

Rezension zu:

Baur, Nina/Kelle, Udo/Kuckartz, Udo (Hg.): *Mixed Methods*. Kölner Zeitschrift für Soziologie und Sozialpsychologie. Sonderheft 57. Wiesbaden: Springer VS, 2017. 481 Seiten, 49,99 EUR, ISBN 978-3-658-19803-9

Marcus Neureiter¹

Mit dem Sonderheft „Mixed Methods“ der *Kölner Zeitschrift für Soziologie und Sozialpsychologie* haben sich die Herausgeber(innen) zum Ziel gesetzt, sowohl die internationale Diskussion um Mixed Methods systematisch aufzugreifen als auch anhand von Beispielen aus empirischen Arbeiten deutschsprachiger Forscher einen Einblick in verschiedene Herausforderungen und Ansätze der Mixed-Methods-Forschung zu geben. Der Band verspricht gerade für die Evaluation – in deren Praxis Mixed-Methods-Ansätze zwar einerseits weitgehend etabliert sind, der jedoch auf der anderen Seite eine systematische methodologische Reflexion und eine gemeinsame Sprache über Mixed Methods (zumindest im deutschsprachigen Raum) weitgehend fehlen – wertvolle Inspiration.

Gegliedert ist das Sonderheft in vier größere Abschnitte und insgesamt 19 Einzelbeiträge von Autor(inn)en aus der internationalen „Mixed-Methods-Community“ einerseits und unterschiedlichen Feldern der deutschsprachigen empirischen Sozialforschung andererseits. In einer Einleitung zum Stand der Debatte und aktuellen Problemlagen machen die Herausgeber(innen) deutlich, dass es – anders als im deutschsprachigen Raum – international inzwischen eine Reihe von Lehrbüchern, umfangreichen Handbüchern, eine eigene Fachgesellschaft und eine eigene Fachzeitschrift für Mixed Methods gibt. International gesehen sei, so die Herausgeber(innen), inzwischen „ein gewisser Grad der wissenschaftlichen Reife erreicht, der sich unter anderem in gefestigten Konzepten und Terminologien niederschlägt“ (S. 31). Die aktuell zentralen Diskussionsstränge innerhalb der Mixed-Methods-Community werden in dem Beitrag pointiert anhand beispiel-

hafter Fragestellungen dargestellt. Aus dieser Darstellung wird der Aufbau des Sonderbandes nachvollziehbar abgeleitet.

Der erste Abschnitt des Sonderheftes, der neben deutschsprachigen auch englischsprachige Beiträge aus der internationalen Mixed-Methods-Community umfasst, beschäftigt sich mit den philosophischen, methodologischen und methodischen Grundlagen von Mixed-Methods-Forschung. Die Herausgeber(innen) möchten damit auf Entwicklungen und Erkenntnisse der internationalen Mixed-Methods-Forschung aufmerksam machen, die im deutschsprachigen Raum bisher kaum rezipiert wurden. Die unterschiedlichen Beiträge machen deutlich, dass es zu vielen Aspekten von Mixed Methods inzwischen zumindest auf internationaler Ebene einen breiten und fundierten Forschungsstand sowie Leitlinien und Anleitungen für die Praxis gibt.

Daran anschließend umfassen die weiteren drei Abschnitte des Sonderheftes Anwendungsbeispiele aus einer Fülle unterschiedlicher Forschungsfelder, die in drei Bereiche methodischer und methodologischer Herausforderungen gegliedert sind: *Designs and Sampling; Formen der Datenkombination und Fragen der Validität; Prozesse, Längsschnittanalyse und Evaluation*. Im Folgenden soll beispielhaft auf einige der Beiträge näher eingegangen werden, die in besonderer Weise Anregungen für die Evaluationspraxis versprechen.

In ihrem Beitrag „How to construct a Mixed Methods Research Design“ gehen Schoonenboom und Burke Johnson auf die sieben primären sowie auf zehn sekundäre Dimensionen ein, welche für die Entwicklung von Mixed-Methods-Designs berücksichtigt und abgewogen werden müssen. Die sieben primär-

¹ Ramboll Management Consulting GmbH, Politikberatung und Evaluation, Berlin

ren Dimensionen (Untersuchungsziel, Rolle von Theorie im Forschungsprozess, Timing, Integrations-Schnittstellen, Systematischer vs. interaktiver Design-Ansatz, geplante vs. emergente Designs, Komplexität des Designs) werden in dem Artikel ausführlich diskutiert. Die Diskussion und die Liste der Dimensionen als Checkliste und Leitfaden können Forschende und Evaluator(inn)en bei der Entwicklung von Mixed-Method-Evaluierungsdesigns unterstützen und dazu anregen, Designentscheidungen systematisch zu hinterfragen und besser zu begründen.

Mit der Bedeutung der Stichprobenziehung und Fallauswahl für die interpretative Konsistenz in Mixed-Methods-Designs beschäftigen sich Onwuegbuzie und Collins in ihrem Beitrag „The role of Sampling in Mixed Methods-Research. Enhancing Inference Quality“. Mit interpretativer Konsistenz ist die Frage gemeint, welche Art von Generalisierung bei gegebener Stichprobenziehung und Fallauswahl gerechtfertigt ist. Diese Frage, so die Autoren, sei in Mixed-Methods-Designs deutlich komplexer als in rein qualitativen oder quantitativen Designs, weil beide methodischen Ansätze darin berücksichtigt werden müssen. Sie unterscheiden daher sechs Generalisierungstypen, von der externen (statistischen) Generalisierung über die analytische Generalisierung bis hin zur pragmatischen Alltagsgeneralisierung. Jede dieser Arten von Generalisierung, so die Autoren, sei nur dann gerechtfertigt, wenn eine interpretative Konsistenz im Hinblick auf dieses Generalisierungsniveau erreicht wurde. Onwuegbuzie und Collins schlagen sechs Gesichtspunkte vor, die bei Stichprobenziehung und Fallauswahl (und in allen weiteren Forschungsphasen) berücksichtigt werden sollten, um interpretative Konsistenz zu erreichen und das Generalisierungsniveau der Ergebnisse rechtfertigen zu können. Dazu zählen unter anderem die *Inside-Outside-Orientierung*, die *probabilistische Orientierung* und die *abduktive Orientierung*. Solche Gesichtspunkte können auch als Leitlinien für Evaluator(inn)en dienen, wenn es um die Frage der Verallgemeinerbarkeit ihrer Ergebnisse geht. Umgekehrt können sie Anregungen für Mixed-Method-Evaluationsdesigns und Stichprobenstrategien liefern, wenn das gewünschte Verallgemeinerungsniveau feststeht. Um eine wirklich gute Leitlinie für die Praxis zu sein, sollte das Konzept allerdings noch weiter ausgearbeitet werden, wie die Autoren es für die Zukunft auch ankündigen.

Ein anschaulicher und für Evaluator(inn)en sehr relevanter Beitrag ist Kuckartz' Artikel zum Thema „Datenanalyse in der Mixed-

Methods-Forschung. Strategien der Integration von qualitativen und quantitativen Daten und Ergebnissen“. Ausgehend von der Feststellung, dass dem Thema „Integration bei der Datenanalyse“ in der Mixed-Methods-Diskussion bislang vergleichsweise wenig Aufmerksamkeit gewidmet wurde, und dass Mixed-Methods-Studien relativ selten qualitative und quantitative *Daten* (statt Ergebnisse) integrieren, fokussiert sein Beitrag auf die Bestimmungsmomente von Integrationsstrategien sowie die Schnittstellen der Methodenintegration. Er beschreibt zehn verschiedene Integrationsstrategien, die sich in resultatsbasierte, datenbasierte und sequenzorientierte Integrationsstrategien unterteilen lassen. Er argumentiert, dass im Wesentlichen drei Determinanten dafür relevant sind, welche Integrationsstrategien verfolgt werden. Neben der Begründung für die Wahl des Mixed-Methods-Ansatzes sind dies das Forschungsdesign und die Sampling-Strategie: Die hier getroffenen Entscheidungen haben einen großen Einfluss darauf, welche Schnittstellen der Integration zielführend und sinnvoll sind. Am Beispiel der zehn Integrationsstrategien geht Kuckartz auch auf das Potenzial von sogenannten ‚Joint Displays‘ für die Methodenintegration ein: Unter diesem Label werden in der jüngeren Mixed-Methods-Diskussion Darstellungsformen diskutiert, welche Daten und/oder Ergebnisse des qualitativen und quantitativen Forschungsstrangs in einer gemeinsamen Darstellung zusammenfassen. Durch die verstärkte Nutzung von Joint Displays könnte sowohl die integrative Analyse, als auch die Darstellung von Evaluationsergebnissen aus Mixed-Methods-Evaluationen verbessert und für die jeweilige (Fach-)Öffentlichkeit verständlicher und nachvollziehbarer gemacht werden. Interessant wäre gewesen, wenn der Autor an dieser Stelle noch einen Schritt weitergegangen wäre und die Konsequenzen der verschiedenen Integrationsstrategien für die ausführliche Ergebnisdarstellung stärker diskutiert hätte. Deutlich wird durch Kuckartz' Argumentation in jedem Fall, dass eine systematische Methodenintegration kaum ohne entsprechende Software-Pakete wie MAXQDA oder NVivo bewältigbar scheint. Diese Pakete, so der Autor, seien in den vergangenen Jahren um spezielle Funktionen zur Unterstützung von Mixed-Methods-Ansätzen erweitert worden.

Neben diesen eher einführenden und überblicksartigen Beiträgen, gehen in den letzten drei Abschnitten des Sonderbandes verschiedene Autor(inn)en anhand eigener Beispiele näher auf spezifische Herausforderungen in der Forschungs- und Evaluationspraxis ein. Hier er-

scheinen die folgenden Beiträge in besonderem Maße inspirierend für Evaluator(inn)en:

Der Beitrag „Integriertes multidimensionales Design“ von Haunss, Schmidtke und Biegon zeigt beispielsweise anschaulich an einer Fragestellung aus der Politikwissenschaft, wie ein Mixed-Methods-Design dabei helfen kann, mehrdimensionale Fragestellungen zu beantworten – eine für Evaluator(inn)en durchaus vertraute Herausforderung.

Hense gibt in ihrem Beitrag „Sequentielles Mixed-Method-Sampling: Wie quantitative Sekundärdaten qualitative Stichprobenpläne und theoretisches Sampling unterstützen können“ einen detaillierten – und in dieser Ausführlichkeit seltenen – Einblick in die empirisch informierte Bildung von Kontrastgruppen für eine qualitative Studie auf Basis einer quantitativen Sekundärdatenanalyse.

Vogl vertieft in ihrem Beitrag „Quantifizierung“ anhand konkreter Beispiele und Zielstellungen (Komplementarität, Systematische Vergleiche und Verständniserweiterung) eine der zehn Integrationsstrategien aus Kuckartz' Artikel. Dabei reflektiert sie ihre jeweiligen Quantifizierungen und deren Verallgemeinerbarkeit systematisch und weist auf dieser Basis nachdrücklich auf die Gefahr der Überinterpretation quantifizierter qualitativer Daten hin.

Langfeldt und Goltz beschäftigen sich in dem Artikel „Die Funktion qualitativer Vorstudien bei der Entwicklung standardisierter Erhebungsinstrumente“ mit dem Mehrwert eines sequenziellen Mixed-Methods-Designs für die Entwicklung eines praxistauglichen Fragebogens im Fall von „Strukturen begrenzter Reichweite“ (S. 317), wie sie im Kontext von Evaluationen häufig sind. Ihr Beitrag verdeutlicht jedoch auch, wie aufwändig solche sequenziellen Designs sind. Nicht in jedem praktischen Evaluationskontext ist dies umsetzbar.

Die Herausforderungen, Chancen und Strategien für Mixed-Methods-Studien, in denen sehr verschiedene Datentypen – und damit auch Forschungstraditionen – miteinander kombiniert

werden, behandeln Baur und Hering in ihrem Beitrag „Die Kombination von ethnografischer Beobachtung und standardisierter Befragung“ anhand zweier Beispielstudien. Dabei diskutieren sie auch, wie damit umgegangen werden kann, wenn beim Vergleich der Ergebnisse Divergenzen festgestellt werden: Welchen Ergebnissen ist eher zu glauben und wie kann dies begründet werden? Die diskutierten Kriterien können auch als Anregung für Evaluator(inn)en dienen, die mit einer ähnlichen Herausforderung konfrontiert sind.

Mayring zeigt in dem Beitrag „Evidenztriangulation in der Gesundheitsforschung“ auf, wie eine Mixed-Methods-Evaluation der Effektivität einer Drogenambulanz aussehen kann. Dafür wurden quantitativ-experimentelle Ansätze mit teilstrukturierten qualitativen Interviews, einer inhaltsanalytischen Dokumentenanalyse und Formen der Selbstevaluation kombiniert. Interessant ist auch sein Vorschlag für eine „dialektische Evidenztriangulation“ der Ergebnisse mehrerer Untersuchungsbausteine, den er mit einem Beispiel veranschaulicht.

Zusammenfassend ist der Band „Mixed Methods“ ein sehr gelungener Beitrag zur Methodendiskussion in Deutschland, der dazu beitragen kann, die Gräben zwischen den unterschiedlichen methodologischen Paradigmen zu überbrücken. Viele der in dem Band enthaltenen Beiträge und die Zusammenstellung des Bandes insgesamt beinhalten vielfältige Inspirationen, Leitlinien und Anleitungen, die auch für die Evaluationspraxis nutzbar gemacht werden können und diese bereichern dürften. So werden Anregungen für methodische Mindeststandards, Gütekriterien und Ablaufmodelle in Mixed-Methods-Evaluationsdesigns gegeben, wie sie unter anderem im Rahmen der Frühjahrstagung 2016 des AK Methoden gefordert wurden.² Der Band könnte ein guter Ausgangspunkt und Anlass dafür sein, diese für die Evaluationspraxis weiterzuentwickeln und noch stärker in die Aus- und Weiterbildung von Evaluator(inn)en zu integrieren.

2 Vgl. Jacobs, Lucas/Schendzielorz, Jessica/Giesen-Thole, Dorothea (2016): Zur Qualität qualitativer Sozialforschung in der Evaluation – Möglichkeiten und Grenzen. Frühjahrstagung des AK „Methoden in der Evaluation“ der DeGEval in Bonn. In: Zeitschrift für Evaluation, 15 (2), S. 358.

Rezension zu:

Befani, Barbara: *Pathways to Change: Evaluating Development Interventions with Qualitative Comparative Analysis (QCA)*. 2016. 242 Seiten, ISBN 978-91-88143-16-7

Verfügbar unter: http://eba.se/wp-content/uploads/2016/07/QCA_BarbaraBefani-201605.pdf [03.01.2018]

*Lisa Ringhofer*¹

The EBA report “Pathways to Change: Evaluating Development Interventions with Qualitative Comparative Analysis (QCA)”, as the title suggests, presents the potential and pitfalls of one specific method for development evaluation: Qualitative Comparative Analysis (QCA). First introduced in 1987, QCA is a method for systematic cross-case comparison to understand which qualitative factors are likely to influence an outcome. Qualitative data is translated into a numerical format and systematically analysed in order to detect any causal patterns in the data and allow for causal claims to be tested without the need of a counterfactual situation. QCA is therefore considered “*at the crossroad*” of the qualitative and quantitative culture, incorporating the “*best of both worlds*” (p. 21)².

The 185-page report (plus 57 pages of annexes and glossary) covers different aspects of QCA, from trying to define its place within the range of useful methods for development evaluation to offering an eight-step how-to handbook built on real-world cases. For its intended audience – evaluators and commissioners of evaluations – it also considers a checklist on how to quality-assure such evaluations and avoid stepping on a territory densely mined with pitfalls and traps.

Overview

The report consists of three main chapters. The **opening chapter** introduces the method and is divided into two parts. The first part (sections 1.1. and 1.2.) attempts to locate QCA within a

broader “*methodological map*” (p. 7) of different rigorous methods for development (impact) evaluations. Part two of this chapter discusses the potential, relevance and limitations of QCA for development evaluation. This is particularly interesting for its use to develop and test a programme theory. Using the fictitious example of evaluating a policy influence programme implemented in eight countries to improve evidence-based policy-making in the health sector, QCA tests the different programme nested theories of change (first and second level) with the aim to identify the conditions³ that need to be in place in order for stakeholders to reach consensus. In sum, QCA can help the evaluator make recommendations on which types of interventions to prioritise in future interventions. This chapter also contains some useful information for commissioners when applying QCA in evaluation practice, providing guidance on when it might be wise to consider it in a Terms of Reference for an evaluation project.

Chapter two illustrates the sequence of eight practical steps that should be taken in a high-quality application of QCA. Each of these steps is described in terms of opportunity (what the step is useful for), challenges and pitfalls to look out for, or more generically the “issues at stake”. The sequence of these steps however is not carved in stone since the order of steps may vary or alternatively not all of these are strictly required. The author argues that while the first three steps (i.e. model specification, ensuring data availability, and calibration) are nec-

1 TripleMinds Consultancy Network, Vienna

2 Historically, the method has always been very popular with political scientists and other scholars interested in cross-country generalisation.

3 The term ‘conditions’ is used with QCA rather than ‘variables’ to emphasise the distinction between QCA and statistics.

essary for a QCA “proper”, the last five steps (i.e. the Venn diagram, the SuperSubset Analysis, the Truth Table, the Boolean minimisation, and the INUS analysis) may be considered optional. However, the author argues “*it is advisable to make the most of the opportunities offered by the approach and complete all the steps*” (p. 49).

Chapter three addresses some crosscutting reflections that are relevant for quality assurance at the multiple stages of the QCA application process. This includes the issues of generalisation, bias, and the dialogue with theory. Since QCA’s ability of generalisation – its key advantage compared to the more mainstream qualitative within-case methods – is seen with some concern by evaluators with a quantitative background, some authors have assigned it the attribute of “*modest*” or “*limited generalisation*” (e.g. Berg-Schlösser/De Meur/Rihoux/Ragin 2009). The chapter further illustrates how QCA can draw on other evaluation approaches (like Contribution Analysis, Realist Evaluation and Process Tracing) to help make sense of the output generated by the QCA analysis. As the demand for inclusion of QCA in development evaluation (or at least some of its components) rises, a final section proposes a quality assurance checklist for good practice. The 19 listed ideas aim to ensure that the potential offered by the method is fully exploited and the pitfalls evaluators can run into are avoided.

The report also entails three **Annexes**. Annex A lists a range of causal frameworks underpinning scientific inference (with a focus on Impact Evaluation). This helps relate QCA to other evaluation methods using different frameworks. Annex B discusses the differences between QCA in relation to Regression Analysis and Annex C comprises some background information about the evaluation cases used in the report.

Discussion

The report is clearly structured and generally well written (with some inconsistencies and misspellings especially in chapter three) and efforts were made to simplify the technical language, such as hyperlinks to a glossary and other sections for non-everyday use terms.

As regards content, the overarching question to be answered by QCA – “*What sets of factors are likely to influence an outcome?*”

– is highly relevant in the development context. The report tackles this question in various breadth and depth through the use of either fictitious, stylised or real-life evaluation examples. The fictitious evaluation, for instance, deals with evidence-based policy evaluation of health systems in eight developing countries. It seeks to understand which conditions have led policy makers to ground their decisions on evidence when legislating on access to the health system for the poor. The three real-life evaluations mentioned for different illustrations in the report include an example of assessing the performance of mobile phones in repairing rural water points (“*Which conditions facilitated the achievement of different outcomes?*”), a gender-sensitive budget support evaluation (“*Which of two policy instruments works best?*”), and a biodiversity evaluation (“*Which conditions facilitate the creation of protected areas?*”). Equally so, QCA can be combined with other, more familiar evaluation methods in development to take on board the perspectives of stakeholders about “*what differences matter and for whom?*”. Among these methods are the Most Significant Change (Davies/Dart 2005) and Outcome Mapping (Earl/Carden/Smutylo 2001).

In sum, QCA offers a refreshing alternative to the widespread ‘counterfactual thinking’ (with Randomized Control Trials hailed as the gold standard) as the only feasible strategy to demonstrate causal connections when assessing the impact of interventions. Since QCA marries the depth of qualitative information (it retains the ability to depict complexity) with the rigour of quantitative methods (it allows for replications), the method also contributes to shortening the distance between qualitative and quantitative methods, sometimes referred to as a divide⁴.

The paper makes a convincing case of why QCA should be an important addition to the development evaluator’s toolkit and has certainly whetted my appetite for applying QCA in evaluation practice. For this to materialise, however, the following four aspects need to be in place: (1) the existence of a theory of change (or several nested theories) to guide the initial selection of conditions and outcomes; (2) the availability of data for all conditions across the cases that are compared; (3) the need for technical and sectoral skills among the evaluation team; and (4) a flexible budget frame, since costs are

4 Despite its name and despite being a case-based method, QCA is not always considered ‘qualitative’, particularly in the academic traditions of some Latin cultures which translate it “Quali-Quantitative Comparative Analysis” (Meur/Rihoux/Yamasaki 2002) because of its mathematical basis.

difficult to calculate in advance (it is not known beforehand how many iterations are needed). In my experience, none of these aspects can be taken for granted in a development programme context. Hence, while QCA's main protocols are well tried and tested in academic research, I agree with the author's statement that the method may "*still be getting its feet wet in evaluation*" (p. 44).

References

- Berg-Schlosser, Dirk/De Meur, Gisèle/Rihoux, Benoît/Ragin, Charles (2009): Qualitative Comparative Analysis (QCA) As An Approach. In: Rihoux, Benoît/Ragin, Charles (Eds.): *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques*. Thousand Oaks: Sage.
- Davies, Rick/Dart, Jess (2005): *The 'Most Significant Change' (MSC) Technique: A Guide to Its Use*.
- Earl, Sarah/Carden, Fred/Smutylo, Terry (2001): *Outcome Mapping: Building Learning and Reflection into Development Programs*. Ottawa: International Development Research Centre.
- Meur, Gisèle/Rihoux, Benoît/Yamasaki, Sakura (2002): *L'Analyse Qualiquantitative Comparée (AQQC-QCA): Approche, Techniques et Applications en Sciences Humaines*. Louvain-la-Neuve: Academia-Bruylant.

Rezension zu:

Vettori, Oliver/Salmhofer, Gudrun/Mitterauer, Lukas/Ledermüller, Karl (Hg.): Eine Frage der Wirksamkeit? Qualitätsmanagement als Impulsgeber für Veränderung an Hochschulen. Bielefeld: UVW UniversitätsVerlagWebler, 2015. 268 Seiten, 43,20 EUR, ISBN 978-3-946017-00-4

*Gianpiero Favella*¹

Es ist die Einschätzung, dass die „Frage nach der Wirkung und Wirksamkeit von Qualitätsmanagement und Qualitätssicherung zunehmend an Bedeutung“ (S. 3) gewinnt, die Befunde zu diesem Themenfeld dringlich machen. Der Band von *Oliver Vettori, Gudrun Salmhofer, Lukas Mitterauer* und *Karl Ledermüller* geht auf eine Tagung zurück, die 2014 vom Netzwerk Qualitätsmanagement und Qualitätsentwicklung an der Wirtschaftsuniversität in Wien gehalten wurde. Im ersten Beitrag stellen die Herausgeber(innen) ein analytisches Modell dar, welches in vier Dimensionen gegliedert ist: Dimension der Perspektivität, Dimension der organisationsspezifischen Rahmenbedingungen, Dimension der disziplinären Zugangsweisen und Dimension der Bewertung. Derart aufgestellt bildet das Modell ein Instrumentarium, um aufzeigen zu können, dass Qualitätsmanagement und dessen Wirksamkeit vom jeweiligen Betrachtungsstandpunkt abhängig ist.

Aus kritischer Perspektive betrachtet *Benjamin Ditzel* zunächst Umsetzungsvarianten des Qualitätsmanagements. Als Anlass für eine organisations- und managementtheoretische Reflexion dient dem Autor die Einschätzung, dass im Qualitätsmanagementdiskurs wenig auf die Erkenntnisse der Organisations- und Managementforschung Bezug genommen wird. Durch die systemtheoretische Perspektivierung auf Qualität ist es gerade die Beobachtung einer dezentralen Struktur von Qualitätssicherungsprozessen und Qualitätsentwicklungsprozessen, die Anlass für eine dezentrale Kontextsteuerung geben. Welche Grenzen und Möglichkeiten von Qualitätssicherung an der Universität Rostock

bestehen, führt *Dorit Sorge* aus. Es wird dabei der Versuch unternommen, aus einer kommunikationswissenschaftlichen Perspektive auf organisationsstrukturelle Probleme zu antworten. Herausgestellt wird die Aufgabe eines Qualitätsbeauftragten an der „Schnittstelle zwischen einer symmetrisch organisierten Fakultät und einem vom Landeshochschulgesetz asymmetrisch gedeuteten Rektorat“ (S. 48). Mit der Beobachtung, einerseits trage die Drittmittelfinanzierung zur wettbewerbsorientierten Forschungsuniversität bei und andererseits seien die Ressourcen je Studierende bei steigender Studienzahl gesunken, stellen *Jan Andrä* und *Nora Krzywinski* kulturelle und wirtschaftliche „Herausforderungen einer wirksamen Forschungsevaluation“ (S. 52) dar. Dies nehmen die Autor(inn)en zum Anlass, universitätsübergreifende Standards als Kriterium dafür zu identifizieren, um die Wirksamkeit von Forschungsevaluation durch Begutachtungsrahmen, Gutachter(innen)qualität, Implementierung und Verfahrensqualität und einen Treiber einer Forschungsevaluation zu erhöhen. In ihrer explorativ-empirisch angelegten Studie stellen *Markus Seyfried* und *Alexa Kristin Brase* die Frage, „wie das Qualitätsmanagement zum individuellen wie organisationalen Lernen in Hochschulen beitragen kann“ (S. 68). Mittels einer Heuristik, die zwischen Single-Loop-Lernen – einem instrumentellen Lernen – und Double-Loop-Lernen² – einem wertverändernden Lernen – unterscheidet, befunden die Autor(inn)en Bedingungen für das organisationale Lernen. Auf der Grundlage einer empirischen Evaluation des Akkreditierungssystems deutscher Hochschulen untersucht *Manuel Pietzonka* Wirkmechanis-

1 TU Kaiserslautern

2 Vgl. Argyris, Chris/ Schön, Donald A. (1999): Die lernende Organisation: Grundlagen, Methode, Praxis. Stuttgart: Klett-Cotta.

men der Akkreditierung und illustriert zunächst das Evaluationsdesign sowie das Vorgehen, welches auf einer Dokumentenanalyse der Akkreditierungsaufgaben, einer Online-Befragung von Hochschulmitarbeiter(inne)n und einer Interviewerhebung von Fachschaftsvertreter(inne)n basiert. Bevor der Autor aus der empirischen Studie Empfehlungen für das Akkreditierungssystem ableitet, werden „Mängel und Schwächen des Akkreditierungssystems“ (S. 91) identifiziert, die auf Inkonsistenzen im Entscheidungsprozess und auf ein Überwachungsverständnis des Akkreditierungsverfahrens zurückzuführen sind. Aus Sicht einer Qualitätsberaterin zeigt *Doris Carstensen* auf, wie eine Rollenklärung und ein Promotorenmodell im Handeln von Qualitätsmanager(inne)n Orientierung geben kann. Dazu beschreibt die Autorin zunächst die Funktionen des QMs, die in der „strategischen Vorgabe“ (S.105), der „generelle[n] Umsetzung“ (ebd.), im „Operationalisieren der Ziele in den Teilsystemen“ (ebd.) und in der „Befähigung und Motivation“ (ebd.) der Mitarbeiter(innen) gründen. Während die Rollenklärung die Erwartungen der QM-Einheiten formuliert und gleichzeitig zu Aushandlungsprozessen anstiftet, beinhaltet das Promotorenmodell eine gestaltende Komponente bei Veränderung bzw. Innovationen. *Alexa Kristin Brase* und *Moritz Ansmann* befinden zunächst aus dem Stand der Forschung zur Qualitätssicherung und -entwicklung, dass die Frage, „ob und inwiefern die Bemühungen um Qualitätssicherung Wirkungen entfalten, bislang wenig adressiert worden“ (S. 117) sei. Wie die Akteurinnen und Akteure den Plan-Do-Check-Act-Zyklus – kurz: ‚PDCA‘ – nutzen, ist ein Aspekt, der in 14 leitfadengestützten Experten(innen)interviews mit QM-Personal sowie Vizepräsident(inn)en thematisch ist. Welche Grenzen und welche Möglichkeiten sich aus einer Lernerzentrierung herausstellen, zeigt *Wolfgang Sucharowski* am Beispiel zweier Studienvorhaben „Lernen durch Projektarbeit“ und „Vorlesung besser verstehen“ an der Universität Rostock auf. Formuliert wird dabei eine Antwort auf die Frage, „ob die lernerzentrierte Lehre ein Qualitätskriterium“ (S. 143) sei und wie diese die Lehre verbessern kann. *Christoph Burger*, *Michaela Pirker*, *Evelyn Maria Bergsmann* und *Petra Winter* illustrieren eine Strategie, wie Qualitätssicherung für eine kompetenzorientierte Lehre vorgenommen werden kann. Hierzu verknüpfen die Autor(inn)en am Beispiel der Vetmeduni Vienna ein kaskadiertes Vorgehen der Evaluation mit einem Prozessmodell kompetenzorientierter Lehre mit dem Ziel, „Vermittlung der Kompetenzen“ (S. 159) zu fördern

und den „Kompetenzerwerb der Studierenden im Verlauf des Studiums zu überwachen“ (ebd.). Wie qualitätsorientiertes Prüfen im Bereich der darstellenden Künste umgesetzt wird, beschreibt *Ester Tomasi-Fumics* im Rückgriff ihrer Erfahrungen in der Arbeitsgruppe Assessment und Standards des Projekts ‚Polifonia‘, welches vom Verband der europäischen Musikhochschulen koordiniert wird. Ihrer Meinung nach sind es insbesondere zwei Fragen, die für die Qualität bedeutsam sind: Inwieweit trägt die Institution Sorge für die Organisation der Prüfung und welche Kriterien und Standards für die Beurteilung der Leistungen sind explizit formuliert? *Karl Ledermüller*, *Michaela Nettekoven* und *Maria Weiler* entwickeln ein Analyseinstrument, ein Reporting-System, für Multiple-Choice-Prüfungen. Dieses System trägt dazu bei, einen Feedbackbericht zu „deskriptive[n] Aussagen über MC-Prüfungen als Ganzes wie auch über die einzelnen Prüfungsfragen“ (S. 188) zu generieren. Am Beispiel der TU Darmstadt untersucht *Tina Klug*, ob Institutionelle Evaluationen eine strukturverändernde Wirkung auf Wissenschaftsorganisationen haben. Als Untersuchungsauslöser dient der Autorin die kritische Einschätzung Institutioneller Evaluation, da deren Ergebnisse „oft in Zielvereinbarungen einfließen, die zwischen politischer Instanz oder Organisationsleitung und evaluierter Einheit“ (S. 194) ausgehandelt werden. *Kalle Hauss*, *Olaf Ratzlaff* und *Frank Niedermeier* schlagen ein Berichtssystem auf der Basis von Indikatoren vor, die sich aus zwei Datengrundlagen herauspeisen und zugleich aufeinander bezogen werden: Studienverlaufsstatistik und Studierenden-Panel. Umgesetzt wird diese Strategie in der Studieneingangsphase anhand eines Vergleichs zwischen Lehramtsstudierenden und Nicht-Lehramtsstudierenden. Standortgebunden für die Wirtschaftsuniversität Wien skizzieren *Julia Zeeh* und *Lukas Dünser* zunächst den Forschungsstand von Studierendenverläufen mit besonderer Berücksichtigung des methodischen Vorgehens. Indem die Clusteranalyse und die Cox-Regression begründet ausgeschlossen werden, zeigen die Autor(inn)en im Rahmen von zwei Teilstudien die Vorteile von n-Gramm – ein Verfahren aus der Computerlinguistik, um Studienverlaufsmuster zu identifizieren – und von Strukturgleichungsmodellen – um auf diese Weise Studien(in)aktivitäten zu befunden – für die Evaluierung von Studierendenaktivitäten auf. Welche Faktoren die Drittmittelwerbung befördern oder begrenzen, zeigen *Marina Zeldovich*, *Christine Zeiller*, *Ursula Leiter-Köhler* und *Andreas Raggautz* exemplarisch für die Universität Graz auf. Mit dieser Frage-

stellung stoßen die Autor(inn)en auf ein bisher wenig bearbeitetes (Forschungs-)Feld wissenschaftlicher Leistungserfassung. Die methodische Anlage besteht in einem Strukturgleichungsmodell, welches „die Möglichkeit bietet, theoretisch begründete ausgewählte Faktorenstrukturen zu bestätigen bzw. zu verwerfen“ (S. 242).

Es ist beeindruckend, wie der Band das komplexe Thema von Wirkungen und Wirksamkeit des Qualitätsmanagements in Hochschulen in knapp zweihundertfünfzig Seiten zusammentragen kann. Wenngleich „die Wirkungsanalysen letztlich doch in der Minderzahl sind“ (S. 16), wird hierdurch zugleich auf ein rar gesätes Forschungsfeld aufmerksam gemacht.

Für den Überblick und die Verortung der Beiträge hätte sich durchaus noch angeboten, die einzelnen Beiträge entlang des zu Beginn des Bandes einleitenden – überzeugenden – analytischen Modells zu ordnen. Durch die zu Beginn dargelegte Dimensionierung von Wirksamkeit können die Autor(inn)en auf Forschungslücken hinweisen und demzufolge Forschungsperspektiven eröffnen. Es ist dem Band zu wünschen, dass er breit respektive disziplinübergreifend gelesen wird – und dies gerade deshalb, um erkunden zu können, welche Konsequenzen sich aus der Überschreibung managerieller Eigenlogiken in und von Hochschulen ergeben.

Rezension zu:

Greve, Bent (Ed.): *Handbook of Social Policy Evaluation*. Cheltenham, UK: Edward Elgar Publishing, 2017. 542 Seiten, £ 175,50, ISBN 978-1-78536-323-8

Frans L. Leeuw¹

The “Handbook of Social Policy Evaluation” counts 542 pages and 26 chapters. It is an interesting and important book, but it also is a bit fragmented. Many topics are covered, dealing with examples of social policies, methods and designs of evaluation, ethical issues and interest issues. However, the Handbook does not address questions like where social policy evaluation currently stands after many decades of work, whether or not evaluations have contributed to the growth of knowledge in this field and which challenges lie ahead, now that societies are confronted with data-ification, digitization and a whole range of other developments that directly or indirectly influence social policies (like the role of social neurosciences in developing and evaluating interventions).

First, the positive aspects of the volume. Next to different designs of evaluation like Randomized Controlled Trials and Quasi-Experimental Designs that both are a pleasure to read, one can find chapters on participatory evaluation, systematic reviews-approaches, cost-benefit analysis and other economic evaluations of social policies. Also, interventions like social impact bonds, labour market interventions, poverty interventions, social integration and family policies are discussed. Several chapters discuss utilization and the impact of evaluations, including evidence-based policy making and the uses and misuses of evaluation in social policies. The chapter on (10) critical perspectives on using evidence in social policy should also be mentioned here.

In the final chapter, the editor, Bent Greve, summarizes his three ‘core’ perspectives of the book: “possible ethical issues related to the use

of evidence; technical issues in undertaking research and analysis and the issue of impact of pressure and interest groups on the gathering of knowledge and the possible interpretation of that knowledge” (p. 516). This threesome offers something of a framework of the book.

Also positive is that the volume contributes to the idea that evaluation is an important aspect of society. In fact, Greve starts his final chapter with such a statement: “Without doubt, evaluating and use of evidence are here to stay”. I will come back to this point, but let me first raise several critical points.

– *Fragmentation*. Although it may be an old-fashioned idea, Handbooks should in my opinion be more than a collection of chapters. Why a chapter on participatory evaluation, but not on theory-driven evaluations? Why a chapter on performance management but not on performance auditing? Why chapters on health (policies/organizations) but not on delinquency or social isolation? Why neglecting digital policies and programs, and big data analytics? The editor’s final words in chapter 26 on the three ‘inter-linked issues’, although interesting as such, do not address this problem in depth. I also raise this point from an educational perspective: students, junior evaluators and junior faculty, commissioners of evaluation may be in need of *some guidance* when doing and/or organizing evaluations. Think of questions like which evaluation problems to tackle first and how, how to deal with (the numerous) types and subtypes of ‘evaluation theories’ that can be found in the literature

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

and to some extent resemble Rolf Klima's (1972) 'pseudo-pluralism' in sociology (Leeuw/Donaldson 2015); how to carry out systematic reviews that also build on the realist evaluator's Context-Mechanisms-Outcomes perspective and how to incorporate Big Data in evaluations? The contributions of the book to these and similar questions are not available, or not very clear.

- *Added value of the book.* Related to the fragmentation issue is the question of the added value of the volume. For sure, there is added value, given its focus on *social policies* and the book's *accessibility*. But focusing more on innovative policies and programs would have increased this value. Think in terms of what social neurosciences/brain and cognition-studies have to offer to developing and evaluating new social policies (for example regarding loneliness, social isolation, resilience of people, attachment, behavior modification of juvenile delinquents, drop-out problems of pupils etc. (Cacioppo/Cacioppo 2013); see also: <https://www.hersenenen-cognitie.nl/contents/1109?locale=en>). And think in terms of the role of Big Data for social policy evaluations (Bamberger 2016; Petersson/Breul 2017).

Finally I would like to point to a few *missed opportunities*. One concerns the issues of growth of knowledge: given the fact that social policies have existed for decades as have evaluations (see the USA's Golden Years when Johnson's Great Society was evaluated), it would have made a fascinating chapter summarizing which (behavioral/other) mechanisms made certain policies/interventions (more) successful than others, taking into consideration the contexts in which they operate. A second one concerns the interesting question Calman/Douglas raise in their chapter on evaluating new medicines: "Whether evaluative activity of medicines and healthcare is any different from any other social policy arena?" (p. 319). Although the authors' sociological analysis on how in different countries (including Cuba!) medicine evaluations are carried out and in which contexts is interesting, the more fundamental question, on which Ray Pawson currently is working, unfortunately remains unanswered. In line with Pawson's work-in-progress, at least one fundamental difference can be mentioned: the timeline for medical-pharmaceutical development and evaluation is often 10 to 15 years and the one for social policy making and evaluation often less than 3 years. A third and last critical point refers to the chapter on systematic reviews in social policy

evaluation. Although it is written in a coherent and elegant way, and it also pays attention to the debate between Systematic Review researchers and Realist synthesis-oriented scholars, the missed opportunity is that Realists not only are working on guidelines how to review & synthesize research (http://www.ramesesproject.org/Home_Page.php) but also that several authors have developed and used approaches combining Systematic Reviews à la Campbell and realist evaluations (e.g. Van der Knaap/Leeuw/Bogaerts/Nijssen 2008; Vaessen et al. 2014). It would have been wise to present these and other examples to the readers, showing them that it is not a yes or no issue.

References

- Bamberger, Michael (2016): Integrating Big Data Into the Monitoring and Evaluation of Development Programmes. New York: Global Pulse/Rockefeller Foundation.
- Cacioppo, John T./Cacioppo, Stephanie (2013): Social Neuroscience. In: Perspectives on Psychological Science, 8 (6), pp. 667-669.
- Klima, Rolf (1972): Theoretical Pluralism, Methodological Dissension and the Role of the Sociologist: The West German Case. In: Social Science Information, 11 (3-4), pp. 69-108.
- Leeuw, Frans L./Donaldson, Stewart (2015): Theory in Evaluation: Reducing Confusion and Encouraging Debate. In: Evaluation, 21 (4), pp. 467-480.
- Petersson, Gustav J./Breul, Jonathan D. (Eds.) (2017): Cyber Society, Big Data and Evaluation. Comparative Policy Evaluation, Volume 24. New Brunswick: Transaction Publishers.
- Vaessen, Jos/Rivas, Ana/Duvendack, Maren/Palmer Jones, Richard/Leeuw, Frans L./Van Gils, Ger/Lukach, Ruslan/Holvoet, Nathalie/Bastiaensen, Johan/Garcia Hombrados, Jorge/Waddington, Hugh (2014): The Effects of Microcredit on Women's Control over Household Spending in Developing Countries: A Systematic Review and Meta-analysis. In: Campbell Systematic Reviews, 8 (1), pp. 1-205.
- Van der Knaap, Leontien/Leeuw, Frans L./Bogaerts, Stefan/Nijssen, Laura T. J. (2008): Combining Campbell Standards and the Realist Evaluation Approach. The Best of two Worlds? In: American Journal of Evaluation, 29 (1), pp. 48-57.