon Moore's Law (computer capacity will double every two years) is a bit of an exaggeration. DNA and Moore's law are of a different nature and work on a different level. I always teach my evaluation students to be strong, will powered and tough on tenacity, but also ... showing a certain humbleness.

References

- Brandon, Paul R./Singh, J. Malkeet (2009): The Strength of the Methodological Warrants for the Findings of Research on Program Evaluation Use. In: *American Journal of Evaluation*, 30(2), pp. 123-157.
- Bunge, Mario (1997): *Philosophy of Science: From Problem to Theory*. New Brunswick: Transaction Publishers.
- Dahler-Larsen, Peter (2000): Surviving the Routinization of Evaluation: The Administrative Use of Evaluations in Danish Municipalities. In: *Administration and Society*, 32, pp. 70-92.
- Leeuw, Frans L. (2010): Benefits and Costs of Evaluation – An Essay. In: *Zeitschrift für Evaluation*, 9 (2), pp. 211-229.
- Tilley, Nick (1999): Evaluation and Evidence-(mis)led Policy. In: *Evaluation Journal of Australasia*, 11, pp. 48-63.