

LITERATURÜBERBLICK UND REZENSIONEN

Zeitschrift für Evaluation | 23. Jahrgang | 2024
Heft 1 | S. 126-139 | <https://doi.org/10.31244/zfe.2024.01.08> |
© 2024 Waxmann

Rezension zu:

Alkin, Marvin C./Christie, Christina A. (Hrsg.). (2023) *Evaluation roots. Theory influencing practice*. 3. Auflage, New York: Guilford Press, 318 Seiten, \$55.00, ISBN 978-1-462-55139-2

- Zugleich: Kurz-Einführung in Theorien der Evaluation -

*Wolfgang Beywl*¹

Entwickelt sich die Evaluationswissenschaft kumulativ weiter, steht sie vor einer neuen Stunde Null oder dämmert ihr das Jüngste Gericht? Zum einen bündelt dieses Buch über bald 100 Jahre publizierte Lehrwerke, Zeitschriftenaufsätze und methodologisch innovative Evaluationsberichte. Es ist damit Ausweis des durch tausende Mitwirkende erreichten hohen wissenschaftlichen Niveaus der Evaluation. Zum anderen markiert es einen Krisenpunkt der nordamerikanischen Evaluationscommunity. Auf deren Theorien beschränkt sich das Buch. Dies ist insofern angemessen, als die Theorien aus den USA und Kanada und die damit verbundenen Lehrwerke das Evaluationsverständnis auf dem gesamten Globus wesentlich geprägt haben. Dabei ist diese Evaluationsgemeinschaft tief verunsichert. Einige fragen sich gar, ob künftig die verbreitete Metapher vom „Baum der Evaluationstheorie“ weiter genutzt werden sollte, bzw. ob eine weitere Auflage des Buches überhaupt angemessen wäre. Vorbehalte betreffen einerseits die Systematik: Klassifizierung in die drei Äste. Andererseits stehen sie im Zusammenhang mit soziokulturellen und -politischen Spannungen. Die lange und breite öffentlich geführte Diskussion darüber belegt gleichzeitig die besondere Stellung dieses Buches für das Selbstverständnis der Evaluationscommunity.

Die ausführlichere Darstellung des Buchinhalts soll dazu beitragen, die besonders im deutschsprachigen Raum oft verengte Rezeption von Evaluationstheorien auf ein breiteres Fundament zu stellen. Zudem könnte ein Impuls sein, aus atlantischer Distanz die sich auch hier abzeichnenden Spannungen frühzeitig anzugehen.

1 Univation – Institut für Evaluation und Fachhochschule Nordwestschweiz, Windisch (Schweiz)

Neben der kurzen Einleitung besteht das Buch aus sechs Hauptteilen: I – Einführung; Teile II– IV mit insgesamt 18 Originalbeiträgen zu den drei „Ästen“, Teil V mit „anderen Theorien“ und Schlussteil VI. Hinzu kommen Personen- und Sachindex sowie Kurzporträts der 29 Autorinnen und Autoren.

Entstehungshintergrund und begleitende Kritik

Die dritte Auflage von „Evaluation Roots“ wurde mit mehrjährigem Verzug publiziert. Dabei waren viele Beiträge schon 2020 fertiggestellt. Es bestand (und besteht) ein tiefgehender Konflikt in Bezug auf die Legitimität der Buchkonzeption. An der Auseinandersetzung darüber haben sich mehrere hundert Personen beteiligt.

Den Beginn markiert der Blog-Beitrag von Vidhya Shanker (2020). Im Gefolge der sich über verschiedene Kanäle ausweitenden, teils hitzig geführten Debatten hat der Vorstand der American Evaluation Association (AEA) 2020 entschieden, die über Jahrzehnte weltweit führende offene elektronische Diskussionsliste *evaltalk* zu schließen und das Awards-Programm (bis 2023) zu sistieren. *Evaltalk* ging Ende 2021 exklusiv für AEA-Mitglieder moderiert wieder online. Der Traffic ist mehr als dezimiert. Es gibt – ähnlich wie im deutschsprachigen *forum-evaluation* – fast nur noch Ankündigungen, kaum Diskussionen.

Den von Vidhya Shanker mitverfassten *Call to action: unearthing evaluation roots*² unterzeichneten über 600 Evaluationsfachleute. Einige Kernsätze: „The field of evaluation’s whiteness is ubiquitous, and harmful. Members of our field have actively oppressed the voices of those whose otherization – through artificial categories of race, gender, sexuality, class, and ability status – it has exacerbated. [...] Now is the time to disrupt this unjust pattern. We have an opportunity to support our field’s ongoing healing and transformation by shaping a more accurate, just, and transformational narrative through the next edition of Evaluation Roots and beyond.“

Die Herausgebenden der dritten Auflage akzeptieren die Kritik (S. 4). Dadurch angeregt nehmen sie drei Beiträge aus der Black/BIPOC/Indigenous Community zusätzlich auf (siehe unten zu Kap. 11 und 13 sowie Fujita-Conrads et al. (2023)). Sie räumen ein, dass noch mehr getan werden müsste, um die Beiträge dieser Communities zur Evaluationstheorie zu würdigen (S. 7).

Dass eine weitere Auflage von Evaluation Roots wünschenswert ist, wird von prominenten Evaluationstheoretikern bezweifelt. Einer teilte mir per E-Mail mit: „It’s a nice collection but probably a good thing that this will be the last edition, at least that is my understanding“. In seinem YouTube-Video sieht Michael Patton (2023a, 18:20) die dritte Auflage bereits mit Erscheinen „out of date“. Er schlägt vor, an die Stelle der Metapher eines Baumes mit Wurzeln die des Waldes mit vielen verschiedenen Bäumen der Evaluationstheorien zu setzen. Montrosse-Moorhead et al. (2024)

2 [https://docs.google.com/forms/d/e/1FAIpQLSeNojhYI2RWIlqaKXfGGe97m5mcCwrywXuedcaO\]lw4_YNagg/viewform](https://docs.google.com/forms/d/e/1FAIpQLSeNojhYI2RWIlqaKXfGGe97m5mcCwrywXuedcaO]lw4_YNagg/viewform) [Abruf: 28.3.2024]

skizzieren mit dem „Garten der Evaluationsansätze“ einen ähnlich gerichteten Vorschlag, der ebenfalls beansprucht, die Engführung auf westliche bzw. nordamerikanische Evaluationstraditionen aufzuheben.

Einführungskapitel – Teil I

Alkin unterscheidet eingangs (S. 6 f.) zwei Theorie-Arten:

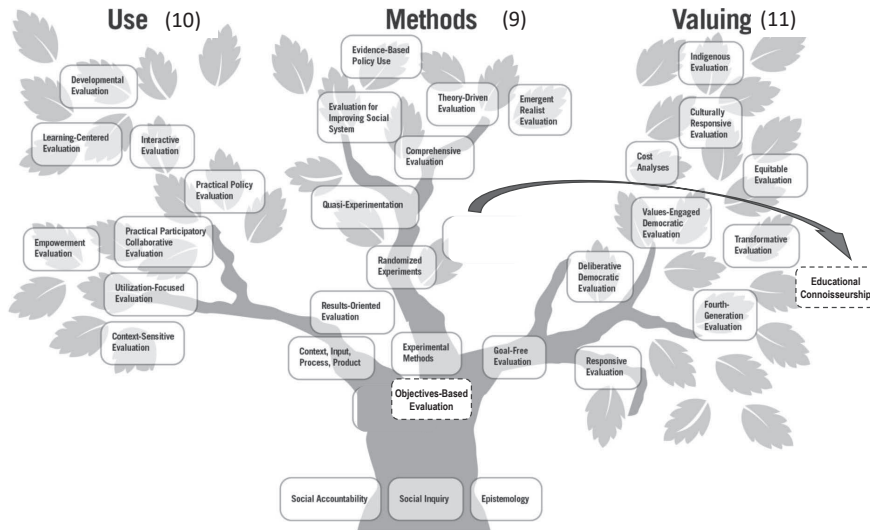
- *Prädiktive* Theorien formulieren Beziehungen oder Prinzipien beobachtbarer Phänomene, die durch Forschung (ansatzweise) verifiziert *sind*. Auf Evaluation angewandt, ließe sich durch solche Theorien erklären oder vorhersagen, inwieweit bestimmte Evaluationsaktivitäten zu bestimmten Folgen führen. Obwohl es Bemühungen gäbe, zu solchen Theorien über Evaluation zu gelangen, lägen bislang keine derartigen vor und sie seien aufgrund der Komplexität von Evaluationen auch schwer erreichbar.
- *Präskriptive* Theorien legen dar, was die Entwickelnden vorschlagen, wie Evaluationen angelegt sein *sollen*, so dass sie die gesetzten/vereinbarten Zwecke erreichen und von hoher Qualität sind.

Im Buch werden insgesamt 30 präskriptive Theorien vorgestellt und durch die Herausgebenden im Theorie-Baum systematisch zugeordnet. Was mit einer solchen „Evaluationstheorie“ gemeint ist, fasst Jennifer Greene in ihrem Beitrag wie folgt zusammen:

... eine normative Aussage einer Autorin oder eines Verfechters einer bestimmten Theorie, die festlegt, was Evaluation sein sollte, insbesondere, welche Fragestellungen und wessen Interessen durch die Evaluation vorangebracht werden sollen, welche Evaluationsansätze und -methoden vertretbar und geeignet sind, um die jeweiligen Fragestellungen anzugehen, welche Ansichten und Sichtweisen der Beteiligten & Betroffenen in die Evaluation einfließen sollen, welche Kriterien zur Beurteilung der Qualität und Effektivität des zu evaluierenden Programms herangezogen werden sollen und, insgesamt über diese gängigen Elemente, *welche Werte durch die Evaluation gefördert werden sollen*.³ (S. 119, Hervorh. im Original)

3 Diese und weitere Übersetzungen erfolgten durch den Verfasser der Rezension.

Abbildung 1: Der Baum der Evaluationstheorie



Quelle: Alkin & Christie, 2023, S. 13; in Absprache mit ihnen revidiert von W. Beywl⁴

Aufgenommen sind Theorien, die folgenden Kriterien entsprechen: Sie beanspruchen feldübergreifende Geltung, sind also nicht z. B. spezifisch auf Bildung oder Entwicklungszusammenarbeit bezogen; sie adressieren (zumindest auch) Programme als Evaluationsgegenstände; sie sind genügend elaboriert und dokumentiert; sie sind englischsprachig und stammen (von Verfassenden) aus Nordamerika (Ausnahme: Fiona Cram, Co-Autorin von Kap. 13, indigene Māori aus Neuseeland). Es gibt Personen, die an der jeweiligen Theorie mitgewirkt haben, und die bereit sind, einen aktuellen Beitrag zu schreiben.

Die 30 in Abbildung 1 enthaltenen Theorien sind in Kapitel 3 durch Christie und Alkin auf je einer Seite knapp beschrieben. Zu 18 von diesen Theorien gibt es je einen Originalbeitrag der jeweiligen Theoretikerinnen bzw. Theoretiker (Kapitel 4–21). Teil V enthält darüber hinaus zwölf zusätzliche Kurzdarstellungen von meist in der zweiten Auflage von 2004 enthaltenen Beiträgen, deren Verfasser und Verfasserinnen nicht erreichbar oder verstorben sind. Von Michael Scriven gibt es sowohl einen Originalbeitrag als auch eine Kurzfassung in Teil V.

Dass nun „Theorien“ dargestellt werden, stellt eine deutliche Perspektivverschiebung gegenüber den beiden vorangegangenen Auflagen von 2004 und 2012 dar: Dort wurden – in den Untertiteln beider Auflagen so angesprochen – „*theorists*“, als die hinter den Theorien stehenden Personen, vorgestellt, und deren Werke zueinander analytisch in Beziehung gesetzt. Der neue Fokus auf Theorien wird wie folgt be-

⁴ Die revidierte Zuordnung zu den Ästen der Abbildung entspricht der in Kap. 3. Durch den beauftragten Grafiker ist diese in der Buchversion für eine Theorie falsch, für die andere unklar umgesetzt worden.

gründet: „This is because our aim in this volume is to show how evaluation practice looks when guided by theory“ (S. 12).

Kurz zu den drei „Wurzeln“ des Theorie-Baums: Es sind die soziale Verantwortlichkeit als Treiber für Evaluationen, das systematische sozialwissenschaftliche Untersuchen als unverzichtbarer Bestandteil sowie die Erkenntnistheorie, die sich mit Zugängen zu und Grenzen der Erzeugung wissenschaftlichen Wissens befasst.

Daraus gehen die drei „Äste“ hervor:

- *Methoden-Ast* – evaluationsgeschichtlich der älteste: Die hier zugeordneten Theorien konzentrieren sich auf möglichst strenges, genaues und gültiges methodisches Vorgehen beim Beschreiben der Evaluationsgegenstände samt deren (Miss-)Erfolgen. So soll subjektiven oder interessen gebundenen Verzerrungen bei Datengewinnung und -interpretation vorgebeugt werden. Diese Theorien stehen denen (angewandter) sozialwissenschaftlicher (meist quantitativer) Forschung nahe und rekurrieren auf die dort entwickelte Begrifflichkeit (z. B. „Randomisierung“; „Generalisierung“).
- *Bewertungs-Ast*: Die hier versammelten Evaluationstheorien fokussieren darauf, wie die (Zusatz-)Funktion der systematischen Bewertung, welche Evaluation von Forschung unterscheidet, gestaltet werden soll. Nahe am Methodenast sind die eher „objektivistischen“ Bewertungsverfahren positioniert (z. B. Kostenanalysen oder Zielfreie Evaluation), während ganz rechts die „subjektivistischen“ Zugänge liegen – mit dem Extrem von Eisners „Kennerschafts-Ansatz“.
- *Nutzen-Ast*: Diese Ansätze sind eine Antwort auf das ab Mitte der 1970er Jahre durch die Forschung über Evaluation konstatierte Nutzungsdefizit der (bis dato vorrangig methodenorientierten) Theorien. Evaluationsschritte werden hier oft von der Überlegung her geplant, wie eine möglichst umfangreiche und inhaltlich angemessene Nutzung von Prozess und Ergebnissen der Evaluation durch die vorgesehenen Nutzenden erreicht werden kann. Folglich müssen sowohl die systematisch-methodische Beschreibung als auch die systematische Bewertung so angelegt sein, dass Nutzung möglichst sicher und wie vorgesehen ausgelöst werden kann.

Alkin betont, dass die zweidimensionale Darstellung des Baumes den Entwicklungen der letzten 20 Jahre nicht mehr gerecht wird. Man müsse sich vielmehr einen dreidimensionalen Baum vorstellen, in dem sich im Laufe seines dynamischen Wachstums Zweige oder Blätter der drei Äste ‚astübergreifend‘ annähern oder – seltener – voneinander entfernen.

These three tree branches are not meant to be viewed as independent from one another; rather, they have been drawn to reflect the relational qualities among them. Thus, if the tree were in a three-dimensional space, the far right of the valuing branch would relate to (or touch) the far-left side of the use branch. Thus, the relational nature of the three branches helps to further depict approaches. (S. 14)

Es mag sein, dass diese Aussage die Kritik von Kapitelverfassenden an der Zweidimensionalität und nur scheinbaren Eindeutigkeit aufnimmt (z. B. Stake & Visse, 2023, S. 107). Michael Pattons (2023a) nach Erscheinen des Buchs erstellte, vielbeachtete kritische Podcasts zum Buch scheinen diese dreidimensionale Ausweitung der Herausgebenden kaum zu würdigen.

Die 18 Originalbeiträge in den Teilen II–IV

Hinter den nachfolgend zusammengefassten Originalbeiträgen stehen je ein oder mehrere Lehrbücher und/oder Zeitschriftenbeiträge der Kapitelverfassenden. Diese sind in kapitelweisen Literaturverzeichnissen aufgeführt.

Methoden-Ast

Robert Boruch und **Herbert Turner** (Kap. 4) sehen Evaluationen nur dann als „wissenschaftlich“ an, wenn sowohl die Resultate der untersuchten Interventionen als auch die Ergebnisse der experimentellen Studien dazu replizierbar sind. Allein so sei Generalisierbarkeit der Befunde über Raum und Zeit erreichbar. Dies sei ein schwer zu erreichender, dabei nicht disponibler Zweck. Sie erörtern ethische und rechtliche Herausforderungen der Methodik sowie typische Probleme wie Stichprobenausfall oder fehlende Werte. Ausführlich schildern sie bezogen auf die USA die zunehmende Verbreitung von *Randomized Controlled Trials* (RCT) und deren Aufbereitung durch 14 *Clearinghouses*.

Melvin Mark und **Gary Henry** (Kap. 5) stehen in der Tradition der Evaluations-Lehrbücher von Peter Rossi und Howard Freeman (Erstauflage 1979; zuletzt Rossi et al., 2019). Evaluation als „unterstützte Sinnggebung“ ist danach auf die Verbesserung der sozialen Lebenslagen (*social betterment*) gerichtet. Gegenüber früheren Evaluationsansätzen aus dem Methodenast betonen sie: Das notwendige „Mehr“ an genauer Information (wesentlich basierend auf – quantitativen – sozialwissenschaftlichen Methoden) sei nicht hinreichend. Zusätzlich sei Aufklärung der Wertgrundlagen (*values inquiry*) erforderlich. Evaluation solle mehrere Zwecksetzungen erwägen und genauer bestimmen, auf welchen Pfaden sie Einfluss nehmen kann. Dieser reflexive Beitrag, der positivistisch-realistisches mit konstruktivistischem Denken verbindet, zeigt exemplarisch die vielfachen Verbindungen zwischen den drei Ästen des Evaluationsbaums.

Huey Chen (Kap. 6) ist wohl wichtigster Promotor der *theory-driven evaluation* (TDE). Eine sowohl für effektive Programme wie auch für deren Evaluation geeignete „Programmtheorie“ umfasse zwei verknüpfte Modelle:

- Das *change model* (deskriptiv) beschreibt den erwünschten kausalen Prozess, der durch das Programm generiert wird. Es enthält Ziele/Outcomes, Interventionen/Treatments sowie kausale Mechanismen/Wirkannahmen.
- Das *action model* (präskriptiv) ist eine Art Programm-/Projektplan, der Ressour-

cen, Personal, tragende/kooperierende Organisationen, Leistungsprozesse und Zielgruppen ausweist und verbindet.

Wie die Stakeholder moderierend unterstützt werden können, eine Programmtheorie zu klären, wird leitfadenartig ausgeführt. Das resultierende *action model-change model-scheme* wird für zwei Outcome-Evaluationen mit Flussdiagrammen visualisiert. Für eine Prozess-Evaluation, in der es um die Art und Treue der Umsetzung der Intervention geht, werden vorgesehene und realisierte Umsetzung in einer Tabelle gegenübergestellt. Chen betont die Bedeutung des (auch kulturellen) Kontexts für die Wirkung/Wirksamkeit der Intervention.

Bewertungs-Ast

Michael Scriven (Kap. 7; 2023 verstorben) charakterisiert mit „*Valuing Science*“ die Evaluation – in Fortführung seiner 60jährigen Argumentation – als „transwissenschaftlich“, also „mehr“ als wissenschaftlich: da sie Kernfragen jeder Wissenschaft übergreifend bearbeite, nämlich die nach „guten“ Theorien, Erklärungen usw.

Robert Stake entfaltet – zusammen mit **Merel Visse** (Kap. 8) – das von ihm seit den 1960er Jahren begründete und die Evaluation maßgeblich prägende Basismodell „Responsive Evaluation“. Analog zu Scriven (dort: „Wissenschaften“) erachtet er alle personengebundenen Dienstleistungen als vom „Bewerten“ durchtränkt (*infused*). Dies erzwingt systematische Evaluation (etwas anderes als Forschung). Evaluation solle sich nicht auf leicht Messbares (z. B. via konsolidierter Item-Skalen) kaprizieren, sondern Daten stets maßgeschneidert erheben. Evaluation müsse emergent vorgehen, auch bzgl. der Bewertungskriterien. „Unethische“ Versprechen, wie das, Kausalität nachzuweisen, seien zu meiden. Vermehrt solle erzählend statt (überwiegend) mit statistischen Maßzahlen kommuniziert werden.

Ernest House (Kap. 9; Hauptwerk: 1999) reagiert mit seinem Beitrag zur aushandelnden (*deliberative*) demokratischen Evaluation auf die vielgestaltigen Spannungsfelder der „postnormalen Zeiten“. Stichworte: Verortung auf der Welt (*region*), soziale Klasse, Ethnizität (*race*), Gender, Identität. Er wünscht sich Evaluation als moralische Treuhänderin. Dabei gelte es, „tiefe Werte“ zu respektieren, die oft durch Familie, Annahmen zu Fairness und In-Group-Loyalität geprägt seien. Dabei seien auch diejenigen einzubeziehen, die sich abseits öffentlich verhandelter Moral bewegten, z. B. die Wählerschaft der Tea-Party-Bewegung.

Jennifer Greene (Kap. 10) betont ebenfalls die Ubiquität von Werten. Diese seien oft implizit, auch in Programmkonzepten: „Werteengagierte demokratische Evaluation“ sei mehr soziale Praxis als Wissenschaft. Deren Aktionsfeld sieht sie in kleinen/mittelgroßen Programmen, besonders in deren *frühen* Phasen.

Rodney Hopson und **Vidya Shankar** (Kap. 11) bekräftigen ihre bereits vor Erscheinen des Buchs geäußerte Kritik an dessen in der „weißen Überlegenheitsüberzeugung wurzelndem“ Evaluationsverständnis. „Kulturell responsive Evaluation“ müsse gelebte Erfahrung integrieren, vorzugsweise von denjenigen, die durch die

sozialen Bedingungen meist benachteiligt sind, und zwar bezüglich Messverfahren, Werkzeugen, Ressourcen, Ansätzen usw.

Donna Mertens (Kap. 12), Verfasserin eines wichtigen aktuellen Evaluationslehrbuchs, plädiert für die „transformative Linse“ in der Evaluationstheorie. Sie sieht sich dabei als Advokatin der Benachteiligten und von Macht Ausgeschlossenen. Sie sucht Verbindungen u. a. zu „kulturell responsiven“ und „indigenen“ Ansätzen. An zwei Beispielen aus Afrika (Aids-Prävention in der LGBTQI-Community/schulische Förderung Armutsbetroffener) arbeitet sie heraus, dass die Methodenwahl in Evaluationen auf die (oft unterprivilegierten) Beteiligten und Betroffenen ausgerichtet werden sollte.

Fiona Cram und **Jill Anne Chouinard** (Kap. 13), die beide mit Donna Mertens veröffentlicht haben, fragen, wie Evaluation eine Kraft sein könnte, um indigene kulturelle Vitalität, Wohlergehen und Nachhaltigkeit zu unterstützen. Es sei wichtig, dass Evaluierende über eine mit solchen Stakeholdern geteilte gelebte Erfahrung verfügen. In zehn Dimensionen seien Verbindungen von Kultur und Evaluation möglich (u. a. ontologische, epistemologische, ökologische, methodologische).

Nutzen-Ast

Eleanor Chelimsky (Kap. 14; 2022 verstorben) hat über Jahrzehnte zentrale Evaluationsstellen in nationalen/internationalen Organisationen geleitet. Sie adressiert Evaluationsverantwortliche, die in ‚heißen‘ politischen Konfliktzonen arbeiten, z. B. Energieversorgung, Verkehr oder aktuell zur Polizeireform im Kontext von „*Black Lives Matter*“. In diesen Zonen sieht sie Evaluationen durch unvorhersehbare politische Interventionen bedroht. Sie plädiert daher – im Unterschied zu vielen anderen Buchbeiträgen – für „*nonadvocacy*“. Sorgsamer Sprachgebrauch, glaubwürdige Objektivität und ein wacher Blick für angrenzende Politikfelder seien essentiell.

Marvin Alkin (Kap. 15) vertritt kompromisslos einen auf instrumentellen und direkten Nutzen fokussierten Evaluationsansatz. Dieser basiert wesentlich auf seinen jahrzehntelangen Forschungen zur (Nicht-) Nutzung von Evaluationen. Schlüssel seien vier Kontexte, für die Evaluationen „sensitiv“ sein müssten: der der (1) Evaluierenden, (2) der Nutzenden, (3) der individuellen und organisationalen Kultur und (4) der Evaluation selbst. In einer mehrseitigen Tabelle listet er für die vier Kontexte je mehrere konkrete Anforderungen an professionelle Evaluationen auf. Konkretisiert werden diese durch beobachtbare Indikatoren angemessenen professionellen Evaluationshandelns.

Michael Quinn Patton (I) (Kap. 16) ist der wohl produktivste (Lehrbuch-)Autor und Theoretiker der Evaluation. Sein erster Beitrag schildert die „Evolution“ der Nutzungsfokussierten Evaluation (U-FE) von der ersten Auflage 1978 bis zur fünften im Jahr 2022. Er belegt deren weite Verbreitung und Anerkennung bis in herausragende internationale Organisationen. Ihr Zweck bestehe darin, „... effektives Handeln und informierte Entscheidungsfindung basierend auf bedeutungsreicher Evidenz zu unterstützen, samt durchdachter Interpretation und engagierter Aushandlung“ (S. 183). Die wichtigsten Weiterentwicklungen in den Auflagen 1–5: 1: Wich-

tigkeit „politischen“ Aushandelns/„personaler Faktor“; 2: „Intendierter Nutzen für intendierte Nutzende“ mit vorab geklärten Bewertungskriterien; 3: Entdeckung des „Prozessnutzens“, der sich im veränderten Denken und Handeln als Lernresultat bei Evaluationsinvolvierten zeige; 4: Globalisierung auch der Evaluationsperspektive und 5: Relevanz der Komplexitätstheorien; Erweiterung des Kriterienkataloges *jeder* Evaluation um „Nachhaltigkeit“ und „Verteilungsgerechtigkeit“ als präskriptive Kriterien. U-FE habe keine fixierte Schrittfolge. Sie sei ein metatheoretisches Denkgerüst, mit dem je nach Evaluationsgegenstand/Kontext ein bestgeeigneter Evaluationsansatz gewählt und (rück-)schleifenhaft feinabgestimmt wird („Adaption“).

David Fetterman (Kap. 17) präsentiert die von ihm maßgeblich geprägte „Empowerment Evaluation“. Sie ist aus der Gemeinwesenpsychologie und -arbeit hervorgegangen und basiert auf Grundannahmen des Pragmatismus (Lewin, Argyris, Schön). Ihr Zweck bestehe darin, kollektives Handeln primär der Programm-Mitarbeitenden zu stärken. Dies gilt auch für deren Zielgruppen, so dass diese „mehr Kontrolle gewinnen über ihr Leben und die damit verbundenen Übergänge“ (S. 193). Evaluierende „arbeiten *mit* Teilnehmenden, anstatt *für* sie anwaltlich tätig zu werden“ (S. 196). Es gehe darum, moderiert durch die Evaluierenden, eine „Theory of Action“ zu entwickeln. Empirie und Daten kämen (später) ins Spiel, die „Theory of Use“ zu untermauern. Beide Theorien werden gegenübergestellt. Aus dem Vergleich Ideales/Reales könnten optimierte Konzepte und Aktionen hervorgehen. Der Beitrag stellt einen dreischrittigen Moderationsablauf (*Mission – Taking Stock – Planning for the Future*), illustriert durch Beispiele, plastisch dar.

Bradley Cousins bietet – zusammen mit **Hind Al Hudib** – (Kap. 18) eine Übersicht zu den von ihm entwickelten und beforschten kollaborativen Ansätzen der Evaluation (CAE). Als deren nutzungsorientierten Zweig stellt er die Praktische Partizipative Evaluation (PPE) detailliert vor. Hier kooperierten Evaluationsprofis mit Beteiligten & Betroffenen des zu evaluierenden Programms auf Augenhöhe. PPE sei besonders für formative Evaluationen geeignet, in denen sich Informationsbedarfe und Fragestellungen emergent entwickeln, und die eine transformative Intention haben. Sechs ethische PPE-Dimensionen sind: Reflexivität, Relationalität, Verantwortung, Anerkennung, Repräsentation, Reziprozität. PPE wird orientiert durch sieben Prinzipien, darunter unabdingbar die geklärte Motivation für Zusammenarbeit, darüber hinaus u. a. sinnstiftende Beziehungen und Förderung evaluativen Denkens. Bemerkenswert sind die Ausführungen zum Verhältnis von Theorie und Praxis (darüber hinausgehend auch: Forschung): Gute Theorien könnten nicht anders als über induktive Arbeit auf Basis realer lebendiger Praxis entwickelt werden. Erst dann sei die gewonnene Theorie sinnvoll überprüfbar.

Hallie Preskill (Kap. 19) stellt Lernen ins Zentrum „Evaluation sollte anhaltend, reflexiv und eingebettet in organisationale Praxis sein [...] integriert in die Alltagspraxis der Organisationsmitglieder“ (S. 229). Sie fungiere als Katalysator, sei immer auch Intervention. Diversität der in die Evaluationsplanung einbezogenen Beteiligten und Betroffenen könne sowohl die Nutzung der Evaluation fördern als auch deren Fähigkeiten und Kompetenzen stärken. Evaluierende nähmen unterschiedliche

Rollen und Funktionen wahr wie Moderatorin, Guide, Ausbilderin, Mentorin, kritischer Freund oder Provokateur. Dabei sei die verbreitete Defizitorientierung zurückzudrängen, zugunsten wertschätzender Sichtweise und Sprache.

Jean King (Kap. 19) skizziert ihre „Theorie“ der interaktiven Evaluationspraxis. Es gehe darum, „Menschen dabei zu helfen, etwas zu systematisieren, was Routinebestandteil ihres (Arbeits- [Zusatz W.B.] Lebens ist“ (S. 234). Entscheidende in Organisationen sollten befähigt werden, das Funktionieren von Programm bzw. Organisation evaluations- und damit auch datenbasiert zu verbessern, im Sinne einer positiven und funktionalen Organisationskultur. Wichtig sei der Aufbau tragfähiger persönlicher Beziehungen zu Beteiligten. Der Ansatz stellt hohe spezifische Anforderungen an die „interpersonalen Fähigkeiten“ der Evaluierenden. Dies konkretisiert King durch vielfachen Bezug auf die Kompetenztaxonomien der AEA.

Michael Quinn Patton (II) (Kap. 20) vertieft den Ansatz der „*Developmental Evaluation*“ (DE; von mir als „Evolutive Evaluation“ bezeichnet (Beywl, 2011)). Diese „stellt evaluative Informationen und Feedback bereit für soziale Innovatorinnen und Innovatoren und deren Finanzierende und Unterstützende, um adaptive Entwicklung von Initiativen des Wandels in komplexen dynamischen Umwelten zu informieren“ (S. 243). Damit könne eine Leerstelle „traditioneller“ Evaluationsansätze geschlossen werden, die sich typischerweise auf klare, spezifische und messbare Outcomes konzentrierten. Nach einem geschichtlichen Abriss der DE werden deren fünf Zwecke und acht Prinzipien erläutert. Herausstechend: die Kritik an mechanistischem Festhalten an Fidelity, Compliance und Umsetzungstreue als Kriterien. Patton steigert die Emphase weltumspannend wirksamer Evaluationswissenschaft, z. B.: „[Erstrebenswerte] Transformation beinhaltet großangelegte und dramatische Entwicklung eines bedeutsam verschiedenen Systems als dem, das in der Ausgangslage existierte“ (S. 250).

Vergleichende Darstellung der 18 Theorien – Teil V

Die Beiträge der Teile II–IV lösen die Vorgabe des Herausgabe-Teams, eine präskriptive Theorie vorzustellen, nicht alle ein. Einige markieren stattdessen eine evaluationspolitische Position (Scriven, Kap. 7; Greene, Kap. 10; Hopson/Shanker, Kap. 11) oder thematisieren vorrangig die persönliche Evaluationsbiographie (Preskill, Kap. 19; King, Kap. 20).

Um dies zu kompensieren, stellen die Herausgebenden – ergänzend zu Kapitel 3 – alle 18 Theorien nach einem einheitlichen Schema auf je einer Druckseite dar. Sie greifen auf weitere Quellen der je Verfassenden zurück und bitten diese um Gegenlesen. Folgende Punkte werden behandelt: (Selbst-)Benennung der Theorie; jeweilige Verfassende (*theorists*), generische Zwecksetzung (*goal*); Basisannahmen (*rationale*); typischer Anwendungskontext (z. B. nach Größe des Evaluationsgegenstands oder Überwiegen der formativen vs. der summativen Rolle); Anwendungshinweise (*prescriptions*); beobachtbare Evaluationsaktivitäten.

Für die Evaluationspraxis bietet dieser Teil einen erheblichen Mehrwert: Evaluierende können in der Angebots- oder Vorplanungsphase gezielt nach geeigneten präskriptiven Theorien suchen oder diese kombinieren. Mit dieser Aufarbeitung steigen die Chancen für eine konsistente Inanspruchnahme der theoretischen Konzepte und Vorgaben, im Sinne einer sich gegenseitig unterstützenden Entwicklung der Theorien und ihrer praktischen Anwendung.

Diskussion

Das in der Einleitung (S. V) genannte Doppelziel, dass das Buch sowohl zum besseren Verständnis der Evaluationstheorie beitragen als auch Evaluierenden ‚praktisch‘ helfen soll, für ein bestimmtes Vorhaben die geeignete Evaluationstheorie zu wählen, wird erfüllt. Nicht wenige Beiträge bieten auch ‚Handwerkliches‘, z. B. Leitfaden- oder checklistenartige Moderationshilfen oder Kompetenztaxonomien.

Bei Unterschieden in Schwerpunktsetzung und Details *konvergieren* die präsentierten Evaluationstheorien:

- Der Einsatz sozialwissenschaftlicher Methoden ist durchgängig selbstverständlich. Das genaue ‚Welche‘ oder ‚Wie‘ wird meist nur ansatzweise erörtert, oft mit Hinweisen auf die erforderliche Passung auf Evaluationszweck, Evaluationsgegenstand und Kontext. Folglich herrscht Methodenpluralismus.
- Dass Evaluationen nützlich sein müssen, steht außer Frage. Meist steht der instrumentelle Nutzen deutlich im Vordergrund, orientiert durch Informationsbedarfe der Stakeholder. Einige betonen das (organisationale) Lernen als zentralen Nutzen. Manche dieser Beiträge schränken ein, dass eine derart ausgeprägte Nutzenfokussierung vorzugsweise für kleine und mittelgroße Evaluationsgegenstände, also nicht solche auf (inter-)nationaler Ebene angemessen ist.
- Die Berücksichtigung von (sozialen) Werten ist ebenfalls fast durchgängig ein wichtiges Element der vorgestellten Ansätze.
- Die Kompetenz des „evaluativen Denkens“ – „für Nicht-Evaluationsfachleute oft befremdlich“ (Patton, 2023b, S. 186) – wird oft auch für Beteiligte und Betroffene als wünschenswert angesehen.

Von diesem Grundmuster setzen sich folgende Theorien in mindestens einem Aspekt deutlich ab:

- Boruch und Turner (Kap. 4) stellen den konzeptionellen Nutzen, das wünschenswerte Mehr an generalisiertem wissenschaftlichem Wissen deutlich in den Vordergrund. Aus ihrer Sicht sind zufallskontrollierte Experimente gegenüber anderen Erhebungsdesigns klar zu bevorzugen. Soziale Werte bleiben implizit.
- Chelimsky (Kap. 14) situiert ihre Theorie auf der politischen Makroebene – vergleichbar dem schweizerischen Konzept der Politikevaluation (Sager et al., 2021; Widmer, 2020) – und fordert kompromisslos die unparteiische Positionierung der Evaluierenden.

Die zu Beginn dieser Besprechung angesprochenen Selbstzweifel von mehreren Beitragenden gegenüber diesem Grundlagenwerk sind einerseits aus europäischer Sicht befremdlich. Sie regen andererseits zum Nachdenken darüber an, wie in der Wissenschaft und Profession der Evaluation Grundsatzdebatten – möglichst präventiv – produktiv geführt werden können. Dies ist auch demokratietheoretisch relevant, denn eine Delegitimierung der Evaluation schwächt die demokratischen Institutionen genauso wie eine Infragestellung ‚der‘ Wissenschaft, etwa im Zusammenhang mit Befunden zur Erdererhitzung oder zu Pandemien. Statt in die Defensive zu gehen, könnte ein offensives Vorgehen darin bestehen, auch originäre Theorien aus Europa (Pawson & Tilley, 1997; Vedung, 1999) oder Neuseeland/Australien (Davidson, 2005; Roorda, 2020) einzubeziehen und weitere aus anderen Weltregionen systematisch zu identifizieren. Dies wie Beywl (1988) deutschsprachig zu tun, wäre eine Option. Hier könnten auch die deutschsprachigen Evaluationstheorien, die qualitative Methoden stärker elaborieren und priorisieren, berücksichtigt werden (z. B. Bohnsack, 2010; Kuckartz et al., 2008). Der Bedarf nach breit ausgreifenden evaluationstheoretischen Diskussionen und Verortungen steigt. Denn auch Evaluationstheorien unterliegen kürzeren Halbwertszeiten, da sich die Kontexte des modernen Lebens in einer turbulenten Welt schnell und unvorhergesehen verändern (Cousins & Al Hudib, Kap. 18).

Fazit und Position

Das Buch bietet einen schnellen Zugang zu aktuellen Evaluationstheorien. Es ermöglicht tiefe Einblicke in Vielgestaltigkeit und Verbindungen zwischen den Theorien, in den Reichtum an Konzepten, Modellen, Prinzipien usw. – und auch in virulente Konfliktlinien. Es ist darin international alleinstehend. Es ist Pflichtlektüre für alle, die sich als wissenschaftlich ausgewiesene Evaluationsprofessionelle sehen (wollen). Es belegt, dass ein evaluatives Wissenschaftsverständnis besteht, das über verbreitete Kategorisierungen z. B. von Gibbons et al. (1994) oder Stokes (1997) hinausgeht, da es die Frage sozialer Werte und darauf bezogener systematischer Bewertung aufgreift. Auch ist es für ihren privilegierten Gegenstandsbereich – „Programme“ – präziser.

Das Buch hilft gegen die häufige Engführung von Evaluationsverständnissen, sei es, dass diese auf disziplinären (z. B. „Methoden und Evaluation“ in der Psychologie-Ausbildung), erkenntnistheoretischen (Neopositivismus, radikaler Konstruktivismus) oder administrativen (Akkreditierungsverfahren/Qualitätsmanagement) Voreingenommenheiten beruht. Es kann die eigene Sicht auf Evaluation erweitern und neu justieren. Es gibt vielfach (zusätzliche) Sicherheit für die professionelle Steuerung auch handwerklich guter Evaluationen. Seine Entstehungsgeschichte und sein offenbar prekärer Status in der nordamerikanischen Evaluations-Community könnten als Impuls aufgenommen werden, sich in den deutschsprachigen Evaluationsgesellschaften intensiv mit den demokratie- und wissenschaftspolitischen sowie den damit verbundenen wissenschaftstheoretischen Grundlagen der Evaluation zu beschäftigen.

Das Buch *wäre* die beste aktuelle Bezugsbasis für alle, die Evaluationswissenschaft auf hohem Niveau studieren wollen. Doch ist bedauernd, dass es im ganzen deutschsprachigen Raum keinen Lehrstuhl gibt, der spezifisch auf die Weiterentwicklung des evaluationswissenschaftlichen Theoriefundus zugeschnitten ist. Wünschenswert wären z. B. ein einschlägiges Graduiertenkolleg, ein Fördergebiet einer großen Stiftung oder ein DFG-Schwerpunktprogramm bzw. Analoges seitens der österreichischen und schweizerischen Wissenschaftsförderung. Viele in den etablierten Sozialwissenschaften verharren in Desinteresse oder gar Aversion gegen Evaluation. Auch für die kleinen Evaluations-Communities im deutschsprachigen Raum drohen die Verbindungen zum Reichtum der Evaluationswissenschaft auszubleichen. Dass – in Folge der vielfach beklagten Praxisferne etablierter qualitativer und quantitativer Sozialwissenschaften – die „Design-based Research“ (Reinmann, 2022) einen Aufschwung erlebt, ist begrüßenswert. Auch dort verblüfft, dass die zu den vielen drängenden Themen bereits in der evaluationstheoretischen Literatur geführten Diskurse kaum rezipiert werden. Schließlich entwickeln Forschende in den Naturwissenschaften, die die Zukunftsrisiken des Planeten Erde besonders im Blick haben, in Reaktion auf die übermächtigen anthropogenen Bedrohungen ein transdisziplinäres Wissenschaftsverständnis (Henze, 2021). So unerlässlich letzteres ist, so bedauerlich ist, dass auch dort viele Räder neu erfunden werden, dass auch dort der Theorie- und Methodenschatz von 100 Jahren Evaluationswissenschaft weitgehend ungenutzt bleibt. Hier könnte diese und eine Folgeauflage von *Evaluation Roots* helfen. Und es braucht deutlich mehr.

Literatur

- Beywl, W. (1988). *Zur Weiterentwicklung der Evaluationsmethodologie. Grundlegung, Konzeption und Anwendung eines Modells der responsiven Evaluation* [Zugl. Diss Univ Köln 1987]. Lang.
- Beywl, W. (2011). Rezension zu: Patton, Michael Quinn (2010): *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use*. Guilford Press. *Zeitschrift für Evaluation*, 10(1), 151–154.
- Bohnsack, R. (2010). Qualitative Evaluationsforschung und dokumentarische Methode. In R. Bohnsack & I. Nentwig-Gesemann (Hrsg.), *Dokumentarische Evaluationsforschung: theoretische Grundlagen und Beispiele aus der Praxis* (S. 9–62). Budrich.
- Davidson, J. E. (2005). *Evaluation methodology basics. The nuts and bolts of sound evaluation*. Sage.
- Fujita-Conrads, E., Christie, C. A. & FastHorse, E. (2023). Evaluation Policy as a Bridge between Evaluation Theory and Practice. The Office of Indian Health Services, a Case Example. In M. C. Alkin & C. A. Christie (Hrsg.), *Evaluation Roots. Theory Influencing Practice* (S. 292–299). Guilford press.
- Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P. & Trow, M. (1994). *The new production of knowledge: the dynamics of science and research in contemporary societies*. Sage.

- Henze, J. (2021). Zur Wissenschaftlichkeit transdisziplinärer Forschung [On the scientificity of transdisciplinary research]. *GAIA – Ecological Perspectives for Science and Society*, 30(1), 35–43. <https://doi.org/https://doi.org/10.14512/gaia.30.1.8>
- House, E. R. & Howe, K. R. (1999). *Values in evaluation and social research*. Sage.
- Kuckartz, U., Dresing, T., Rädiker, S. & Stefer, C. (2008). *Qualitative Evaluation. Der Einstieg in die Praxis* (2. Aufl.). Springer VS.
- Montrosse-Moorhead, B., Schröter, D. & Becho, L. W. (2024). The Garden of Evaluation Approaches. *American Journal of Evaluation*, 0(0). <https://doi.org/10.1177/10982140231216667>
- Patton, M. Q. (2023a, 13. Juni). *From evaluation tree to evaluation forest. Review of evaluation roots* [Video]. Youtube. [https://www.youtube.com/watch?v=pqB\]xm6Dn7k](https://www.youtube.com/watch?v=pqB]xm6Dn7k)
- Patton, M. Q. (2023b). The Ongoing Evolution of Utilization-Focused Evaluation. Practice Informed by Theory, Research, and Reflection. In M. C. Alkin & C. A. Christie (Hrsg.), *Evaluation Roots. Theory Influencing Practice* (S. 183–192). Guilford press.
- Pawson, R. & Tilley, N. (1997). *Realistic evaluation*. Sage.
- Reinmann, G. (2022). Replik und Revision: Standards für Design-Based Research. *EDeR. Educational Design Research*, 6(2). <https://doi.org/10.15460/eder.6.2.1973>
- Roorda, M. B. S. (2020). *Developing defensible criteria for public sector evaluations*. The University of Melbourne. <https://minerva-access.unimelb.edu.au/handle/11343/239232>
- Rossi, P. H., Lipsey, M. W. & Henry, G. T. (2019). *Evaluation: a systematic approach* (8. Aufl.). Sage.
- Sager, F., Hadorn, S., Balthasar, A. & Mavrot, C. (2021). *Politikevaluation. Eine Einführung*. Springer VS.
- Shanker, V. (2020). *The Invisible Labor of Women of Color and Indigenous Women in Evaluation by Vidhya Shanker*. AEA365Blog, 2. März 2020. <https://aea365.org/blog/the-invisible-labor-of-women-of-color-and-indigenous-women-in-evaluation-by-vidhya-shanker/>
- Stake, R. E., & Visse, M. (2023). Responsive Evaluation. In M. C. Alkin & C. A. Christie (Hrsg.), *Evaluation Roots. Theory Influencing Practice* (S. 100–109). Guilford press.
- Stokes, D. E. (1997). *Pasteur's quadrant basic science and technological innovation*. The Brookings Institution.
- Vedung, E. (1999). *Evaluation im öffentlichen Sektor*. Böhlau.
- Widmer, T. (2020). Switzerland. In R. Stockmann, W. Meyer & L. Taube (Hrsg.), *The Institutionalisation of Evaluation in Europe* (S. 199–225). Palgrave Macmillan.

Prof. Dr. Wolfgang Beywl | Univation – Institut für Evaluation | Köln und
 Fachhochschule Nordwestschweiz | Bahnhofstrasse 6 | CH-5210 Windisch |
 E-Mail: wolfgang.beywl@univation.org