



Reinhard Stockmann, Wolfgang Meyer (Hrsg.)

Die Zukunft der Evaluation

Trends, Herausforderungen, Perspektiven

Sozialwissenschaftliche Evaluationsforschung

Herausgegeben von
Reinhard Stockmann

Band 13

Reinhard Stockmann, Wolfgang Meyer
(Hrsg.)

Die Zukunft der Evaluation

Trends, Herausforderungen, Perspektiven



Waxmann 2017
Münster • New York

Bibliografische Informationen der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.dnb.de> abrufbar.

Print-ISBN 978-3-8309-3708-1

E-Book-ISBN 978-3-8309-8708-6

First published in English by Palgrave Macmillan, a division of Macmillan Publishers Limited under the title *The Future of Evaluation* by R. Stockmann and W. Meyer. This edition only includes selected chapters from the above work and has been translated and published under license from Palgrave Macmillan. The authors have asserted their right to be identified as the author of this work.

© Waxmann Verlag GmbH, 2017

www.waxmann.com

info@waxmann.com

Umschlaggestaltung: Pleßmann Design, Ascheberg

Umschlagbild: Ulrich Thul, Ludwigshafen

Satz: Stoddart Satz- und Layoutservice, Münster

Druck: Hubert & Co., Göttingen

Gedruckt auf alterungsbeständigem Papier,
säurefrei gemäß ISO 9706



Printed in Germany

Alle Rechte vorbehalten. Nachdruck, auch auszugsweise, verboten.

Kein Teil dieses Werkes darf ohne schriftliche Genehmigung des Verlages in irgendeiner Form reproduziert oder unter Verwendung elektronischer Systeme verarbeitet, vervielfältigt oder verbreitet werden.

Inhalt

<i>Reinhard Stockmann und Wolfgang Meyer</i> Vorwort.....	7
<i>Reinhard Stockmann und Wolfgang Meyer</i> Einleitung	9
<i>Peter Dahler-Larsen</i> Die sich verändernde Rolle der Evaluation in einer sich verändernden Gesellschaft	21
<i>Reinhard Stockmann</i> Die Zukunft der Evaluation in modernen Wissensgesellschaften	35
<i>Pablo Rodríguez-Bilella und María Alejandra Lucero</i> Evaluation als globales Phänomen: Die Entwicklung transnationaler Netzwerke.....	53
<i>André Martinuzzi und Wolfgang Meyer</i> Evaluation nachhaltiger Entwicklung in einer globalen Gesellschaft	71
<i>Jan-Eric Furubo</i> Evaluation – ihr Erbe und ihre Zukunft	99
<i>Verena Friedrich</i> Evaluationsstudiengänge an europäischen Hochschulen: Merkmale und Herausforderungen	111
<i>Christine Nolte und Sandra Schopper</i> Moderne Lehr- und Lernmethoden in der Evaluation	125
<i>Frans L. Leeuw</i> Cyber Society und die „Verdschungelung“ von Regeln. Neue Herausforderungen für die Evaluation im 21. Jahrhundert	141
<i>Reinhard Stockmann und Wolfgang Meyer</i> Evaluation zwischen Wissenschaft und Praxis.....	155
<i>Wolfgang Meyer, Evelyn Funk und P. Linh Nguyen</i> Partizipation und Bewertung: Weiterbestehende methodologische Herausforderungen	177

Christoph Emanuel Müller
Kausale Wirkungsevaluation zwischen methodischem Anspruch
und empirischer Praxis..... 205

Brad Astbury
Von der Evaluationstheorie zum Testen von Evaluationstheorien? 223

Wolfgang Meyer und Reinhard Stockmann
Schlussfolgerungen: Gemeinsame Perspektiven für die
Institutionalisierung von Evaluation?..... 243

Vorwort

Jedes Buch hat seine Entstehungsgeschichte. Die des vorliegenden beginnt 2002 mit der Gründung des Centrums für Evaluation (CEval) an der Universität des Saarlandes. Es wurde aus der Not geboren und seine Zukunft war mehr als ungewiss. Nachdem die Soziologie als Fach (gemeinsam mit der Volks- und Politikwissenschaft) abgeschafft worden war, entstanden Freiräume, die wir nutzten. In der Lehre schufen wir für den abgewickelten Studiengang Soziologie einen neuen, den Master für Evaluation und für die Forschung gründeten wir das Centrum für Evaluation, dessen zehnjähriges Jubiläum wir zum Anlass nahmen, nicht nur zurück, sondern vor allem nach vorne – in die Zukunft – zu blicken. Den Geburtstag feierten wir mit einer internationalen Konferenz, die so viele wegweisende Beiträge erbrachte, dass wir uns entschlossen, diese in einem Buch mit dem Titel „The Future of Evaluation: Global Trends, New Challenges, Shared Perspectives“ bei Palgrave Macmillan (2016) zu publizieren. Nachdem sich das Buch einer hohen Nachfrage und großer internationaler Aufmerksamkeit erfreut, kamen wir auf die Idee, das in diesem Jahr anstehende Jubiläum – 20 Jahre DeGEval – sowie das aktuelle Tagungsmotto „Evaluation (in) der Zukunft“ zu nutzen, um erneut einen Blick in die Zukunft zu wagen.

Das vorliegende Werk folgt dem Konzept der englischsprachigen Vorlage und fokussiert auf die zukünftige Entwicklung der Evaluation in der Gesellschaft, die daraus resultierenden Herausforderungen für Forschung und Lehre sowie für die Evaluationstheorien und -methoden in der Zukunft. Hierfür wurden einige der Beiträge aus dem Englischen ins Deutsche übertragen, aber auch einige Artikel zusätzlich geschrieben. Deshalb gebührt unser Dank nicht nur den Autoren und Autorinnen dieses Bandes, sondern auch den Übersetzern und Übersetzerinnen, die feststellen mussten, dass es gar nicht so einfach ist, einen für sie fremden englischsprachigen Text in ein lesbares Deutsch zu übertragen. Wir danken deshalb besonders Gareth Bartley, Friederike von Cölln, Angelika Nentwig, Christine Nolte, Selina Röhrig, Tanja Stockmann und Sophie Wannemacher für ihre Übersetzungsleistungen.

Damit Texte zu einem Buch werden, braucht es fleißige Helfer und Helferinnen, die sich um Formatierungen, Grafiken, Korrekturen etc. kümmern. Hierfür danken wir Fabian Otto und vor allem Selina Röhrig, die dabei nie die Übersicht verlor.

Damit ein Buch auch publiziert wird, braucht es zudem einen Verlag. Deshalb danken wir Ursula Heckel dafür, dass sie diesen Band in die Reihe „Sozialwissenschaftliche Evaluationsforschung“ aufgenommen hat und ganz herzlich auch an dieser Stelle für die jahrelange, stets erfolgreiche und immer erfreuliche Zusammenarbeit.

Bürstadt und Saarbrücken im Juni 2017

Reinhard Stockmann und Wolfgang Meyer

Einleitung

Anhand einer Vielzahl von Belegen lässt sich zeigen, dass Evaluation eine globale Erfolgsgeschichte darstellt. Ihr Anwendungsbereich reicht von der Rechenschaftslegung über den Einbezug in Planungs- und Durchführungsprozesse von Projekten und Programmen bis hin zu reflexiven Einschätzungen gesellschaftlicher Prozesse. Man wird kaum einen anderen wissenschaftlichen Ansatz finden, der in so kurzer Zeit einen ähnlich erfolgreichen Diffusionsprozess vorzuweisen hat. Daran konnte auch die Vielfalt weltweiter und regionaler Konflikte, Krisen, die globalen ungelösten, ökologischen, sozialen und ökonomischen Probleme sowie die Fundamentalkritik am westlichen Konsummodell als Leitbild einer modernen Entwicklung (vgl. Stockmann, Menzel & Nuscheler 2016) nichts ändern. Allerdings stellen diese „turbulent times“, wie Furubo, Rist und Speer (2013) sie nennen, die Evaluation vor neue Herausforderungen, sodass nicht automatisch davon ausgegangen werden kann, dass sich diese Erfolgsgeschichte einfach bruchlos fortsetzen wird. Deshalb wird in diesem Buch Fragen nachgegangen, wie: Gibt es aus praktischer oder theoretischer Perspektive Herausforderungen, die diesen Prozess schwächen, stoppen oder sogar umkehren könnten? Gibt es einen gemeinsamen globalen Pfad der Evaluation oder wird die kulturelle Vielfalt zu einer Vielfalt von unterschiedlichen Richtungen führen? Auch wenn die Zukunft der Evaluation sicherlich nicht vorhersehbar ist, so lassen sich doch einige gute Argumente für zukünftige Trends, Herausforderungen und gemeinsame Perspektiven finden.

Der Ausgangspunkt für das Abenteuer, die Zukunft der Evaluation zu erforschen, ist ihre Rolle in der Gesellschaft. Auf der Grundlage der Annahme, dass soziale Strukturen gezielt verändert und geformt werden können, stellt Evaluation ein zentrales Instrument für die Planung, Implementation und Wirkungsmessung von Politiken, Strategien, Programmen, Projekten und Maßnahmen dar, die sowohl von Regierungs- als auch Nichtregierungsorganisationen durchgeführt werden. Infolgedessen kann Evaluation gesellschaftliche Entwicklungen beeinflussen, so wie sie umgekehrt auch selbst durch solche Entwicklungen beeinflusst wird.

Wie bereits erwähnt, erfolgt Evaluation durch die Anwendung einer breiten Palette von Grundkonzepten mit einer beträchtlichen Anzahl unterschiedlicher Methoden für viele und verschiedenartige Zwecke. Angefangen mit einer Reihe von englischsprachigen Ländern, vor allem den USA und einigen wenigen europäischen Pionieren, hat sich die Idee der Evaluation im Zuge der sozialen Modernisierungsprozesse rund um den Globus verbreitet. Das Schlüsselmerkmal dieser Entwicklung ist die Überzeugung, dass soziale und politische Prozesse verwaltet und gesteuert werden können.

Allerdings ist der Einsatz und die Verbreitung der Evaluation an einige Rahmenbedingungen geknüpft. In korrupten, repressiven und / oder diktatorischen Regimen, in denen jegliche Form von Kritik, Transparenz und Offenheit verhindert wird, kann Evaluation nicht gedeihen. Ebenso wenig kann Evaluation in (staatlichen und nicht staatlichen) Organisationen oder Gruppen, in denen konstruktive Kritik unmöglich ist, wirk-

lich als Instrument für rationale Entscheidungsfindung oder Qualitätsverbesserung genutzt werden. Die Bereitschaft und die Fähigkeit, aus Fehlern zu lernen, konstruktive Kritik zuzulassen und Evaluation als Lernressource zu nutzen, sind notwendige, aber nicht ausreichende Bedingungen, um sicherzustellen, dass die durch die Evaluation offenbaren Defizite und Verbesserungspotentiale nicht in Schuldzuweisungen enden, sondern konstruktiv für die Reform von Strukturen und Prozessen oder Verhaltensänderungen genutzt werden. Die Evaluation braucht Freiräume, um sich zu entwickeln, wenn sie nicht zu einem starren Kontrollinstrument entarten soll.

Diese Voraussetzungen finden sich in einer wachsenden Zahl von Ländern, so dass Evaluation nicht nur zu einem „booming business“ (Leeuw 2009: 3) geworden ist, sondern auch immer mehr Lebensbereiche umfasst. Doch dieses quantitative Wachstum sagt uns natürlich nichts über die Qualität von Evaluation oder über deren Nutzung. Darüber, wie es darum bestellt ist, gibt es keine Einstimmigkeit, auch nicht unter den Autorinnen und Autoren dieses Buches. Unumstritten ist jedoch, dass der weltweite Boom der Evaluation und die damit einhergehende Professionalisierung empirisch aufgezeigt werden können (vgl. Furubo, Rist & Sandahl 2002; Fitzpatrick, Sanders & Worthen 2004: 50f.; Dahler-Larsen 2006: 141ff., Smith et al. 2011; Smith & Brandon 2011: 566; Jacob, Speer & Furubo 2015). Hierfür werden zwar eine Vielzahl unterschiedlicher Indikatoren verwendet, ohne sie jedoch theoretisch zu begründen oder ihre Auswahl zumindest explizit zu machen.

Dies soll hier zumindest insoweit erfolgen, dass für die Beschreibung der Institutionalisierung der Evaluation bewusst drei zentrale Perspektiven eingenommen werden:

Erstens werden die Rolle der Evaluation innerhalb der Gesellschaft und die damit verbundenen Aufgaben und Funktionen dieser Form angewandter Sozialforschung thematisiert. Zweitens stehen die Verankerung der Evaluation innerhalb des Wissenschaftssystems und damit der Beitrag zur Aus- und Weiterbildung von Evaluatorinnen und Evaluatoren sowie die Forschung über Evaluation im Fokus der Betrachtung. Im dritten und letzten Teil schließlich werden die Theorien und Methoden der Evaluation ins Zentrum gerückt und die damit verbundenen Spezifika von Evaluation, die diese von anderen wissenschaftlichen Konzepten und praktischen Anwendungen unterscheiden. Die verschiedenen Autorinnen und Autoren haben dabei auf Grundlage vergangener Entwicklungstendenzen einen Blick in die Zukunft gewagt und verschiedene Möglichkeiten einer Weiterentwicklung der Evaluation zur Diskussion gestellt.

1. Die Institutionalisierung von Evaluation in der Gesellschaft

Eine wichtige Voraussetzung für die Entwicklung eines Themengebietes zu einer Profession ist die Nachfrage nach den Produkten und Dienstleistungen, die diese Disziplin bereitstellt. Für die Evaluation ist zu beobachten, dass sie in den letzten Jahrzehnten zu einem Schlüsselkonzept für gesellschaftliche Entwicklung avanciert ist. Auch wenn dieser Trend differenziert nach Ländern und Politikfeldern in unterschiedlichem Tempo voranschreitet, so ist doch unstrittig, dass es übergreifende Ursachen dafür gibt. Ein wesentlicher Motor für diesen Evaluationsboom hat mit dem Bestreben öffentlicher Verwaltungen zu tun, ihre Dienstleistungen kundenfreundlicher und effizienter anzubieten.

Der Transfer von Qualitätsmanagementideen aus dem privaten Sektor in die öffentliche Verwaltung hat weltweit zu Prozessen der Verwaltungsmodernisierung geführt. Neue Konzepte wie New Public Management, Performance Management oder Wirkungsorientierte Verwaltungsführung erfordern neue Steuerungsinformationen, die durch die Instrumente des Monitorings und der Evaluation generiert werden können.

Ein weiterer Motor für den Antrieb des Evaluationsbooms ist ebenfalls im Bereich des staatlichen Handelns zu sehen. In Zeiten knapper werdender finanzieller Ressourcen steigt der Rechtfertigungsdruck auf die Politik. Sie muss nicht nur öffentliche Güter und Dienstleistungen möglichst effektiv und effizient bereitstellen, sondern ihre Prioritätensetzungen und Selektionsentscheidungen legitimieren. Auch hierfür wird Evaluation mittlerweile weltweit eingesetzt. In einer Reihe von Ländern ist die National Evaluation Policy legislativ geregelt (z.B. in den USA, Kanada, Kolumbien, Mexiko, Schweiz, Österreich) oder es existieren elaborierte Richtlinien zur Evaluation (England, Finnland, Dänemark, Norwegen, Schweden, Niederlande, Australien, Neuseeland usw.) (vgl. Rosenstein 2013). Auch viele internationale Zusammenschlüsse wie die OECD, die Europäische Union, die Weltbank und die Vereinten Nationen verlangen, dass alle von ihnen finanzierten Programme evaluiert werden.

Zivilgesellschaftliche Organisationen treten bisher weniger als Nachfrager von Evaluation auf. Viele Nichtregierungsorganisationen nutzen das Potential von Monitoring und Evaluation für die Steigerung der Effektivität, Effizienz und Wirksamkeit ihrer Arbeit noch nicht ausreichend. Eine starke Rolle spielen zivilgesellschaftliche Organisationen hingegen bei der Verbreitung und Promotion von Evaluation. Dennoch steckt im zivilgesellschaftlichen Bereich für Evaluation noch ein ungeheures, bisher kaum genutztes Potential, um Demokratisierungsprozesse und partizipative Beteiligungsverfahren zu unterstützen.

Welche Rolle Evaluation in der Gesellschaft spielen kann und wie sich die Funktionen und Aufgaben von Evaluation in der Zukunft verändern werden, sind die Themen dieses Kapitels.

Peter Dahler-Larsen untersucht die sich verändernde Rolle der Evaluation in einer sich verändernden Gesellschaft aus einer konstruktivistischen Perspektive. Die Definition von Evaluation sowie ihre Durchführung werden als Ergebnisse sozialer Konstruktion begriffen, welche selbst wiederum dazu beitragen, gesellschaftliche Phänomene zu konstruieren. In dieser Perspektive sind Evaluation und ihre Entwicklung in die gegenwärtigen und zukünftigen Gesellschaftsprozesse bzw. ihre Konstruktionen eingebettet.

Der Wandel der Evaluation im Laufe der Zeit wird als Ergebnis eines breiten Spektrums gesellschaftlicher Kräfte angesehen. Der Autor analysiert fünf Themen: Popularisierung, Systematisierung der Evaluation, Wirkungsorientierung, sich verändernde Beziehungen zwischen Evaluation und Forschung sowie Veränderung von Nutzungsmustern. Dahler-Larsen gibt zu bedenken, dass die Popularisierung von Evaluation und die Integration von Evaluation in die verschiedensten Lebensprozesse und organisatorischen Routinen zwar zu einem weitverbreiteten Bestandteil moderner Gesellschaften geworden ist, an das hohe Erwartungen geknüpft sind, dass der Evaluation aber gerade deshalb auch die Gefahr droht, zu bürokratischen „Evaluationsmaschinen“ (evaluation machines) zu verkrusten.

Die Modernisierungsprozesse der öffentlichen Verwaltung, die sich durch eine Abkehr von der kameralistischen Steuerung und eine Hinwendung zur Wirkungsorientierung auszeichnen, haben die Evaluationsentwicklung stark belebt. Dahler-Larsen hinterfragt das diesem Modell zugrundeliegende Kausalitätsverständnis und sieht eine zukünftige Aufgabe von Evaluierenden darin, komplexere und dynamischere Formen des Denkens über Kausalität und Verantwortung zu definieren.

Abschließend behandelt Dahler-Larsen die sich verändernde Beziehung zwischen Evaluation und Forschung und die sich verändernden Nutzungsmuster von Evaluation. Die oft beklagte Tatsache, dass Evaluation nicht genutzt wird, sieht er durch neue Trends weitgehend überwunden.

Reinhard Stockmann folgt in seinem Beitrag ebenfalls der Vorstellung, dass die Entwicklung der Evaluation stark von der Entwicklung der Gesellschaft beeinflusst wird. Da Evaluation größtenteils aus staatlichen oder nicht-staatlichen Aufträgen resultiert, ist sie weitgehend abhängig von Auftraggebern und ihren Interessen. Diese fokussieren auf die Fragen, die sie im jeweiligen gesellschaftlichen Kontext für am relevantesten halten. Um Vorhersagen über die Zukunft der Evaluation zu machen, so die These des Autors, ist es notwendig, zu wissen, was die Auftraggeber von Evaluationen in Zukunft interessieren wird.

Ausgehend von der Prämisse, dass sich Evaluation im Gefolge der Entstehung moderner Gesellschaften entwickelt hat, werden drei Funktionen der Evaluation unterschieden, die in modernen und postmodernen (nachhaltigkeitsorientierten) Gesellschaften mit der Fähigkeit zur Selbstreflexion existieren: (1) Auf der Mikroebene der Gesellschaft: die Managementfunktion, (2) auf der Mesoebene: die Legitimitäts- oder Rechenschaftsfunktion und (3) auf der Makroebene: die gesellschaftliche Aufklärungsfunktion. Der Autor untersucht in seinem Beitrag, inwieweit die gesellschaftliche Nachfrage nach Evaluation die Bedeutung der drei Evaluationsfunktionen verändert. Er kommt zu dem Schluss, dass die Managementfunktion bislang am weitesten verbreitet ist, und die Etablierung von wirkungsorientierten Managementkonzepten und Qualitätsmanagementansätzen, die Organisationen als lernende Akteure betrachten, dazu führen wird, dass diese Funktion der Evaluation ihre Signifikanz in Zukunft behalten oder sogar ausbauen kann. Die Bedeutung der Legitimitäts- oder Rechenschaftsfunktion von Evaluation hat sich im Laufe der Zeit ebenfalls erheblich erhöht. Dafür, dass dieser Trend auch in Zukunft anhalten wird, spricht die anhaltende Knappheit finanzieller Ressourcen. In solchen Zeiten steigt der öffentliche Druck, Entscheidungen zu legitimieren.

Die Aufklärungsfunktion von Evaluation ist nach wie vor am wenigsten entwickelt, obwohl gerade diese in reflexiven, auf Nachhaltigkeit ausgerichteten Gesellschaften am bedeutendsten ist. Dies liegt nach Auffassung des Autors u.a. an dem mangelnden politischen Willen und dem fehlenden gesellschaftlichen Druck, die dafür notwendigen, unabhängigen Evaluationseinrichtungen zu schaffen, die frei von administrativen Vorgaben und ausgestattet mit ausreichenden finanziellen Mitteln, agieren können. Um auf zukünftige Herausforderungen vorbereitet zu sein, wird empfohlen, den weiteren Professionalisierungs- und Kapazitätsaufbau von Evaluation international voranzutreiben.

Während die Entwicklung der Nachfrage nach Evaluation bisher vor allem staatlich generiert ist, erfolgt der Ausbau der Evaluationskapazitäten genauso wie die Entwicklung transnationaler Evaluationsnetzwerke vor allem über zivilgesellschaftliche Orga-

nisationen. Der Beitrag von *Pablo Rodríguez-Bilella* und *María Alejandra Lucero* über „Evaluation als globales Phänomen“ fokussiert auf „die Entwicklung transnationaler Netzwerke“.

Neben den nationalen Evaluationsgesellschaften, die in den letzten beiden Jahrzehnten von fünf im Jahr 1995 explosionsartig auf fast 120 (2014) angewachsen sind, hat sich auch die Zahl der transnationalen Evaluationsnetzwerke spaltpilzartig vergrößert. Neben kontinentalen Evaluationsgesellschaften (z.B. AES, EES, AfrEA, EvalMENA, IPEN, CoE, APEA, ReLAC)¹ erfreuen sich zunehmend auch große nationale Evaluationsgesellschaften einer internationalen Mitgliedschaft (z.B. die AEA). Daneben sind eine Reihe von Organisationen entstanden, die zum Kapazitätsaufbau von Evaluation in Entwicklungsländern beitragen sollen, wie z.B. die International Organization for Cooperation in Evaluation (IOCE) oder die International Development Evaluation Association (IDEAS). Hinzuzuzählen wäre EvalPartners, die keine eigene Organisation darstellt, sondern eine internationale Kooperationspartnerschaft. Weiterhin gibt es Zusammenschlüsse, die bestimmte Evaluationsthematiken bearbeiten, z.B. das Network of Networks for Impact Evaluation (NONIE) und die International Initiative for Impact Evaluation (3ie). In dem Beitrag werden die Aufgabenstellungen und Angebote der Transnational Evaluation Networks vorgestellt und zukünftige Entwicklungschancen und -herausforderungen thematisiert.

Mittlerweile gibt es eine solche Vielfalt an transnationalen Organisationen und Netzwerken, dass diese zu einer wachsenden Unübersichtlichkeit, Zersplitterung und Duplizierung von Aktivitäten führt, die die schlagkräftige Bündelung von Aufgaben und Ressourcen beeinträchtigen könnte. Eine Situation, die nach Ansicht der Autoren bisher jedoch noch nicht eingetreten ist. Im Gegenteil, sie konstatieren eher Prozesse der Integration und Konvergenz. Damit dies auch in Zukunft so bleibt, sind sicherlich sorgfältige Abstimmungsprozesse und Profilschärfungen notwendig.

André Martinuzzi und *Wolfgang Meyer* befassen sich mit einem brandaktuellen Thema. Vor dem Hintergrund der im Jahr 2015 verabschiedeten Sustainable Development Goals (SDGs) stellt sich ihre Frage nach der „Evaluation nachhaltiger Entwicklung in einer globalen Gesellschaft“ mit besonderer Dringlichkeit. Die Autoren möchten aus europäischer Perspektive einen Überblick zu allen Aktivitäten des Monitorings und der Evaluation von nachhaltiger Entwicklung geben. Im Unterschied zu den, den SDGs vorangegangenen Millennium Development Goals (MDGs: 2000 – 2015) beziehen sich diese nicht nur auf die Entwicklungsländer in einigen ausgewählten Politikfeldern, sondern auf alle Länder in der gesamten Breite der Politik. Ebenfalls anders als die MDGs sollen die Umsetzungserfolge nicht nur gemonitort, sondern auch evaluiert werden. Dies stellt die Evaluation vor große Herausforderungen, da die 17 SDG-Ziele und 169 Zielvorgaben, die mit insgesamt 231 Indikatoren messbar gemacht werden sollen, ein äußerst komplexes Gebilde darstellen.

Die Autoren kommen zu dem Ergebnis, dass die bisher existierenden Monitoring- und Evaluationssysteme viel zu schwach entwickelt sind, um diesen neuen Anforderungen zu genügen und formulieren deshalb Zukunftsaufgaben zur Weiterentwicklung dieser Systeme. Dabei versuchen sie herauszufinden, ob es überhaupt ein gemeinsames

1 Zu den Abkürzungen siehe den Beitrag der genannten Autoren.

Monitoring- und Evaluationsverständnis gibt, ob Institutionen existieren, die eine Führungsrolle in der Entwicklung eines globalen Konzepts für das Monitoring und die Evaluation nachhaltiger Entwicklung übernehmen könnten und ob Faktoren vorhanden sind, die eine solche Entwicklung fördern oder hemmen könnten.

Wie schon die Autorinnen und Autoren vor ihm, geht auch *Jan-Eric Furubo* von der Feststellung aus, dass Evaluation allgegenwärtig ist – aber im Unterschied zu diesen bemängelt er, dass Evaluation kaum kritisch hinterfragt werde. Dies möchte er in seinem Beitrag nachholen, in dem er aufzeigt, dass sich die gesellschaftlichen Rahmenbedingungen, unter denen sich die Evaluation entwickelte, heute komplett geändert haben. Dabei postuliert er ein inkrementelles Zeitalter, in dem sich die Evaluation hauptsächlich mit der Verbesserung bereits existierender Programme beschäftigte. Dies kam der Politik – so Furubo – sehr gelegen, da die schrittweise Veränderung vorhandener Strategien einfacher zu bewerkstelligen ist, als einen grundlegenden politischen Wandel herbeizuführen. Heutzutage sei eine solche Politik nicht mehr möglich, da fundamentale gesellschaftliche Veränderungen unabdingbar geworden seien. In diesem nicht-inkrementellen Zeitalter werde auch die Evaluation vor neue Herausforderungen gestellt. Zum einen benötigten die Entscheidungsträger ein umfassenderes Wissen, zudem müssten sich Evaluationsergebnisse an eine breitere Öffentlichkeit als bisher richten.

2. Die Institutionalisierung von Evaluation im Wissenschaftssystem

Zu den Eigenschaften von Professionen gehört es, dass sie Elemente des Wissenschaftssystems sind und als solche in die universitäre Praxis von Forschung und Lehre eingebunden sind. Die Etablierung eines neuen Fachs kann dabei auf verschiedene Arten vorstattengehen. Es kann sich beispielsweise als Ausdifferenzierung einer bestehenden Disziplin zunächst als eigenständiges Teilelement etablieren und so z.B. den Status eines Schwerpunktfachs erhalten. Forschungsarbeiten werden in den bestehenden Medien dieser Disziplin publiziert und diskutiert. Selbst wenn neue Zeitschriften, Arbeitsgruppen, Kommunikationsforen usw. entstehen, werden diese ausschließlich von Expertinnen und Experten mit einer entsprechenden Fachausbildung der Ursprungsdisziplin getragen. Dieser Weg kann schließlich zu einer Abspaltung und eigenständigen Profilierung mit entsprechenden Ausbildungsgängen und Forschungseinrichtungen führen. Die Entstehung der Soziologie als Abspaltung von der Volkswirtschaftslehre mag als Beispiel für eine solche Entwicklung dienen.

Der Weg der Evaluation an den nordamerikanischen Universitäten folgte allerdings einem leicht abweichenden Muster: Evaluation hat sich nicht nur in einem, sondern praktisch parallel in zwei unterschiedlichen Fächern (Erziehungswissenschaften und Psychologie) verankert und später sind noch weitere Sozialwissenschaften hinzugekommen. Evaluation ist keine Spezialisierung eines Fachs, sondern ein aus der Praxis entstandenes neues interdisziplinäres Anwendungsfeld. Hieraus ergeben sich bei der Institutionalisierung im Wissenschaftssystem verschiedene Schwierigkeiten, die von den Autorinnen und Autoren aus unterschiedlichen Blickwinkeln beleuchtet werden.

Hier knüpft der Beitrag von *Verena Friedrich* an, die mit der Frage beginnt, wie man eigentlich Evaluatorin bzw. Evaluator wird. Der Besuch evaluationsspezifischer Aus- und

Weiterbildungen ist hierfür keineswegs selbstverständlich oder gar obligatorisch. Die Autorin hat ermittelt, dass es in Europa 13 Evaluationsstudiengänge gibt, die sich im Hinblick auf Lernziele und -inhalte, Dauer, Intensität und Preis erheblich voneinander unterscheiden. Die meisten Studiengänge befinden sich in Süd- und Westeuropa. Nord- und Osteuropa erscheinen unterrepräsentiert. Zudem gibt es in dem Angebot eine erhebliche institutionelle Fluktuation.

Obwohl die Studiengänge einen wichtigen Beitrag zur Professionalisierung der Evaluation leisten, müssen die akademischen Programme nach wie vor eine Reihe von Schwierigkeiten überwinden. Da Evaluation aufgrund ihrer Nähe zur Praxis häufig auf Weiterbildungsstufe unterrichtet wird, ist sie kostspielig und steht in der Regel nicht im Fokus der akademischen Lehre. Wird Evaluation im Rahmen grundständiger sozialwissenschaftlicher Studiengänge unterrichtet, dann wird sie häufig nur als methodisches Thema abgehandelt. Dadurch werden Themen, die in der Evaluationspraxis wichtig sind, wie die besonderen Bedingungen bei der Durchführung von Evaluationen, Umgang mit Stakeholdern, partizipative Ansätze oder das Berichtswesen, nicht behandelt. Da die Evaluation keine klare disziplinäre Zugehörigkeit aufweist, hat sie oft auch keine stabile nachhaltige Verankerung in den Hochschulen. Dies hat darüber hinaus zur Folge, dass die Forschung über Evaluation zu kurz kommt. Nach wie vor gilt, dass die meisten Evaluierenden nicht in einem Studiengang Evaluation, sondern in anderen Fächern ausgebildet werden. Unter professionspolitischen Gesichtspunkten ein bedenklicher Befund.

Der Beitrag von *Christine Nolte* und *Sandra Schopper* schließt an die Ausführungen von Verena Friedrich an, denn er beschäftigt sich ebenfalls mit dem Aus- und Weiterbildungsangebot in der Evaluation, allerdings aus dem Blickwinkel moderner Lehr- und Lernmethoden. Konkret geht es um das Potential von E-Learning und Blended Learning. Da Evaluation vor allem ein Weiterbildungsfach ist, eignen sich Blended-Learning-Formen ganz besonders für das berufsbegleitende Training, denn sie ermöglichen ein zeitlich und räumlich flexibles, individuelles Lernen. Da Blended Learning eine Zusammensetzung von Präsenz- und Online-Lernen darstellt, werden die didaktischen Vorteile beider Modalitäten genutzt. Im Unterschied zu reinem Online-Lernen schaffen die Präsenzanteile im Blended-Learning-Format für die gerade bei der Evaluation so wichtigen Themen wie Teambuilding, das Einüben von Konfliktlösungen in Rollenspielen, den persönlichen Umgang mit Stakeholdern oder die Präsentation von Evaluationsergebnissen die für die Einübung notwendigen Kontextbedingungen.

Ein besonders wichtiger Teil des Beitrags ist die Aufstellung über universitäre und außeruniversitäre Aus- und Weiterbildungsangebote sowie über E-Learning-Materialien. Die Übersicht zeigt, dass es weltweit nur wenige, in Europa gar keine, Blended-Learning-Angebote gibt. Dies soll sich ab dem Wintersemester 2017/18 ändern. Ab diesem Zeitpunkt wird gemeinsam von der Universität des Saarlandes und dem Fernstudienzentrum DISC (Distant and Independent Study Center) der Technischen Universität Kaiserslautern ein „Master of Evaluation“ (MABLE) in englischer Sprache im Blended-Learning-Format angeboten, der auf den Inhalten des Präsenz-Masterstudiengangs der Universität des Saarlandes und der Hochschule für Technik und Wirtschaft basiert. Die gemeinsam mit der Deutschen Gesellschaft für Internationale Zusammenarbeit (giz) entwickelten Online-Module (SCORM-Pakete) liegen auch Angeboten in Entwicklungs-

ländern zugrunde, sodass sich hier hervorragende Kooperations- und Netzwerkchancen bieten, die in den nächsten Jahren ausgebaut werden sollen. Ziel ist, einen interkulturellen und regional unabhängigen Kern gemeinsamen Evaluationsverständnisses zu entwickeln, der jeweils durch länder- und kulturspezifische Besonderheiten angereichert wird.

Ebenfalls in die Zukunft gerichtet ist der Beitrag von *Frans Leeuw*, der auf zwei Entwicklungen aufmerksam macht, deren Konsequenzen für die Evaluation bisher erst wenig diskutiert wurden. Dabei handelt es sich zum einen um das „neue“ Leben in der „web-society“, zum anderen um das mit der Globalisierung sich ausbreitende Phänomen der „Transnational Governance“.

The Digitalisierung der Gesellschaft ist eigentlich kein brandneues Phänomen, allenfalls die zunehmende Geschwindigkeit mit der sie das persönliche Leben, aber auch die Gesellschaft durchdringt. Angesichts der sich ausbreitenden virtuellen Lebenswelt, die sich u.a. in sozialen Netzwerken, Informationsportalen für alle erdenklichen Dienstleistungen, Märkten für jegliche Arten von Produkten, Datenbanken für alle Fragen des wissenschaftlichen und persönlichen Lebens etc. erstreckt, ist es eher erstaunlich, dass die Herausforderungen, die damit für Evaluation verbunden sind, bisher kaum thematisiert werden. Gerade die Konsequenzen der Cyber Society für die politische Mobilisierung, das Eröffnen breiter bürgerlicher Partizipationschancen und bürgernaher Administration – für die Frage, wie solche Prozesse adäquat evaluiert werden können, bieten ein breites Betätigungsfeld. Frans L. Leeuw geht in seinem Beitrag der Frage nach: Was sind die herausfordernden Fragen für Evaluation in dieser Welt?

Eng verbunden mit dieser Entwicklung sieht der Autor die wachsende Bedeutung transnationaler Akteure, deren Regelsetzung (rulemaking) nationalen Parlamenten Vorgaben macht, sie beeinflusst oder gar entmachtet. Leeuw geht von der These aus, dass „rulemaking is becoming less formal, more private, less hierarchical, more competitive and more contractual“ und formuliert Fragen, was dies für die traditionellen Evaluationskonzepte bedeutet und wie sie sich diesen neuen Entwicklungen anpassen müssen.

Der Beitrag von *Reinhard Stockmann* und *Wolfgang Meyer* unterscheidet sich von den anderen Artikeln dieses Kapitels durch seine Fokussierung auf die Wissenschaft. Es wird der Konflikt zwischen Praxis und Wissenschaft thematisiert, der auch für die Ausbildung von Evaluierenden eine Rolle spielt. Die Autoren halten das Spannungsverhältnis zwischen diesen beiden Bereichen für strukturell bedingt und damit unausweichlich, da die aus der Praxis stammende Forderung nach Nützlichkeit der Evaluation und das Streben der Evaluierenden (sofern sie sich als angewandte Sozialforscherinnen und Sozialforscher verstehen) nach Wissenschaftlichkeit aus unterschiedlichen logischen Systemen stammen.

Während die Praxis vor allem an der Beantwortung von „Wie“-Fragen interessiert ist – wie ein als defizitär erkannter Zustand verändert werden kann – ist die Wissenschaft vor allem an „Warum“-Fragen interessiert, um Kausalitätshypothesen zu überprüfen. In der Evaluation müssen beide Prinzipien miteinander verknüpft werden. Evaluation als angewandte Wissenschaft im Auftrag und zum Nutzen von diversen Stakeholdern liefert deshalb nicht nur möglichst objektive, reliable und valide Befunde dazu, wie etwas funktioniert, sondern auch, warum. Inwieweit dieses Verhältnis zwischen Praxis und Wissenschaft austariert ist und wie es sich in Zukunft verändern wird, wird an-

hand der vier gesellschaftlichen Funktionen von Evaluation untersucht (vgl. Stockmann 2007: 36ff.; Stockmann 2013: 73ff.): Dabei zeigt sich, dass die Funktion des „Erkenntnisgewinns“ zu Wissenschaftszwecken eher unterentwickelt ist, da Evaluation nach wie vor ein Auftragsmarkt ist, in dem primär andere Ziele verfolgt werden, wie z.B. die Initiierung von Lernprozessen. Bei dieser Art von Evaluation, die immer mehr an Bedeutung gewinnt, steht klar der Nutzungsaspekt im Vordergrund.

Performance Evaluationen zu Kontrollzwecken, die sich häufig durch einfache Soll-Ist-Vergleiche und einfache Methodiken auszeichnen, sind eher auf dem Rückzug und machen komplexeren Wirkungsevaluationen Platz, die sich durch umfassendere wissenschaftliche Ansprüche und aufwändigere methodische Verfahren auszeichnen und vermehrt extern durchgeführt werden, um ihre Legitimitätsfunktion erfüllen zu können.

3. Die Institutionalisierung von Evaluation als eigenständige Profession

Die Tatsache, dass Evaluation an Hochschulen gelehrt wird, bedeutet noch lange nicht, dass es sich auch um eine eigenständige Profession mit einem entsprechenden Profil handelt. Zur Kennzeichnung einer neuen Wissenschaftsdisziplin werden beispielsweise als Kriterien die Entstehung einer eigenständigen Begrifflichkeit, die Entwicklung spezifischer Theorien und der Einsatz eines besonderen methodischen Instrumentariums herangezogen, durch die sich eine Einzeldisziplin von anderen abgrenzen lässt (vgl. Rothland 2008). Insofern können auch die mehr oder weniger formalisierten Regeln, wie ein bestimmtes Fach betrieben werden soll, als Institutionalisierung innerhalb der Fächerlandschaft begriffen werden. Dementsprechend stellt sich hier insbesondere die Frage, inwieweit Evaluation eine eigenständige theoretische Grundlage und ein spezifisches methodisches Vorgehen entwickelt hat, welches sie von anderen anwendungsbezogenen Sozialwissenschaften abgrenzt. Die vorgestellten Beiträge in diesem Band heben hierzu einige interessante Aspekte hervor und nehmen bezüglich ihrer zukünftigen Entwicklung vorsichtige Einschätzungen vor.

Der erste Beitrag dieses Kapitels von *Wolfgang Meyer, Evelyn Funk* und *P. Linh Nguyen* behandelt einen weiteren, im Rahmen von Evaluationen vorkommenden Spannungsbogen, den zwischen Partizipation, Bewertung und Wissenschaftlichkeit. Ähnlich dem Konflikt zwischen Wissenschaftlichkeit und Nützlichkeit ist auch das Verhältnis zwischen wissenschaftlichen Prinzipien, die auf neutrale, objektive, nicht wertende Ergebnisse, auf die Suche nach „Wahrheit“ ausgerichtet sind und der Beteiligung von Stakeholdern am Forschungsprozess, auszutarieren.

Partizipation dient nicht nur dazu, die Interessen und spezifischen Fragestellungen der verschiedenen Stakeholder zu integrieren, sondern auch deren internes Wissen zu erschließen, ihre Beteiligungsbereitschaft zu erhöhen, um schließlich auch die Akzeptanz der Bewertungsergebnisse und damit die Nützlichkeit der Evaluation sicherzustellen. Mit der Frage, wie Beteiligungs- und Bewertungsprozesse organisiert sein müssen, damit die Wissenschaftlichkeit einer Evaluation nicht verloren geht, beschäftigt sich dieser Beitrag. Zunächst werden die methodologischen Anforderungen eines wissenschaftlichen Evaluationsansatzes diskutiert, dann die bisher üblichen Beteiligungsverfahren

klassifiziert und deren Verbreitung untersucht, bevor zentrale Herausforderungen für die Zukunft abgeleitet werden.

Das Autorenteam kommt zu dem Ergebnis, dass die Auseinandersetzung mit diesen Themen bisher erst am Anfang steht, dass wissenschaftliche Methoden, die einen Bewertungsprozess korrekt anleiten können, fehlen und dass es zu wenige Forschungsarbeiten zur Umsetzung von Partizipation in Evaluationen gibt. Zwar wird eine Vielfalt von Partizipationsansätzen mit steigender Tendenz angewendet, doch scheint dies zu einem Abdriften, weg von wissenschaftlichen Anforderungen, zu führen.

Während sich der Beitrag von Meyer et al. mit der Methodik und dem Selbstverständnis von Evaluation zwischen den prinzipiell antagonistischen Ansprüchen möglichst weitgehender Einbeziehung von Stakeholdern in den Forschungs- und Bewertungsprozess von Evaluation bei gleichzeitiger Aufrechterhaltung wissenschaftlicher Standards beschäftigt, setzt sich *Brad Astbury* mit dem theoretischen Selbstverständnis der Profession auseinander. Dabei stellt er fest, dass sich die „Evaluationstheorie“ in den letzten 50 Jahren enorm entwickelt und diversifiziert hat. Im Unterschied dazu gibt es jedoch kaum Forschung über Evaluation. Dies ist eigentlich nicht so erstaunlich, da Evaluation zumeist eine Auftragsarbeit darstellt und die Auftraggeber Forschung über Evaluation kaum interessiert. Da zudem Evaluationen vor allen von Consulting-Firmen durchgeführt werden, die sich an den Bedürfnissen ihrer Kunden orientieren und wohl auch eher an Gewinnmaximierung als an der Weiterentwicklung von Theorien der Evaluation interessiert sind, gibt es wenige Anreize, sich mit dem Thema auseinanderzusetzen. Hinzu kommt drittens, dass es auch an entsprechenden Forschungsgeldern fehlt. Aus diesen Gründen ist kaum bekannt, welche Evaluationstheorie unter welchen Kontextbedingungen und für welche Evaluationsziele und -aufgaben am besten geeignet ist.

Auch wenn jedem klar ist, dass es keinen „einen besten Weg“ für die Durchführung von Evaluationen gibt, so wäre es doch sehr hilfreich, zu wissen, welcher Ansatz oder welche Kombination von Ansätzen in einer bestimmten Situation welche Vor- und Nachteile mit sich bringt. Dabei wäre es schon hilfreich, geeignete Klassifizierungskriterien oder -schemata zu haben, um den Wildwuchs an „neuen“ Ansätzen, die sich oft nur als neues Label auf altbekannten Ansätzen entpuppen, zu strukturieren. Das ist aus einer Reihe von Gründen gar nicht so einfach, denn es gibt in der Evaluationscommunity noch nicht einmal einen Konsens darüber, ob es so etwas wie eine Evaluationstheorie gibt. Jedenfalls nicht im klassischen wissenschaftstheoretischen Sinne als ein System von aufeinander bezogenen Aussagen zur Erklärung eines Phänomens. Stattdessen werden im Evaluationsbereich schon Anleitungen zur Praxis, also wie eine Evaluation durchzuführen ist, mitunter als „Evaluationstheorie“ bezeichnet.

Astbury differenziert zwischen drei sich überschneidenden Kategorien von „Arten“ der Evaluationstheorie: Evaluationsparadigmen, Taxonomien und Ansätzen, die er als Ansatzpunkte für Forschung über Evaluation betrachtet. Die Frage, warum man sich damit überhaupt beschäftigen soll, beantwortet er u.a. damit, dass das gründliche Studium der Evaluation für den Erwerb von Wissen und die Förderung der Disziplin entscheidend ist, und dass schlecht oder unzureichend entwickelte Evaluationstheorien in der Praxis irreführen können, was zu ernsthaften negativen Konsequenzen führt. Darüber hinaus erwartet der Autor von Forschung über Evaluation ein wachsendes empirisches Wissen darüber, welche Evaluationsansätze am besten funktionieren; dass die Komplexi-

tät des Bezugs zwischen Praxis und Theorie deutlicher wird; dass der Status der Professionalisierung gestärkt wird und dass der Beitrag der Evaluation zur gesellschaftlichen Entwicklung besser herausgearbeitet werden kann. Abschließend zeigt Astbury einige Optionen zur Erforschung der Evaluation in der Zukunft auf.

Der Beitrag von *Christoph Müller* befasst sich mit den methodischen Ansprüchen an die kausale Wirkungsevaluation und die empirische Praxis. Auch bei diesem Thema ist ein Spannungsverhältnis zwischen methodologischen Standards und der Evaluationsrealität festzustellen. Häufig lassen sich die hohen methodischen Anforderungen, die mit für kausale Wirkungsevaluationen tauglichen Untersuchungsdesigns verbunden sind, aus einer Reihe unterschiedlicher Gründe nicht umsetzen. Randomized Controlled Trials (RCTs), die von einigen als „Goldstandard“ für Kausalanalysen propagiert werden (vgl. Duflo, Glennerster & Kremer 2007), sind nicht nur per se mit einer Reihe von Problemen behaftet, sondern lassen sich zudem häufig im Evaluationskontext nicht umsetzen. Das Vergleichsgruppendesign schwächelt zwar bei der Kontrolle von konfundierenden Drittvariablen, ist aber schon eher in der Evaluation einsetzbar. Häufig müssen jedoch alternative Designs, wie das Pretest-Posttest-Design, mitunter mit retrospektiven Pretests, verwendet werden. Ein wichtiger Schritt in der Weiterentwicklung der kausalen Wirkungsevaluation ist die Verwendung von theoriegeleiteten Evaluationsansätzen, die einen breiten Einsatz von quantitativen und qualitativen Datenerhebungs- und Analysemethoden ermöglichen.

Mit Blick auf die Zukunft kommt Müller zu dem Schluss, dass sich die Zahl und die Qualität von Wirkungsevaluationen erhöhen werden. Insbesondere die zunehmende Bedeutung von Legitimitätsevaluationen und die damit verbundenen Ansprüche an wissenschaftliche Unabhängigkeit und methodische Angemessenheit (vgl. Stockmann sowie Stockmann & Meyer in diesem Band) lassen eine weitere Methodenentwicklung sowie einen Bedeutungszuwachs von Methodenkompetenzen bei Auftraggebern und vor allem Evaluierenden erwarten.

Den Abschluss dieses Buches bildet ein reflektierendes Kapitel von *Wolfgang Meyer* und *Reinhard Stockmann* zur weiteren Institutionalisierung der Evaluation entlang der hier gewählten drei Entwicklungslinien: in der Gesellschaft, im Wissenschafts- und Professionalisierungssystem. In einer Synopse werden die zentralen Befunde zur Zukunft der Evaluation zusammengefasst, indem Zukunftstrends herausgearbeitet und ihre Bedeutung für die Institutionalisierung der Evaluation bewertet werden.

Literatur

- Dahler-Larsen, P. (2006). Evaluation after Disenchantment? Five Issues Shaping the Role of Evaluation in Society. In: Shaw, I. F.; Greene, J. C.; Mark, M. M. (Hrsg.). *The SAGE Handbook of Evaluation*. London: Sage, S. 141–160.
- Duflo, E.; Glennerster, R.; Kremer, M. (2007). Using Randomization in Development Economics Research: A Toolkit. In: Schultz, T.; Strauss, J. (Hrsg.). *Handbook of Development Economics*. 4. Auflage. Amsterdam, New York: North Holland, S. 3895–3962.
- Fitzpatrick, J. L.; Sanders, J. R.; Worthen, B. R. (2004). *Program Evaluation. Alternative Approaches and Practical Guidelines*. 3. Auflage. Boston: Pearson & AB.

- Furubo, J.-E.; Rist, R. C.; Sandahl, R. (Hrsg.) (2002). *International Atlas of Evaluation*. New Brunswick: Transaction Publishers.
- Furubo, J.-E.; Rist, R. C.; Speer, S. (Hrsg.) (2013). *Evaluation and Turbulent Times. Reflections on a Discipline in Disarray*. New Brunswick: Transaction Publishers.
- Jacob, S.; Speer, S.; Furubo, J.-E. (2015). The Institutionalization of Evaluation Matters: Updating the International Atlas of Evaluation 10 Years Later. In: *Evaluation*, 21(1), S. 6–31.
- Leeuw, F. L. (2009). Evaluation: A Booming Business but is it Adding Value? In: *Evaluation Journal of Australasia*, 9(1), S. 3–9.
- Rosenstein, B. (2013). Mapping of National Evaluation Policies. Commissioned by Parliamentarians Forum on Development Evaluation in South Asia jointly with EvalPartners (im Internet unter http://api.ning.com/files/saEvSFZMI0naAMuIhi89Gvi-w8lUqOKc*yzLZZ4Hif3BpGzHtm3IXVES8mjmCWstNTOTFWVCvysHjO2kOuLPLlcd5a3BmzOh/MappingNEPReportDecember2013.pdf, zuletzt eingesehen am 06.06.2017).
- Rothland, M. (2008). *Disziplingeschichte im Kontext. Erziehungswissenschaft an der Universität Münster nach 1945*. Bad Heilbrunn: Klinkhardt.
- Smith, N. L.; Brandon, P. R. (2011). If not to Predict, at Least to Envision, Evaluation's Future. In: *American Journal of Evaluation*, 32(4), S. 565–566.
- Smith, N. L.; Brandon, P. R.; Hwalek, M.; Kistler, S. J.; Labin, S. N.; Rugh, J.; Thomas, V.; Yarnall, L. (2011). Looking Ahead: The Future of Evaluation. In: *American Journal of Evaluation*, 32(4), S. 565–599.
- Stockmann, R. (2007). *Handbuch zur Evaluation. Eine praktische Handlungsanleitung*. Münster: Waxmann.
- Stockmann, R. (2013). The Evaluation Process. In: Stockmann, R.; Meyer, W. (Hrsg.). *Functions, Methods and Concepts in Evaluation Research*. Basingstoke, New York: Palgrave Macmillan, S. 175–211.
- Stockmann, R.; Menzel, U.; Nuscheler, F. (2016). *Entwicklungspolitik. Theorien – Probleme – Strategien. 2. Auflage*. Berlin, Boston: De Gruyter.

Die sich verändernde Rolle der Evaluation in einer sich verändernden Gesellschaft¹

Übertragen aus dem Englischen von Selina Röhrig

Wie bereits die Überschrift dieses Kapitels nahelegt, ist Evaluation ein sich veränderndes, bewegendes Phänomen, das sich am besten durch Einbezug seines soziologischen Kontexts verstehen lässt. Konkret werden im Folgenden fünf Diskussionsthemen aufgezeigt, die, wie von mir jeweils dargelegt, für dieses Verständnis zentral sind. Was ein Diskussionsthema kennzeichnet, ist schlichtweg, dass es wichtig für Evaluation und Gesellschaft ist und dass es wahrscheinlich umstritten bleibt, da hierzu unterschiedliche Auffassungen bestehen. Bevor ich jedoch zu den Diskussionsthemen komme, bedarf es einer näheren Erläuterung meiner Perspektive und daraus folgender Überlegungen zur Definition von Evaluation.

Meine Perspektive ist konstruktivistisch. Dies impliziert nicht mehr und nicht weniger als drei Dinge. Erstens: Ich sehe Evaluation als Ergebnis einer sozialen Konstruktion. Evaluation ist kein natürliches, sondern unterstütztes Sensemaking (vgl. Mark, Henry & Julnes 2000), das heißt, eine Art der Weltsicht, die von Menschen bewusst zu bestimmten Zwecken konstruiert wird. Die Konstruktion der Evaluation hängt von Aktivitäten, Institutionen, politischen Ansichten, Normen, Werten und Erfahrungen ab, zusammengekommen von dem, was Schwandt (2009) eine Evaluationsvorstellung (engl. „evaluation imaginary“) nennt. Kein Verständnis von Evaluation ist allgemeiner als die ihm zugrundeliegende Evaluationsvorstellung. Evaluation als konstruiert zu begreifen, bedeutet, ihren sozialen und kulturellen Anker zu betrachten.

Zweitens: Evaluation hilft etwas zu konstruieren. Sie beeinflusst Policies, Entscheidungen, Praktiken und Denkstile. Je „fragiler“ die heutige Gesellschaft durch die Einbindung von Wissen in die soziale Ordnung wird (vgl. Stehr 2001), desto wahrscheinlicher wird Evaluation ein „Produktionsfaktor“, obwohl Evaluation natürlich nicht immer in linearer und direkter Weise genutzt wird.

Und drittens: Der Begriff „Evaluation“ ist selbst ein gedankliches Konstrukt. Folglich ist die Frage, wie Evaluation zu definieren sei, politisch und historisch sensibel. Konzepte prägen die Zukunft, sagt Koselleck (2004). Wenn wir Evaluation definieren, wählen wir eine Art der Begriffsklärung auf Kosten alternativer Varianten. Wir beschreiben tendenziell nicht wie Evaluation in der Praxis genutzt wird, sondern wie wir uns wünschen, dass sie genutzt würde. Mit anderen Worten: Die Definition von Evaluation ist selbst Teil eines größeren Konstruktionsprozesses.

1 Erschienen in Stockmann, R.; Meyer, W. (Hrsg.) (2016). *The Future of Evaluation. Global Trends, New Challenges, Shared Perspectives*. Basingstoke, New York: Palgrave Macmillan.

1. Die Definition von Evaluation

Sollen wir im internationalen Kontext von Evaluation, Evaluierung, Evaluación, Utvärdering und / oder Naliliineq sprechen? Selbstverständlich braucht es einen gemeinsamen Nenner für internationale Kommunikation, aber heißt das, dass wir semantische Unterschiede ignorieren können, die möglicherweise für kulturelle oder politische Unterschiede stehen? Oder, schwerwiegender noch, können wir annehmen, dass Begriffe einer vorherrschenden Sprache einen unproblematischen universellen Begriff abbilden? Ferner, welcher Personenkreis würde Tätigkeiten ausüben, die in unser Konzept passen? Zählen wir nur die „Großen“ der Evaluationstheorie (vgl. Alkin 2013), oder auch die Entwickler², Lehrer, Berater, Zeitschriftenherausgeber, Leiter von Evaluationszentren, Sprecher von Evaluationsgesellschaften und Autoren von Richtlinien, kurz jene, die Stame (2013) die „theory weavers“ der Evaluation nennt, die in der Tat einen maßgeblichen Beitrag zum Feld geleistet haben, ganz zu schweigen von den Evaluationspraktikern, deren Arbeitspraxis eine Collage von Aktivitäten sein kann, die nur lose mit irgendeiner formalen Evaluationsdefinition verbunden ist?

In den meisten Fällen ignorieren wir diese Fragen und springen direkt zur formalen Definition. Diese umfasst typischerweise vier Komponenten, als da wären

- systematische Untersuchung;
- Kriterien, die gemäß eines expliziten, legitimen Wertesystems gerechtfertigt sind;
- Evaluationsgegenstände wie z.B. Policies, Programme, Projekte und Interventionen;
- und einen intendierten Zweck oder Nutzen wie z.B. Rechenschaftslegung oder Optimierung / Lernen / Entwicklung.

Alles in allem verweist jede solche Definition, einschließlich ihrer Variationen entlang der vier Dimensionen auf eine Relevanzstruktur (vgl. Berger & Kellner 1982) oder Evaluationsvorstellung (vgl. Schwandt 2001), die expliziert, warum es sich lohnt, systematisches Wissen über einen Evaluationsgegenstand unter Bezugnahme auf Kriterien zu generieren, die in einer gegebenen Situation demokratisch und / oder ethisch relevant erscheinen. Eine solche Relevanzstruktur oder Evaluationsvorstellung kann am besten im Kontext einer demokratischen Gesellschaft verstanden werden, die sich in expliziter und intelligenter Weise zu organisieren und zu verändern sucht und aus ihren Fehlern lernen will. Mit anderen Worten steht die Definition von Evaluation in unmittelbarem Zusammenhang mit dem, was Stehr (2001) als die Selbstaneignung der modernen Gesellschaft bezeichnet.

Unsere Evaluationsdefinition auf der abstrakten, konzeptuellen Ebene, hat natürlich auch Implikationen auf der konkreten Ebene. In der heutigen Gesellschaft gibt es etliche benachbarte Felder mit praktischen und funktionellen Überschneidungen zur Evaluation, z.B. anwendungsbezogene Forschung, Akkreditierung, Auditierung, Organisationsentwicklung, Politikanalyse und Performance-Management. Einige befürworten klare Unterscheidungen zwischen Evaluation und diesen benachbarten Feldern, vielleicht um Komplementaritäten zu bestimmen oder um die Unterscheidbarkeit und Autonomie ei-

2 Bei geschlechtsspezifischen Begriffen wird aus Gründen der besseren Lesbarkeit die männliche Form verwendet. Die weibliche Form ist dabei jeweils mit eingeschlossen.

nes gegebenen Feldes zu bewahren. Möglicherweise wird zu viel Energie in Grenzscharmützel benachbarter Felder investiert, die lediglich für die beteiligten Konfliktparteien von Interesse sind.

Analytisch bevorzuge ich eine breite und allgemeine Perspektive auf Evaluation, aus der heraus wir beobachten können, dass sich Evaluation im Laufe der Zeit aufgrund umfassender gesellschaftlicher Kräfte verändert. Analog könnten wir mit einer auf unseren Erfahrungen aus den 1950ern basierenden Definition von „Medien“ oder „Familie“ nicht die heutige Wirklichkeit erfassen. Mit der Zeit entwickelt sich auch die Grenzlinie zwischen Evaluation und einem gegebenen benachbarten Feld kontingent mit den Formen und Funktionen von Evaluation. So kann es passieren, dass Aufgaben oder Funktionsbereiche, die früher zur Evaluation gehörten, von benachbarten Feldern übernommen werden (oder umgekehrt einstige Aufgaben und Funktionsbereiche benachbarter Felder unter den Evaluationsbegriff fallen). Z.B. kann Evaluation insgesamt expandieren, aber zugleich auch einige Bereiche an das Performance-Management verlieren.

Obwohl kein Grund besteht, eine „imperialistische“ Sichtweise auf Evaluation im Verhältnis zu anderen Tätigkeitsfeldern zu befürworten, hat die der Evaluation inhärente breite Relevanzstruktur etwas zu bieten. Insoweit Performance-Management sich auf systematische Untersuchung, explizite, auf gerechtfertigten Werten basierende Kriterien, eine eindeutige Definition eines Evaluationsgegenstands und ein explizites Interesse an demokratischer, zielgerichteter und intendierter Nutzung stützt, besteht überhaupt keine Notwendigkeit Performance-Management aus der Definition von Evaluation auszuschließen. Stattdessen könnten wir fragen, wie gut Performance-Management in unserem Fall verglichen mit oder in Verbindung mit anderen Evaluationsformen funktioniert.

Mit anderen Worten: Ich will nicht behaupten, dass Evaluation all den oben genannten, benachbarten Praxisfeldern technisch, institutionell oder funktionell äquivalent ist. Alles was ich sagen will ist, dass die in Evaluation angedeutete Relevanzstruktur so zentral für eine moderne, demokratische Gesellschaft ist, dass ebenjene Relevanzstruktur uns als Fenster dienen kann, durch das wir Diskussionsthemen, Kontroversen und Veränderungen betrachten können, die relevant für eine Reihe von Phänomenen sind, die mehr umfassen als das, was wir in der Praxis unmittelbar als „Evaluation“ bezeichnen. Ich veranschauliche das etwas konkreter in den folgenden Abschnitten, in denen ich fünf Diskussionsthemen behandle. Unter einem Diskussionsthema (engl. „issue“) verstehe ich „a question, point or concern to be disputed or decided; a main matter of contention; a sticking point or grievance; a belief at variance“ (Stake 2004: 89). Solche Themen sind im Evaluationsbereich mit hoher Wahrscheinlichkeit zu erwarten, da Evaluation konstruiert ist und wiederum selbst etwas konstruiert. Sie sind Angelegenheiten, derer sich Evaluatoren und Nutzer von Evaluation annehmen müssen, wobei sie ihre Interpretationen und Vorgehensweisen weiterentwickeln. Nachstehend werde ich fünf Diskussionsthemen unter den folgenden Überschriften anschneiden: „Popularisierung“, „Systematisierung“, „Outcome-Orientierung“, „Beziehungen zwischen Evaluation und Forschung im Wandel“ und „Nutzungsmuster im Wandel“.

2. Diskussionsthema 1: Popularisierung

Unter Popularisierung verstehe ich, Evaluation den Menschen nahezubringen, sie zu Gegenständen, Konsumenten oder Teilnehmern im Feld der Evaluation zu machen, Evaluation populär zu machen und mit Normen, Werten sowie institutionellen Mechanismen zu unterstützen. Durch Popularisierung wird Evaluation, im Einklang mit gesellschaftlichen Vorstellungen, Teil des modernen Lebens. Evaluation wird ein kulturelles Phänomen, das nicht in jedwedem Fall unter Begründungszwang steht. Durch Popularisierung wird Evaluation eine weit verbreitete und als selbstverständlich betrachtete Tätigkeit – in einem Ausmaß, das in einer nicht modernen Gesellschaft unvorstellbar gewesen wäre. Ohne Popularisierung hätte Evaluation niemals ihren heutigen, hohen Verbreitungsgrad erreicht.

Wenn Evaluation tatsächlich durch modernisierende Kräfte geformt wird, die das allgemeinere gesellschaftliche Klima verändern, muss es möglich sein, verschiedene Modernisierungsphasen mit jeweils anders gearteten, charakteristischen Evaluationsvorstellungen zu unterscheiden.

Lassen Sie mich idealtypisch zwischen Moderne, Reflexiver Modernisierung und der Auditgesellschaft unterscheiden. Die durch Fortschritts- und Rationalitätsglauben geprägte Moderne, hat selbst den Weg für die Evaluation als rationales Instrument zur Verbesserung sozialer Initiativen geebnet. Die Vorstellung der Moderne stützte frühe Formen der Evaluation, wie z.B. das Experiment, in Kombination mit rationalen Annahmen über den instrumentellen Nutzen von Daten bei der Entscheidungsfindung. Die Vorstellung der Moderne vertrat zudem eine sehr rationale Herangehensweise an die Evaluation als solche. Z.B. glaubte man den potentiellen Beitrag von Evaluation durch Anwendung eines Verfahrens, des sogenannten Evaluability Assessments, im Vorhinein bestimmen zu können. In einem Evaluability Assessment wurden eine Reihe von Fragen über die Ziele, den Evaluationsgegenstand und seine Implementierung, sowie potentielle Nutzer der Evaluation gestellt. Auf dieser Grundlage sollte ermittelt werden, ob Evaluation im konkreten Einzelfall ein rational nutzbares Instrument war.

Seit der Phase der Reflexiven Modernisierung lässt die Planungsgläubigkeit jedoch nach. Es wird deutlich, dass keine Maßnahme völlig frei von Problemen ist (vgl. Beck 1992). Soziale Komplexität und Bedingtheit führen zu einem neuen Interesse an Begleiteffekten. Die Berücksichtigung von Begleiteffekten, auch als Lernen bekannt, wird wichtiger als rationale Planung. Entsprechend diesem kognitiven Mentalitätswandel, verändert sich auch die soziale Mentalität zur Steuerung von Evaluationsprozessen. Rationale Planung und wahrgenommener sozialer Konsens werden durch ein neues Interesse an sozialem Lernen, Dialog und Reflexivität ersetzt. Vielfältige kulturelle und ethische Perspektiven werden anerkannt. Auf Grundlage dieser Vorstellung entstehen responsive, partizipative und empowerment-orientierte Evaluationsmodelle. Der allumfassende Sozialmythos ist nicht mehr der des Fortschritts, sondern der der Entwicklung. Entwicklung erfordert keinen rationalen Konsens, nur irgendeine Form kontinuierlicher sozialer Veränderung, unterstützt durch ständigen Dialog und laufende Anpassung.

Im Rahmen der Reflexiven Modernisierung entpuppt sich Evaluation schon bald als mit Widersprüchen und Schwächen behaftet. Ihr fehlt eine koordinierte gesellschaftliche Agenda. Jede Evaluation ist so unbeständig und perspektivenabhängig wie die Wirk-

lichkeit, die sie beschreibt. Daher ist es schwierig Erkenntnisse und Ergebnisse mehrerer Evaluationen zu bündeln und zusammenzufassen.

Manager bleiben ohne Aufsichts- und Kontrollinstrumente. Somit bereitet die Reflexive Modernisierung einer neuen Art sozialer Vorstellung den Weg, der der Auditgesellschaft. In Folge des 11. Septembers bilden Sicherheit und Prävention einen neuen Fokus. Die von einem Absicherungsmythos dominierte Auditgesellschaft, richtet neue Kontroll-, Dokumentations- und Berichtsverfahren ein (vgl. Power 1997). Ihr Standardwerkzeug sind „evaluation machines“, das heißt umfassende, bindende Berichtsverfahren, die dazu beitragen, die Rechenschaftslegung herauszustellen.

Mit anderen Worten: Obwohl die Popularisierung der Evaluation ein immerwährender Aspekt der Moderne zu sein scheint, wird Evaluation in verschiedenen Phasen durch sehr verschiedene soziale Vorstellungen gestärkt. Ein Skeptiker könnte den Standpunkt vertreten, dass sich die Unterscheidung in Moderne, Reflexive Modernisierung und Auditgesellschaft ausschließlich auf abstrakter Ebene bewegt und wenig mit realer Evaluation zu tun hat. Das ist absolut zutreffend, aber selbst Teil des Problems. Die bedeutenden sozialen Zusagen der jeweiligen Epochen können möglicherweise nur schwer in jeder einzelnen Evaluation eingehalten werden. Es ist schwierig für Evaluation so viel Fortschritt, Reflexion und Gewissheit wie versprochen zu erbringen (vgl. Dahler-Larsen 2012). Evaluation wird zwar durch bedeutende Zusagen bekannter, aber jeder der führenden Mythen in den drei Phasen der Modernisierung hat eine Schattenseite. Vor allem im vorherrschenden System der Auditgesellschaft können wir eine Tendenz zu großangelegten, bindenden Evaluationssystemen beobachten. Diese bezwecken dem Eintreten unliebsamer Vorkommnisse gänzlich vorzubeugen, nicht Reflexion und Lernen zu fördern.

3. Diskussionsthema 2: Systematisierung

Unter der Systematisierung von Evaluation verstehe ich das Anliegen, Evaluation geplant, koordiniert, kontinuierlich und oftmals obligatorisch, in größerem Rahmen sowie zeit- und raumübergreifend durchzuführen (vgl. Leeuw & Furubo 2008). Systematisierung bedeutet auch, Evaluation stärker in Organisationsabläufe und Managementprozesse zu integrieren. Die Systematisierung von Evaluation wird gefördert, wenn sich die Popularisierung von Evaluation auf die systemische Ebene bewegt, auf der die sie prägenden Bedingungen festgelegt werden. Die Fähigkeit zu evaluieren, steht nun auf der Tagesordnung. Evaluationspolicies, Evaluationsstrategien, Evaluationskulturen und Evaluation Capacity Building werden zu Schlüsselbegriffen (vgl. Stockdill, Baizerman & Compton 2002).

Mit Blick auf Erfahrungen mit einzelnen, gescheiterten Projekten und mit einzelnen Evaluationen ohne Follow-up ist das ein sinnvoller Gedanke. Kontinuierliche Evaluation muss vielmehr systematisch in die normale Organisationspraxis eingegliedert werden. Inzwischen steht nicht mehr jede einzelne Evaluation unter Begründungszwang. Evaluability Assessment kommt aus der Mode. Die systematische Fähigkeit zu evaluieren, zählt mehr als das prognostizierte Ergebnis einer bestimmten Evaluation.

Auf dem Höhepunkt der Systematisierung von Evaluation steht das, was ich als „evaluation machines“ bezeichne. „Evaluation machines“ sind relativ beständig und basieren auf abstrakten und allgemeinen Vorgängen. Im Prinzip lassen sie das menschliche Element in Evaluation unsichtbar werden. Richtlinien, Indikatoren, Handbücher, Verfahren und Computertechnik definieren die Operationen von „evaluation machines“ und die Dokumentation liefert den Treibstoff, mit dem sie betrieben werden. „Evaluation machines“ verkörpern die Automatisierung von Evaluation. Sie sind der Gegenpol zu allem was die Reflexive Modernisierung an Evaluation kritisierte: Lokalität, Unbeständigkeit, Dialog und Subjektivität. Ihre Stärke liegt stattdessen in ihrer Objektivität und Zuverlässigkeit.

Aus der Distanz operierende „evaluation machines“ verändern das Berufsbild eines lokalen, internen Evaluators. Das Kernelement dieser Tätigkeit ist nicht mehr die Durchführung lokaler, maßgeschneiderter Evaluationen, sondern die Versorgung mit der Dokumentation, die eine externe „evaluation machine“ (sei es ein Monitoringsystem, ein Akkreditierungssystem oder ein Audit) am Laufen hält (vgl. Mathison 2011).

Die phänomenologische Erfahrung Gegenstand statt Gestalter oder Teilnehmender einer Evaluation zu sein, bildet die Grundlage für eine neue soziale Kritik an diesen Evaluationssystemen. Statt Führungsverantwortung zu übernehmen, etablieren Topmanager und Politiker „evaluation machines“, die Verantwortung die Umsetzungskette hinab verschieben. Allerdings sind viele soziale Probleme der heutigen Zeit schlichtweg zu komplex, um nur durch diejenige Ebene, die von einem solchen System kontrolliert wird, eingestanden und bewältigt zu werden. Es heißt „evaluation machines“ hätten den Verantwortungsgedanken getrübt, da Risiken innerhalb der Sozialordnung von einer Ebene zur nächsten verlagert werden (vgl. Rothstein, Huber & Gaskell 2006). Unter dieser Kontrolle zu stehen, bringt eine Reihe neuer Risiken mit sich, denen sich niemand gerne aussetzen möchte.

Auch heißt es, die Evaluationskosten würden defensive Qualität auf Kosten offensiver Qualität fördern, dass zu viel Gewicht auf Mindeststandards und bürokratische Abläufen liegt, während „evaluation machines“ faktisch nicht dazu ermutigen, mit dem, was in Zukunft Qualität darstellen könnte, zu experimentieren.

Darüber hinaus verursachen „evaluation machines“ wachsende Kosten, obgleich deren genaue Höhe oft schwer zu ermitteln ist, da ein Großteil der zu ihrem Betrieb erforderlichen Ressourcen von der Arbeitszeit der Praktiker abgeht, die die nötige Dokumentation im Rahmen ihrer täglichen Pflichten bereitstellen müssen.

Alles in allem, scheint die Schaffung dominanter Systeme von „evaluation machines“ in Reaktion auf die wohl durchaus aufrichtige Besorgnis über „unsystematische“ und „subjektive“ Evaluation demokratisch höchst strittig.

4. Diskussionsthema 3: Outcome-Orientierung

Eine der recht deutlichen, aktuellen Tendenzen im Evaluationsbereich besteht nicht nur in wachsenden Erwartungen hinsichtlich der Frage, wie Evaluation gestaltet werden sollte, sondern auch darüber, was gemessen werden sollte. Ein Schwerpunkt sollte auf dem durch die Intervention bewirkten Unterschied liegen, oder, wenn man so will, auf den