

Potenziale und Grenzen von Big Data in Evaluationen

Frühjahrstagung 2018 des Arbeitskreises „Methoden in der Evaluation“ der DeGEval – Gesellschaft für Evaluation e.V.

*Simone Ullrich*¹

Keinem anderen Ansatz wird in der Evaluierungsforschung ein ähnlich großes Potenzial zugesprochen wie Big Data. Dennoch werden in Evaluationen bis heute nur selten solche neue Datenquellen als empirische Basis genutzt. Die Klärung dieser Diskrepanz war das Ziel der Frühjahrstagung des Arbeitskreises „Methoden in der Evaluation“ der DeGEval – Gesellschaft für Evaluation e.V., die am 22. und 23. Juni 2018 im Gustav-Stresemann-Institut in Bonn stattfand. Gastgeber der Veranstaltung war das Nationale Zentrum für Kriminalprävention (NZK). Insgesamt konnten vier international bekannte und renommierte Referent(inn)en für die Frühjahrstagung gewonnen werden, die mit ihren Vorträgen ein weites Spektrum an Feldern abdeckten, in denen Big Data gesammelt werden und zur Anwendung kommen.

Nach der Eröffnung der Frühjahrstagung durch die Sprecher des Arbeitskreises, Prof. Alexandra Caspari (Frankfurt University of Applied Sciences) und Dr. Tobias Polak (Austrian Development Agency, Wien) erörterte Herr Dr. Andreas Armbrorst, Leiter des NZK, in seinem Grußwort die Relevanz von Big Data für die Prävention kriminellen Verhaltens. Obwohl Big Data am NZK bislang nicht zum Einsatz kommen, machte Dr. Armbrorst am Beispiel des „Predictive Policing“ deutlich, wie solche Datenmengen hilfreich bei der Vorhersage von Straftaten, z.B. Einbruchdiebstählen sein können. Mittels Big Data können Örtlichkeiten spezifiziert werden, an denen das Risiko der Verübung solcher Delikte erhöht ist, und operative Maßnahmen und Personaleinsatz der Polizei können somit besser gesteuert werden.

Im Anschluss machte Herr Dr. Gustav Jakob Petersson den Auftakt der wissenschaftlichen Beiträge am Freitag. Dr. Petersson ist Senior Analyst beim Schwedischen Research Council und referierte zu dem Thema „Cyber-Gesellschaft, Big Data und Evaluation“. Der Titel seines einführenden Vortrages korrespondierte dabei mit einem von ihm und Jonathan D. Breul im Jahr 2017 publizierten Buch, in dem die Implikationen von Big Data für Evaluatoreninnen und Evaluatoren in unserer neuen Cyber-Gesellschaft herausgearbeitet werden.² Mit seinen Ausführungen zielte Dr. Petersson darauf ab, u.a. die folgenden Fragen zu beantworten: Was sind Big

1 Nationales Zentrum für Kriminalprävention, Bonn

2 Petersson/Breul 2017.

Data und worin liegen die Vorteile und Herausforderungen für den Gebrauch in Evaluationen? Wie kann der Gebrauch von Big Data die Kompetenzen der Evaluator(inn)en komplementieren? Dr. Petersson betonte, dass einer der Gründe, warum Big Data so wertvoll im Feld der Evaluationen sind, der Fakt ist, dass sie viele Aspekte der Gesellschaft widerspiegeln, die unter anderen Umständen schwer zu erforschen sind. Darüber hinaus ist solch ein Datenvolumen extrem wertvoll für komplexe statistische Analysen, und die Datensammlung ist im Vergleich zu anderen Methoden nicht sonderlich kostspielig. Trotz dieses großen Potenzials wurde von Dr. Petersson jedoch mehrfach auf ein potenzielles Problem hingewiesen: die Verwendung und Analyse von Big Data ohne zugrundeliegende Theorie. Eine derartige Vorgehensweise kann zu falschen Schlussfolgerungen führen und sollte somit unbedingt vermieden werden. Nach Dr. Petersson liegen besondere Herausforderungen für Evaluatorinnen und Evaluatoren im Erlernen neuer Designs und Instrumente, die bereits in anderen Bereichen für evaluative Zwecke verwendet werden, dem Erwerb neuer Kompetenzen im Datenmanagement, sowie in Kollaborationen mit Gruppierungen, die bereits große Erfahrung im Bereich Big Data haben.

Nach einer Kaffeepause folgte der Vortrag von Herrn Dr. Dirk Tunger, wissenschaftlicher Mitarbeiter im Forschungszentrum Jülich, Themenfeld Evaluation und Wirkungsanalyse. Dr. Tunger fokussierte seine Ausführungen auf das Thema Big Data und „Altmetrics“. Hierunter versteht man alternative Metriken zu den klassisch bibliometrischen Indikatoren, da letztere aufgrund des veralteten Konzeptes des Peer Review und der Zitationsanalyse nicht mehr als zeitgemäß gelten. Altmetrics messen nicht nur die Reichweite von Publikationen in der Fachgemeinschaft, sondern vor allem die Reichweite in einer breiteren Öffentlichkeit. Wie die klassisch bibliometrischen Indikatoren zielen Altmetrics dabei auf die Messung des Einflusses einer wissenschaftlichen Publikation hin, berücksichtigen hierbei aber auch Einflussfaktoren wie Nennungen, Bookmarks, Verlinkungen oder Download-Zahlen. Dr. Tunger berichtete dann von einer „Machbarkeitsstudie Altmetrics“, in der den Fragen nachgegangen wurden, wie weit die Forschung zu Altmetrics bereits fortgeschritten ist, ob gegenwärtig bereits Anwendungsszenarien denkbar sind, worüber in der Altmetrics-Gemeinschaft diskutiert wird, und wo die Engpässe liegen. Das Untersuchungsdesign beinhaltete eine quantitative Erhebung, Experteninterviews, Austausch mit der Anbieterseite (Altmetric.com) sowie Berücksichtigung der Sekundärliteratur. Die Schlussfolgerungen, die von Dr. Tunger aufgrund der Ergebnisse gezogen wurden, waren vielversprechend. Zwar muss ein tiefergehendes Verständnis hinsichtlich der Heterogenität und Aussagekraft von Altmetrics noch stattfinden; es ist aber auf jeden Fall möglich, diese Daten als Proxi für die Diskussion von Wissenschaft in der Gesellschaft zu sehen. Der größte Vorteil im Vergleich zu klassischer Bibliometrie liegt in dem deutlich geringeren Zeitaufwand der Datenerhebung.

Der dritte Vortragende am Freitag war Renzo Giudice, Biologe für Naturschutz und derzeit Doktorand am Zentrum für Entwicklungsforschung an der Universität Bonn. Herr Giudice führte das Auditorium in die Welt natürlicher Ressourcen, geographischer Informationssysteme und ökologisch-ökonomischer Analysen. Geodaten basieren auf einer Vielzahl von Informationen wie z.B. biophysikalischen Daten (z.B. Biomasse, Temperatur), Landbedeckung (z.B. Wälder, Agrikultur, Flüs-

se), Landnutzung (z.B. Naturschutzgebiete, urbane Entwicklung), Infrastruktur (z.B. Straßennetz, Häfen) oder sozioökonomischen Daten (z.B. Bevölkerungsdichte, Armut). Sie werden durch Satellitenbilder, Messungen der Luminosität, hochfrequente Wetterdaten oder detailliertes Luft-Monitoring erfasst. Satellitendaten mit hoher räumlicher und zeitlicher Auflösung bezüglich Landnutzung, Änderungen in der Landnutzung und Bewaldungsdaten sind erst seit kurzem verfügbar und ermöglichen die empirische Evaluation von Programmen und Richtlinien zum Naturschutz. Bei der Nutzung solcher Geodaten in Evaluationen ist jedoch zu berücksichtigen, dass die Daten von einem Maßstabeffekt abhängig sind, der auf den verschiedenen Ebenen der Aggregation des Outcomes variieren kann. Höhere Ebenen der Aggregation sind mit einem Verlust an Variation verbunden. Dies führt zu einer Verzerrung der Schätzwerte. Um dem entgegenzuwirken empfiehlt Herr Giudice die Wahl eines den räumlichen Parametern (unabhängige Variable) angemessenen Maßstabs zur statistischen Modellierung des Prozesses. Gleichmaßen wichtig ist jedoch auch der Maßstab für die abhängige Variable. Hierbei ist zu berücksichtigen, wie Entscheidungen getroffen werden, z.B. wird Entwaldung aufgrund von Neigung in Zentimeter-Auflösung oder Hektar-Auflösung entschieden? In seinem Ausblick forderte Herr Giudice mehr interdisziplinäre Zusammenarbeit, insbesondere mit Geograf(inn)en und räumlichen Statistiker(inne)n, die mit Geodaten schon sehr viel länger arbeiten als Evaluatorinnen und Evaluatoren. Darüber hinaus besteht Bedarf an intelligenten Datenanalysemethoden, um nützliche von nutzlosen Daten zu trennen.

Zum Abschluss des ersten Tages der Frühjahrstagung wurden drei Arbeitsgruppen (Auftraggeber(innen), Auftragnehmer(innen) und Wissenschaftler(innen)) gebildet, in denen der praktische Nutzen von Big Data, deren Möglichkeiten und Grenzen in Evaluationen diskutiert wurde. Darüber hinaus sollten die Ergebnisse der Kleingruppenarbeit für die Podiumsdiskussion am Folgetag zusammenzufasst werden.

Nach einer kurzen Begrüßung am Samstag durch Prof. Caspari und Dr. Polak erhielt Frau Dr. Vera Allmanritter, selbständige Kulturmanagerin aus Berlin, das Wort. Der Schwerpunkt ihres Vortrages lag auf Besucherevaluationen von Kultur- und Freizeiteinrichtungen mit besonderem Augenmerk auf dem Besucher-Forschungsprojekt „Kulmon“. Zu Beginn ihrer Ausführungen beschrieb Frau Dr. Allmanritter die Gründe für die Durchführung von Besucher-Monitoring in Kultur- und Freizeiteinrichtungen. Hierunter fallen an erster Stelle der Wunsch nach Verbesserung der Arbeit durch Überprüfung von vorab gesetzten Zielen, Vergleiche mit anderen Institutionen sowie die Generierung von Ideen für zukünftige Strategien und Prognosen zur Nachfrageentwicklung. Das Projekt „Kulmon“ war auf Initiative der Berliner Kulturverwaltung und Tourismusmarketing entstanden und zusammen mit der Freien Universität Berlin in den Jahren 2008 und 2009 entwickelt worden. Bislang wurden mehr als 50 Kulturinstitutionen innerhalb und außerhalb von Berlin eingeschlossen; über 250.000 Einzelbefragungen wurden durchgeführt. Die Befragung folgte einer einheitlichen Methode: Jeder Befragte erhielt einen Satz an Kernfragen, die modular aufgebaut waren. Bereiche, die abgedeckt wurden, umfassten u.a. Informationsverhalten, Mediennutzung, Besuchsgründe und -motivation sowie die Zufriedenheit mit dem Angebot, der Vermittlung und dem Service.

Die Befragung wurde durch ein professionelles Umfrageinstitut durchgeführt, und die Interviewer(innen) gaben die Daten direkt über ein Smartphone in eine Datenbank ein. Dieses System ermöglichte die schnelle Verfügbarkeit einer umfassenden Datenbank mit Online-Auswertung und SPSS/Excel-Export der Auswertungen und Grafiken. Für 2018/2019 geplante Ausbaustufen umfassen Ergänzungsmodule zu den klassischen Besucherbefragungen in Institutionen, Verknüpfung der Kulmon-Datenbank mit Daten aus Ticketing-Systemen und statistischen Vergleichsdaten sowie eine regionale Erweiterung.

Im Anschluss an Frau Dr. Allmanritters Vortrag folgte eine kurze Einführung in Informationsportale für evidenzbasierte Kriminalpräventionsprogramme durch Herrn Dr. Armbrorst, NZK. Wissenschaftliche Erkenntnisse zu Sicherheits- und Präventionsmaßnahmen durch Evaluation werden im NZK durch das Portal WESPE vermittelt und umfassen zurzeit die Bereiche Islamistische Radikalisierung, Sexualdelikte, Mehrfachtäter und Gewalt im Fußball. Weiterführende Informationen finden sich im Beitrag von Armbrorst/Walsh in dieser Ausgabe.

Abschluss und Highlight der Arbeitstagung war die Podiumsdiskussion, an der auch alle Referent(inn)en teilnahmen. Eröffnet wurde diese durch eine Zusammenfassung der Arbeitsgruppen vom Vortrag zum praktischen Nutzen von Big Data.

Von Seiten der Wissenschaftler(innen) wurde ein Mehrwert von Big Data für Evaluationen anerkannt, da diese für die Identifikation von Zielgruppen hilfreich sein können. Hier sind insbesondere Geo- und sozioökonomische Daten von Relevanz. Kritisch hinterfragt wurden jedoch die Form der Datenerhebung, Kodierung und Auswertung von Big Data. Weiterhin wurde ein Selektions-Bias thematisiert, da soziale Medien von einer bestimmten (eher jungen) Benutzergruppe frequentiert werden, die nicht die Allgemeinbevölkerung repräsentiert. Auch wurde der Nutzen für die Sozialwissenschaften kritisch reflektiert, da der Wahrheitsgehalt der Informationen nicht verifiziert werden kann. Weiterhin ist fraglich, ob die technische Kompetenz im Umgang mit den Daten gewährleistet werden kann. Ein finaler Punkt der Diskussion war die fehlende Möglichkeit der Verknüpfung von Daten, die für Analysen auf Individualniveau ausschlaggebend sind.

Seitens der Auftraggeber(innen) standen insbesondere Datenschutz und ethische Vorgaben im Vordergrund. Weiterhin wurden die wissenschaftliche Fundierung und das Vertrauen in die Daten thematisiert. Auch stellte sich die Frage, wie mögliche Datenmanipulationen erkennbar gemacht werden können. Kritisch angemerkt wurden auch die „Black-Box“-Algorithmen sowie der Tatbestand, dass erst die Daten erhoben werden und dann die Fragen gestellt werden, ein Prozedere, das wissenschaftlichem Arbeiten widerspricht. Ein Mehrwert von Big Data in der Evaluation wurde hingegen in der Kombination von zugrunde liegender Theorie und Methodik gesehen, in der Geschwindigkeit der Datenerhebung und in der Möglichkeit zur Bildung von Kontrollgruppen. Auch wurden die vergleichsweise geringen Kosten der Datenerhebung thematisiert.

Die Diskussion verdeutlichte große Überlappungen zwischen Auftraggeber(inne)n und Auftragnehmer(inne)n hinsichtlich deren Beurteilung von Big Data. Ein deutlicher Mehrwert seitens der Auftragnehmer(innen) wurde in der Verlässlichkeit der Daten gesehen, der Möglichkeit des Einsatzes von Mixed Methods sowie der Tri-

angulation. Auch hier wurde jedoch die Black Box problematisiert, da die Intentionen der Auftraggeber der Datenerhebung und zugrundeliegende Algorithmen oftmals nicht transparent sind. Weiterhin wurden Datenschutzaspekte problematisiert, da es für manche Nutzer(innen) durchaus möglich wäre, Personen auf der Grundlage der gesammelten Informationen zu identifizieren.

Im Anschluss an die Vorstellung der Arbeitsgruppen wurde die Frage thematisiert, ob es sich bei Big Data um einen ‚Hype‘ handelt. Dem entgegengehalten wurde die Geschwindigkeit, in der sich Begrifflichkeiten ändern, und dass in zehn Jahren der Terminus ‚Big Data‘ zwar nicht mehr existieren mag, die Daten per se jedoch nach wie vor von großer Relevanz sind. Zwar wurde die Überlegenheit von Primärdaten nicht in Frage gestellt; es wurde jedoch auf das Thema abnehmender Rücklaufzeiten aufmerksam gemacht und die Option in den Raum gestellt, dass Sekundärdaten die ‚neue Zukunft‘ darstellen könnten.

Eines der Themen, das sich wie ein roter Faden durch die ganze Tagung zog, war die Warnung vor Datenanalysen ohne Theorie. Auch in der Podiumsdiskussion wurde mehrfach die Relevanz theoretischer Fundierung betont, da ansonsten das Risiko von Fehlschlüssen deutlich erhöht ist. Weiterhin wurde auf die Möglichkeit der Theoriebildung und Theorietestung mittels Big Data hingewiesen, beispielsweise im Hinblick auf eine Theory of Change mittels Mustererkennung.

Insgesamt wurden Big Data als zukunftsweisendes Thema beschrieben. Zwar sind die Begrifflichkeiten noch diffus und benötigen Klärung, ein großer Vorteil dieser Daten liegt jedoch in dem Tatbestand, dass es sich hierbei um nicht reaktive Daten handelt, die somit unverfälscht sind.

Zum Ende der Podiumssitzung betonte Herr Dr. Polak als Sprecher des AK Methoden drei essenzielle Punkte für den zukünftigen Umgang mit Big Data:

1. Die Notwendigkeit von Entmystifizierung und Klärung der Begrifflichkeiten;
2. Das Herausarbeiten von Potenzialen – wie z.B. die Möglichkeit der Erhebung neuer Indikatoren;
3. Die Güte neuer Indikatoren – erste Ergebnisse mit Altmetrics belegen, dass, basierend auf Big Data, qualitativ hochwertige Indikatoren gewonnen werden können.

Mit diesen Take-Home-Messages endete eine weitere erfolgreiche Frühjahrstagung des Arbeitskreises „Methoden in der Evaluation“ der DeGEval.

Literatur

Petersson, Gustav J./Breul, Jonathan D. (2017): Cyber Society, Big Data, and Evaluation. Comparative Policy Evaluation Series, Volume 24. New Brunswick: Transaction Publishers.

Evaluation deradikalisierender Maßnahmen – Wirkungen der Prävention gegen gewaltbereiten Extremismus bestimmen

Stefanie Reiter,¹ Frank Buchheit,² Rainer Strobl³

1. Besondere Herausforderungen in der Evaluation deradikalisierender Maßnahmen – Ausgangslage und Rahmung der Frühjahrstagung 2018 des AK Soziale Dienstleistungen der DeGEval mit dem KPEBW

Was funktioniert, was sicher nicht und was ist vielversprechend? Die Frage, welche Ansätze der (Kriminal-)Prävention ‚funktionieren‘ und welche nicht, bestimmt seit 20 Jahren die fachliche Diskussion. Die methodische Fragestellung, wie Wirkungen im Bereich der Prävention und insbesondere im Bereich der Extremismusdistanzierung oder Deradikalisierung methodisch zu fassen sind, umfasst auf den ersten Blick widersprüchliche Ansätze. Während einerseits von einigen Akteuren rigorose Verfahren zur Bewertung und Metaevaluationen gefordert werden, sind für viele Programme und Projekte eher Selbstevaluationen oder kleinere externe Evaluationen mit hohem formativem Anteil realisierbar. Ist dies angemessen, ‚besser als nichts‘ oder als untauglicher Versuch anzusehen?

Deradikalisierende Maßnahmen sind zudem ein komplexes Handlungsfeld. Schon die verwendeten Begriffe sind in der Lage, kontroverse Diskussionen auszulösen. Bei der Zieldefinition kommen mehrere Dimensionen in Betracht, bspw. die Fähigkeit der Programme, Zielgruppen zu erreichen und sie in Maßnahmen zu binden, eine reduzierte Rückfallquote, Prozesse der Distanzierung aus einschlägigen Kontexten oder gar ein Einstellungswandel, dessen Messung weitere methodische Fragen aufwirft.

Weitgehende Einigkeit besteht hingegen in der Defizitbeschreibung: Bislang gab es in diesem Bereich zu wenige Studien und Metaevaluationen, zumal nur wenige veröffentlicht wurden. Die Verbindungen zwischen ‚großen‘ Programm- und ‚kleinen‘ Projektevaluationen werden als mangelhaft beschrieben. Der Austausch über Evaluationen, auf dem Weg zu einer konstruktiven Fehlerkultur und zu gegenseitigem Lernen, scheint auch in diesem Feld noch verbesserungswürdig.

1 Deutsches Jugendinstitut e.V., Außenstelle Halle (Saale)

2 Landeskriminalamt Baden-Württemberg, Stuttgart

3 proVal, Gesellschaft für sozialwissenschaftliche Analyse – Beratung – Evaluation, Hannover

Die gemeinsame Tagung des AK Soziale Dienstleistungen in der DeGEval⁴ in Kooperation mit dem Kompetenzzentrum zur Koordinierung des Präventionsnetzwerks gegen Extremismus in Baden-Württemberg (KPEBW)⁵ am 19./20.04.2018 in Stuttgart führte vor diesem Hintergrund im Handlungsfeld tätige Evaluierende sowie Programmverantwortliche zusammen. Das Ziel war, Möglichkeiten auszuloten, um mithilfe evaluativer Verfahren zu größerer Wirkung und Effizienz in der Prävention und Intervention von Radikalisierungen zu gelangen.⁶ Hierbei wurden Ansätze aufgezeigt und kontrovers diskutiert, übertragbare methodische Zugänge gesucht, Forschungslücken umrissen und weitere Diskussion angestoßen.

Bereits im Vorfeld der Tagung war ein großes Interesse an diesem Thema und den damit aufgeworfenen Fragen spürbar, was dazu führte, dass die aus Kapazitätsgründen festgelegte Beschränkung der Teilnehmenden auf 60 Personen vollständig ausgeschöpft wurde und nicht alle Interessentinnen und Interessenten an der Veranstaltung teilnehmen konnten.⁷

Der erste Tag umfasste drei Beiträge von Referierenden, die aus unterschiedlichen Perspektiven Begrifflichkeiten und Evaluationsansätze schilderten und die damit verbundenen Grenzen und Potenziale reflektierten. Anschließend wurden die Perspektiven in einer Podiumsdiskussion mit den Referierenden zusammengeführt und kontrovers erörtert. Im Folgenden werden die Inhalte des ersten Tages zusammengefasst.

2. Was ist gegenstandsangemessen? Unterschiedliche Perspektiven und Ansätze der Evaluation deradikalisierender Maßnahmen

Der erste Beitrag von Andreas Armbrorst⁸ und Maria Walsh⁹ vom Nationalen Zentrum für Kriminalprävention trug den Titel „Evidenzbasierte Extremismusprävention“. Die Referierenden wiesen einleitend darauf hin, dass die Bundesregierung laut

4 Der Arbeitskreis „Soziale Dienstleistungen“ der Gesellschaft für Evaluation – DeGEval e.V. greift methodische Fragen der Evaluation im Handlungsfeld auf, das durch spezifische Leistungscharakteristiken und über spezielle Zugänge zu Fragen der Evaluation gekennzeichnet ist (vgl. www.degeval.org/de/arbeitskreise/soziale-dienstleistungen/). Die Autorin und die Autoren stellten das Sprecherteam des Arbeitskreises.

5 Das Zentrum wurde in der Zwischenzeit umbenannt in „Kompetenzzentrum gegen Extremismus in Baden-Württemberg“ (KONEX, vgl. <https://www.konex-bw.de/>).

6 Besonderer Dank für die organisatorische Unterstützung der Tagung geht an das Team des KPEBW sowie an Edith Halves von der Hochschule für Angewandte Wissenschaften Hamburg für die Leitung der Fishbowl-Diskussion.

7 Weitere Informationen zur Tagung sowie die Liste der Teilnehmenden sind einsehbar unter <https://www.degeval.org/arbeitskreise/soziale-dienstleistungen/bisherige-aktivitaeten/>.

8 Dr. Andreas Armbrorst ist Leiter des Nationalen Zentrums Kriminalprävention (NZK). Seine Schwerpunkte liegen auf der Kriminologie, Methoden der empirischen Sozialforschung, dem Dschihadismus, Salafismus und der islamistischen Ideologie sowie der Sicherheitsforschung.

9 Dr. Maria Walsh ist stellvertretende Leiterin des NZK. Sie beschäftigt sich vor allem mit Jugendkriminalität, Jugendstrafvollzug, Bewährungshilfe, kriminologischer Lebenslauforschung sowie Methoden der empirischen Sozialforschung.

Koalitionsvertrag vom 14.03.2018 für eine evidenzbasierte Kriminalprävention eintritt, bei der „Evidenzen sowohl bei der Erarbeitung von Gesetzentwürfen als auch bei deren Evaluation berücksichtigt werden“. Auch die Prävention extremistisch motivierter Gewalt sei zunehmend der Forderung ausgesetzt, ihre Wirksamkeit und (intendierte sowie unintendierte) Effekte anhand von Evaluationen nachvollziehbar zu belegen. Vor diesem Hintergrund stellte der Vortrag das Nationale Zentrum Kriminalprävention mit seinen Arbeitsschwerpunkten in diesem Bereich und insbesondere das Internetportal „WESPE – Wissenschaftliche Erkenntnisse zu Sicherheits- und Präventionsmaßnahmen durch Evaluation“¹⁰ vor. Die Referierenden skizzierten Merkmale und idealtypische Abläufe evidenzbasierter Kriminalprävention (mit den Phasen Grundlagenforschung, evidenzbasierte Entwicklung von Präventionsansätzen sowie Entwicklung und Anwendung von Evaluationsansätzen) und diskutierten deren Implikationen sowie gegenstandsspezifische Weiterentwicklungsbedarfe (bspw. in Bezug auf Indikatoren) für den speziellen Bereich der Extremismusprävention als einem gesellschaftlich normativ aufgeladenen Feld. Zur Illustration wurde die konzeptionelle Anlage eines laufenden Evaluationsprojekts vorgestellt, bei der die Untersuchung der Wirkung einer schulbasierten Primärpräventionsmaßnahme durch einen Aussteiger aus dem Bereich Rechtsextremismus u.a. mithilfe eines kontrollierten Experiments mit geclusterter Zufallsstichprobe angestrebt wird.

Christian Lüders¹¹ vom Deutschen Jugendinstitut e.V. (DJI) berichtete im Anschluss über die „Evaluation von Modellprojekten zur Deradikalisierung. Erfahrungen aus den Bundesprogrammen“. Zum Verständnis der Evaluation von komplexen Bundesprogrammen wie „Demokratie leben!“¹² stellte der Beitrag zunächst die Vorgeschichte, den Aufbau, die Teilprojekte und die Förderlogik des Programms mit den damit verbundenen Implikationen vor. Aufgrund der zeitlich begrenzten Förderung und des Entwicklungs- und Erprobungscharakters spielten übergreifende Leitfragen bei der Evaluation der Gesamtarchitektur des Bundesprogramms sowie formative Elemente eine wichtige Rolle. Die Evaluation der Programmbereiche starte meist mit der Rekonstruktion der Programmtheorie, (z.B. mithilfe von logischen Modellen). Im Kontext dynamischer und emergenter „Experimentalprogramme“ kristallisiere sich eine Programmtheorie dabei oftmals erst im Programmverlauf heraus. Sie stelle einen wichtigen Ausgangspunkt und eine unverzichtbare Voraussetzung für die Befassung mit Wirkungs- und Zielerreichungsfragen und der Entwicklung von angemessenen Erhebungsdesigns dar, die in Abgrenzung von ‚Black-Box-Ansätzen‘ nicht nur auf die bloße Erfassung einer Veränderungsmessung abzielten. An der Stelle seien darüber hinaus auch konzeptionelle Klärungen (u.a. Vermeidung eines verkürzten Wirkungsbegriffs) sowie eine differenzierte Betrachtung, welche Art von Wirkung erzielt werden soll, vonnöten. So gehe es bspw. bei der Förderung von Modellprojekten im Bundesprogramm um eine Anregungsfunktion für die Fachpraxis, bei der entsprechend das Interesse an der Entwicklung

10 Vgl. <https://www.nzkrim.de/wespe/>.

11 Dr. Christian Lüders ist Leiter der Abteilung Jugend und Jugendhilfe des Deutschen Jugendinstituts e.V. Seine Schwerpunkte liegen auf der Kinder- und Jugendhilfeforschung, der Sozialpädagogik, der qualitativen Sozialforschung und der Evaluationsforschung.

12 Vgl. <https://www.demokratie-leben.de/>.

von Praxiskonzepten im Mittelpunkt stehe und bei der im günstigsten Fall „qualitative Evidenz“ erzeugt werden könne. Die Vielschichtigkeit und Komplexität des Programms, seiner Bestandteile und der Einzelprojekte ergäben somit spezifische Herausforderungen (bspw. im Umgang mit einer oftmals hohen internen Komplexität und Dynamik, dem Fehlen einer kohärenten Programmtheorie, Diskrepanzen zwischen Projektprogrammatik und Realität, schwierige Zugangsbedingungen zu Modellprojekten etc.), welche auf den Gegenstand und die jeweilige Betrachtungsebene zugeschnittene Ansätze der Evaluation erfordern.

Im dritten Beitrag mit dem Titel „Wirkungsanalysen im Rahmen formativer Projektevaluation? Möglichkeiten und Grenzen wirkungsbezogener Untersuchungen“ stellte Kurt Möller¹³ (Hochschule Esslingen) ausgehend von Begriffsklärungen und vor dem Erfahrungshintergrund von selbst durchgeführten Studien im Bereich der Extremismusprävention vor, wie und inwieweit wissenschaftliche Evaluationen von themenbezogenen Projekten gewinnbringend erfolgen können, welcher Nutzen aus ihnen für wen zu ziehen ist und welche Rolle dabei die Identifizierung von Wirkungszusammenhängen spielt. Diesbezüglich ging der Beitrag auf ein formatives Evaluationsverständnis ein, beschrieb Radikalisierung als Prozess und beleuchtete die Voraussetzungen für deradikalisierende Maßnahmen. Evidenzorientierung sei in diesem Forschungsverständnis als wissenschaftliches, methodisch-systematisches, empirisch-datengestütztes Vorgehen, jedoch nicht im Sinne einer wissenschaftlichen Objektivierbarkeit und einer methodischen Engführung auf randomisierte Kontrollgruppendesigns, quasi-experimentelle Designs und Kosten-Nutzen-Analysen zu verstehen. Die Evaluation deradikalisierender Maßnahmen wird aus dieser Perspektive vor allem im Sinne des Lernens und der Entwicklung der Projekte verstanden: Aus der multiperspektivischen Identifikation von Wirkannahmen, Wirkungsweisen und Schwierigkeiten unter Einbeziehung der Fachpraxis (als Ko-Kreations-, Reflexions- und Modifikationsprozess auf Augenhöhe im „dritten Raum“), könne sich ein vielschichtiger Verstehensprozess ergeben, der zu einer kommunikativen Entwicklung zum beiderseitigen Nutzen führe. Nicht die Dissemination einmal gesicherten Wissens, sondern die gestaltungsorientierte Transformation als Ziel der Evaluation stände dabei im Vordergrund.

In der anschließenden Podiumsdiskussion mit den Referierenden unter der Leitung von PD Dr. Rainer Strobl (proVal Hannover) wurde Unterschiedliches und Verbindendes der drei Inputvorträge herausgearbeitet.¹⁴

In der Diskussion wurde der Thematik eines „Wirkungshypes“ im Feld, im Sinne von wachsenden Forderungen nach Wirkungsnachweisen und damit verbundenen Zielstellungen nachgegangen: Hier wurde einerseits konstatiert, dass Wirkungsanalysen gesellschaftliche Relevanz besitzen und es je nach Kontext sinnvoll und

13 Prof. Dr. phil. habil. Kurt Möller ist Lehrstuhlinhaber zu Theorien und Konzepten Sozialer Arbeit, Jugendarbeit, Jugendpolitik, Kultur- und Bildungsarbeit an der Hochschule Esslingen. Seine Forschungsschwerpunkte liegen auf Gewalt, Rechtsextremismus, Fremden- und Menschenfeindlichkeit, männlicher Sozialisation, Jungen- bzw. Männerarbeit, politischer Sozialisation sowie auf Jugendkulturen.

14 Die Diskussion orientierte sich an den Standards der DeGEval (vgl. DeGEval – Gesellschaft für Evaluation e.V. 2017) sowie an Fragen der Programmsteuerung (z.B. evidenzbasierte oder -orientierte Praxis) und ging zudem auf Fragen der Teilnehmenden ein.

erstrebenswert sein kann, auch Veränderung bei den Zielgruppen zu erfassen. Dies kann aus Sicht von Andreas Armbrorst bspw. notwendig sein, um einer Instrumentalisierung von Befunden durch politische Akteure gesellschaftlich entgegenzusteuern. Maria Walsh verwies hier auf das Potenzial der Wissenschaft, zu Veränderungen beizutragen, bspw. mit Blick auf Gesetzesänderungen. Von weiteren Teilnehmenden wurde zudem angemerkt, dass insbesondere in einem auch hoch sicherheitsrelevanten Feld gesicherte Aussagen notwendig seien. Vertreterinnen und Vertreter der Fachpraxis argumentierten nicht nur aus gesellschaftlicher Perspektive und betonten den Wunsch, über Evaluationen Auskunft über die Wirkung der eigenen Arbeit zu erhalten, um bspw. die Weiterführung von Maßnahmen zu legitimieren. Thematisiert wurde auch der Bedarf, vorhandene Ergebnisse politisch anschlussfähig zu kommunizieren.

Gleichzeitig wurde die Diskussion um den „Wirkungshype“ – mit Rückblick auf die unterschiedlichen Inputs – auch mit der Frage verbunden, welche Kriterien ein Programm, ein Projekt oder eine Maßnahme für Wirkungsmessungen erfüllen muss und für welche Fragestellungen und Zielsetzungen von Evaluation diese geeignet sind. Christian Lüders betonte, dass der „Reifegrad“ von Projekten und ein fundiertes Verständnis der eigenen Praxis bzw. das Vorhandensein einer Programmtheorie dafür maßgebliche Faktoren seien. Weitgehende Einigkeit herrschte dahingehend, dass klare, überprüfbare Zielvorstellungen als eine Voraussetzung zwingend vorhanden sein müssen. Überdies erinnerte Kurt Möller daran, dass Evaluationen nicht zentral wirkungsorientiert angelegt sein müssten, sondern auch andere Funktionen erfüllten und Schwerpunkte (bspw. auf Prozesse, Strukturen oder Konzepte) setzen könnten, die dann symmetrisch und dialogisch mit der Praxis bearbeitet werden könnten.

Darüber hinaus entfachte sich eine kontroverse Diskussion darüber, welche Ansätze und Methoden angemessen seien, um Wirkungen aufzuspüren und nachzuweisen. Andreas Armbrorst sprach sich hierbei dafür aus, bei der Forderung nach Wirkungsnachweisen durch Auftraggebende möglichst randomisierte Kontrollgruppendesigns als „Goldstandard“ umzusetzen oder sich dem zumindest mit einem Vergleichsgruppendesign anzunähern. Von den anderen Podiumsteilnehmern wurde dies als gegenstandsunsensibles Vorgehen kritisch in Frage gestellt. Ein solches Vorgehen sei insbesondere in pädagogischen Kontexten mit komplexen Prozessen nicht tragfähig und könne zudem zu negativen Nebeneffekten beitragen, bspw. zu einer durch Manualisierung und Standardisierung vorangetriebenen fehlenden Flexibilisierung pädagogischer Praxis. Ferner wurde die Frage aufgeworfen, ob objektivierte Wirkungsfeststellungen hierbei letztendlich nicht Konstruktionen seien. Aus Sicht von Christian Lüders besteht der Dissens dabei nicht darin, ob Wirkungsanalysen an sich nötig seien, sondern gehe über ein Verständnis von „richtiger Wissenschaft“ mit – von einer Teilgruppe vermeidlich gesetzten – Standards, die dann aber der Praxis und den Bedingungen in diesem Feld nicht gerecht würden.

3. Wo treffen sich Praxis, Politik und Evaluationsforschung?

Der zweite Tagungstag brachte im Rahmen von drei parallel stattfindenden Workshops zunächst Evaluierende, Programmgestaltende und Praktiker(innen) zu ausgewählten Themen des Vortrags in den näheren Austausch. Die Workshops waren durch Kurzbeiträge strukturiert, die als Input für die weitere Diskussion dienten.

Im ersten Workshop (Workshopleitung Stefanie Reiter) zur „Evaluation von Präventionsprojekten“ berichtete Maria Kamenowski (Zürcher Hochschule für Angewandte Wissenschaften) von der „Evaluation von Pilotprojekten zu Gegennarrativen und alternativen Narrativen zur Prävention von Radikalisierung im Netz (Schweiz)“. Das Schweizer Bundesamt für Sozialversicherungen hat hierzu vier Projekte ausgewählt, in denen Gegennarrative bzw. alternative Narrative zur Prävention von Radikalisierung im Netz erarbeitet werden sollen. Die Projekte zeichnen sich dadurch aus, dass partizipativ mit den Jugendlichen bzw. jungen Erwachsenen Narrative ausgearbeitet werden. Für die Untersuchung sowohl der Projektprozesse als auch der Wirkung der Narrative wurden sieben Module entwickelt, anhand derer die Projekte evaluiert werden. Die methodischen Module und auch die Herausforderungen der Evaluation standen im Fokus des Vortrags. In einem zweiten Input stellten Josephina Schmidt und Athanasios Tsirikiotis (Institut für partizipative Evaluation und Sozialforschung, IPES) die „Formative Evaluation des Präventionsprogramms ‚ACHTUNG?!‘ – Qualitative Methoden und Zielentwicklung“ vor. Das Programm des Polizeipräsidiiums Ludwigsburg möchte mit mehreren, durch unterschiedliche Akteure gestalteten Modulen einen Beitrag zur Radikalisierungsprävention leisten. Bei der formativen Evaluation kam der Zielfindung eine zentrale Rolle zu. Hieran wurden die unterschiedlichen Perspektiven der Stakeholder des Projekts veranschaulicht.

Im zweiten Workshop „Aussteigerprogramme für Rechtsextremist*innen“¹⁵ (Workshopleitung Frank Buchheit) referierten Dr. Frank Greuel und Frank König (DJI Halle) zum Thema „„doing Ausstieg“ – Gegenstandskonstruktionen in der Ausstiegsarbeit und die Rolle von Evaluation“. Im Beitrag wurde thematisiert, dass das Ziel von Ausstiegsangeboten meist klar scheine: Personen bei ihrem Ausstieg zu unterstützen und so viele wie möglich zum Ausstieg zu bewegen. Jede Aussteigerin und jeder Aussteiger gelten damit als Projekterfolg. Die genaue Bestimmung des Ziels „Ausstieg“ erweise sich in der Praxis jedoch als überaus schwierig und unterscheide sich zwischen den verschiedenen Angeboten. Während teilweise bereits das Lösen von problematischen Szenezugehörigkeiten als Ausstieg gelte und als (abschließender) Erfolg gewertet werde, gelten in anderen Angeboten erst die dauerhafte Abkehr von der Szene und demokratische Haltungen als Ende eines gelingenden Unterstützungsprozesses. Im Workshop wurde diskutiert, wie Evaluation im Feld

15 Ein zweiter, für diesen Workshop geplanter Beitrag zur Evaluation des Aussteigerprogramms EXTRA von Florian Neuscheler (Hochschule Esslingen) musste kurzfristig entfallen. Das beim Landesamt für Verfassungsschutz in Sachsen-Anhalt angesiedelte Aussteigerprogramm für Rechtsextremisten wurde durch die Hochschule Esslingen im ersten Schritt formativ evaluiert. Aktuell findet der zweite, summative Evaluationsschritt statt. Der Referent wollte über methodische Herausforderungen und die Lösungsansätze dieses Evaluationsprojekts berichten.

der Deradikalisierung mit den zugrunde liegenden verschiedenen Gegenstandskonstruktionen umgehen und Erfolgskriterien verbindlich formulieren kann.

Im dritten parallel stattfindenden Workshop zur „Evaluation von Programmen“ (Workshopleitung Rainer Strobl) stellte Verena Fiebig (KPEBW) „Strukturelle Qualitätsstandards als Ausgangspunkt der Bewertung von Deradikalisierungsprogrammen“ zur Diskussion. Sie erläuterte, dass Programme zur Deradikalisierung und Extremismusprävention in den letzten Jahren einen festen Platz in vielen nationalen und internationalen Strategien zur Terrorismusbekämpfung gefunden haben. Doch die Frage, wie sich der Erfolg und die Qualität von Deradikalisierungsprogrammen einschätzen lässt, sei bisher weitestgehend offen. Wissenschaftler(innen) kritisieren seit Jahren die mangelhafte konzeptionelle Klarheit und massive Intransparenz der überwiegenden Mehrheit dieser Programme, obwohl der grundsätzliche Wert der Deradikalisierung nicht bestritten werde. Mit den von ihr vorgestellten strukturellen Qualitätsstandards in der Interventions- und Präventionsarbeit gegen gewaltbereiten Extremismus wird eine erste Grundlage für die strukturellen Mindestanforderungen für Programme, Träger und Initiativen vorgeschlagen, welche als Leitfaden für eine mehrstufige Evaluation und Qualitätssicherung der Extremismusprävention fungieren könne. Neben der Vorstellung der strukturellen Qualitätsstandards wurden die Chancen und Grenzen dieser Evaluationsmethode diskutiert.

Die Weiterführung der Workshopkenntnisse erfolgte im Anschluss im Format einer Fishbowl-Diskussion (Innenkreis-/Außenkreis-Methode) mit den Workshopleitenden Frank Buchheit, Stefanie Reiter und Rainer Strobl als feste Diskutant(inn)en sowie weiteren Teilnehmenden. Edith Halves von der Hochschule für Angewandte Wissenschaften Hamburg leitete die abschließende Diskussion. Diese diente einerseits der Darstellung der Workshopergebnisse und bot andererseits einen abschließenden Diskussionsraum, die Erträge der Tagung und ihre Übertragbarkeit in die Praxis zu bestimmen. Auch in diesem Format wurden die Diskussionen über Begrifflichkeiten und unterschiedliche Ansätze fortgeführt. Vielfach wurde auf die Potenziale der Zusammenarbeit zwischen Wissenschaft und Praxis im Rahmen einer formativ ausgerichteten Evaluation und den sich daraus ergebenden Mehrwert für die Projekte rekurriert. In diesem Zusammenhang wurde betont, dass die gezielte Einplanung von Ressourcen und eine frühzeitige Einbindung von möglichst allen relevanten Stakeholdern wünschenswert sei. Hier gelte es, auch proaktiv auf Auftraggebende und Politik (Ministerien/Entscheidungsträger in Ausschüssen) zuzugehen, um deren Vorverständnis von Phänomenen und Feldern im Bereich der Extremismusprävention und Deradikalisierung zu beeinflussen und sie für gegenstandsgemessene Evaluationen zu sensibilisieren.

4. Fazit und Ausblick

Die mit über 60 Teilnehmenden ausgebuchte Frühjahrstagung führte im Handlungsfeld tätige Evaluierende sowie Programmverantwortliche zusammen. An zwei Tagen wurden intensiv Möglichkeiten diskutiert, wie mithilfe von evaluativen Ver-

fahren eine größere Wirkung und Effizienz in der Prävention und Intervention von Radikalisierungen zu erreichen ist. Hierbei wurden unterschiedliche Perspektiven, Ansätze und Methoden mit ihren Potenzialen und Grenzen aufgezeigt. Die Podiumsdiskussion und die Fishbowl-Diskussion verdeutlichten zum Abschluss der beiden Tagungstage Unterschiedliches und Verbindendes der Inputvorträge und der Workshopergebnisse. Erwartungsgemäß zeigte sich, dass nicht nur die Breite unterschiedlicher Perspektiven und Ansätze, sondern auch die verwendeten Begrifflichkeiten im komplexen Handlungsfeld „deradikalisierende Maßnahmen“ kontroverse Diskussionen anregen und an vielen Stellen Anschlussfragen aufwerfen und Forschungslücken aufzeigen. Dank der regen Beteiligung und Offenheit der Tagungsteilnehmenden kam ein spannender Austausch zustande, der als ein kleiner Schritt auf dem Weg zu einer Intensivierung des Erfahrungsaustausches über Evaluationen in diesem Feld, zu einer konstruktiven Fehlerkultur und zu gegenseitigem Lernen gesehen werden kann. Es gab ein starkes Interesse, den Diskurs sowohl über Begrifflichkeiten als auch über die Gegenstandsangemessenheit von Methoden in diesem Feld weiterzuführen.

Literatur

DeGEval – Gesellschaft für Evaluation e.V. (Hg.) (2017): Standards für Evaluation. Erste Revision 2016. Mainz.

Nachwuchspreis der DeGEval 2018

Edith Halves, Christiane Kerlen, Philipp Mayring, Jessica Prigge, Sonja Sheikh, Kerstin Wilhelm, Brigitte Wotha

Der DeGEval-Nachwuchspreis wurde auf der 21. Jahrestagung der DeGEval in Dresden zum inzwischen 12. Mal vergeben. Ziel des Preises ist es, jährlich eine Nachwuchsevaluatorin/einen Nachwuchsevaluator oder auch eine Gruppe von Nachwuchsevaluatoren und Nachwuchsevaluatorinnen auszuzeichnen und damit eine herausragende Arbeit im Bereich Evaluation im deutschsprachigen Raum zu würdigen. Aus Sicht der DeGEval soll dieser Preis die Bedeutung des wissenschaftlichen Nachwuchses auf dem Gebiet der Evaluation sowie der Nachwuchsförderung insgesamt hervorheben, der sich die DeGEval in besonderem Maße verpflichtet fühlt.

Für den Nachwuchspreis 2018 wurden insgesamt fünf Beiträge eingereicht: zwei Dissertationen, eine Masterarbeit und zwei Praxisarbeiten. Die Entscheidung war wie so oft keine einfache. Schließlich hat sich die Jury für die Masterarbeit entschieden, die sie auch deshalb so überzeugt hat, weil sie weit über die Ansprüche an eine Maserarbeit hinausgeht.

In der Arbeit wird untersucht, welche Bedeutung Evaluation als Policyinstrument auf kommunaler Ebene hat und damit eine sehr interessante Forschungslücke behandelt. Es gibt bisher noch keine Arbeiten, die sich mit dem Zusammenspiel verschiedener evaluativer Instrumente innerhalb einer Kommune beschäftigen. Die ausgezeichnete Masterarbeit untersucht dies am gewählten Fall der Stadt Köln und trägt damit in beeindruckender Weise zur Weiterentwicklung der Evaluationspraxis im deutschsprachigen Raum bei.

Einen besonderen Mehrwert für die Evaluierungscommunity stellt dabei das im Rahmen der Arbeit entwickelte Analyseraster dar, um die Relevanz von Evaluierungen zu erfassen. Es ermöglicht die systematische Einschätzung der Bedeutung von Evaluierung auch wenn die entsprechende Daten- und Informationslage stark begrenzt ist, was am Beispiel der Stadt Köln sehr eindrücklich gezeigt wird.

Die Arbeit sticht sowohl in der Herleitung des Ansatzes als auch in dessen methodischer Umsetzung hervor. Die Triangulation verschiedener Methoden und Perspektiven sowie der Umfang der innerhalb der Arbeit erhobenen und analysierten Daten sind beeindruckend, ebenso wie die abschließend praxisnah, realistisch und präzise formulierten Empfehlungen an die Stadt Köln.

Die Autorin leistet mit ihrer Arbeit einen wichtigen Beitrag zur systematischen Einschätzung der Bedeutung von Evaluierung in Kommunen, der sich jedoch auch auf andere Anwendungsbereiche außerhalb der Kommunalpolitik übertragen lässt.

Damit ging der 12. DeGEval Nachwuchspreis 2018 an Frau Barbara Flatters für ihre Masterarbeit mit dem Titel „The Relevance of Evaluation in German Municipalities“. Wir gratulieren Frau Flatters ganz herzlich zu dieser hervorragenden Forschungsarbeit und hoffen auf zahlreiche weitere Beiträge von ihr zur Praxis und Theorie der Evaluation.

Wir freuen uns, auch im Jahr 2018 den Preis wieder an eine Masterarbeit vergeben zu können und hoffen damit auch weitere Studierende für eine Einreichung ihrer Masterarbeiten für den nächstjährigen DeGEval-Nachwuchspreis zu motivieren.

Herzlichen Glückwunsch an Frau Flatters!

Die Jury des Nachwuchspreises 2018

DeGEval-Nachwuchspreis zur Auszeichnung von Beiträgen zur Evaluationsforschung oder Evaluationspraxis

Der Preis wird einmal jährlich für eine herausragende Arbeit im Bereich Evaluation im deutschsprachigen Raum an eine/-n Nachwuchsevaluator/-in oder eine Gruppe von Nachwuchsevaluatoren/-innen vergeben (bei Gruppenarbeiten muss die Gesamtgruppe aus Nachwuchsevaluator/-innen bestehen). Die prämierte Arbeit soll einen fundierten Beitrag zur fachlichen und professionellen Weiterentwicklung von Evaluation leisten. Einschlägig sind sowohl Beiträge zur Theorie und Methodik der Evaluation (referierte Zeitschriftenartikel, Dissertationen oder Masterarbeiten) als auch beispielhafte Evaluationsberichte.

**Die Höhe des Preisgeldes beträgt Euro 1.000,-
Die Vergabe des Preises erfolgt im Rahmen der
22. Jahrestagung der DeGEval**

Die DeGEval dankt dem Waxmann Verlag für die Spende. Eine zusätzliche Ehrung erfolgt über eine Meldung in der Zeitschrift für Evaluation (ZfEv) und eine Vorstellung der Preistragenden auf der Homepage der DeGEval. Ein Abstract der Arbeit wird in der ZfEv veröffentlicht. Weitere Formen der Veröffentlichung, z.B. als Originalartikel in der ZfEv oder als Download auf der Internetseite der DeGEval werden mit den Preistragenden vereinbart.

Jeder Bewerbung müssen zwei einseitige schriftliche Befürwortungen renommierter Evaluatoren/-innen beiliegen. Alternativ kann eine der beiden Befürwortungen auch von Auftraggebenden erstellt werden. Zum Zeitpunkt der Preisverleihung sollen die Preistragenden das 38. Lebensjahr noch nicht vollendet haben und keine Lebenszeitprofessur oder vergleichbare Position innehaben.

Einreichungen werden **bis zum 1. Juli 2019** erbeten. Sie sollen – neben der vorgeschlagenen Arbeit – ein Schriftenverzeichnis und Curriculum Vitae der vorgeschlagenen Person enthalten. Im Falle der Fremdnomination soll auch ein Curriculum Vitae der vorschlagenden Person beigelegt werden.

Die Vorschläge sind per E-Mail zu richten an: info@degeval.de.

Bitte beachten Sie für Ihre Einreichung unbedingt unsere Datenschutzbestimmungen (<https://www.degeval.org/de/kontakt/datenschutz/>) sowie die Hinweise zur Einreichung für den Nachwuchspreis auf unserer Homepage (<https://www.degeval.org/veranstaltungen/jahrestagungen/bonn-2019/degeval-preise/>).

Die Auswahl der Preistragenden erfolgt unter Ausschluss des Rechtsweges durch eine Jury sowie den Vorstand der DeGEval auf Basis folgender **Bewertungskriterien**:

- A) Wissenschaftliche Arbeit (Dissertation, wissenschaftliche Artikel über Evaluation)
- Bearbeitung und Weiterentwicklung von innovativen Themenfeldern der Evaluation
 - Theoretische Fundierung sowie Weiterentwicklung theoretisch-konzeptioneller Diskussionslinien
 - Nachvollziehbar sinnvolle Verknüpfung von theoretischen und empirischen Ansätzen
 - Kritische Reflexion/Diskussion der Ergebnisse samt ihrer Aussagekraft mit explizitem Rückbezug auf die aktuelle wissenschaftliche Diskussion
 - Eigenständige Vorschläge zur Weiterentwicklung theoretischer und/oder praxisorientierter Ansätze
- B) Qualifikationsarbeit (Masterarbeit)
- Bearbeitung von relevanten Fragestellungen der Evaluation oder eines Evaluationsgegenstandes
 - Darstellung theoretisch-konzeptioneller Diskussionslinien
 - Nachvollziehbarkeit der methodischen Umsetzung der Untersuchung
 - Darstellung der Relevanz der Ergebnisse für die wissenschaftliche Diskussion oder den Evaluationsgegenstand
 - kritische Reflexion/Diskussion der erzielten Ergebnisse und ihrer Aussagekraft
- C) Praxisarbeit (Evaluationsbericht, Fachaufsatz über eine Evaluierung)
- Verknüpfung wissenschaftlicher Qualität mit Relevanz für Evaluationsgegenstand oder -feld
 - Nutzbarmachung von Evaluationsansätzen/-theorien und Methodologie für die Praxis
 - Klarer methodischer Zugang und adäquater Methodeneinsatz
 - Reflexion der Ergebnisse sowie ein klarer Rückbezug auf theoretisch-konzeptionelle Grundlagen
 - nachvollziehbar systematische Bewertung der Ergebnisse und Handlungsempfehlungen

Für Rückfragen wenden Sie sich bitte an die Geschäftsstelle der DeGEval
(E-Mail: info@degeval.de; Tel.: +49 6131 / 39-26869).

„Evaluation und Nachhaltigkeit“

**22. Jahrestagung 2019 der DeGEval –
Gesellschaft für Evaluation e.V.
vom 11. bis 13. September 2019 in Bonn**

**Lokale Organisation: Deutsches Evaluierungsinstitut
der Entwicklungszusammenarbeit**

Tagungsort: Universität Bonn

„Nachhaltigkeit“ ist ein Begriff, der aus politischen Diskursen, Positionspapieren oder Projektanträgen kaum mehr wegzudenken ist. Ursprünglich aus dem Umweltbereich stammend, wo er sich auf eine Ressourcennutzung bezieht, die Rücksicht auf die Regenerationsfähigkeit eines Systems nimmt, wird er heute viel breiter verwendet. Nachhaltigkeit dient auch alltagssprachlich oft als Synonym für die Dauerhaftigkeit von Wirkungen und Veränderungen. Die aktuelle Debatte um Nachhaltigkeit bezieht sich zudem auf das Zusammenspiel und die Wechselwirkungen zwischen ökologischer, ökonomischer und sozialer Entwicklung. Da für die Beurteilung von Maßnahmen deren Wirksamkeit ein wichtiges Bewertungskriterium ist, stellt Nachhaltigkeit für die Evaluation eine besondere Herausforderung dar.

Denn nachhaltige Wirkungen sind definitionsgemäß erst mit zeitlichen Verzögerungen nach Maßnahmenbeendigung zu beobachten, zu einem Zeitpunkt also, zu dem viele Evaluationen bereits abgeschlossen sind. Erschwerend kommt hinzu, dass der oft gewünschte Nachweis kausaler Zusammenhänge, der schon für kurzfristige Effekte herausfordernd ist, umso schwieriger wird, je längerfristig nachhaltige Veränderungen nachgewiesen und je stärker auch multidimensionale Wechselwirkungen erfasst werden sollen.

Neben dem Umweltbereich hat Nachhaltigkeit auch im Bereich der Entwicklungszusammenarbeit eine lange Tradition als Bewertungskriterium. Dort wurde sie durch die OECD Anfang der 1990er Jahre zu einem einschlägigen Bewertungskriterium für sämtliche Evaluierungen erhoben. In Deutschland kommt dies seit Beginn der 2000er Jahre flächendeckend zur Anwendung. Das Verständnis und der Umgang mit Nachhaltigkeit wird im Zuge der globalen Agenda 2030 für Nachhaltige Entwicklung noch bedeutsamer. Erstens weil die Agenda 2030 die Evaluationsfunktion als genuine Bestandteil nachhaltiger Entwicklung begreift und zudem, weil die Agenda einen universellen, politikfeldübergreifenden Anspruch erhebt.

Vor diesem Hintergrund ist Nachhaltigkeit und ihre Rolle für und in Evaluationen das zentrale Thema der 22. Jahrestagung der Gesellschaft für Evaluation. Diskutieren wollen wir u.a. die folgenden Fragen:

- **Nachhaltigkeit in unterschiedlichen Evaluationsfeldern**

- Wie wird Nachhaltigkeit in unterschiedlichen Politik- und Anwendungsfeldern der Evaluation definiert bzw. interpretiert und welche Rolle spielt sie jeweils als Bewertungskriterium? Gibt es ein übergreifendes Verständnis?
- Welche übertragbaren Erfahrungen mit Nachhaltigkeit gibt es im Bereich der Entwicklungszusammenarbeit im Kontext der DAC-Kriterien?
- Welche Erfahrungen liegen aus dem Umweltbereich vor, von wo der Begriff der Nachhaltigkeit ursprünglich stammt, und welche Implikationen ergeben sich aus dem breiteren Begriffsverständnis, das heute üblich ist?
- Zunehmend verfolgen auch privatwirtschaftliche Akteure in ihren Strategien ambitionierte Beiträge mit Blick auf nachhaltige Entwicklung. Welche Monitoring- und Evaluationsinstrumente werden hierbei genutzt, um Effektivität und Nachhaltigkeit von privatwirtschaftlichen Maßnahmen zu messen?

- **Untersuchung von Nachhaltigkeit**

- Wie kann die Nachhaltigkeit in Evaluationen methodisch angemessen untersucht werden? Inwiefern und unter welchen Bedingungen ist hier der Anspruch eines Nachweises kausaler Zusammenhänge erfüllbar insbesondere mit Blick auf die Mehrdimensionalität und Wechselwirkungen von Nachhaltigkeit?
- Gibt es verlässliche Proxy-Variablen, die noch während oder mit Ende der Laufzeit von Maßnahmen als Indikatoren für die spätere Nachhaltigkeit dienen können?
- Welches Potenzial haben Ex-ante-Evaluationen, um die voraussichtliche Nachhaltigkeit von geplanten Vorhaben prognostizierend abzuschätzen?
- Welche Rolle können die Sustainable Development Goals (SDG) und der Indikatorenkatalog der UN für die Evaluation von Nachhaltigkeit einnehmen?
- Welche Möglichkeiten bieten Querschnittsauswertungen zum Thema Nachhaltigkeit und auf welche Weise ermöglichen Auswertungen über Einzelfallstudien hinaus systematisches Lernen im Sinne der gemeinsamen Verantwortung?

- **Nachhaltigkeit und Programmevaluation**

- Braucht es ganz andere Formen und Konstellationen als die klassische, meist begleitend angelegte Programmevaluation, um Nachhaltigkeit angemessen in den Blick nehmen zu können?
- Inwiefern sind Ex-post-Evaluationen angemessen, um den Herausforderungen bei der Evaluation von Nachhaltigkeit zu begegnen?
- Welche Rolle spielen Konzepte wie Implementation, Diffusion oder Transfer bei der Förderung von Nachhaltigkeit und wie sind diese angemessen in Evaluationen zu adressieren?
- Wie erfolgt eine angemessene Abgrenzung von Outcome-, Impact- und Nachhaltigkeitsmessung in Evaluationen?

- **Beiträge der Evaluation zur Nachhaltigkeit**

- Was kann Evaluation zur Nachhaltigkeit von Wirkungen der evaluierten Maßnahmen beitragen?
- Welche inhaltsspezifischen Kenntnisse brauchen Evaluierende, um gute Empfehlungen und Prognosen zur Nachhaltigkeit von Maßnahmen geben zu können?
- Unter welchen Bedingungen können Evaluationen selbst nachhaltig wirken?

Nähere Informationen zur Tagung finden Sie unter
<http://www.degeval.de/veranstaltungen/jahrestagungen/bonn-2019>.

Kontakt:

DeGEval – Gesellschaft für Evaluation e.V.
Frau Mareike Hinz
Johannes Gutenberg-Universität Mainz
c/o Zentrum für Qualitätssicherung und -entwicklung (ZQ)
Colonel-Kleinmann-Weg 2, SBII, 03-246
D-55099 Mainz
Tel.: +49 (0) 6131 / 39-2 68 69
Fax: +49 (0) 6131 / 39-2 68 68
E-Mail: info@degeval.org
<http://www.degeval.org>