

Rezension zu:

Giel, Susanne/Klockgether, Katharina/Mäder, Susanne (Hg.): *Evaluationspraxis. Professionalisierung – Ansätze – Methoden*. Münster: Waxmann, 2015 (1. Auflage). 300 Seiten, 34,90 EUR, ISBN 978-3-8309-3345-8

*Petra Wagner*¹

Der vorliegende Sammelband verfolgt die Zielsetzung, Entscheidungen, die sich in der Evaluationspraxis stellen, sichtbar zu machen. In der Einleitung betonen die Herausgeberinnen, dass sie einerseits für die Vielschichtigkeit und Herausforderungen in der Evaluationspraxis sensibilisieren und andererseits Anregungen für einen kompetenten Umgang mit den damit einhergehenden Entscheidungen liefern möchten. Der fachliche Bezugspunkt des Sammelbands stellt Univation dar. Dabei handelt es sich um ein Institut für Evaluation, das 1997 von Wolfgang Beywl und Berthold Schobert gegründet wurde. Sowohl die Herausgeberinnen als auch die meisten Autor(inn)en der Beiträge sind oder waren feste oder freie Mitarbeiter(innen) in diesem Institut, weshalb die Inhalte und Themen des Sammelbands stark von der Evaluationspraxis von Univation gekennzeichnet sind.

Die Themen selbst sind in drei Abschnitte gegliedert: (1) Professionalisierungsinstrumente, (2) Evaluationsansätze, (3) Evaluationsmethoden.

Der erste Abschnitt zu Professionalisierungsinstrumenten umfasst fünf Beiträge. Im ersten Beitrag beschreiben Melanie Niestroj, Wolfgang Beywl und Berthold Schobert, wie in den letzten Jahrzehnten bei Univation kontinuierlich und systematisch ein Glossar für Evaluation entwickelt wurde und wie dieses genutzt werden kann. Das Glossar steht sowohl in einer Print- als auch Online-Version zur Verfügung. Die Online-Version (vgl. <http://www.eval-wiki.org/glossar>) umfasst aktuell mehr als 400 evaluationsrelevante Begriffe, mehr als 100 Begriffe befinden sich im Rahmen einer Glossar-Werkstatt in Bearbeitung.

Katharina Klockgether befasst sich in ihrem Beitrag mit Checklisten zur Planung und Steuerung von Evaluationen. Der Beitrag gliedert sich entlang folgender Fragen: Welche Arten von Checklisten gibt es? Wie verbreitet sind Checklisten? Was sind die Vorteile und Tücken von Checklisten? Welche Eigenschaften hat eine gute Checkliste? Im Fazit empfiehlt die Autorin, bei der Erstellung und Durchführung von Evaluationen dann auf Checklisten zurückzugreifen, wenn man vor Hürden steht oder Unsicherheiten bestehen. Ganz generell können sie als Erinnerungshilfe dienen, um wesentliche Evaluations-schritte nicht zu vergessen. Zu beachten ist jedoch ein kritisches Hinterfragen der Checklisten selbst.

Marc Jelitto diskutiert das Thema der Informationssicherheit in der Evaluation entlang folgender Fragen: Welche Gefahren sehen Expertenteams? Welche Informationen sind bei Evaluationen in Gefahr? Welche Ziele werden beim Informationsschutz verfolgt? In welchen Bereichen ist ein Schutz besonders wichtig? Welche Faktoren spielen eine wichtige Rolle? Daran anknüpfend werden grundlegende Empfehlungen für den Evaluationsalltag mit einem speziellen Fokus auf Datenschutz im IT-Bereich (Computer, Internet) herausgearbeitet.

Im vierten Beitrag thematisiert Schahrzad Farrokhzad die interkulturellen Kompetenzen als Baustein für professionelles Handeln im Evaluationsbereich. Die Autorin stellt eine Verbindung zwischen den Empfehlungen für die Aus- und Weiterbildung in der Evaluation der DeGEval (2008) und den Standards für Evaluation der DeGEval (2008) auf der einen Seite und der Bedeutung der interkulturellen Kompetenzen auf der anderen Seite her. Darauf aufbau-

¹ FH Oberösterreich, Fakultät für Gesundheit und Soziales, Linz

end geht sie der Frage nach, inwiefern interkulturelle Kompetenzen in Evaluationen hilfreich sein können. Entlang eines Modells zur interkulturellen Kompetenzentwicklung werden verschiedene Dimensionen interkultureller Kompetenzen und ihre Bedeutung für die professionelle Evaluationspraxis anhand konkreter Beispiele veranschaulicht. Der Beitrag kommt zu dem Schluss, dass interkulturelle Kompetenzen für Evaluator(inn)en keine „Sonderkompetenzen“ darstellen sollten, sondern eine Erweiterung der ohnehin für Evaluator(inn)en notwendigen allgemeinen Selbst-, Sozial- und Methodenkompetenzen sowie Feldkenntnisse um eine interkulturelle und migrationssensible Perspektive.

Dörte Schott und Wolfgang Beywl beschreiben im fünften Beitrag den Begriff der Schwarm-Evaluation als Steuerungsansatz. Die Autor(inn)en bedienen sich der Metapher des Schwarms, um das Potenzial mehrerer parallel laufender Evaluationen sichtbar und nutzbar zu machen. Als Anschauungsbeispiel dazu wurden 14 Small-Scale-Evaluationen herangezogen, welche im Auftrag der zweiten europäischen EQUAL-Initiative zwischen 2005 und 2008 durch und in Kooperation mit Univation geplant und umgesetzt wurden. Konkret arbeiteten 12 Evaluierende an Evaluationsaufgaben in einem netzwerkartigen Austausch bezüglich Ideen, Informationen und Arbeitsergebnissen. Der Beitrag thematisiert die Möglichkeiten und Grenzen des Ansatzes der Schwarm-Evaluation.

Der zweite Abschnitt „Evaluationsansätze“ gliedert sich in drei Beiträge. Im ersten Beitrag wird der Programmbaum als Evaluationsinstrument vorgestellt. Samera Bartsch, Wolfgang Beywl und Melanie Niestroj erläutern die Entstehungsgeschichte, die Elemente und die Nutzung des von Univation entwickelten Programmbaums. Dies geschieht anhand von konkreten Beispielen. Der Programmbaum stellt ein Instrument dar, um den Evaluationsgegenstand systematisch zu veranschaulichen und damit sowohl für Programmverantwortliche als auch für Evaluator(inn)en besser verständlich zu machen (siehe <http://www.univation.org/programmbaum>). Die Autor(inn)en verdeutlichen, dass der Programmbaum die gedachte Logik eines Programms textlich und visuell darstellt. Er verkettet die Bedingungen eines Programms in logischer Wiese mit dem darauf bezogenen Programmkonzept, welches als zentrales Element Programmziele enthält sowie die Beschreibung der zur Zielerreichung geplanten Interventionen samt Zeitplänen.

Susanne Giel stellt in ihrem Beitrag einen theoriebasierten Evaluationsansatz vor. Die Au-

torin begründet die Anwendung von theoriebasierten Evaluationsansätzen (z.B. Programmbaum) damit, dass diese, um Wirkungen und Wirkzusammenhänge von Programmen, Projekten und Maßnahmen feststellen zu können, eine Alternative zu experimentellen und zielbasierten Zugängen eröffnen. Sie macht jedoch auch deutlich, dass die Frage, wie Evaluator(inn)en zu nachvollziehbaren Programmtheorien kommen, eine zentrale Herausforderung aller theoriebasierten Evaluationsansätze ist. Zur Veranschaulichung eines theoriebasierten Evaluationsansatzes illustriert die Autorin eine entsprechende Umsetzung am Beispiel einer Evaluation von Modellprogrammen im Bereich der Prävention von Rechtsextremismus, Antisemitismus und Rassismus. Die qualitative und diskursorientierte Evaluation wird anhand von fünf Phasen beschrieben.

Maria Gutknecht-Gmeiner skizziert in ihrem Beitrag den Ansatz der Developmental Evaluation nach Michael Patton. Mit Blick auf den Evaluationszweck führt die Autorin aus, dass die Rolle der Developmental Evaluation über die von Scriven getroffene Unterscheidung zwischen formativer und summativer Evaluation hinausgeht. Darüber hinaus zielt Developmental Evaluation auf Evaluationsgegenstände ab, die sich durch ein hohes Ausmaß an Komplexität auszeichnen. Developmental Evaluation ist daher weder ein Evaluationsmodell noch eine Evaluationsart, sondern ein Evaluationsansatz, der dem Evaluationszweck der Entwicklung sowie der Komplexität des Evaluationsgegenstandes Rechnung trägt. Basierend auf den theoretischen Ausführungen zur Developmental Evaluation veranschaulicht die Autorin die Umsetzung dieses Ansatzes anhand eines Fallbeispiels aus ihrer eigenen Evaluationspraxis – einer Evaluation des Ausbildungsprogramms „Computer-Science-Management“. Methodologisch verwendete die Evaluation den Programmbaum als programmtheoriebasierten Ansatz. Anhand dessen wird die thematische Nähe aller drei Beiträge in diesem Abschnitt deutlich.

Im dritten Abschnitt werden sieben Beiträge zum Themenbereich der Evaluationsmethoden zusammengefasst. Stefan Schmidt beschreibt den Nutzen einer dreistufigen wirkungsorientierten Zielsystematik. Er leistet damit einen Beitrag zu einer systematischen Zielexplikation als Bindeglied zwischen Evaluation und Programmsteuerung. Im Zuge der Zielsystematik unterscheidet der Autor drei Aspekte: (1) hierarchische Ordnung von Zielen, (2) Entwicklung von Outcome-Zielen, (3) Differenzierung von Zielgruppen. Alle drei Aspekte werden um-

fassend dargestellt und exemplarisch anhand der Erfahrungen aus einem Evaluationsprojekt im Bildungsbereich (Bildungsinitiative „Lernen Erleben In Freiburg – LEIF“) veranschaulicht.

Katharina Klockgether beschreibt die Entwicklung von Monitoring-Instrumenten im Rahmen von Evaluationen. Unter Monitoring wird in diesem Beitrag die wiederholte Erfassung von Daten zu gleichen Merkmalen eines Programms, in der Regel dessen Outputs, zum Zweck der Steuerung und Rechenschaftslegung verstanden. Aufbauend auf diese grundsätzliche Begriffsklärung werden Beispiele von Monitoring-Instrumenten präsentiert, die in Evaluationen von Univation angewandt wurden: Output- und Bedingungs-Monitoring im Rahmen der Evaluation von EQUAL-Entwicklungspartnerschaften, Output-, Outcome- und Impact-Monitoring im Förderprogramm „Integration durch Qualifikation“ (IQ).

Matthias Sperling widmet sich in seinem Beitrag der praktischen Durchführung von Face-to-Face Interviews mit Jugendlichen im Rahmen von Evaluationen. Er verdeutlicht, welche Herausforderungen und Problemlagen mit der Befragung von Jugendlichen im Rahmen von Evaluationen verbunden sein können und wie diesen in der Evaluationspraxis begegnet werden kann. Konkret befasst sich der Beitrag mit dem Zugang und der Erreichbarkeit von Jugendlichen, der Sprache in Interviews mit Jugendlichen, dem Verhalten von Jugendlichen in Interviews, der Wahl einer passenden Interviewform, dem Ablauf eines Interviews, der Rolle von Leitfäden zur Interviewgestaltung, der Dokumentation von Interviews, dem Datenschutz und Einverständnis, der Interviewatmosphäre, der Reflexion der Interviewsituation sowie der Kommunikation mit den Jugendlichen im Anschluss an das Interview.

Ute B. Schröder beschäftigt sich in ihrem Beitrag mit der Methode der Gruppendiskussion. Anhand von Beispielen aus der Evaluationspraxis wird verdeutlicht, inwieweit die Rekonstruktion von kollektiven Orientierungen und implizitem Wissen für Evaluationen von Bedeutung sein kann.

Marianne Lück-Filsinger thematisiert in ihrem Beitrag die Anwendung der Grounded Theory im Evaluationskontext. Die Grounded Theory versteht sich als qualitativer Forschungszugang, der durch die Verschränkung von empirischer Forschung und Theoriebildung die Generierung von gegenstandsbezogenen, d.h. in den Daten verankerten Theorien im Sinne einer Strukturgeneralisierung erlaubt. Daran anknüpfend umreißt die Autorin in groben Zügen die theoretischen Grundlagen der Grounded Theory und

veranschaulicht das damit einhergehende methodische Vorgehen im Rahmen einer Evaluationsstudie (Evaluation der Übergangsgestaltung vom Kindergarten in die Grundschule): (1) Offenes Kodieren zum Finden von Konzepten und ersten Ideen für Kategorien, (2) Axiales Kodieren zum Herstellen von Verbindungen zwischen den Kategorien, (3) Selektives Kodieren zum Generieren einer gegenstandsbezogenen Theorie.

Elitsa Uzunova zeigt in ihrem Beitrag auf, in welcher Form neue Technologien bei der Durchführung von Evaluationen dienlich sein können und inwieweit sie den Evaluationsprozess in sinnvoller Weise unterstützen können. Der Beitrag ist nach Anwendungsbereichen gegliedert. Es werden webtechnologische Möglichkeiten zur Datenerhebung, zur Datenauswertung und zur Präsentation von Ergebnissen vorgestellt. Dazu werden jeweils die Vor- und Nachteile, die mit jeder technologischen Lösung verbunden sind, diskutiert. Darüber hinaus werden Technologien erläutert, die nicht direkt im Evaluationsprozess eingesetzt werden, sondern für den Wissenstransfer in der Evaluationscommunity förderlich sein können.

Basierend darauf, dass es eine wichtige Aufgabe von Evaluationen ist, zu Bewertungen zu gelangen, widmet sich Susanne Mäder in ihrem Beitrag Fragen zur systematischen Bewertung durch Evaluationen. Vor diesem Hintergrund ergeben sich einige wesentliche Fragestellungen: Wer übernimmt die Bewertung (Stakeholder und/oder Evaluierende?) Anhand welcher Kriterien erfolgt die Bewertung? Wer legt diese Bewertungskriterien fest? Inwieweit sind diese Bewertungskriterien allgemein akzeptiert? Wird der Bewertungsvorgang systematisch und nachvollziehbar angelegt? Anhand von zwei in der Praxis entwickelten und erprobten Bewertungsverfahren, die in diesem Beitrag vorgestellt werden, werden diese Fragen kritisch reflektiert.

Bei dem vorliegenden Sammelband handelt es sich um eine praxisorientierte Handreichung für Evaluierende und an Evaluation interessierte Personen. Eine erkennbare Gemeinsamkeit der Beiträge ist, dass die dargestellten Evaluationsansätze, Instrumente oder Methoden mehrheitlich anhand von Praxisbeispielen veranschaulicht werden. Diese Form der Darstellung erleichtert es, einen direkten Bezug zur Evaluationspraxis herstellen zu können und unterstützt die Nachvollziehbarkeit der Inhalte. Die Beiträge weisen einheitlich eine grundlegende Qualität und ansprechende Lesbarkeit auf. Es wird deutlich, dass es sich bei den Autor(inn)en um ausgewiesene Expert(inn)en in ihren Bereichen handelt, die be-

reits selbst Evaluationen erfolgreich durchgeführt haben.

Darüber hinaus liefert der Sammelband einen beeindruckenden Einblick in die Entwicklungen und Aktivitäten rund um Univation als Institut für Evaluation. Da sich die inhaltliche Ausrichtung der Beiträge primär an den vorhandenen personellen und fachlichen Ressourcen von Univation orientieren, stellt dieser Sammelband stückweit eine Leistungsschau dieses Instituts dar. Diese Ausrichtung bringt es jedoch auch mit sich, dass die ausgewählten Beiträge in der inhaltlichen Zusammenstellung eher heterogen wirken. Auch ist die Zuordnung der Beiträge zu den drei Abschnitten (Professionalisierungsinstrumente, Evaluationsansätze, Evaluationsmethoden) nicht ganz schlüssig nachvollziehbar. Es stellt sich z.B. die Frage, ob es sich bei der interkulturellen Kompetenz wirklich um ein Pro-

fessionalisierungsinstrument handelt, ob es sich bei der Schwarm-Evaluation nicht auch um einen Evaluationsansatz handeln könnte oder die wirkungsorientierte Zielsystematik und Bewertungsverfahren in der Evaluation nicht eher oder auch Professionalisierungsinstrumente darstellen. Insbesondere bei den Evaluationsmethoden fällt die heterogene Zusammenstellung der Beiträge auf, der Beitrag zu Interviews mit Jugendlichen reiht sich direkt an den Beitrag zu Monitoring in der Evaluation, zwei Themen, die auf ganz unterschiedlichen Ebenen der Evaluation angesiedelt sind.

Zusammenfassend liegt ein inhaltlich bunt zusammengestellter Sammelband vor, der innovative und zugleich brauchbare Konzepte, Anregungen und Ideen für konkrete Evaluationsprojekte enthält.

Rezension zu:

Pietsch, Marcus/Scholand, Barbara/Schulte, Klaudia (Hg.): *Schulinspektion in Hamburg. Der erste Zyklus 2007-2013: Grundlagen, Befunde, Perspektiven*. Münster: Waxmann, 2015. 458 Seiten, 39,90 EUR, ISBN 978-3-8309-3278-9

*Michael Fraiss*¹

Der Sammelband „Schulinspektion in Hamburg. Der erste Zyklus 2007-2013: Grundlagen, Befunde, Perspektiven“ bietet der Leserin/dem Leser einen vielperspektivischen Blick auf die Hamburger Schulinspektion, auch im Sinne einer Bilanz acht Jahre nach ihrer Einführung.

Der Grundaufbau in die vier Teile „Grundlagen“, „Arbeit“, „Befunde“ und „Zukunft“ bietet eine umfassende Gesamtschau auf die Hamburger Schulinspektion – dies auch im Sinne einer ersten Bilanz hinsichtlich des Standes ihrer Umsetzung und der Implementierung im Schulsystem.

Teil 1

Die Grundlagen der Schulinspektion

Norbert Maritzen greift in seinem Beitrag auf einen älteren Aufsatz (2008) zurück. In dem ergänzten und überarbeiteten Nachdruck widmet er sich Aspekten der Transformation von Governance-Strukturen. So wird im ersten Abschnitt auf die grundlegende Veränderung der Rolle von Schulaufsicht mit der Einführung von Schulinspektionen (oder anderen externen Evaluationsverfahren)² verwiesen. Die Vielfalt möglicher Funktionszuweisungen an die Schulinspektionen kann dabei zu jeweils unterschiedlichen Dynamiken in den jeweiligen bildungspolitischen Kontexten führen. Im Zusammenhang mit der Darlegung verschiedener Modelle schulischer Steuerung stellt Maritzen fest, dass die unverbundenen und unsystematisch verteilten Wissensbestände im Zusammenhang mit der Einführung von Schulinspektion eine nach wie vor ungelöste Entwicklungsaufgabe in den Bemühungen um die Umgestaltung schulischer Steu-

erung sind. In diesem Zusammenhang stehen auch die Ausführungen hinsichtlich der Frage der passenden Modellvarianten von Schulinspektionsverfahren im Hinblick auf ihre institutionelle Anbindung. Maritzen bilanziert u.a., dass es nun wichtig sei, verloren gegangenes Vertrauen in die Steuerfähigkeit zurückzugewinnen. Ohne Berücksichtigung neuerer evaluations-theoretischer Ansätze sieht Maritzen dabei drei wesentliche Anforderungen: Nutzung wissenschaftlicher Verfahren, Evidenzbasierung politischer und administrativer Entscheidungen sowie Transparenz über Rahmenbedingungen.

Der folgende Beitrag beinhaltet den Abdruck eines Gespräches zwischen *Herbert Altrichter* und *Norbert Maritzen*. Dabei wird Rückschau auf den Prozess gehalten, der von den ersten Ideen, Impulsen und Startbedingungen bis zur offiziellen Einrichtung der Hamburger Schulinspektion im Jahr 2007 reicht und abschließend noch Fragestellungen ihrer Weiterentwicklung berührt. Mit diesem Gespräch wird die Governance-Thematik im Vergleich zum ersten Beitrag noch einen Zeitabschnitt weiter zurückverfolgt. Norbert Maritzen schildert dabei als Mitbeteiligter der ersten Stunde, wie bei der Einrichtung der Hamburger Schulinspektion verschiedenste Entscheidungen getroffen und systemische Spannungen zwischen den Akteursgruppen im Hamburger Schulsystem austariert wurden. Wichtige Aspekte dabei waren und sind die Klärung und die Definierung der verschiedenen Rollen. Abschließend wird u.a. darauf hingewiesen, dass eine tatsächliche Nutzung der Ergebnisse einer Schulinspektion ein sehr wich-

¹ Pädagogische Hochschule Zürich

² Der Autor dieser Rezension betrachtet Schulinspektion als eine Verfahrensweise von externer Schulevaluation.

tiger Aspekt für die Legitimation des Verfahrens ist.

Über Aufbau und Rolle der Schulinspektion Hamburg berichtet *Martina Diedrich*. Sie erläutert grundlegende Funktionen sowie die Perspektive auf Inspektion als besondere Form der Evaluation sowie ihr Selbstverständnis und ihre Rolle im Kontext. Dies erfolgt als Versuch einer programmtheoretischen Einordnung von Evaluation wie z.B. das Entwicklungs-, Kontroll- und Forschungsparadigma. Sie kommt zu dem Fazit, dass mit dem Anspruch an Multiperspektivität, was die Unterschiedlichkeit der verschiedenen Funktionen von Inspektion betrifft, auch weiterhin verschiedene offene und latente Konflikte vorprogrammiert sind. Diese werden vermutlich auch in Zukunft nicht gelöst werden können. Die arbeitsteilige Trennung von Schulaufsicht und Schulunterstützung bei Entwicklungsaufgaben im zweiten Zyklus der Hamburger Schulinspektion, zusätzlich erschwert durch die Veröffentlichung aller Inspektionsberichte im Internet, beeinträchtigt eine konfliktfreie und konstruktive Nutzung der Inspektionsberichte.

Thomas Brüsemeister versucht mit seinem Beitrag eine Annäherung an eine später einmal auszuarbeitende Programmtheorie der Schulinspektion Hamburg darzustellen. Sein Verständnis von Programmtheorie bezieht sich auf Ehren et al. (2005) und somit auf die Frage nach (oft impliziten) Unterstellungen, die Gesetze und das Qualitätsmanagement des Inspektorats umfassen. Er plädiert dafür, Programmtheorie noch viel weiter zu fassen, als nur nach Interaktionseffekten des Programms mit der Schule zu fragen. Es drängt sich damit eine Sichtweise auf, die Schulinspektion stärker im Kontext des Gesamtsystems Schule betrachten sollte. Er schlägt vor, diese Blickerweiterung mit der Governance-Perspektive vorzunehmen. Brüsemeister nutzt für seine explorative Analyse der Hamburger Schulinspektion acht Interviews mit Expert(inn)en des Instituts für Bildungsmonitoring und Qualitätsentwicklung (IfBQ). Eine wesentliche Erkenntnis seiner Analysen ist, dass die Erfahrungsbereiche des IfBQ in verschiedenen Handlungsbereichen (Wissenserschließung, Zusammenarbeit, politische Verankerung, Positionsklärung, Sichtbarmachung) unterschiedlich stark ausgeprägt sind. Durch stabile Datengrundlagen könne eine Versachlichung des politischen Blicks auf Schulen unter Berücksichtigung des sozialen Wandels möglich sein.

Einen Einblick in das Qualitätsleitbild der Hamburger Schulinspektion gibt *Susanne Ulrich*. Dieses in einem partizipativen Prozess erarbeitete Leitbild ist integraler Bestandteil des

Qualitätskonzepts der Schulinspektion Hamburg. Neben einer detaillierten Darstellung der verschiedenen Qualitätsdimensionen und Qualitätsbereiche werden Entstehung und Struktur des Leitbildes beschrieben. Zudem wird ein Ausblick über die weitere Entwicklung des Qualitätsleitbildes gegeben. So können beispielsweise im zweiten Inspektionszyklus die Schulen ihre Ergebnisse mit ihren eigenen Wahrnehmungen und Erkenntnissen vergleichen.

Teil 2

Die Arbeit der Schulinspektion

Pietsch, van den Ham und Köller stellen einen konzeptionellen Rahmen zur Evaluation von Schulinspektionswirkungen vor. Dies erfolgt vor dem Hintergrund der Erfahrung, dass aktuelle Veränderungen in den Verfahren von Schulinspektion eher auf anekdotischer Evidenz beruhen denn auf empirischer Evidenz. Gleichzeitig steigt insbesondere in deutschen Ländern der Anspruch an Schulinspektion, diese als ein Unterstützungs- und Dienstleistungsangebot für die Schulen zu sehen. Zudem wird sie zunehmend als Grundlage und Auslöser für Schulentwicklungsprozesse betrachtet.

Die Autor(inn)en schlagen vor, an die aktuelle Validitätsdebatte in der empirischen Bildungsforschung anzuknüpfen und das *Interpretation/Use Argument* von *Kane* im Rahmen einer theoriegeleiteten Evaluation von Schulinspektionswirkungen anzuwenden.

Der Frage, wie Expertenurteile bei der Schulinspektion einzuschätzen sind, ob als Achillesferse oder Trumpf, geht *Moritz G. Sowada* nach. Kritisch und vertiefend betrachtet werden dabei die von Schulinspektionen eingesetzten Bewertungsverfahren und deren zugrundeliegende Wertordnungen. Hierbei wird deutlich, dass Berufserfahrung und der professionelle Hintergrund Bewertungen von eingesetzten Evaluationsexpert(inn)en bei externen Evaluationsverfahren beeinflussen können. Sowada empfiehlt den fachlichen Austausch zwischen den Inspektorinnen und Inspektoren zu fördern und eine professionelle Lerngemeinschaft entstehen zu lassen. Wesentlich ist es jedoch nicht, die wissenschaftliche Frage nach Validität in den Vordergrund der Bemühungen zu stellen, sondern die Entwicklungsfunktion und damit den Nutzen externer Schulevaluationsverfahren.

Interpretationen und Fehlinterpretationen von Daten aus der Schulinspektion sind eine Thematik, der häufig viel zu wenig Beachtung geschenkt wird. *Knut Schwippert* widmet sich in seinem Beitrag „Daten für die Schulentwicklung – auf die Perspektive kommt es an“ dieser Pro-

blematik. Anhand verschiedener Analysebeispiele aus Daten der ersten Inspektionswelle der Hamburger Schulinspektion sensibilisiert Schwippert die LeserIn/den Leser für die Tragweiten unterschiedlicher Analyseansätze. Dabei geht er insbesondere auf die sogenannten ‚Klumpungseffekte‘ ein und warnt vor der Gefahr von Übergeneralisierungen bei der Nutzung von hochaggregierten Daten. Er zeigt zudem auf, dass es bei gleichen Datenlagen zu unterschiedlichen Befunden und Differenzierungen kommen kann. Die Auswirkung des Führungshandelns von Schulleitungen wird als weiteres Beispiel für eine differenzierte Betrachtung verschiedener Analyseebenen herangezogen. Er empfiehlt die Frage, was die verschiedenen Akteure in Bezug auf die Inspektion erwarten, in Zukunft stärker ins Blickfeld zu nehmen.

„Wie Schulinspektor/innen fragen und Grundschüler/innen antworten“, so lautet der Beitrag von *Barbara Scholand* unter Mitarbeit von *Stephanie Graw*. Die Eingangs- und Abschlussfragen von sechs Interviews aus dem ersten Zyklus der Hamburger Schulinspektion werden dabei einer detaillierten Analyse unterzogen. Es geht dabei um die grundlegende Fragestellung, welche Formulierungen für die Evaluation von Schulqualität geeignet sind. Neben einer Beschreibung des Interviewablaufs wird die Analyse der Fallbeispiele ausführlich dargestellt. Die Autorinnen kommen unter anderem zu dem Schluss, dass die Beteiligung von Grundschüler(inn)en an Interviews außer Frage steht, dass jedoch fragwürdig sei, ob Aufwand und Ertrag von Schülerinterviews im Rahmen von Schulinspektion in einem vertretbaren Verhältnis stehen.

Peter Schulze und *Christiane von Schachtmeyer* stellen das Hamburger Modell der Schulformexpert(inn)en vor. Schulformexpert(inn)en sind Mitglieder einer Schulleitung, die bei der Inspektion das Team von zwei hauptamtlichen Inspektor(inn)en ergänzen. Sie üben als Schulleitung die gleiche Funktion in derselben Schulform aus, wie die zu evaluierende Schule. Neben einer Einführung in das Modell der Schulformexpert(inn)en werden u.a. Vorzüge und Herausforderungen dieses Modells beschrieben. Schulformexpert(inn)en erleben diese Aufgabe als persönliche Bereicherung, auch wenn unter Umständen Rollenkonflikte auftreten können.

Teil 3

Befunde der Schulinspektion

Kontrolle als zentrales Ziel von Schulinspektion stellen *Melanie Ehren* und *Jaap Scheerens* mehrfach in den Mittelpunkt ihrer Betrachtungen über

„Evidenzbasierte Referenzrahmen zur Schulqualität als Grundlage von Schulinspektion“. Zudem, so ihre Empfehlung, sollte der zu beobachtende verstärkte Fokus auf dem sogenannten Output von Schulen auf Messungen des *Value-Added* beruhen. Diese Verfahrensweisen versuchen, den Beitrag der Schule zur Lernleistung getrennt von allen anderen Faktoren zu betrachten.

Im Weiteren beschreiben sie verschiedene neuere Erkenntnisse der Schuleffektivitäts- und Schulwirksamkeitsforschung, da viele Inspektionssysteme ihre Zielsetzung in den vergangenen Jahren dahin gehend geändert haben, dass sie auf eine Verbesserung von Lehren und Lernen hinarbeiten.

Sie kommen bei ihrem Vergleich von sechs Inspektionskonzepten aus sechs europäischen Ländern abschließend zu dem Schluss, dass Unterrichtsqualität zwar als Hauptbedingung für Schulqualität angesehen wird, jedoch die wesentliche Wirkebene von Unterricht außer Acht gelassen wird, da Daten zur Unterrichtsqualität in der Regel auf Schulebene aggregiert werden.

Der Zusammenhang von Schul- und Unterrichtsqualität und schulischen Rahmenbedingungen findet zurzeit noch viel zu wenig Beachtung. *Holger Gärtner* stellt dies anhand einer Sekundäranalyse mit den Daten der Hamburger Schulinspektion (2007 bis 2013) dar. Er skizziert zunächst Modellannahmen und geht dabei insbesondere auf kompositorische (soziale Zusammensetzung der Schülerschaft) Effekte ein. Im Weiteren wird das methodische Vorgehen unter besonderer Berücksichtigung des Hamburger Sozialindex beschrieben. Abschließend werden Ergebnisse zu Zusammenhängen zwischen Rahmenbedingungen und der Ausprägung beurteilter Aspekte von Schul- und Unterrichtsqualität, bezogen auf die einzelnen Schularten, dargestellt. Grundlegend dabei ist u.a. die Fragestellung, wie relevante Informationen zu den sozialen Kontextbedingungen von Schule systematisch in die Berichtslegung und Bewertung der Inspektionsteams einfließen können. Leider können Aussagen zum Zusammenhang zwischen Kontextbedingungen und Unterrichtsqualität auf Klassenebene nicht analysiert werden, da die Inspektionsdaten nur auf Schulebene vorliegen.

„Die Qualität des Unterrichts an SINUS-Grundschulen“ lautet der Beitrag von *Marcus Pietsch* und *Timo Ehmke*. Anhand der Daten der Hamburger Schulinspektion (2007 bis 2013) wird überprüft, inwieweit sich SINUS-Grundschulen von Grundschulen unterscheiden, die nicht am SINUS-Programm teilgenommen haben. Dies erfolgt in Bezug auf unterrichtliche, erzieherische und organisatorische Prozessmerk-

male sowie auf Unterschiede in der Unterrichtsqualität. Mittels der Methode des *Propensity Score Matchings* wurden adäquate Vergleichsgruppen zwischen den Schulen gebildet. Die Autoren können u.a. nachweisen, dass an SINUS-Grundschulen die Qualität des Unterrichts auf Schulebene homogener, und extrem schwacher Unterricht vergleichsweise seltener zu finden ist als an Schulen, die nicht am SINUS-Programm teilgenommen haben. Die Autoren kommen zu dem Schluss, dass das Programm SINUS vor allem das intendierte Ziel einer verbesserten Unterrichtsentwicklung erreiche.

Klaudia Schulte und *Markus Lücken* untersuchen den Einfluss schulischer Prozesse auf die Lernentwicklung der Schülerinnen und Schüler an weiterführenden Schulen in Hamburg. Sie vergleichen dafür die zehn Schulen mit den größten Lernzuwächsen mit den zehn Schulen mit den niedrigsten Lernzuwächsen zwischen der fünften und siebten Klasse auf Grundlage der Hamburger Vergleichsarbeiten (KERMIT). Die Ergebnisse der Hamburger Schulinspektion werden dabei unter Berücksichtigung der Kontextmerkmale (Hamburger Sozialindex) sowie der Schulform regressionsanalytisch überprüft. Die plausible Annahme „Je besser der Unterricht, desto größer die Lernentwicklung der Schüler(innen)“ wird durch die vorliegende Analyse nicht bestätigt.

„Eltern als Stakeholder von Schule“, so lautet der Beitrag von *Susanne Peters*. Die Notwendigkeit, Eltern als Stakeholder von Schule zu betrachten, ist mittlerweile theoretisch gut abgestützt und lässt sich aus unterschiedlichen theoretischen Ansätzen herleiten. Die Autorin gibt Einblick in Erkenntnisse über die Sichtweise von Eltern, auf der Grundlage von Daten der Hamburger Schulinspektion. Fokussiert wird dabei auf den Informationsfluss zwischen Schule und Eltern bezogen auf das eigene Kind. Wesentlich ist dabei nicht nur der Aufbau effizienter und hinreichender Kommunikationsstrukturen, sondern dabei auch die Kommunikation als gelingenden Interaktionsprozess zu betrachten.

Teil 4

Die Zukunft der Schulinspektion

Welche Trends und Perspektiven sind für die Zukunft der Schulinspektion in Deutschland erkennbar? Dieser Frage geht *Sabine Müller*, mit Blick auf die inzwischen fest etablierte Schulinspektion oder externe Schulevaluation in den meisten deutschen Bundesländern, nach. Einer Skizzierung von Funktionen, Zielen und Wirkungserwartungen von Schulinspektionen in Deutschland folgt eine Bilanz bisheriger Erfahrungen und Er-

kenntnisse. Die Autorin hält als langfristigen Trend eine dynamische oder flexible Balance von interner und externer Evaluation für denkbar, die sie als „*Proportionalität*“ bezeichnet. Eine enge Verzahnung von interner und externer Evaluation scheint dabei ein vielversprechender Ansatz zu sein, um Schulentwicklung zu fördern.

Der Beitrag von *Hans-Jürgen Kuhn* lautet: „Rechtliche, strukturelle und politische Rahmenbedingungen für externe Evaluation in Deutschland“. Der Autor analysiert darin diese Rahmenbedingungen und zeigt beispielhaft auch Konsequenzen, die getroffene Grundsatzentscheidungen auf Wirkungen und Wirksamkeit externer Evaluationsverfahren haben können. Inputvariablen sowie das Verhältnis des Evaluationspersonals zur operativen Schulaufsicht sind für die jeweilige Ausprägung externer Schulevaluation strukturentscheidend. Der Autor zeigt an vielen Beispielen die starke Verflechtung zwischen Bildungspolitik und externen Evaluationsverfahren. Dabei wird deutlich, dass in vielen Punkten noch dringender Klärungsbedarf besteht. Zu Recht weist er daraufhin, dass es „keinen Hebel im Schulsystem gibt, den man nur ergreifen und umlegen muss“. Dementsprechend ist externe Schulevaluation auf langfristige und stabile bildungspolitische Rahmenbedingungen angewiesen, um das ‚Gesamtsystem Schule‘ weiterentwickeln zu können.

Martina Diedrich gibt einen Ausblick auf den zweiten Zyklus der Schulinspektion Hamburg. Dieser ist auf Grundlage der Erfahrungen und Rückmeldungen zum ersten Zyklus weiterentwickelt worden. Neu sind dabei unter anderem eine Ausdifferenzierung des Referenzrahmens Schulqualität, dies auch auf Grundlage wissenschaftlicher Erkenntnisse, sowie die auszugswise Veröffentlichung von Inspektionsergebnissen im Internet. Grundlegend ist die Veränderung des Selbstverständnisses der Hamburger Schulinspektion: „*Vom Objektivitätsanspruch zum Deutungsangebot*.“ Ein Schritt in den evaluationstheoretisch als partizipative Richtung zu beschreibenden Wandel, der sich mit vielen Erfahrungen über Schulevaluation als Impuls von Schulentwicklung im deutschsprachigen Sprachraum der letzten Jahre deckt, ohne dass dies explizit ausformuliert wird. Folgerichtig wird mit dieser Neuausrichtung ein größerer Beitrag der Inspektion für die Schulentwicklung erwartet. Um die Schnittstelle zwischen Schulinspektion und Schulentwicklung konstruktiv zu gestalten, zieht die Autorin das Fazit: „*Inspektionstätigkeit ist Beziehungsarbeit*.“

Abschließend widmet sich *Martin Spiewak* in einem Interview mit *Angela Hellberg*, Clau-

dia Lemke, Gerrit Petrich, Norbert Rosenboom und Mina Zulal der Perspektive der Stakeholder. Im Fokus stehen dabei Rückmeldungen der verschiedenen schulischen Interessengruppen im Hinblick auf die Entwicklungsperspektive von Schulen sowie auf die Verbesserungsperspektiven des Verfahrens der Hamburger Schulinspektion insgesamt. Die Stakeholder empfehlen u.a. die Hamburger Schulinspektion stärker partizipativ auszurichten. Sie spiegeln damit eine Reihe von Schlussfolgerungen, die die Autor(inn)en in diesem Band auf Grundlage ihrer Analysen gezogen haben.

Fazit

Der Band „Schulinspektion in Hamburg“ bietet der Leserin/dem Leser eine umfassende Sammlung von Perspektiven auf die Hamburger Schulinspektion an. Dabei werden viele grundlegende Aspekte und Fragestellungen aus dem Bereich externer Schulevaluationsverfahren angesprochen. Das Hamburger Verfahren wird dabei facettenreich und kritisch betrachtet. Dabei werden Entwicklungskorridore für die Weiterentwicklung der Hamburger Schulinspektion sowie für externe Schulevaluationsverfahren insgesamt erkennbar.

Einige Fragen bleiben bei der Lektüre jedoch offen:

Die Bedeutung der Schnittstelle zwischen Schulinspektion und Schulentwicklung wurde immer wieder thematisiert, ohne dass jedoch explizit eine evaluationstheoretische Verortung vorgenommen wurde. Die Umsetzung des Qua-

litätsmanagements der Hamburger Schulinspektion wurde lediglich durch eine Zufriedenheitsbefragung seitens der Schulleitungen dokumentiert. Von Interesse wäre, wie das umfangreiche Konzept des internen QM auf Grundlage empirisch erhobener Daten, welche die unterschiedlichen Interessengruppen berücksichtigen, evaluiert wird oder werden soll. Zu dem sehr differenzierten Qualitätsleitbild hätte man sich gewünscht, ob und inwieweit dieses tatsächlich umgesetzt und von den Akteur(inn)en mitgetragen und gelebt wird. Hier beschleicht die Leserin/den Leser ein skeptisches Gefühl.

Nicht ganz nachvollziehbar ist, dass Programmtheorie auf den Inhalt des Programms sowie ihre Umsetzung beschränkt wird. Die Entwicklungsperspektive der Programmtheorie selbst, wie von Giel (2013) vorgeschlagen, wird nicht explizit thematisiert. In diesem Zusammenhang wäre noch das Verhältnis von theoriebasierter Evaluation und Governance für die Leserin/den Leser von Interesse: Wo gibt es Unterschiede, wo Überschneidungen?

Insgesamt betrachtet handelt es sich um einen lesenswerten Sammelband, durch den ein hohes Maß an professioneller Selbstreflexion der Hamburger Schulinspektion unter Beweis gestellt wird. Vorbildlich ist auch, dass explizit Kontextfaktoren und die bildungspolitische Ebene thematisiert werden. Es entsteht der Eindruck, dass die Hamburger Schulinspektion sich am Puls der Zeit und auf gleicher Höhe mit dem aktuellen Fachdiskurs aus wissenschaftlicher und praxisbezogener Perspektive befindet.

Rezension zu:

Balzer, Lars/Beywl, Wolfgang: *evaluiert. Planungsbuch für Evaluationen im Bildungsbereich*. Bern: hep verlag, 2015 (1. Auflage). 200 Seiten, 27,00 EUR, ISBN 978-3-0355-0339-5

*Maria Gutknecht-Gmeiner*¹

1. Autoren, Buch und Zielgruppe

Wolfgang Beywl und Lars Balzer sind beide bekannt als Evaluationsforscher. Ihr neues gemeinsames Buch „evaluiert. Planungsbuch für Evaluationen im Bildungsbereich“ basiert daher auf jahre- bzw. jahrzehntelanger Beschäftigung mit dem Thema Evaluation sowie auf einschlägiger Lehrtätigkeit. Es ist aus Veranstaltungsskripten und Studienbriefen hervorgegangen, hat mehrere Vorgängerversionen und wurde immer wieder in der Praxis erprobt und weiterentwickelt. Auch Evaluationsforscher wie Michael Patton und Bob Stake haben, wie die Autoren selbst sagen, „Spuren“ hinterlassen – dies wird im Buch auch an dem stark nutzenorientierten Zugang zu Evaluation, der Einbeziehung von Stakeholdern sowie der gleichwertigen Behandlung qualitativer Methoden erkennbar.

Das Buch versteht sich nicht nur als Einführung in die Evaluation sondern „ist ein praxisnaher Leitfaden für die Planung und Durchführung einer Bildungsevaluation und bietet Ihnen eine systematische Anleitung zum Erwerb bzw. Ausbau der genannten [Evaluations] Kompetenzen“ (S. 7). Die Autoren haben sich also einen expliziten Bildungsauftrag gegeben, sie wollen einschlägiges Evaluationswissen und -können vermitteln, wobei es um die fachlichen Grundlagen geht (Selbst- und Sozialkompetenzen z.B. werden nicht angesprochen). Das gesamte Buch ist didaktisch aufgebaut, weshalb auch empfohlen wird, die Kapitel in der angebotenen Reihenfolge zu lesen oder besser gesagt: durchzuarbeiten. Das Buch richtet sich v.a. an Studierende sowohl im Erststudium als auch in Weiterbildungsstudiengängen, in denen es einen Bezug zur Evaluation gibt, aber auch an

Praktiker(innen) in Bildungsorganisationen, die ihr Wissen zu Evaluation erweitern wollen.

2. Inhalte des Buchs/Überblick

Das Buch ist in 13 Kapitel gegliedert, wobei die ersten beiden Kapitel eine Sonderstellung einnehmen: als Auftakt wird in Kapitel 1 mit einer aus der Praxis von Bildungseinrichtungen gegriffenen *Fallbeschreibung*² das Feld aufgemacht. Es geht dabei um die Evaluation eines Kurses, der Basiswissen zur Erstellung von Internetseiten vermittelt. Dabei wird rasch ersichtlich, wie komplex Evaluation sein kann und welche Überlegungen bereits im Vorfeld angestellt werden müssen, um zweckmäßige und brauchbare Evaluationen zu gestalten. Im Dialog zwischen den für den Kurs Verantwortlichen werden die wichtigsten Fragestellungen (in den ersten Schritten angelehnt an die 4-Ebenen-Systematik von Kirkpatrick (Kirkpatrick/Kirkpatrick 2006) angesprochen, denen sich die Evaluation von Programmen im Bildungsbereich stellen muss. Auf die Fallbeschreibung wird in weiterer Folge immer wieder Bezug genommen.

Kapitel 2 bietet eine recht ausführliche *Einführung*, in der Evaluation definiert und das (Be-)Werten als das besondere Charakteristikum von Evaluation herausgestrichen wird. Es wird wissenschaftliche Evaluation einerseits von Alltagsbewertungen, andererseits von (Grundlagen-)Forschung unterschieden. Dabei werden auch die Begrifflichkeiten geklärt, wobei die Autoren den Begriff ‚Evaluationsforschung‘, der lange als Synonym für Evaluation galt, nun als Forschung über/zu Evaluation definieren. Der Dienstleistungscharakter, der Nützlichkeitsprimat sowie die gesellschaftliche

1 Impulse – Evaluation und Organisationsberatung, Wien

2 Das Fallbeispiel in Kapitel 1 findet sich ähnlich in Balzer 2005: 1ff.

Relevanz von Evaluation werden hervorgehoben. Von Qualitätsmanagement und Controlling wird Evaluation als eigener Ansatz abgegrenzt.

In den nächsten 11 Kapiteln wird dann der gesamte Evaluationsprozess nachgezeichnet. Dies beginnt in Kapitel 3 mit der *Auftrags- und Planungsphase*. Die Autoren argumentieren für „maßgeschneiderte“ Vorgehensweisen anstatt standardisierter Evaluationen „von der Stange“, um dem jeweiligen Evaluationsgegenstand und den Informationsbedürfnissen der Betroffenen und Beteiligten gerecht werden zu können. Sie empfehlen eine gründliche Klärung des Evaluationsauftrags sowie eine systematische Abfolge von zehn Schritten für die Planung und Umsetzung einer Evaluation, die in den darauffolgenden Kapiteln näher erläutert werden.

Kapitel 4 widmet sich der *Bestimmung des Evaluationsgegenstandes*, d.h. seiner Eingrenzung (nicht zuletzt in Abhängigkeit von den Fragestellungen), der Klärung der Ziele bzw. des Zielsystems des Evaluationsgegenstands sowie der Erarbeitung und Nutzung eines logischen Modells. Hier wird auch der Programmbaum als anschauliches und einfach einzusetzendes Modell dargestellt.

Um die *Bestimmung der interessierten Akteure und die Rolle der Evaluierenden* geht es in Kapitel 5, ein wichtiger Schritt, um der Einbettung von Evaluation in einen sozialen Zusammenhang gerecht zu werden. Das Erkennen der verschiedenen Betroffenen und Beteiligten, die Interessens- und Rollenklärung zwischen den verschiedenen Stakeholdern sowie die Notwendigkeit der (möglichst frühen und aktiven) Einbindung des Evaluationsumfeldes sind daher wichtige Themen des Kapitels. Ausführlich wird auf die verschiedenen Aufgaben und Rollen der Evaluierenden eingegangen sowie auf die sich daraus ergebenden (Kompetenz-)Anforderungen. Dabei werden die verschiedenen Evaluationsarten und ihr Einfluss auf die Position und den Auftrag der Evaluierenden vorgestellt.

Als nächster Schritt kommt es in Kapitel 6 zur *Bestimmung der Evaluationszwecke und -fragestellungen*. Es wird ausführlich auf Evaluationszwecke sowohl in zeitlich-funktionaler Hinsicht als auch in Bezug auf den Verwendungszusammenhang – hier wird die Trias Verbesserung, Entscheidung, Wissensgenerierung erläutert – eingegangen und zum besseren Verständnis das Zusammenspiel dieser verschiedenen Funktionen, Zwecke und ‚Rollen‘ (vgl. Definition im „Glossar der Evaluation“, verfügbar unter: <http://eval-wiki.org/glossar/Evaluationsrolle>, letzter Zugriff: 01.07.2016) dargestellt, wobei ‚formativ‘ und ‚summativ‘ als Schlüssel-

begriffe nur en passant eingeführt werden. Dann wird das Finden und Formulieren von Fragestellungen als zentrales Element einer Evaluation thematisiert. Als Anregung und erste Orientierung gibt es eine Übersichtstabelle mit ‚generischen‘ Fragestellungen für Bildungsprogramme. Sehr positiv ist der Hinweis auf die verschiedenen Arten von Fragestellungen – von rein deskriptiven, über evaluative zu attributiven. Diese Unterscheidung fehlte bislang in deutschsprachigen Einführungswerken zu Evaluation, trägt jedoch zur Klarheit im Evaluationsauftrag und -prozess entscheidend bei.

Der *Festlegung von Bewertungskriterien* ist ein eigenes Kapitel (7) gewidmet. Dies verdeutlicht, wie wichtig den Autoren die Bewertungsfunktion von Evaluation ist: Sie sehen Bewertung als den „professionelle[n] Kern der Evaluation und auch ihr Alleinstellungsmerkmal gegenüber anderen Ansätzen wie dem Controlling oder dem Qualitätsmanagement“ (S. 82). Kriterien sowie Kriterienpunkte (Schwellenwerte) sind daher aus Sicht der Autoren „unverzichtbar“ (S. 76) und sollten im Vorfeld festgelegt werden. Die Autoren nehmen keine normative Setzung vor, wer im Evaluationsprozess Kriterien bestimmen und/oder die Bewertung vornehmen darf/soll, sondern zeigen die verschiedenen Möglichkeiten auf – von der alleinigen Verantwortlichkeit der Evaluierenden über den Einbezug von Betroffenen und Beteiligten des evaluierten Programms bis zur Ausklammerung der Bewertung aus dem eigentlichen Evaluationsprozess. Es wird mit dem nicht einlösbaren Anspruch der „absoluten Objektivität“ aufgeräumt und die Notwendigkeit eines planvollen, systematischen, methodisch kontrollierten und nachvollziehbaren Vorgehens bei der Bewertung eingefordert.

Das Buch kommt ohne detaillierte Beschreibungen von Datenerhebungsmethoden und -instrumenten aus (siehe unten). Sehr wohl aber wird in Kapitel 8 die *Auswahl von Erhebungsdesigns und -methoden*, die jeder Evaluationsplanung zugrunde liegt, ausführlich erläutert. So werden die verschiedenen quantitativen Erhebungsdesigns – mit Ein-Gruppen- und Mehr-Gruppen-Designs und verschiedenen Messzeitpunkten – beschrieben und ihre Aussagekraft und Einsatzmöglichkeiten diskutiert. Dann gehen die Autoren auf qualitative und quantitative Erhebungsmethoden ein. Sie legen jeweilige Stärken und Schwächen dar und – daraus folgend – ihre Eignung für verschiedene Evaluationszwecke und Fragestellungen. Entscheidend ist es, aus Sicht der Autoren, eine sorgsame, für

den jeweiligen Kontext passende Design- und Methodenwahl zu treffen.

Der *Durchführung der Erhebungen* ist Kapitel 9 gewidmet, dabei wird vor allem auf eine gute Einbindung in den Alltag des Bildungsprogramms/der Bildungseinrichtung hingewiesen sowie auf den Umgang mit Störungen und Konflikten. In Kapitel 10 geht es um die *Datenauswertung, Interpretation und Bewertungssynthese*. In Bezug auf quantitative und qualitative Auswertungsmethoden verweisen die Autoren wieder auf vorhandene Literatur. Sie widmen sich zuerst den Grundlagen jeder Analyse: der Auswertungsplanung sowie der Beschreibung von Datenquellen, Stichprobe(n) und Erhebungsdurchführung. Den folgenden Prozess der eigentlichen Auswertung und Interpretation gliedern sie in elf an Carol Weiss (1998) angelehnte Analyseschritte, die für qualitative und quantitative Verfahren gelten, verschiedene Analysegesichtspunkte beschreiben und in denen das Beschreiben und ‚Zählen‘ nur den Anfang darstellen. Das Kapitel endet mit einer kurzen Behandlung der Bewertungssynthese, die es ermöglicht, aus den Einzelbewertungen von Kriterien eine Gesamtbewertung eines Evaluationsgegenstands zu ziehen.

Es folgt in Kapitel 11 die *Berichterstattung*, die auf die verschiedenen Personen(gruppen), an die sich die Berichterstattung wenden kann („Adressierte“, S. 117), zugeschnitten werden muss und rechtzeitig zu erfolgen hat. Verantwortung und Zuständigkeiten werden noch einmal geklärt, insbesondere was das Schlussfolgern und Bewerten angeht sowie das Formulieren (möglicher) Empfehlungen und die daraus folgenden Entscheidungen über die weitere Programmumsetzung/-entwicklung (Letztere werden eher nicht bei den Evaluierenden verortet.) Schließlich werden (und auch das wurde in der bisherigen deutschsprachigen Literatur kaum behandelt) die verschiedenen Berichtsformate, unter denen schriftliche Berichte (für die Qualitätsanforderungen definiert werden) nur eine Möglichkeit darstellen, ausführlich beschrieben.

Der *Nutzung der Ergebnisse* ist Kapitel 12 gewidmet. Die Autoren gehen kurz auf den Prozessnutzen von Evaluation ein, um sich dann v.a. der instrumentellen Nutzung von Evaluationen zuzuwenden. Eine erfolgreiche Nutzung der Ergebnisse wird im Sinne der Nutzungsorientierung als Teil des Evaluationsprozesses gesehen, auf den auch die Evaluierenden einen Einfluss haben – und nicht als etwas, was rein nachgelagert erfolgt. Sie erfordert eine vorausschauende Planung, zentraler Faktor ist der Einbezug der Betroffenen und Beteiligten. Verbreitung und Rezeption werden als notwendige, aber nicht

hinreichende Bedingungen für die Nutzung gesehen. Auch auf die Gefahren der Nichtnutzung bzw. des Missbrauchs weisen die Autoren hin.

Mit der geforderten Dokumentation der Nutzung leiten die Autoren über zu Kapitel 13, das sich der *Evaluation der Evaluation* annimmt. Nach einer Unterscheidung zwischen Evaluations-synthese, Meta-Analyse und Meta-Evaluation werden die „Standards für Evaluation“ als Kriterien für die Meta-Evaluation präsentiert. Die Meta-Evaluation als Schlussphase des Evaluationszyklus wird v.a. in Hinblick auf die Qualitätssicherung argumentiert, hier hätten die Autoren noch stärker auf die interne oder externe Weiterentwicklung des Vorgehens bei der Evaluation verweisen können.

Im Anhang des Buches finden sich ein ausführliches Literaturverzeichnis sowie Hinweise zu Onlineresourcen, Musterlösungen zu den Übungsaufgaben, ein Glossar, das auf dem von Univation erstellten Online-Glossar (vgl. eval-wiki.org/glossar/Eval-Wiki:_Glossar_der_Evaluation) basiert, sowie ein Stichwortverzeichnis.

3. Diskussion

Das Buch besticht durch seine gute theoretische Fundierung, die auch die internationalen Status quo der Evaluationsforschung berücksichtigt, und die praxisorientierte und eingängige Aufbereitung. Es zeichnet sich dadurch aus, dass die Spezifika der Evaluation sehr klar herausgearbeitet und die komplexen Anforderungen an Evaluation und Evaluator(innen) in gut verständlicher Art und Weise präsentiert werden.

Als roter Faden ziehen sich das Bewerten und die Nutzungsorientierung als grundlegende Merkmale von Evaluation sowie die Einbettung von Evaluation in einen sozialen Kontext durch das Buch. Darin hebt es sich positiv ab von anderen, ähnlichen Einführungen in die Evaluation, in denen diese Dimensionen oft wenig oder gar nicht thematisiert werden. Auffällig ist auch, dass es kein Kapitel mit Beschreibungen gängiger Methoden gibt. Die Autoren selbst geben als Begründung an, „dass es sehr viel gute Literatur gibt, die die unterschiedlichen Erhebungsmethoden ausführlich beschreibt. Diese Publikationen können für den Evaluationsprozess verwendet werden“. Stattdessen werden Erhebungsdesigns und Auswahl von Methoden intensiver behandelt, die aus Sicht der Autoren von entscheidender Bedeutung sind, oft aber „zu wenig Beachtung finden“ (S. 84). Dieser Fokus ist eine sehr sinnvolle Entscheidung, vermeidet er doch Redundanzen zu bestehender Methodenliteratur (die zudem ausführlicher sein kann als ein Kapitel in einem Evaluationslehrbuch), während er Lücken

in Bezug auf die grundlegenden Fragen der empirischen Vorgehensweise schließt.

Die Entscheidung, die Methoden auszuklammern, hat aber aus Sicht der Rezensentin zusätzlich Signalwirkung: Sie macht deutlich, dass Evaluation mehr ist als sozialwissenschaftliche Methoden anzuwenden.³ Mit dem Planungsbuch von Balzer und Beywl gibt nun ein Werk, in dem es hauptsächlich darum geht, darzulegen, was Evaluation – über die Anwendung sozialwissenschaftlicher Methoden hinaus – ausmacht. Und das ist doch einiges: immerhin 142 Seiten haben die beiden Autoren dazu gefüllt. Gleichzeitig wird damit, so das Buch als Evaluationsleitfaden genutzt wird, der gerade bei Evaluationsnoviz(inn)en beobachtbaren Fixierung auf Erhebungsinstrumente und (unbedachte) Aktion vorgebeugt, zugunsten eines planvollen, analytischen und systematischen Vorgehens.

Was das Buch noch bereichern könnte, wäre eine kurze Abhandlung zu qualitativen Designs. Auch die Verwendung eines Mixed-Methods-Ansatzes ist aus Sicht der Rezensentin eine Designfrage (und nicht nur eine Methodenfrage) und sollte dort diskutiert werden.

Die didaktische Aufbereitung des Buches ist mit Lernzielen zu Beginn jedes Kapitels, vielen Beispielen (insgesamt 42) und Übungsaufgaben (mindestens eine pro Kapitel) sowie Angaben zu vertiefender Literatur sehr gut gelungen. Es ist ein Lehrbuch, wie man es sonst eher aus dem angloamerikanischen Raum kennt: gut verständlich und leicht lesbar geschrieben, mit anschaulichen Fallbeispielen aus der Praxis, Zusammenfassungen und Aufbereitung von zentralen Informationen in abgesetzten Kästen. Die Übungsbeispiele regen zur Reflexion und zur Diskussion an, sie fördern und fordern die gedankliche Durchdringung des ‚Stoffes‘. Das Layout könnte noch verbessert werden. So ist die helltürkise Schrift, die für die Beispielsbeschreibungen verwendet wird, aus Sicht der Rezensentin nicht sehr gut lesbar; Schlüsselaussagen – wie z.B. die Definition von Evaluation (S. 24) sind im Schriftbild weniger stark hervorgehoben als z.B. die Übungsaufgaben, wodurch sie weniger gut sichtbar werden.

Das Buch eignet sich zum Selbststudium oder als Basisliteratur für Evaluationsseminare. Es kann auch Fachpersonal aus dem Bildungsbereich, die Evaluationen durchführen (müssen) als Begleitbuch dienen. Allerdings geht es inhalt-

lich und auch vom Umfang her über einen reinen Praxisleitfaden (wie von Wolfgang Beywl oder seinem Umfeld bereits mehrfach publiziert: vgl. z.B. Beywl/Schepp-Winter 2000; Keller-Ebert/Kißler/Schobert 2005; Beywl/Kehr/Mäder/Niestroj 2008) deutlich hinaus. Das Buch ist nicht nur ausführlicher, sondern auch anspruchsvoller, es geht auf dahinter liegende theoretische Konzepte ein und verbindet Theorie mit Praxis. Es empfiehlt sich also für Personen, die durchaus tiefer in die Materie eindringen wollen und nicht allein eine praxisorientierte Anleitung suchen (auch wenn das Buch als solche genutzt werden kann). Lehrende der Evaluation finden einen Fundus an gut aufbereiteten Schlüsselinformationen, Beispielen und Übungsaufgaben, die entweder direkt übernommen oder dem jeweiligen Kontext angepasst werden können.⁴

Das Buch spricht vor allem Personen aus der Weiterbildung/Erwachsenenbildung direkt an: Obwohl Balzer und Beywl sich vorgenommen haben, den gesamten Bildungsbereich abzudecken, ist das Buch doch vor allem auf diesen speziellen Bildungssektor ausgerichtet. Das merkt man deutlich an den Beispielen, aber auch an fehlenden Hinweisen auf die speziellen institutionellen Bedingungen in anderen Sektoren wie Schule oder Hochschule. Das widerspricht aber nicht unbedingt der Eignung des Buches für eine breitere Verwendung – außer man pocht auf eine sektorspezifische Einführung. Die Autoren selbst sehen auch „benachbarte Felder“, wie z.B. soziale Arbeit, Gesundheitsbereich, Konsumentenschutz, Beratung, Coaching, als in ihrem Buch angesprochen, da auch dort „gelernt“ wird. Dem ist zuzustimmen: die inhaltlichen Grundlagen passen prinzipiell auf jede Form von Bildungsprogramm (und mit ein wenig Flexibilität lassen sich auch die Beispiele in andere Bildungskontexte ‚übersetzen‘). Dazu gehören aus Sicht der Rezensentin auch alle Evaluationsgegenstände/-felder, in denen der „Programmarchetyp ‚Beratung, Information und Bildung‘“ (Funnell/Rogers 2011: 352ff.) umgesetzt wird – das Buch kann damit auch über den eigentlichen Bildungssektor hinaus Anwendung finden.

4. Fazit

Das Buch bietet eine theoretisch fundierte, aber gleichzeitig praxisnahe Einführung in die Evaluation für Studierende, ist aber gleichermaßen für

3 Nach wie vor gibt es ja einige ‚Evaluationslehrbücher‘, in denen es (fast) ausschließlich um (vornehmlich quantitative) Erhebungsmethoden geht.

4 Die Rezensentin hat selbst schon an Fallbeispielen und Übungen aus Vorversionen des Buches für Informationsangebote oder Schulungen zu Evaluation Anleihen genommen – mit gutem Erfolg bei den Zielgruppen (siehe z.B. <http://www.sqa.at/mod/page/view.php?id=525>, 01.07.2016).

interessierte Praktiker(innen) sehr gut geeignet und auch für Lehrende als Grundlagentext und/oder Inspiration für die eigene Lehrtätigkeit interessant. Es kann als erste Einführung empfohlen werden – oder auch, da es eigene, neue Akzente setzt, als ergänzender Text zu bereits vorhandenen Einführungswerken.

5. Literatur

- Balzer, Lars (2005): *Wie werden Evaluationsprojekte erfolgreich?* Landau: Verlag Empirische Pädagogik.
- Beywl, Wolfgang/Kehr, Jochen/Mäder, Susanne/Niestroj, Melanie (2008): *Evaluation Schritt für Schritt: Planung von Evaluationen.* Münster: hiba.
- Beywl, Wolfgang/Schepp-Winter, Ellen (2000): *Zielgeführte Evaluation von Programmen: Ein Leitfaden.* Berlin: Bundesministerium für Familie, Senioren, Frauen und Jugend, Referat 501.
- Funnell, Sue C./Rogers, Patricia (2011): *Purposeful Program Theory: Effective Use of Theories of Change and Logic Models.* San Francisco, CA: Jossey-Bass.
- Keller-Ebert, Cornelia/Kiöbler, Mechthilde/Schober, Berthold (2005): *Evaluation praktisch! Wirkungen überprüfen, Maßnahmen optimieren, Berichtsqualität verbessern.* Darmstadt: hiba.
- Kirkpatrick, Donald/Kirkpatrick, James D. (2006): *Evaluating Training Programs: The Four Levels.* San Francisco, CA: Berrett-Koehler (3. Aufl.).
- Weiss, Carol H. (1998): *Evaluation – Methods for Studying Programs and Policies.* Upper Saddle River, N.J.: Prentice Hall (2. Aufl.).
- Glossar der Evaluation. Verfügbar unter: http://eval-wiki.org/glossar/Eval-Wiki:_Glossar_der_Evaluation [01.07.2016].