

Positionspapier des DeGEval-Vorstands zur Zukunft der Evaluation

Die 20. Jahrestagung der DeGEval – Gesellschaft für Evaluation e.V. stand unter dem Motto „Evaluation (in) der Zukunft“. Die Jubiläumstagung wurde genutzt für einen Blick zurück auf 20 Jahre DeGEval, in der die Gesellschaft ihren Teil zur jüngeren Erfolgsgeschichte der Evaluation im deutschsprachigen Bereich beigetragen hat. Gleichzeitig zeigte der Blick nach vorn, dass eine Fortsetzung dieser Erfolgsgeschichte kein Selbstverständnis ist, obwohl gute Evaluation mehr denn je in verschiedensten gesellschaftlichen Handlungsfeldern gebraucht würde.

Der DeGEval-Vorstand greift in den folgenden Positionen die Diskussionen und Ergebnisse der Tagung auf.

Evaluation und Zivilgesellschaft

- In jüngerer Zeit werden bewusste Fehlinformationen in bisher nicht gekanntem Ausmaß als Instrument der öffentlichen Auseinandersetzung genutzt. Gleichzeitig wird von Teilen der Öffentlichkeit wissenschaftliche Evidenz in substanziellem Ausmaß nicht mehr nur skeptisch hinterfragt, sondern grundsätzlich abgelehnt und durch selbstrekursiv erzeugte ‚Wirklichkeiten‘ ersetzt. Als Fachgesellschaft für Evaluation beobachten wir diesen Trend mit großer **Aufmerksamkeit und Sorge**.
- Wir setzen dagegen, dass gute Evaluation ein starkes Werkzeug für mehr Transparenz staatlichen und nicht staatlichen Handelns sein kann. Durch die systematische Überprüfung von Maßnahmen gelangt sie in nachvollziehbarer Weise zu fundierten Bewertungen der Planung, Umsetzung und Wirkung von Maßnahmen. Evaluation verstehen wir somit als ein wichtiges Instrument zur **Stärkung der zivilgesellschaftlichen Auseinandersetzung und Entscheidungsfindung**.
- In diesem Sinne fordern wir alle, die gleichermaßen an einer Stärkung der Zivilgesellschaft interessiert sind, dazu auf, verstärkt die **Evaluation öffentlichen Handelns einzufordern und zu unterstützen**. Dazu gehört,
- Politik, Verwaltung, öffentliche Institutionen, Stiftungen und andere Akteure staatlichen und nicht staatlichen Handelns nach der nachweisbaren Fundierung von Entscheidungsprozessen zu fragen,
- zu verlangen, dass Maßnahmen auf Wirkungen und Wirkweise hin überprüft werden und

- dass vorhandene Evidenz grundsätzlich in politischen Entscheidungsprozessen berücksichtigt wird.

Politik und Verwaltungen

- In der Politik zeigt sich in den vergangenen Jahren eine stetige Zunahme der **Verwendung des Evaluationsbegriffs**, beispielsweise in der parlamentarischen Arbeit. Dieser Zuwachs verbleibt bisher allerdings vorwiegend im Nominellen, da eine gewachsene Nachfrage nach Evaluationen etwa aus den Parlamenten bisher kaum zu verzeichnen ist. Für die Evaluation in der Politik ist also nach wie vor ein Umsetzungsdefizit zu konstatieren.
- Gleiches gilt für die Verwaltung. Auch wenn in der Bundeshaushaltsordnung festgeschrieben ist, dass Zielerreichungs-, Wirkungs- und Wirtschaftlichkeitskontrollen durchzuführen sind, wird dies keineswegs **flächendeckend und kontinuierlich** umgesetzt. Es mangelt an einer systematischen Vorbereitung, an der Definition messbarer Ziele, an Ressourcen und Kompetenzen zur Durchführung von Effektivitäts- und Effizienzbetrachtungen. Der Bundesrechnungshof weist in seinen jährlichen Gutachten regelmäßig auf dieses Manko hin.
- Auch die Politik ist also dazu aufgefordert, Evaluation stärker als bisher **im politischen Prozess** zu nutzen. Insbesondere kann Evaluation hier im Zuge der parlamentarischen Kontrolle zur Überprüfung von Maßnahmen hinsichtlich ihrer Umsetzung und Wirkungen dienen. Durch ihre explizite Bewertungsperspektive ist sie nur beschreibend angelegten Steuerungsinstrumenten wie Auditing und Monitoring hierbei überlegen.

Evaluation in Organisationen und Institutionen

- Viele öffentliche, nicht kommerzielle und privatwirtschaftliche Organisationen und Institutionen nutzen Evaluation. Zu ihnen gehören etwa Verwaltungen, Bildungs- und Gesundheitseinrichtungen, Stiftungen oder Unternehmen, die ihre Aktivitäten und Maßnahmen intern evaluieren oder Evaluationen extern als Dienstleistung beauftragen.
- In Zukunft ist hier mit einer **wachsenden Bedeutung interner Evaluationen** zu rechnen, bei denen Untereinheiten der Organisation Evaluationsaufgaben wahrnehmen. Wichtig ist, dass auch diese den professionellen Ansprüchen guter Evaluation unterliegen, wie sie in den *Standards für Evaluation* der DeGEval formuliert sind.
- In verschiedenen Bereichen, wie etwa dem Hochschulsektor, wird die **Sinnhaftigkeit und der Nutzen insbesondere von verstetigten Evaluationssystemen** immer wieder kritisch hinterfragt. Wir halten dieses Hinterfragen für legitim, da die Standards für Evaluation *Nützlichkeit* als erstes Kriterium guter Evaluation definieren.

- In Organisationen und Institutionen kann die Nützlichkeit von Evaluationen am ehesten gewährleistet werden, wenn Evaluation eng in organisationale Strukturen und Prozesse eingebettet ist. Dies kann am besten durch eine **Evaluation Policy** gewährleistet werden. Diese klärt für einen bestimmten Geltungsbereich, zu welchen Zwecken, wann, wie oft, von wem was evaluiert wird und legt Ort und Verantwortung für die Verwendung der Ergebnisse vorab fest; gleichzeitig sorgt sie für einen bewussten Ressourceneinsatz und beugt ritualisierten Evaluationsformen vor.
- Wir ermuntern alle Organisationen und Institutionen, die Evaluation nutzenbringend einsetzen wollen, zur **Entwicklung einer Evaluation Policy** in diesem Sinne.

Evaluation und Wissenschaft

- Es ist als Erfolg zu werten, dass Evaluation zunehmend als akademisches Forschungsfeld vertreten ist. Dies äußert sich u.a. in einer Vielzahl von Professuren, die Evaluation als Teil ihrer Denomination tragen, oft in Ergänzung zu Forschungsmethoden oder etwa fachspezifischen Themenfeldern.
- Wie u.a. die *Standards für Evaluation* der DeGEval belegen, hat Evaluation aber eine eigene **transdisziplinäre Fachlichkeit**. Diese lässt sich nicht alleine auf methodische Aspekte reduzieren, denn gute Evaluation besteht nicht alleine in der Anwendung sozialwissenschaftlicher Methoden.
- Für eine Weiterentwicklung der Evaluation ist daher die Einrichtung dezidierter **Professuren für Evaluation** an Hochschulen in Deutschland und Österreich erforderlich, die dieser transdisziplinären Fachlichkeit Rechnung tragen. Diese sind auch eine wichtige Voraussetzung, damit die vielen Nachwuchsforscherinnen und -forscher, die sich im Bereich der Evaluation qualifizieren, dieses Forschungsfeld auch als expliziten Karriereweg weiterverfolgen können.
- Wie andere Tätigkeitsfelder auch, ist Evaluation auf eine empirische Wissensbasis für ihr professionelles Handeln angewiesen. Bisher erfolgt die entsprechende Forschung allerdings meist im Rahmen fachbezogener Untersuchungen (z.B. Nutzung von Evaluationsergebnissen in Schulen), was zu einer fragmentierten Wissensbasis führt. Notwendig ist daher eine verstärkte explizite **Forschung über Evaluation**, die sich transdisziplinär versteht und den Austausch zur Evaluationsforschung in verschiedenen Handlungsfeldern sucht.

Professionalisierung der Evaluation

- Evaluation ist ein **ungeschützter Begriff**. In der Praxis werden viele sehr unterschiedliche Aktivitäten Evaluation genannt. Da das Feld der Evaluation noch relativ jung ist und dynamischen Veränderungen unterliegt, erscheinen uns Bemühungen zu einer Schließung des Zugangs zur Tätigkeit in der Evaluation als verfrüht.

- Dieser offene Zugang sollte aber nicht als Freifahrtschein zur Beliebigkeit missverstanden werden. Wichtigster **fachlicher Bezugspunkt** für Evaluation sind **professionelle Standards** guter Evaluation. Die Standards der DeGEval fordern, dass gute Evaluationen sich nicht auf ein reines Messen beschränken, sondern **nützlich, fair, durchführbar** und **genau** sein sollen. Wir rufen alle auf, die Evaluationen durchzuführen und diese als Dienstleistung anbieten, sich der Einhaltung der DeGEval-Standards für Evaluation zu verpflichten.
- Ebenso sind auch jene, die Evaluationen ausschreiben und in Auftrag geben, aufgerufen, sich ihrer Verantwortung für die Ermöglichung guter Evaluationen bewusst zu sein. Auch **Auftraggebende von Evaluationen** können durch realistische Erwartungshaltungen, rechtzeitige Beauftragungen und angemessene Ausstattung zur Qualität von Evaluationen beitragen.

Evaluationspraxis

- Die Evaluation hat als junge, fächerüberspannende Disziplin bisher nur in Ansätzen eine einheitlich akzeptierte, eindeutige Fachterminologie herausgebildet. Für die Professionalisierung einer Tätigkeit ist die Entwicklung einer **eindeutigen Fachsprache** eine wichtige Voraussetzung der internen und externen Verständigung. Mit den neu revidierten Standards für Evaluation hat die DeGEval ein **Glossar zentraler Evaluationsbegrifflichkeiten** verabschiedet, das im Zweifel als Bezugspunkt für die Klärung von terminologischen Fragen herangezogen werden sollte.
- Bei zunehmend eingesetzter Evaluation, wie sie in wenigen Bereichen bereits beobachtet werden kann, ist darauf hinzuwirken, dass keine negativen Sättigungseffekte eintreten. Evaluation ‚von der Stange‘ läuft Gefahr, zum gut ausgebauten Monitoring zu werden und den Blick auf das Nichtintendierte, das Ungewöhnliche zu behindern. Dieses Risiko speist sich sowohl aus Routineevaluationen der Evaluierenden als auch aus detaillierten und umfangreichen Leistungsbeschreibungen auf der Seite von Auftraggebenden für Evaluation.

Die Diskussionen und Debatten der 20. Jahrestagung der DeGEval zeigten auch, dass Werkzeuge und Methoden für eine Evaluation der Zukunft zur Verfügung stehen. Entwicklungen lassen sich systematisch extrapolieren, es lassen sich Szenarien entwickeln, Trendbrüche identifizieren und mit Wahrscheinlichkeiten des Eintretens hinterlegen. Die beste Voraussetzung, die Zukunft aktiv zu gestalten, besteht darin, gemeinsam über Vergangenes und Bestehendes zu debattieren und Wünsche und Ansprüche auszuformulieren. Hierzu wird die DeGEval auch in den kommenden Jahren beitragen, indem sie durch Newsletter, Zeitschrift, Tagungen und Workshops die Möglichkeit bietet, in Austausch zu treten. Dazu sind alle an Evaluation Interessierten herzlich eingeladen: externe und interne Evaluatorinnen und Evaluatoren, Auftraggebende von Evaluation, Betroffene von Evaluation und Entscheidungsträger in Verwaltung und Politik. Wir würden uns freuen, wenn Sie das vorliegende Positionspapier zum Anlass nehmen, mit uns ins Gespräch zu kommen!

DeGEval-Vorstand

Prof. Dr. Jan Ulrich Hense, Vorstandsvorsitzender
Dr. Christiane Kerlen, stellvertretende Vorstandsvorsitzende
Dr. Marianne Lück-Filsinger
Dipl.-Soz.-Arb. Stefan Schmidt
Dr. Sonja Sheikh

<https://www.degeval.de/de/ueber-uns/vorstand/>

Evaluation (in) der Zukunft. Bericht zur 20. Jahrestagung der DeGEval

Uwe Schmidt

20 Jahre DeGEval markieren eine eindrucksvolle Geschichte des Aufbaus einer Fachgesellschaft, die ihren Anfang in den 1990er Jahren auf Initiative einer Handvoll Interessierter an Fragen der Evaluation nahm und zur Gründung der seinerzeit noch auf Deutschland beschränkten Deutschen Gesellschaft für Evaluation führte. Die spätere Umbenennung in ‚Gesellschaft für Evaluation‘ und die damit verbundene Erweiterung auf den deutschsprachigen Raum charakterisiert eine Entwicklung, die mit Blick auf die Mitgliederzahl, aber auch die Rolle, die der DeGEval im Fachdiskurs zukommt, als beeindruckend zu werten ist. Die Jahrestagung der DeGEval, die vom 20. bis zum 22. September an der Johannes Gutenberg-Universität Mainz stattfand und in Kooperation mit dem Zentrum für Qualitätssicherung und -entwicklung (ZQ) ausgerichtet wurde, widmete sich mit dem Thema der „Evaluation (in) der Zukunft“ dieser Historie, wie auch dem Ausblick auf die Weiterentwicklung nicht nur der DeGEval, sondern der Evaluation insgesamt.

Am Vortag der Tagung wurden, wie in den vergangenen Jahren auch, Weiterbildungsseminare für Evaluatorinnen und Evaluatoren angeboten, die sowohl methodische Aspekte zur Gestaltung von Online-Befragungen und Kausalanalysen als auch grundsätzliche Fragestellungen zu Möglichkeiten und Grenzen von Wirkungsanalysen in den Blick nahmen. Hieran anschließend fand die Mitgliederversammlung der DeGEval statt, in deren Rahmen unter anderem auch ein neuer Vorstand gewählt wurde mit Jan Hense als neuem Vorstandsvorsitzenden und Marianne Lück-Filsinger sowie Stefan Schmidt als neue Mitglieder des Vorstands, die das bestehende Team um Christiane Kerlen und Sonja Sheikh ergänzen.

Im Anschluss an die Eröffnung der Tagung befasste sich Christina A. Christie, Professor for Education an der University of California in Los Angeles, im Rahmen der ersten Keynote mit Fragen der Theoriebildung in der Evaluation. Ausgehend davon, dass eine zentrale Frage von Evaluation jene nach der Messung von Wirkungen ist, konstatierte sie ein Theoriedefizit, das die Ausbildung von systematischen und theoretisch fundierten Wirkannahmen und -messungen zumindest nicht fördert. Unter dem Titel „Future of Evaluation – the Role of Theory“ stellte sie insbesondere darauf ab, dass eine „Evaluation Policy“ definiert werden solle, um einen Bezugspunkt für das konkrete Evaluationsdesign zu erhalten. Soweit Christie mit Blick auf die häufig induktive Entwicklung von Evaluationsverfahren zuzu-

stimmen ist und vielen Evaluationen dieser Bezugspunkt fehlt bzw. dieser im Nachhinein rekonstruiert wird, so deutlich wird auch, dass Christies Theoriebegriff von einem ‚klassischen‘ Theorieverständnis, beispielsweise im Sinne des kritischen Rationalismus, abweicht.

In der Keynote am zweiten Tag warf Reinhard Stockmann, Professor für Soziologie an der Universität des Saarlandes und Leiter des Centrums für Evaluation (CEval), einen Blick auf die bisherige und künftige Entwicklung der Evaluation. Er resümierte einen Bedeutungsgewinn von Evaluation in den vergangenen Jahrzehnten, sieht aber auch latente Herausforderungen, die bislang nur zum Teil gelöst seien. Unter Zugrundelegung eines differenzierungstheoretischen und strukturfunktionalistischen Ansatzes, unterscheidet Stockmann mit Blick auf die unterschiedlichen Herausforderungen an Evaluation nach politischem, sozialem und wissenschaftlichem System. Diese drei Systeme charakterisieren in unterschiedlichen Ausprägungen die Aufgaben an das Evaluationssystem. Als besondere Aufgaben zur Weiterentwicklung der Evaluation sieht er zum einen die Sensibilisierung des politischen Systems für Evaluation, was u.a. durch die Implementierung eines parlamentarischen Forums und eines Nationalen Evaluationsrates geschehen könne. Zum anderen sieht er in der Förderung der Bürgerpartizipation an politischen Entscheidungen durch Evaluation eine wesentliche Herausforderung. Darüber hinaus seien ein Zertifizierungssystem für Evaluatord(inn)en wie auch die Forcierung der Grundlagenforschung im Bereich der Evaluation vonnöten.

Neben diesen beiden Hauptvorträgen bot die Tagung ein breites Spektrum an Sessions, die in vier Blöcken mit jeweils sechs bis sieben Parallelveranstaltungen durchgeführt wurden. Die Themen spiegelten die unterschiedlichen Felder der Evaluation, die durch die Arbeitskreise der DeGEval repräsentiert werden, wobei auch vielfach feldübergreifende Sessions angeboten wurden, die insbesondere auf Fragen der theoretischen Implikationen und methodischen Herausforderungen von Evaluation abstellten. So wurde in der Session „Das Potenzial von Kausalmechanismen in komplexen Evaluierungen“ der Frage nachgegangen, inwieweit die Nutzung von Kausalmechanismen im Rahmen von Wirkungsanalysen einen spezifischen Beitrag zur Erklärung der Effekte von Interventionen leisten und welche Herausforderungen an das Evaluationsdesign damit – sowohl unter der Perspektive summativer als auch formativer Evaluationen – verbunden sind. In der von Alexandra Caspari und Tobias Polak moderierten Session „Signifikanzrituale, Netzwerkanalyse und Kompetenzmessung: Methodische Herausforderungen und Innovationen“ wurden methodische Aspekte erörtert, die feldübergreifend von Relevanz sind und beispielsweise mit einem verstärkten Einsatz netzwerkanalytischer Ansätze neue Wege für die Evaluation eröffnen könnten.

Zum Teil fokussierten die Sessions auch ausdrücklich auf das Tagungsthema, so in der durch Philipp Mayring moderierten Veranstaltung zur Zukunft der DeGEval, in der u.a. mit Günter Tissen und Christiane Spiel langjährige Vorstandsvorsitzende zu Wort kamen. Doch auch weitere Sessions widmeten sich in der Regel mit einem feldspezifischen Bezug der Frage, wie Evaluation in der Zukunft gestaltet sein sollte, um den veränderten Bedarfen und Rahmenbedingungen Rechnung zu tragen – so u.a. in der Session zur Evaluation in Organisationen, in der die Frage diskutiert

wurde, inwieweit die zum Teil in entsprechenden Evaluationen dominierende Innensicht genügt, um angemessene Evaluationsverfahren durchzuführen und die Trennung zwischen der Entwicklung und Realisierung von Maßnahmen einerseits und deren Bewertung andererseits zu gewährleisten.

Grundsätzlicher mit Blick auf die nach wie vor vergleichsweise geringe Relevanz von Evaluation im Kulturbetrieb widmete sich der Arbeitskreis Kultur und Kulturpolitik den Fragen, ob Kunst und Kultur überhaupt evaluierbar sind, welchen Nutzen sie aus Evaluationen ziehen können und inwieweit und unter welchen Rahmenbedingungen eine stärkere Implementierung von Evaluation im Kulturbetrieb möglich erscheint.

Der Blick wurde auch auf die internationale Perspektive zur Zukunft der Evaluation gerichtet. So befasste sich die Session „The Future of Evaluation in Costa Rica: Demand, Supply, Sustainability of Evaluation Capacity Development“ mit der Entwicklung der Evaluation in den vergangenen Jahren in Costa Rica, insbesondere aber auch mit Herausforderungen in der Zukunft in Anbetracht wechselnder politischer Konstellationen und einer mehr oder weniger stabilen Implementierung von Evaluation in einzelnen Handlungsfeldern.

Neben der Zukunft der Evaluation klang in wenigen Sessions auch die Frage an, ob und wie eine Evaluation der Zukunft möglich erscheint – so in der Session „Ex Ante Evaluation: Wann gelingt der Blick in die Zukunft?“. Damit rückt der Beitrag von Evaluation für die Prognose künftiger Entwicklungen und die Relevanz für Entscheidungen in Organisationen und Politikfeldern in den Vordergrund, was als konstitutiv für die Relevanz von Wissenschaft insgesamt und den Sozial- und Wirtschaftswissenschaften im Besondern gesehen werden kann.

Neben einer geführten Posterausstellung, die nochmals einen Blick auf die Vielfalt von Evaluation und Evaluationsfeldern erlaubte, und vielfältigen Möglichkeiten zum informellen Austausch schloss die Tagung mit einem Resümee in Form einer Podiumsdiskussion, in der abschließend zentrale Herausforderungen für die Zukunft der Evaluation resümiert wurden.

Fasst man die Eindrücke der 20. Jahrestagung der DeGEval in Mainz zusammen, so zeigte sich, dass die Präferenz, einen Blick in die Zukunft der Evaluation zu werfen anstatt sich umfassend der Vergangenheit zu widmen, als gelungener Zugang gewertet werden kann. Erfreulich unaufgeregt und an den vielfältigen Herausforderungen der Evaluation orientiert gelang eine Jubiläumstagung, die mit dem Verzicht auf eine Nabelschau wohltuend zurückhaltend und inhaltlich ergiebig war.

Nachwuchspreis der DeGEval 2017

*Pirmin Bundj, Edith Halves, Christiane Kerlen,
Philipp Mayring, Sonja Sheikh*

Der DeGEval-Nachwuchspreis wurde auf der 20. Jahrestagung der DeGEval in Mainz zum inzwischen 11. Mal vergeben. Ziel des Preises ist es, jährlich eine Nachwuchsevaluatorin/einen Nachwuchsevaluator oder auch eine Gruppe von Nachwuchsevaluator(inn)en 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.

Im vergangenen Jahr wurden insgesamt leider nur zwei Beiträge für den Nachwuchspreis 2017 eingereicht: eine Praxisarbeit und eine Masterarbeit. Und trotzdem ist die Wahl nicht leicht gefallen. Die Jury hat sich aber gefreut, den Preis wieder an eine Masterarbeit vergeben zu können!

In der Masterarbeit wird zum ersten Mal systematisch ein deutsches Bundesland analysiert und der Frage nachgegangen, wie Evaluation als Instrument der Analyse und Entscheidungshilfe von der Politik genutzt wird und welche Rolle sie bei Entscheidungs- und Planungsprozessen im politisch-administrativen Bereich spielt. Die Arbeit stellt insofern eine Pionierleistung dar, als weder auf Bundes- noch auf Landesebene vergleichbare Arbeiten vorliegen.

Die Daten der Masterarbeit von Frau Kerstin Wilhelm sind aus Experteninterviews mit Vertreter(inne)n der obersten saarländischen Verwaltungsbehörden gewonnen worden. Die Widerstände, die dabei zu überwinden waren, und die Anstrengungen, die notwendig waren, um die Ministerien zu einer Mitarbeit zu motivieren, dürften enorm gewesen sein. Aber das Engagement und die Hartnäckigkeit der Autorin haben sich ausgezahlt und in der Folge als überaus nützlich erwiesen, denn die Landesregierung hat ein Weiterbildungsprogramm für Ministerialbeamte/Ministerialbeamtinnen im Bereich Evaluation veranlasst.

Auch wissenschaftlich ist die Arbeit hervorragend und sehr ambitioniert, umfangreich im Datenmaterial, theoretisch gut begründet und gut in den Ergebnissen zusammengefasst und interpretiert. In einer Diskussion wird zudem kritisch infrage gestellt, welchen Limitierungen die Arbeit unterliegt.

Die Autorin leistet mit ihrer Arbeit insgesamt einen wichtigen Beitrag zur Weiterentwicklung der Evaluationspraxis in den obersten Verwaltungsbehörden.

Und damit ging der 11. DeGEval-Nachwuchspreis 2017 an Frau Kerstin Wilhelm für ihre Masterarbeit mit dem Titel „Wie nutzen die Ministerien des Saarlandes Evaluation zur Politiksteuerung?“. Wir gratulieren Frau Wilhelm ganz herzlich zu dieser hervorragenden Forschungsarbeit und hoffen auf zahlreiche weitere Beiträge von ihr zur Praxis und Theorie der Evaluation!

Wir freuen uns auch, dass wir damit erneut einen Preis an eine Absolventin des Masterstudienganges Evaluation der Universität des Saarlandes vergeben konnten. Dieser Studiengang erfüllt eine wichtige Funktion in der Ausbildung des Nachwuchses im Bereich der Evaluation – wir sind daher sehr froh, dass sein Fortbestand gesichert werden konnte.

Herzlichen Glückwunsch an Frau Wilhelm!!!
Die Jury des Nachwuchspreises 2017

Revision der Standards für Evaluation

*Jan Hense,¹ Wolfgang Böttcher,² Alexandra Caspari,³
Michael Kalman,⁴ Wolfgang Meyer⁵*

Die Standards für Evaluation der Gesellschaft für Evaluation (DeGEval) sind ein zentraler, fachlicher Bezugspunkt für alle, die in unterschiedlichen Rollen mit Evaluation befasst sind. Nach einem mehrjährigen Revisionsprozess wurden die ‚Standards‘, wie sie in der Regel nur verkürzt genannt werden, im November 2016 in ihrer revidierten Fassung von der Mitgliederversammlung verabschiedet. Sie stehen allen Interessierten zum Download auf der Homepage der DeGEval (www.degeval.org) zur Verfügung und sind seit Kurzem auch in einer neuen Druckauflage über die Geschäftsstelle der DeGEval zu beziehen. Dies nehmen wir zum Anlass, um an dieser Stelle noch einmal auf den Stellenwert der Standards aufmerksam zu machen und über den Revisionsprozess und die wichtigsten Änderungen zu informieren.

Schon die seit 2001 gültige Ursprungsfassung der Standards für Evaluation ist auf Basis umfassender Fachdiskussionen und intensiver Abstimmungsprozesse entstanden. Sie formuliert vier zentrale Ansprüche an die Qualität von Evaluationen. Gute Evaluationen sind demnach *nützlich*, *durchführbar*, *fair* und *genau*. Dabei ist eine wichtige, aber gelegentlich vergessene Botschaft der Standards, dass nicht alleine Evaluatorinnen und Evaluatoren für gute Evaluationen verantwortlich sind. Vielmehr sollten alle, die an einer Evaluation beteiligt sind, sei es als Evaluierende, als Auftraggebende oder als in anderer Rolle Beteiligte im Rahmen ihrer Möglichkeiten und Verantwortlichkeiten das Ihre tun, um nützliche, durchführbare, faire und genaue Evaluationen zu ermöglichen, beispielsweise indem Auftraggebende Evaluation rechtzeitig veranlassen oder Beteiligte Datenerhebungen in angemessener Weise unterstützen.

Die vier genannten Attribute guter Evaluation, die auch nach der Revision Bestand haben, dienen in den Standards als Gliederungsschema für insgesamt 25 Einzelstandards. Die *Nützlichkeitsstandards* fordern in insgesamt acht Einzelstandards, dass Evaluationen immer danach streben sollten, einen Mehrwert für die Praxis zu erbringen, z.B. indem konkrete Verbesserungsmöglichkeiten aufgezeigt werden oder die Entscheidungsfindung über einen Evaluationsgegenstand erleichtert wird. Die

1 Justus-Liebig Universität, Gießen

2 Westfälische Wilhelms-Universität Münster

3 Frankfurt University of Applied Sciences

4 Kalman Consult, Berlin

5 Centrum für Evaluation, Saarbrücken

Gruppe der *Durchführbarkeitsstandards* trägt dem Sachverhalt Rechnung, dass Evaluationen oft mit verschiedenen Herausforderungen bezüglich ihrer Realisierbarkeit konfrontiert sind. Die drei Durchführbarkeitsstandards sollen daher sicherstellen, dass Evaluationen dennoch angemessene Verfahren einsetzen, bewusst und sensibel im Handlungsfeld agieren und dabei Aufwand und Nutzen in einem angemessenen Verhältnis stehen. Die fünf *Fairnessstandards* sollen sicherstellen, dass Bewertungen im Rahmen von Evaluationen transparent zustande kommen und die Interessen von Beteiligten und Betroffenen angemessen geschützt werden. Schließlich definieren die neun Einzelstandards der Gruppe der *Genauigkeitsstandards* zentrale Anforderungen hinsichtlich der Gültigkeit und Belastbarkeit von Evaluationsbefunden, die sich inhaltlich mit allgemeinen Standards guter wissenschaftlicher Praxis und der empirischen Forschung überschneiden. Besonders zu erwähnen ist hierbei der abschließende Standard „G9 Meta-Evaluation“, der fordert, dass auch Evaluationen selbst im Hinblick auf ihre Qualität beurteilt werden können und sollten.

Für die 2001 in Kraft getretenen Standards für Evaluation hatte sich im Laufe der Jahre aus unterschiedlichen Perspektiven Handlungsdruck in Richtung einer Revision ergeben. Mit Blick auf das sich dynamisch entwickelnde Tätigkeitsfeld der Evaluation war bereits mit der ersten Fassung selbst beschlossen worden, diese nach einiger Zeit einer Revision zu unterziehen. Dazu wurde 2005 eine Mitgliederbefragung durchgeführt. Diese ergab zwar zum damaligen Zeitpunkt eher eine Bestätigung der gültigen Fassung, so dass die weitere Dissemination und Umsetzung der Standards als wichtiger angesehen wurde. Klar war aber, dass nach einer angemessenen Frist die Frage der Revision neu gestellt werden muss. Zusätzlicher externer Handlungsdruck ergab sich vor allem daraus, dass das englischsprachige Pendant und Vorbild, die Standards des Joint Committee on Standards for Educational Evaluation, 2011 in einer neuen, dritten Auflage erschienen war. Daher wurde 2012 durch die Mitgliederversammlung ein Revisionsprozess initiiert, der auf Basis der vorhergehenden Diskussionen zu möglichen Änderungsbedarfen nur eine moderate Überarbeitung zum Ergebnis haben sollte.

Diesem Auftrag folgend nahm die Revisionsgruppe in einem mehrjährigen Entwicklungs-, Kommentierungs- und Abstimmungsprozess mit der Mitgliedschaft nur moderate Änderungen an den Standards selbst vor. Vor allem die elegante Grundstruktur mit den vier griffigen Qualitätsmerkmalen „nützlich, durchführbar, fair und genau“ blieb erhalten. Die Ergänzung eines fünften Bereichs „Meta-Evaluation“, wie sie das Joint Committee vorgenommen hatte, wurde als inhaltlich nicht begründet verworfen. Auch die Option einer Umstellung auf eine prozedurale Gliederung, in der Standards je Phase eines Evaluationsprojekts definiert werden, wurde v.a. aus drei Gründen verworfen. Erstens bedingt eine solche Gliederung immer Vorannahmen über ein Evaluationsprojekt, die nicht in allen Fällen zutreffen, was dem generischen Anspruch der Standards, für alle Varianten, Konstellationen und Einsatzfelder der Evaluation zu gelten, zuwiderläuft. Zweitens sollte die etablierte ‚Marke‘ der vier griffigen Standardbereiche nicht leichtfertig aufgegeben werden. Drittens besteht eine wichtige Funktion der Standards nicht nur darin, Orientierung für die Praxis zu geben, was vielleicht mit einer prozeduralen Gliederung leichter sein könnte, sondern auch Kriterien für die Meta-Evaluation bereitzustellen, also für

die Beurteilung von Güte und Nutzen von Evaluation selbst, was nur die bisherige Struktur ermöglicht. Als Kompromiss enthält auch die revidierte Fassung ein funktionales Inhaltsverzeichnis, das für die Anwendung der Standards in verschiedenen Evaluationsphasen Orientierung geben soll. Inhaltliche Änderungen wurden somit eher auf Detailebene vorgenommen. Dabei reagieren viele Überarbeitungen und Ergänzungen auf fachliche Entwicklungen der jüngeren Zeit. Dazu gehört etwa die gewachsene Bedeutung von Wirkmodellen und Programmtheorien in der Evaluation, auf deren Rolle jetzt in verschiedenen Einzelstandards verwiesen wird. Ähnliches gilt für den Datenschutz, der an relevanten Stellen stärker betont wurde. Ein drittes Beispiel ist das Thema Implementationstreue, das aus guten Gründen in jüngerer Zeit vermehrt Aufmerksamkeit in der Evaluationsliteratur erfahren hat.

Neben diesen inhaltlichen Aktualisierungen und Verbesserungen ‚unter der Motorhaube‘ wurde viel Aufwand in eine leichtere Verständlichkeit und Zugänglichkeit investiert. Sie soll vor allem durch eine klarere und konsistenter umgesetzte Strukturierung der Erläuterungen zu den Einzelstandards erreicht werden. Zusätzlich wurden die konzeptionellen und begrifflichen Grundlagen in einem eigenen Abschnitt klarer herausgestellt. Eine weitere wichtige Weiterentwicklung war die erstmalige Ergänzung eines Glossars der Standards für Evaluation, das für ein konsistentes Begriffsverständnis, vielleicht auch über die Standards hinaus, sorgen sollte.

Wir möchten alle, die die neu revidierten Standards noch nicht zur Kenntnis genommen haben, einladen, diese Neuerungen einer eigenen Prüfung auf Nützlichkeit, Umsetzbarkeit, Ausgewogenheit und Genauigkeit zu unterziehen. Wie eingangs erwähnt, finden Sie die revidierten Standards zum Download auf der Homepage der DeGEval, ebenso wie Angaben zum Bezug der Druckausgabe. Wir hoffen, dass die Revision dazu beiträgt, weiter das Bewusstsein für Standards guter Evaluation zu schärfen und deren Umsetzung in der Evaluationspraxis voranzubringen.

Aktuelle Entwicklungen im Politikfeld FTI und Konsequenzen für Evaluationsansätze und -methoden: Neue und alternative Förderungsformate in der FTI-Förderung und Wirkungsorientierung in der FTI-Politik

Veranstaltungsbericht zum Frühjahrstreffen des Arbeitskreises Forschungs-, Technologie- und Innovationspolitik der Gesellschaft für Evaluation e.V. am 19. Mai 2017 in Wien

Iris Fischl,¹ Laurenz Wolf,¹ Jan Wessels,² Marianne Kulicke³

Am 19. Mai 2017 fand in der Austria Wirtschaftsservice (aws) in Wien das diesjährige Frühjahrstreffen des Arbeitskreises Forschungs-, Technologie- und Innovationspolitik (AK FTI) statt.

Die DeGEval feierte 2017 ihr 20-jähriges Jubiläum und stellte ihre Jahrestagung daher unter das Motto „Zukunft der Evaluation“. Das Frühjahrstreffen 2017 des AK FTI griff dieses Thema auf und richtete seinen Fokus auf den aktuellen Rand des Evaluationsgeschehens im Politikfeld FTI.

Das Treffen widmete sich aktuellen Entwicklungen im Politikfeld FTI und entsprechenden Konsequenzen für Evaluationsansätze und -methoden. Im Vordergrund standen dabei neue, angepasste bzw. alternative Förderungsformate und die Frage, wie Evaluator(inn)en diesen Anforderungen mit ihren Evaluationsansätzen und -methoden begegnen. Welche neuen methodischen Zugänge gibt es, welche bisherigen Erfahrungen wurden gemacht? Und was heißt das für die Zukunft der Evaluation im Politikfeld FTI?

Darüber hinaus waren auch Anforderungen und Erwartungen Thema des Treffens, die sich aus der Wirkungsorientierten Folgenabschätzung (WFA) in Österreich und dem deutlich gestiegenen Stellenwert von Wirkungskontrollen zu Fördermaßnahmen in Deutschland für externe Evaluationen ergeben.

Das Treffen des Arbeitskreises wurde von rund 25 Evaluator(inn)en und Auftraggebern besucht und fand erstmals in den Räumlichkeiten der aws (Austria Wirt-

1 KMU Forschung Austria

2 Institut für Innovation und Technik – iit Berlin

3 Fraunhofer ISI

schaftsservice Gesellschaft mbH) statt. Da das Treffen in Wien stattfand, hatte die Tagung sowohl von Seiten der Teilnehmer(innen) als auch der präsentierten Inhalte eine österreichische Prägung.

Vortragende und Themen (in chronologischer Reihenfolge) waren dabei:

- Sascha Ruhland (KMU Forschung Austria), *Evaluation des aws-Garantie-Instrumentariums*
- Brigitte Ecker (WPZ Research), *Evaluation der Forschungsprämie*
- Marianne Kulicke (ISI Fraunhofer), *Evaluation der Nationalen Forschungsstrategie BioÖkonomie 2030*
- Friedemann Call (DLR), *Evaluation im DLR Projektträger am Beispiel einer Maßnahme im Bereich Klimaanpassung*
- Rupert Pichler, Mario Steyer (BMVIT), *Wirkungsorientierte Folgenabschätzung in Ministerien – aktuelle Einsichten*
- Michael Dinges (AIT), *Indikatoren zur Erfassung der Wirkungen von Förderungen der FFG*
- Norbert Knoll (aws), *Was bedeutet Innovationsförderung 4.0 für Evaluierung 4.0?*

Die Sprecher(innen) des AK FTI Iris Fischl (KMU Forschung Austria) und Jan Wessels (iit Berlin) eröffneten die Veranstaltung mit einem Kurzeinblick in aktuelle Entwicklungen und Trends der FTI-Evaluation in Österreich und Deutschland. Aus österreichischer Sicht hervorgehoben wurde dabei die geplante Umsetzung des *OECD Review of Innovation Policy*, die eine Beurteilung der heimischen Innovationspolitik verspricht. Für die deutsche Seite schilderte Jan Wessels in vielen Punkten Kontinuität in der Evaluationspraxis, jedoch auch einige interessante Veränderungstrends. So finden sich verstärkt Evaluationen von Rahmenprogrammen mit Maßnahmenbündeln, ebenso gibt es verstärkt Ausschreibungen von Begleitforschung, die auch einen Evaluationsteil einschließen. Ebenfalls zugenommen haben Ausschreibungen von Projektträgerschaften mit Monitoring-/Evaluationsaufgaben. Zu beobachten ist zudem ein höherer Konkretisierungsgrad von Leistungsbeschreibungen. Zum Teil werden relativ konkrete Vorgaben zur Methodik gemacht.

Der erste Vortragsblock umfasste zwei Berichte aus der Praxis die sich mit *alternativen Förderinstrumenten* in Österreich auseinandersetzten. Aufgrund der impliziten Abweichung von der klassischen FTI-Förderung rückte der Themenblock das Erfordernis komplementärer Evaluationsmethoden in den Vordergrund. Sascha Ruhland (KMU Forschung Austria) präsentierte dazu die kürzlich abgeschlossene *Evaluation der Garantieinstrumente der aws* und erläuterte insbesondere die methodischen Herausforderungen des gewählten Kontrollgruppenansatzes (Propensity Score Matching). Möglichkeiten und Herausforderung der methodischen Vorgehensweise gegenüberstellend, wurde den Anwesenden vor allem die Problematik eingeschränkter Datenverfügbarkeiten und -qualitäten dargelegt. Zusätzlich zu dem Umstand, dass die Vorgehensweise nur gewählt werden konnte, weil die Evaluationsseinrichtung (KMU Forschung Austria) zur Auswahl der Kontrollgruppe auf einen langjährigen und breiten Datensatz zurückgreifen konnte, kam der hohe Zeit-

aufwand entlang sämtlicher Schritte der Vorgehensweise hinzu. Dies ist auch vor dem Hintergrund zu betrachten, dass die Methode bereits in einem Vorgängerprojekt herangezogen wurde, abgeleitete Lerneffekte bzw. bereits modellierte Bestandteile aber nur geringfügig zu einer Aufwandsreduktion geführt haben. Basierend auf diesem Umstand stellte der Referent die generelle Frage in den Raum, wann Ergebnisse den Aufwand für technisch anspruchsvolle bzw. zeit- und kostenintensive Methoden rechtfertigen, speziell wenn Resultate nicht umfassend in der Lage sind, komplexe Wirkungszusammenhänge abzubilden. Ergiebiger war nach Einschätzung des Referenten der Zugang über eine standardisierte Befragung und Interviews, die so designed wurden, dass mit den erzielten Resultaten des Matchings Wirkungszusammenhänge des Programms aufgearbeitet werden konnten.

Brigitte Ecker (WPZ Research) berichtete von der ebenfalls erst kürzlich abgeschlossenen *Evaluation der steuerlichen F&E-Förderung (Forschungsprämie) in Österreich*. Diese war bereits vor und während ihrer Umsetzung Thema des öffentlichen Diskurses, sowohl auf politischer Ebene hinsichtlich des Nutzens und der Ausgestaltung des Instruments im Allgemeinen als auch von Seiten der Evaluationscommunity in Bezug auf die gesetzten Rahmenbedingungen der Wirkungsanalyse. Auch in Deutschland wird das Thema derzeit wieder intensiv diskutiert, weshalb ein Blick in den Evaluationsbericht sicherlich lohnenswert wäre. Die Autorin stellte im Vortrag einerseits die methodischen Herausforderungen in den Vordergrund, die beispielsweise aus den restriktiven Datenschutzbestimmungen in Österreich entstanden (Einsehbarkeit und Verknüpfbarkeit von Förderdaten), und andererseits das Erfordernis, die Forschungsprämie im Wechselspiel mit direkten Förderangeboten zu analysieren bzw. die gesamtheitliche Betrachtung des österreichischen Innovationsystems nicht aus den Augen zu verlieren. Letztlich kommt die Evaluation zu dem Ergebnis, dass die steuerliche F&E-Förderung tendenziell in die Tiefe wirkt, jedoch nicht zu einer Verbreiterung der Innovationsbasis. Das bedeutet, dass die Prämie bei bereits forschungsintensiven und regelmäßig F&E-treibenden Unternehmen weiter unterstützend wirkt. Eine Anreizwirkung zur Ausweitung der F&E bei Unternehmen mit bislang geringer bzw. keiner F&E ist aufgrund der Forschungsprämie hingegen kaum gegeben. Diese Unternehmen werden mehr durch die direkte Forschungsförderung angesprochen.

Der zweite Vortragsblock des Frühjahrestreffens widmete sich den sogenannten *missionsorientierten Programmen* und ihrer Evaluation. Die übergreifende Frage war hier, ob es Unterschiede zu klassischen Programmen z.B. der Technologieförderung gibt und welche Auswirkungen dies auf die Evaluationspraxis hat. Marianne Kulicke (Fraunhofer ISI) schilderte dabei ihre Erfahrungen aus der *Evaluation der „Nationalen Forschungsstrategie BioÖkonomie 2030“*, die für sie das Betreten methodischen Neulandes bedeuteten. Die Evaluation umfasste 36 Fördermaßnahmen bzw. rund 1.800 Projekte und sollte die Zielerreichung der Forschungsstrategie unter Berücksichtigung der Maßnahmenkonzeption untersuchen. Als ein wesentliches Problem der missionsorientierten Programme stellte sich heraus, dass Ziele relativ vage formuliert und kaum mit Hilfe vorhandener Leistungsindikatoren zu operationalisieren sind. In der Teilnehmerdiskussion sprach die Referentin von einem „Überstülpen der Missionsorientierung“ über die klassische, technologieorientierte F&E-Förde-

rung, was u.a. auch darauf zurückzuführen ist, dass einige der betrachteten Fördermaßnahmen bereits vor der Entwicklung der Forschungsstrategie gestartet wurden. Eine weitere Problematik ist, dass die Ziele missionsorientierter Fördermaßnahmen einem langfristigeren Zeithorizont unterliegen, sodass sie in Hinblick auf die Zielerreichung kaum evaluierbar sind. Als zu komplex erwies sich die Entwicklung und Abbildung eines realen Wirkungsmodells, wie die missionsorientierte F&E-Förderung zur Erreichung übergeordneter, gesellschaftlicher Ziele beitragen soll, weshalb die Referentin von einem „virtuellen Wirkungsmodell“ sprach. Als Fazit formulierte die Referentin folgende Punkte: eine Missionsorientierung von Fördermaßnahmen stellt besonders herausfordernde Ansprüche an Evaluationen (hinsichtlich des Konkretisierungsgrads des Zielsystems, den Leistungsindikatoren, Kausalitäten von Aktivitäten und Wirkungen, dem Zeithorizont des Eintretens messbarer Wirkungen, der Mehrdimensionalität von Wirkungen usw.). Weiters müssen zeitliche Spielräume gegeben sein, um jeweils auf den Evaluationsgegenstand zugeschnittene Evaluationsmethodiken entwickeln zu können. Ebenso bestehen missionsorientierte Förderansätze meist aus einem Maßnahmenbündel, hier ist eine dezidierte Portfolioanalyse notwendig und umsetzbar, um das Zusammenwirken der Einzelmaßnahmen zu bewerten, Lücken aufzuzeigen, Überschneidungen zu identifizieren usw. Schließlich erfordert eine solche Evaluation auch eine Kombination aus fachlicher Expertise (Bioökonomie) und Evaluationsexpertise.

Friedemann Call vom Projektträger DLR präsentierte eine *Evaluation zu Maßnahmen im Bereich der Klimaanpassung*. Dieser Bereich der Evaluation ist mit Herausforderungen in mehrfacher Hinsicht entlang der Umsetzung konfrontiert. Neben dem Wechselspiel und der Betroffenheit verschiedenster Akteure, die ein trans- und interdisziplinäres Forschungsverständnis erfordern, ist die Klimaanpassung mit der fehlenden/ingeschränkten Messbarkeit von Anpassungserfolgen konfrontiert sowie mit Kriterien und Indikatoren in der Anpassung. Die beiden letzten Punkte werfen die Frage auf, wann eine Maßnahme gefördert werden sollte und wann nicht. Am Beispiel einer umgesetzten Maßnahme der Klimaanpassung in Baden-Württemberg wurden diese Herausforderungen verdeutlicht. So zeigte sich, dass eine ganze Reihe von Maßnahmen schon seit längerer Zeit gefördert worden sind, während eine übergreifende Strategie auf Landesebene erst nachträglich beschlossen wurde. Entsprechend schwierig war es auch hier wieder, den Bezug zwischen übergreifender politischer Zielsetzung im Bereich gesellschaftlicher Herausforderungen und konkreter Projektförderung zu operationalisieren.

Im dritten und abschließenden Vortragsblock wurden Praxisberichte zu den Themen der *Wirkungsorientierten Folgenabschätzung sowie Wirkungsindikatoren* behandelt. Einen in Deutschland eher unbekanntem Sachverhalt schilderte dabei der Beitrag von Rupert Pichler und Mario Steyer (bmvit) zur wirkungsorientierten Verwaltung (WOV) und Wirkungsorientierten Folgenabschätzung (WFA). Über alle Ressorts hinweg muss in Österreich seit ein paar Jahren ein auf verschiedenen Hierarchieebenen gegliedertes System an Zielformulierungen und Indikatoren zur Überprüfung der Zielerreichung formuliert werden. Dabei werden sowohl längerfristige Indikatoren erfasst, die sich aus der FTI-Strategie ableiten lassen, als auch kurzfristigere Indikatoren niedrigerer Hierarchieebenen. Wie auch in klassischen Programm-

evaluationen ist die Zuordnung von Effekten zu einzelnen Maßnahmen dabei eine große Herausforderung. Des Weiteren kam es bisher zu einem relativ hohen Koordinationsaufwand zwischen einzelnen Ressorts, der der Querschnittsmaterie FTI geschuldet ist. Immer wieder wird in der Verwaltung die Frage aufgeworfen, wozu überhaupt noch evaluiert werden soll. Die Referenten kommen diesbezüglich zu dem Schluss, dass wesentliche Erkenntnisse zu Maßnahmen nach wie vor auf externen Evaluationen basieren, nicht zuletzt, da die WFA sehr quantitativ orientiert ist und auf wenige Indikatoren fokussiert. Für die Frühjahrstagung war insbesondere von Interesse, ob dieses übergreifend implementierte System Rückwirkungen auf die Evaluationspraxis im Politikfeld hat. Die knappe Antwort lautete, dass dies kurzfristig noch nicht der Fall ist und es dazu noch einer gewissen Reifung bedarf. Während Monitoringdaten, die z.B. die FFG in ihrem Wirkungsmonitoring erheben lässt, auch in das Indikatorensystem der WOV einfließen (können), sind Evaluationsergebnisse von klassischen Maßnahmeevaluationen eher losgelöst von diesem System. Ein systematischer, wechselseitiger Austausch zwischen WOV/WFA und externen Evaluationen findet bisher noch nicht statt.

Im Themenblock der Wirkungsindikatoren präsentierte Michael Dinges (AIT) eine Studie zur *Erfassung von Wirkungen der von Förderungen der FFG bis zum Ende der Projektlaufzeit*. Zielsetzung des Projekts war die Bildung eines Sets an Indikatoren, das aussagekräftig und valide sowie über alle Programme der FFG hinweg anwendbar sein sollte. Unter Berücksichtigung von Rechtsgrundlagen und Zielvorgaben wurde dabei ein Set von 30 Indikatoren aus acht Datenquellen gebildet, die den gesetzten Kriterien entsprachen. In Frage gestellt wurde dabei der Sinn subjektiver Einschätzungsfragen zu Gunsten einer Konzentration auf Indikatoren zu prüfbareren Fakten. Weiters wurde in der Diskussion erörtert, dass die Nutzung der Indikatoren Vergleichbarkeiten ermöglichen sollen, keinesfalls jedoch ein Benchmarking-System darstellen. Für die Evaluation ergibt sich längerfristig die Möglichkeit, auf bessere Daten und Zeitreihen zurückgreifen zu können.

Der letzte Beitrag von Norbert Knoll (aws) beschäftigte sich mit den Folgen der *Digitalisierung im Bereich der Förderungsdienstleistungen* und der Frage, ob es die Evaluation 4.0 gibt bzw. wie diese aussehen könnte. Angelehnt an die Digitalisierungsbestrebungen im Bereich der Finanzdienstleistungen geht es hier vor allem um Prozessoptimierung und Effizienzsteigerung. Da Förderprozesse eine Vielzahl von Aktivitäten mit hohem Informations- und Interaktionsbedarf umfassen, erscheint die Nutzung dezentral generierter Informationen für die Steuerung von Prozessen unausweichlich. Das Bestreben, alles zu messen und zu vermessen, eröffnet dabei auch Chancen für die Evaluation, wird doch ein nie dagewesener Datenreichtum erzeugt. Darüber hinaus können Evaluator(inn)en als „Hüter“ nicht digitalisierten Wissens agieren, bspw. durch die verfügbare Methodenvielfalt oder die Nutzung unstrukturierter und nicht digitalisierter Daten. Inwiefern allerdings mittelfristig neue Monitoringdaten erzeugt werden, die langfristig bestimmte Evaluationsinhalte ersetzen, konnte in der Diskussion nur umrissen werden.

In der abschließenden Diskussion brachten die deutschen Teilnehmer(innen) zunächst einmal den Eindruck ein, dass vieles in Österreich dennoch anders läuft, vermutlich auch bedingt durch die relativ zentralen Strukturen, die sich aus den beiden

Förderungsagenturen FFG und aws ergeben. Andererseits ist die Evaluationspraxis und sind die Evaluationsberichte in Österreich trotzdem nicht so unterschiedlich zu jenen in Deutschland. Bezüglich der Wirkungsorientierten Verwaltung wurde abschließend herausgearbeitet, dass die ursprünglichen Erwartungen an ebendiese etwas überzogen waren, vor allem aber ist diese mit all ihren Beteiligten in einem immer noch andauernden Entwicklungsprozess zu sehen. Im Vergleich zur WOV/WFA in Österreich gibt es in Deutschland nur die Gesetzesfolgenabschätzung, die allerdings nicht einzelne Maßnahmen und Vorhaben betrifft. Weiters wurde der relativ entkoppelte Status des FFG-Wirkungsmonitorings diskutiert und die abnehmende Trennung zwischen Evaluation und Wirkungsmonitoring aufgegriffen, was eine immer stärkere Verflechtung der beiden Aspekte bedingt.

Mechanismen als Alternative zu kausaler Attribution – Frühjahrstagung 2017 des Arbeitskreises „Methoden in der Evaluation“ der DeGEval – Gesellschaft für Evaluation e.V.

Heike Steckhan¹, Sabine Brüntrup-Seidemann¹, Ida Verspohl¹

Was ist ein Mechanismus und welche Rolle spielt er in Evaluierungen? Diese Frage stand im Mittelpunkt der Frühjahrstagung 2017 des Arbeitskreises „Methoden in der Evaluation“ der DeGEval – Gesellschaft für Evaluation e.V. in Bonn. In den Räumlichkeiten des Wissenschaftszentrums Bonn konnte der Arbeitskreis gemeinsam mit der gastgebenden Organisation, dem Deutschen Evaluierungsinstitut der Entwicklungszusammenarbeit (DEval), 90 Teilnehmende begrüßen, die an der Beantwortung dieser Frage interessiert waren. Dr. Sven Harten, stellvertretender Direktor des DEval, kündigte in seinen einleitenden Worten einen Einblick in die Analyse von Mechanismen als einen Schlüssel zur Öffnung der ‚Black Box‘ der Wirkungsweise von Evaluierungsgegenständen an.

Im Verlauf der Tagung wurde immer wieder deutlich, dass eine Vielzahl von unterschiedlichen Definitionen des Begriffs ‚Mechanismus‘ existiert: James Mahoney hat in einer Sammlung – die keinen Anspruch auf Vollständigkeit erhebt – 24 unterschiedliche Definitionen gesammelt (vgl. Mahoney, 2001: 579f.). Auch die Vorträge der fünf hochkarätigen Referentinnen und Referenten beleuchteten unterschiedliche Sichtweisen auf den Begriff. In der abschließenden Podiumsdiskussion glichen sie ihre Verständnisse miteinander ab und arbeiteten Gemeinsamkeiten und Unterschiede heraus.

Den Auftakt der Beiträge machte Prof. Dr. Frans L. Leeuw von der Maastricht University bzw. vom Research and Documentation Centre (WODC) des Ministry of Security and Justice der Niederlande (Den Haag) mit seinem anspruchsvollen und unterhaltsamen Vortrag „Mechanismen und generative Kausalität: Mechanismen in theoriebasierten Evaluierungen“. Er nahm sein Publikum mit auf eine rasante Reise durch die Zeit mit einem bunten Strauß an Beispielen, die zurückgingen bis zum Einfluss des Konsums von Zitronen auf Skorbut bei Matrosen. Er definierte Mechanismen als „cogs and wheels“, also als ineinandergreifende Zahnräder, die zwischen Input und Output liegen. Sie werden von Interventionen initiiert und führen schließ-

¹ Deutsches Evaluierungsinstitut der Entwicklungszusammenarbeit – DEval, Bonn

lich zur Wirkung. Dabei unterschied er – basierend auf dem Mikro-Makro-Modell der soziologischen Erklärung von Coleman (1990), auch genannt die „Coleman’sche Badewanne“ – Makro-Mikro-Mechanismen (situational mechanisms), Mikro-Mikro-Mechanismen (action-formation mechanisms) und Mikro-Makro-Mechanismen (transformational mechanisms). Zur Rolle von Mechanismen in theoriebasierten Evaluierungen führte er die Begriffe der „Unvollendeten“ und der „fake handbags“ ein. Wirkungslogiken in theoriebasierten Evaluierungen wiesen nach seiner Einschätzung häufig Schwachstellen auf: Einige Wirkungslogiken beziehen sich zwar auf Mechanismen oder nutzen sogar den Begriff, z.B. in ihren Annahmen über Wirkungszusammenhänge, vergessen aber wichtige Aspekte oder Dimensionen des Mechanismus oder lassen sie bewusst außen vor („die Unvollendeten“). Andere Wirkungslogiken stellen die Intervention, das Programm oder die Maßnahme wie einen Mechanismus dar, indem z.B. die Aktivitäten in Boxen abgebildet und durch Linien und Pfeile ergänzt werden. So entsteht der Eindruck, es handele sich um eine Kausalkette oder einen Mechanismus. Es ist aber eigentlich eine rein tautologische Darstellung des Evaluierungsgegenstands („fake handbags“), denn: „Boxes and arrows are not causal explanations“. Leeuw verwies auf den Ansatz der ‚Realist Evaluation‘ als eine Herangehensweise, in der CMO-Beziehungen (context-mechanism-outcome) expliziert werden. Zentraler Gedanke des CMO-Konzepts ist es, dass Mechanismen sowie Effekte bzw. Wirkungen auch von Kontextfaktoren abhängig sind. Um diese zu untersuchen, empfahl er die Auswertung von wissenschaftlichen und evaluationsspezifischen Datenbanken, um bereits vorhandenes Wissen einzubeziehen, sowie die Integration von „little quick experiments“ in das Evaluierungsdesign, um Mechanismen beispielhaft und ohne ausufernden Ressourceneinsatz zu identifizieren.

Der nächste Vortrag „Soziale Mechanismen, Mediation und Phantomvariablen“ von Prof. em. Dr. Peter Schmidt von der Justus-Liebig-Universität Gießen bewegte sich in der statistischen Welt der Hypothesen und Variablen. Er definierte soziale Mechanismen als „Erklärung von Hypothesen durch andere Hypothesen“ z.B. mittels ‚Phantomvariablen‘ in latenten Variablenmodellen. Dabei unterschied er zwischen der ‚Aktionshypothese‘ und der ‚Mediationshypothese‘: der Mediationshypothese zufolge tritt eine Wirkung (indirekt) über Mediation ein, nach der Aktionshypothese kann von direkteren Wirkungen ausgegangen werden. Die Mediationshypothese geht bei Unterstellung von ‚full mediation‘ davon aus, dass Interventionen oder Programme keine direkte Veränderung hinsichtlich Outcomes (z.B. Verhaltensänderung) bewirken, sondern dies ausschließlich indirekt, z.B. über sich verändernde Werte/Haltungen/Einstellungen geschieht. Es wird wiederum davon ausgegangen, dass sich Werte nur verändern, wenn eine Intervention auf Offenheit (‚openness to change‘) trifft. Der Mechanismus ‚openness to change‘ ist hier also eine zugrunde liegende Hypothese, welche erst ermöglicht, dass die darüberliegende Hypothese „die Intervention bewirkt eine Verhaltensänderung“ greift. Auch Schmidt wies in diesem Zusammenhang auf die Bedeutung von Kontextfaktoren und deren Einbezug z.B. durch Multi-Level Regressions hin. Als Analysemethoden empfahl er quasi-experimentelle Designs, aber auch – wie sein Vorredner – die Integration von Experimenten in Evaluationsdesigns sowie Meta-Analysen.

Zudem verwies er auf die Möglichkeiten der Bayes'schen Statistik sowie der Faktorenanalyse, aber auch auf die Vorzüge von kognitiven Interviews, Längsschnittuntersuchungen (z.B. Wachstumsmodellen) und nicht zuletzt auf neue Möglichkeiten der Verhaltensmessung über Smartphones und andere mobile Endgeräte.

Im nächsten Vortrag mit dem Titel „Mechanismen in Evaluierungen: Mechanismen als kausaler Prozess“ beleuchtete Dr. Johannes Schmitt vom Deutschen Evaluierungsinstitut der Entwicklungszusammenarbeit (DEval) Mechanismen aus einem weiteren Blickwinkel. Schmitt bezog sich auf eine prozessuale Definition von Mechanismen als System, das kausale Kräfte von der Intervention zum Outcome transferiert. Er widersprach seinem Vorredner mit der Aussage, Mechanismen seien keine intervenierenden Variablen, sondern ineinandergreifende ‚Teile‘ bestehend aus ‚Einheiten‘ und ‚Aktivitäten‘, die gemeinsam die Kausalkraft übertragen. Sein Vortrag konzentrierte sich auf die Methode des ‚Process Tracing‘ als innovative Möglichkeit und geeignete Analysemethode, die so definierten Mechanismen zu untersuchen. Die Methode ‚Process Tracing‘ kann in folgende fünf Schritte untergliedert werden: Zuerst wird ein hypothetischer Kausalmechanismus entwickelt, der in einem zweiten Schritt operationalisiert wird. Dabei macht das Evaluierungsteam Vorhersagen bezüglich der zu erwarteten Evidenzen. Es gilt also im Vorfeld zu überlegen, welche empirisch zu erfassenden Spuren oder ‚Fingerabdrücke‘ jeder Teil des Kausalmechanismus hinterlassen haben könnte. Gleichzeitig wird die Aussagekraft der Einzelevidenz bewertet, wobei insbesondere auch untersucht wird, ob es plausible alternative Erklärungen gibt. Nach der Datenerhebung wird in einem vierten Schritt die Aussagekraft der Gesamtevidenz geprüft, um dann in einem letzten fünften Schritt die Ergebnisse für die einzelnen Teile des Kausalmechanismus zusammenzuführen und zu bewerten, ob der hypothetisierte Kausalmechanismus im untersuchten Fall funktioniert hat oder nicht. Process Tracing kann sowohl theoriegestützt, als auch theoriebildend oder Outcome-erklärend angewandt werden. Die Methode ist laut Schmitt vor allem zur Beantwortung von Ob- und Wie-Fragen in Einzelfallstudien geeignet. Der Vorteil besteht in der hohen internen Validität. Eine Generalisierung bzw. Übertragung auf andere Kontexte (externe Validität) wird allerdings erst durch die Kombination mit anderen Methoden wie z.B. vergleichenden Fallstudien, QCA oder statistischen Verfahren möglich. Als zentrale Herausforderungen nannte er die Erfordernis eines umfangreichen Kontextwissens, eines guten Datenzugangs bereits in der Vorbereitung und die zeitintensive Vorbereitung von Feldstudien. Außerdem testet die Methode nur hypothetisierte Kausalmechanismen, alternative Erklärungen müssen vorab mitgedacht werden.

Zum Abschluss des ersten Tages zogen sich die Teilnehmenden in Kleingruppen zurück. Diese Gruppen wurden nach der Funktion der Teilnehmenden als Auftraggeber(in), Auftragnehmer(in) und Wissenschaftler(in) eingeteilt. In ihnen wurden die Vorträge diskutiert und Fragen für die Podiumsdiskussion der Referentinnen und Referenten am nächsten Tag herausgearbeitet.

Der zweite Tagungstag begann mit dem Vortrag „Wirkmechanismen auf der Spur ... Theoretische Überlegungen und praktische Anwendungen am Beispiel der Realistic Evaluation“ von Dr. Sigrid Haunberger von der Fachhochschule Nordwestschweiz, Hochschule für Soziale Arbeit. Auch sie, wie Leeuw am Vortag, sah im

Ansatz der Realistic Evaluation (synonym verwendet zur Realist Evaluation) eine geeignete Möglichkeit, Mechanismen zu identifizieren. Die zentrale Frage der Realistic Evaluation „Wie wirkt eine Intervention für wen und unter welchen Bedingungen in welchem Umfang?“ deckte Mechanismen auf. Haunberger definierte Mechanismen als verbindende, kontextabhängige Elemente zwischen Input und Outcome.

Sie stellte eine Studie vor, in der sie der Frage nachging, wie der Ansatz der Realistic Evaluation als Methode der Wirkungsevaluation in der Sozialen Arbeit sinnvoll angewendet werden kann. Dafür wurde zunächst ein Systematic Review von Realistic Evaluations in diesem Bereich im deutschsprachigen Raum durchgeführt. Der Review zeigte zum einen, dass unterschiedliche methodische Herangehensweisen genutzt werden, um Realistic Evaluations durchzuführen. Insbesondere wurde auf qualitative Ansätze und Mixed Methods zurückgegriffen. Zum anderen wurde deutlich, dass die zentralen Begriffe ‚Kontext‘, ‚Mechanismus‘ und ‚Outcome‘ sehr unterschiedlich verstanden werden. Als nächsten Schritt führte Haunberger in der Studie eine Qualitative Comparative Analysis (QCA) einer betrieblichen Sozialberatung durch. Sie sah in der QCA Parallelen zur Realistic Evaluation und schätzte sie als geeignet ein, Wirkmechanismen aufzudecken. In der QCA stieß sie auf Zusammenhänge, die dem in Realistic Evaluations genutzten Muster des CMO widersprachen, wie CO-, MO-, CCO-, MMO-Konfigurationen. In ihrem Fazit beschrieb sie es als Herausforderung, Kontexte, Mechanismen und Outcomes „eindeutig“ zu identifizieren. Zudem werde die Beziehung zwischen Kontext, Mechanismus und Outcome eher interpretativ überprüft als systematisch getestet. Das führe zu dem Problem der Generalisierbarkeit der Ergebnisse. CMO-Beziehungen seien daher „höchstens Theorien mittlerer Reichweite“ und Realistic Evaluation keine exakte wissenschaftliche Methode, sondern eher ein heuristisches Rahmenwerk bzw. eine Methode, Wissenschaft für die Öffentlichkeit zugänglich zu machen.

Der Vortrag des nächsten Referenten, Frank Willemsen, wie Leeuw vom Research and Documentation Centre des Ministry of Security and Justice der Niederlande (Den Haag), vertrat eine ganz andere Sichtweise als seine Vorrednerinnen und -redner. Er begann seinen Vortrag mit den Worten: „We don't need theory. We look at our data and know everything“. Anschließend eröffnete er seinem Publikum einen Blick in eine digitale Welt der Zukunft: Willemsen beschäftigt sich mit Big Data als Quelle zur Untersuchung von Mechanismen. Dabei nutzt er die Datenmengen, die bei Google und Social-Media-Anbietern, wie Twitter oder Facebook, vorliegen und kombiniert sie mit anderen Methoden (z.B. Experteninterviews), um Vorhersagen über zukünftige Entwicklungen wie z.B. Wirkungen von Politikmaßnahmen zu treffen. Dabei wies er explizit darauf hin, dass eine Validierung mit anderen Methoden nötig sei, da Google seine Suchergebnisse manipuliere und nicht nachvollzogen werden könne, wie dies geschehe. Als Beispiele für überprüfte Vorhersagen nannte er unter anderem die Vorhersage des Verlaufs einer Grippeepidemie anhand von häufig genutzten Suchbegriffen bei Google oder der Route von Flüchtlingen anhand der Auswertung von Anfragen an eine arabischsprachige Suchmaschine. Big Data wird von ihm auch in Evaluationen genutzt, z.B. um Reaktionen der Bevölkerung auf geänderte Gesetzeslagen zu analysieren. Durch eine Gegenüberstellung seiner

Ergebnisse mit amtlichen Aufzeichnungen konnte er die Treffsicherheit seiner Analysen veranschaulichen.

In der anschließenden Podiumsdiskussion, an der alle Referentinnen und Referenten teilnahmen, entspann sich ein lebhafter Austausch über Unterschiede und Gemeinsamkeiten zwischen den Verständnissen von Mechanismen. Vor allem Leeuw und Schmidt sahen viele Gemeinsamkeiten in ihren Sichtweisen: Soziale Mechanismen sind ihrer Ansicht nach verhaltenswissenschaftliche Phänomene bzw. intervenierende Variablen, die nicht beobachtbar sind. Johannes Schmitt hingegen sieht Mechanismen aus der Warte des Process Tracing als beobachtbare Teile an, die in der Summe den kausalen Prozesses bilden. Für ihn sind aber Parallelen zur Kontextabhängigkeit von Outcomes und zu Strukturgleichungsmodellen erkennbar, auf die sich Peter Schmidt bezieht. Im Process Tracing würde man allerdings weniger in der Tiefe untersuchen als vielmehr den gesamten Prozess des Mechanismus betrachten. Schlussendlich wurden sich die Referentinnen und Referenten aber über einige Gemeinsamkeiten einig, die Dr. Tobias Polak als Sprecher des Arbeitskreises Methoden in seiner Zusammenfassung zum Abschluss der Frühjahrstagung aufgriff:

- 1) Eine Theorie oder Wirkungslogik sollte auf Mechanismen Bezug nehmen.
- 2) Von Beginn der Evaluierung an sollten „richtige Handtaschen“ und vollendete Wirkungslogiken erstellt werden. Die üblichen genutzten Pfeildiagramme können einen Ausgangspunkt dafür bieten, wenn sie weiter ausdifferenziert und mit schlüssigen Konzepten hinterlegt werden.
- 3) Bereits vorliegendes Wissen aus Evaluierungen und wissenschaftlichen Studien sollte in Theorien/Wirkungslogiken integriert werden.
- 4) Es existieren unterschiedliche Möglichkeiten, Mechanismen empirisch zu untersuchen, z.B. Realist/Realistic Evaluation, experimentelle Designs oder Process Tracing.
- 5) Einen Mechanismus kann man als eine „Erklärung von Zusammenhängen“ definieren. Die Beschäftigung mit Mechanismen trägt dazu bei, das Bewusstsein für die Bedeutung dieser Zusammenhänge zu schärfen.

Mit dieser Kondensierung der komplexen Diskussionen der Tagung erhielten die Teilnehmenden handhabbare ‚take home messages‘, auf die sie in ihrer weiteren Arbeit zurückgreifen können.

Literatur

- Coleman, James S. (1990): *Foundations of Social Theory*. Cambridge, MA: Belknap Press of Harvard University Press.
- Mahoney, James (2001): *Beyond Correlational Analysis: Recent Innovations in Theory and Method*. In: *Sociological Forum*, 16 (3), pp. 575-593.

Methodische Standards der Evaluation zwischen Schema F und Innovation – Weiterführung einer politikfeldübergreifenden Diskussion

Franziska Heinze¹, Stefanie Reiter¹

Im Rahmen der 20. Jahrestagung der Gesellschaft für Evaluation (DeGEval) zum Thema „Evaluation (in) der Zukunft“ in Mainz (20.-22.09.2017) organisierte der Arbeitskreis (AK) *Methoden in der Evaluation* in Kooperation mit weiteren Arbeitskreisen in der DeGEval eine Session als Beitrag zur Anregung einer politikfeldübergreifenden Diskussion zu methodischen Standards. Der AK *Methoden in der Evaluation* setzte damit seine auf der DeGEval-Jahrestagung 2016 begonnene Veranstaltungsreihe zur Bestandsaufnahme von Evaluationsansätzen und methodischen Standards in verschiedenen Arbeitsfeldern fort. Die Veranstaltungsreihe wurde im Format einer Fishbowl-Diskussion (Innenkreis-/Außenkreis-Methode) durchgeführt, an der sich jeweils drei weitere Arbeitskreise der DeGEval beteiligten. Diese berichteten zunächst über typische und innovative Evaluationsansätze und Methoden für ihr Feld und diskutierten darauf aufbauend politikfeldübergreifend methodische Standards sowie Innovationspotenziale von Evaluationsmethoden.

Bei der Auftaktveranstaltung 2016 begannen die Arbeitskreise *Kultur und Kulturpolitik*, *Entwicklungspolitik* sowie *Forschungs-, Technologie- und Innovationspolitik* den Austausch von Erfahrungen hinsichtlich der Anwendung von Evaluationsmethoden in den jeweiligen Arbeitsfeldern und diskutierten Gemeinsamkeiten sowie Unterschiede der Ansätze (siehe Altenburg 2017). Die intensive Diskussion brachte als Ergebnis, dass die evaluative Arbeit in den drei beteiligten Themenbereichen maßgeblich von einem Spannungsfeld zwischen Qualitätsanforderungen an Methoden und zur Verfügung stehenden Ressourcen geprägt ist. Die resümierende Feststellung, dass die Wahl der Evaluationsmethoden sich weitgehend an der vorherrschenden Evaluationspraxis im eigenen Handlungsfeld orientiere und selten als Anregung ein Blick auf in anderen Evaluationsfeldern verbreitete Ansätze und Methoden erfolge, bestätigte die Organisatoren aus dem AK *Methoden in der Evaluation*, die 2016 begonnenen Diskussionen mit dem gleichen Format unter Beteiligung weiterer Arbeitskreise fortzusetzen. Entsprechend wurde

¹ Deutsches Jugendinstitut e.V., Halle

auf der DeGEval-Jahrestagung 2017 eine gemeinsame Session mit den Arbeitskreisen *Soziale Dienstleistungen*, *Hochschulen* und *Gesundheitswesen* veranstaltet.

Als Diskussionsteilnehmende waren folgende Personen im Innenkreis des Fishbowl-Formats gesetzt:

- PD Dr. Rainer Strobl – Sprecher des AK *Soziale Dienstleistungen*, (proVal – Gesellschaft für sozialwissenschaftliche Analyse – Beratung – Evaluation, Hannover, strobl@proval-services.net),
- Prof. Dr. Philipp Pohlenz – Sprecher des AK *Hochschulen*, (Otto-von-Guericke-Universität Magdeburg, philipp.pohlenz@ovgu.de),
- Marcus Capellaro – von 2011 bis 2017 Sprecher des AK *Gesundheitswesen*, (Capellaro GmbH – Konzeption & Evaluation kommunikativer Maßnahmen, M@Capellaro.de).

Ihre Statements und Diskussionen wurden durch Wortbeiträge weiterer Personen aus dem Außenkreis ergänzt, die in abwechselnder Beteiligung die freien Stühle des Innenkreises besetzten und so zur Diskussion beitrugen. *Prof. Dr. Alexandra Caspari* (Evaluationsforschung, Methoden der empirischen Sozialforschung und Statistik; Frankfurt University of Applied Sciences, Fachbereich 4, Soziale Arbeit und Gesundheit) und *Dr. Jan Tobias Polak* (Evaluator, Evaluierung zivilgesellschaftliche Entwicklungszusammenarbeit, entwicklungspolitische Bildungsarbeit; DEval – Deutsches Evaluierungsinstitut der Entwicklungszusammenarbeit gGmbH Bonn) aus dem Sprecherteam des AK *Methoden in der Evaluation* führten wie bereits 2016 die leitfragengestützte Diskussion.

Typische und innovative Evaluationsmethoden – eine Bestandsaufnahme

Rainer Strobl erörterte in seinem Eingangsstatement zunächst die große Heterogenität des Feldes der *Sozialen Dienstleistungen* und eine damit einhergehende Vielfalt von Evaluationsmethoden, welche die Identifizierung von innovativen Ansätzen und typischen Vorgehensweisen ‚nach Schema F‘ erschwere. Zu den Besonderheiten des Feldes der *Sozialen Dienstleistungen* zählten der Aspekt der Koproduktion² sowie eine oftmals sehr geringe finanzielle Ausstattung der Maßnahmen und Evaluationen.³ Die Ressourcenknappheit trage zu einem hohen Anteil an Selbstevaluationen in diesem Politikfeld bei.

Im Bereich dieser *Selbstevaluation* sind die Möglichkeiten zum Einsatz eines breiten methodischen Spektrums eher begrenzt. Es werden häufig in erster Linie Methoden zur Feststellung von Zufriedenheit mit und Akzeptanz von Maßnahmen eingesetzt, um diese Ergebnisse in die Qualitätssicherung und Qualitätsentwicklung einspeisen zu können. Methodische Designs, welche darüber hinaus Effekte aufzei-

2 Dies bedeutet, dass die Dienstleistungen in enger Kooperation mit den Klientinnen und Klienten produziert werden. Zu den Implikationen von Koproduktionsprozessen siehe auch Halves/Lück-Filsinger/Schmidt 2014.

3 Zu den Besonderheiten des Feldes ist ein Kurzbericht des Arbeitskreises *Soziale Dienstleistungen* der DeGEval veröffentlicht worden (Reiter/Strobl/Buchheit 2017).

gen und Wirkungen nachweisen können, sind hingegen selten, wengleich die Frage der Wirkungen in diesem Feld häufig debattiert und auch vonseiten der Auftraggeber zusehends Nachweise für Wirksamkeit eingefordert werden. In der Konsequenz wird in einigen Selbstevaluationen zumindest versucht, Wirkungsziele und Indikatoren für Wirkungen festzulegen, anhand derer wahrgenommene Effekte sich bspw. im Rahmen einer (Selbst-)Beobachtung einordnen lassen.

Der sehr heterogene Bereich der *Fremdevaluation* ist nach Einschätzung von *Rainer Strobl* durch den Einsatz qualitativer und quantitativer Methoden gekennzeichnet, die sich an sehr unterschiedlichen Zielgruppen orientieren. Beispielsweise ist im Bereich der frühkindlichen Bildung die Datenerhebung bei Kindern oftmals nur mittels überwiegend qualitativer Verfahren, wie teilnehmender Beobachtungen und Videoanalysen oder unter Einsatz bestimmter Gesprächsformen, möglich. Aus dem Bereich quantitativ ausgerichteter Verfahren werden randomisierte kontrollierte Studien selten eingesetzt, wengleich es in einigen Feldern wie der Kriminalprävention immer wieder Plädoyers dafür gibt. Möglichkeiten des Einsatzes von (quasi-)experimentellen Designs werden unter Evaluierenden diskutiert und zum Teil realisiert.⁴ Schließlich lasse sich als Gemeinsamkeit vieler Ansätze von Fremdevaluation in diesem Arbeitsfeld identifizieren, dass der Einsatz qualitativer und quantitativer Methoden weit verbreitet sei, das Spektrum an triangulativen Verfahren im Rahmen von ‚echten‘ Mixed-Methods-Designs jedoch – mitunter aus Ressourcen Gründen – selten gänzlich ausgeschöpft wird.

Philipp Pohlenz kündigte in seinem Eingangsstatement eine kritische Perspektive auf das Feld der *Evaluation im Hochschulbereich* an. Er berichtete eingangs über die Bemühungen in den späten 1990er Jahren, Evaluation in Lehre und Hochschule als Instrument der Legitimation zu verankern, und damit einhergehenden, weit verbreiteten (argumentativen) Abwehrreaktionen seitens Lehrender, welche „durch systematisiertes Beurteilen der Qualität von Lehre und Studium“ die Autonomie ihres Lehrhandelns bedroht sahen.⁵ Dies führte zu immer wiederkehrenden Diskussionen um den Einsatz gegenstandsangemessener Verfahren, in denen auch die Frage der Repräsentativität von Daten aufgeworfen wurde. Akzeptanzprobleme (auch von qualitativen Daten) haben zur Weiterentwicklung von messbaren Indikatoren und zu nach ‚Schema F‘ reproduzierbaren Befragungen v.a. von Studierenden „zum Zwecke des Vergleichs einzelner Lehrleistungen“ (bspw. im Rahmen der zumeist summativen Evaluation von Lehrveranstaltungen) beigetragen.

Zum jetzigen Zeitpunkt sieht *Philipp Pohlenz* – vor dem Hintergrund von Gewöhnungseffekten und tendenziell geringeren Akzeptanzproblemen von Evaluation – Potenziale zur Etablierung von Innovationen. Entsprechende Innovationsbedarfe von Verfahren und Nutzungsszenarien ergeben sich dabei auch aus Entwicklun-

4 Beispielsweise wurde auf der Frühjahrstagung 2015 des AK *Soziale Dienstleistungen* zum Thema „Methodische Herausforderungen der Wirkungsanalysen bei knappen Ressourcen“ intensiv diskutiert, welche ethnischen Bedenken es bei der Bildung von Kontrollgruppen in ausgewählten Settings gibt und welche Alternativen (z.B. Wartegruppen, Cross-Over-Designs) möglich sind (vgl. Reiter et al. 2015).

5 Eine Kurzdarstellung von wesentlichen Entwicklungslinien der Evaluation an Hochschulen ist in einem Bericht des AK *Hochschulen* der DeGEval in der ZfEv 2/2017 erschienen (vgl. Mittauer/Pohlenz/Harris-Huermann 2017: 275-276).

gen jenseits des Qualitätsdiskurses im engeren Sinne, die von neuen Zielgruppen in der Lehre⁶ und sich verändernden gesellschaftlichen Erwartungen an Hochschullehre geprägt seien. Erste innovative Schritte in Richtung neuer Wege der Hochschul-evaluation sieht er in den an vielen Orten erkennbaren, gemeinsamen Bemühungen von Evaluierenden und der oftmals an völlig unterschiedlichen Orten verankerten und institutionalisierten Hochschuldidaktik. Diese Wege beziehen sich tendenziell nun „nicht so sehr auf die Frage nach dem Vergleich zwischen Lehrenden X und Lehrender Y“, sondern „viel unmittelbarer auf das konkrete Lehrhandeln“ (bspw. Überprüfung des eigenen Handelns vor dem Hintergrund des eigenen Professionsverständnisses in der *hochschulinternen Selbstevaluation*).

Mit Blick auf weitere, historisch gewachsene Verfahren der Qualitätsbegutachtung von Lehre und Studium im Rahmen von *hochschulexternen Akkreditierungen* (siehe auch Bologna-Prozess) konstatiert er unterschiedliche Herangehensweisen (u.a. Peer-Review-Verfahren).

Bei den Qualitätsanforderungen an Hochschulen, die durch die verschiedenen Verfahren, sowohl intern im Sinne von Selbstevaluation als auch extern durch Akkreditierungsverfahren betrachtet werden, gibt es ein einigermaßen konsentiertes Qualitätsverständnis, wodurch in der Evaluation zusehends inhaltliche Kriterien von Qualität eine Rolle spielen. So lassen sich laut *Philipp Pohlenz* bspw. in neueren, von außen vorgegebenen Standards für die externe Qualitätssicherung der Hochschullehre, z.B. in den 2015 überarbeiteten *Standards and Guidelines for Quality Assurance* (vgl. European Association for Quality Assurance in Higher Education 2015), bindende inhaltliche Qualitätskriterien finden. Da zuvor Qualitätssicherung und Evaluation viel stärker nur auf Prozesse in der Lehre bezogen waren, könnte dieser *Trend in Richtung Inhaltsqualität* bei aller Vorsicht als Innovation bezeichnet werden.

Marcus Capellaro stützte sein Eingangsstatement neben eigenen Erfahrungen als freier Evaluator auf Ergebnisse einer Onlinebefragung der Mitglieder des AK *Gesundheitswesen*.⁷ Er erläuterte zunächst die unterschiedlichen Bereiche des Politikfeldes, die sich von Gesundheitsförderung und Prävention bis hin zu Therapie, Krankheitsbewältigung, Rehabilitation und Pflege erstrecken, sowie die Zusammensetzung des AK, da sich diese Aspekte auf die Befragungsergebnisse auswirkten. Im AK *Gesundheitswesen* sind Politik, öffentliche Verwaltung, Fachhochschulen, Universitäten und nicht universitäre Forschungseinrichtungen relativ gut vertreten, wohingegen Kostenträger (Krankenkassen und Pflegeversicherungen) sowie Leistungserbringende (Ärztinnen und Ärzte, Gesundheits- und Krankenpflegerinnen und -pfleger sowie andere Pflege- und Gesundheitsberufe) nicht repräsentiert sind. Die Onlinebefragung brachte hervor, dass einige Mitglieder die im Politikfeld nach Ein-

6 Zur ansteigenden Heterogenität der Studierendenschaften siehe auch Harris-Huermann/Mittauer/Pohlenz 2015 sowie Mittauer/Pohlenz/Harris-Huermann 2017: 275.

7 An der Onlinebefragung 2017 beteiligten sich 29 von 100 auf der Verteilerliste des AK *Gesundheitswesen* eingetragene Personen.

schätzung von *Marcus Capellaro* relativ gut verfügbaren *Sekundärdaten*⁸ nutzen, insgesamt jedoch weitaus häufiger Primärdaten erhoben werden. Zu den am häufigsten angewandten Datenerhebungsmethoden zählen hierbei Interviews und standardisiertere Befragungen, des Weiteren werden Beobachtungen im Feld oder im Labor unter den Befragten häufig durchgeführt. Medizinische Untersuchungen (physiologische Messungen etc.) und psychologische Testverfahren werden hingegen von den Befragungsteilnehmenden relativ wenig genutzt.⁹

Die auch in der Gesundheitsförderung und Prävention weit verbreitete *Forderung nach evidenzbasierter Medizin* gehe mit einem Appell zur Anwendung methodisch hochwertiger und randomisierter, kontrollierter Studien einher. Der daraus entstehende Druck auf die Forschenden und Studien anbietenden führe zu Diskussionen hinsichtlich des Nutzens solcher Designs.¹⁰ Eine im Politikfeld kontrovers diskutierte These besagt, dass der Nutzen von diesen randomisierten, kontrollierten Studien aufgrund der zu großen Unterschiede in den Populationen der Intervention und der Kontrollgruppen überschätzt werde. Vor allem im Bereich der präventiven Verhaltensänderung, aber auch bei Therapie müssten vielmehr individuelle Lebensumstände und sozio-kulturelle Erfahrungen („Wir müssen für den Erfolg von Prävention und Therapie den Mensch ganzheitlich betrachten.“) berücksichtigt werden. Innovationen seien dabei neben der *Anpassung von Verfahren und Designs an den Gegenstand* auch in den Bereichen *Digitalisierung* und intersektorielle Gesundheitspolitik (*Health in All Policies*) notwendig.

Politikfeldübergreifende Diskussion zu methodischen Standards

Die anschließende Diskussion rückte verschiedene Aspekte in den Mittelpunkt: Neben der Suche nach *innovativen Evaluationsansätzen* in den thematisierten Politikfeldern bildeten angemessene und realisierbare Verfahren zur Erfassung von *Wirkungen* in Evaluationen sowie Fragen zu (weiteren) *methodischen Standards* die drei zentralen Diskussionsachsen. Hinsichtlich der Frage nach Innovationen von Evaluation zeigten sich in den verschiedenen Politikfeldern recht unterschiedliche Anknüpfungspunkte. Im Bereich Soziale Dienstleistungen wurden hier vor allem methodische Innovationsbedarfe und -potenziale zur Aufklärung von komplexen, nicht linearen Zusammenhängen zwischen Intervention und Outcome benannt, welche von Evaluierenden geleistet werden können. Im Hochschulbereich ging es bei der Identifizierung von Innovationsbedarfen primär um die Neubestimmung bzw. Verän-

8 So gibt es bspw. Abrechnungsdaten von den niedergelassenen Ärztinnen und Ärzten mit Leistungen und Diagnosen, Qualitätsberichte der Krankenhäuser, Daten der epidemiologisch arbeitenden Krebsregister sowie die Studien des Robert Koch-Instituts zur Gesundheit von Kindern und Erwachsenen, die auch zur öffentlichen Nutzung freigegeben sind.

9 Nach Einschätzung von Marcus Capellaro könnte dieser Befund mit der AK-Struktur zusammenhängen, da u.a. Kostenträger und Leistungserbringende nicht vertreten sind.

10 Der AK *Gesundheitswesen* lässt dieser Frage besondere Aufmerksamkeit zukommen und veranstaltete bspw. 2015 eine Frühjahrstagung mit dem Titel „Evidenzbasierung in der Gesundheitsförderung: Anspruch, Wirklichkeit und der Beitrag der Evaluation“, um anhand konkreter Beispiele das Spannungsfeld zwischen Anforderungen und realer Evaluationspraxis aufzuzeigen und zu diskutieren (vgl. Wirtz/Capellaro/Spiel 2017).

derung von Nutzungsszenarien von Lehr- und Studiumevaluationen¹¹, weg von reinen Kontroll- bzw. Legitimationsinstrumenten, hin zu einem „Kulturwandel“, der auf die Integration von Evaluation in das professionelle Selbstverständnis von Lehrenden zielt. Jedoch stelle sich die bislang eher getrennte Zuständigkeit für Evaluation, Hochschuldidaktik und Qualitätsentwicklung im Hochschulbereich als eher hinderlich für Innovationen in diesem Bereich dar. Im Gesundheitswesen speisen sich Innovationen im Bereich Evaluation u.a. aus Fragen der Gegenstandsangemessenheit bzw. aus der Arbeit mit spezifischen Zielgruppen (z.B. Demenzerkrankte). Auch die Einbeziehung von sozialen Medien bzw. die Nutzung von Big Data bergen hier sowie in anderen Bereichen Innovationspotenziale, wenngleich noch wenige Erfahrungen diesbezüglich vorhanden sind. Aus dem erweiterten Diskussionskreis vertrat *Christa Peinhaupt* (EPiG GmbH) darüber hinaus die Auffassung, im Gesundheitswesen werde zunehmend verstärkt in Programmen gedacht, d.h. in mehrteiligen, intentional aufeinander bezogenen Aktivitäten bzw. Interventionen. Daraus ergeben sich ihrer Auffassung nach Innovationen für und in der Gestaltung von Evaluationen, da diese nunmehr mehrere, auf ausgewiesene Ziele ausgerichtete Interventionen parallel in den Blick nehmen muss(t)en. *Marcus Capellaro* verdeutlichte diesen Gedanken am Beispiel des politikfeldübergreifenden „Health in All Policies“-Ansatzes.

Dieser Ansatz diene darüber hinaus als Beispiel, um über Fragen von Wirkungsmodellierung bzw. -messung im Rahmen von Evaluationen zu diskutieren und deren Nützlichkeit sowie Durchführbarkeit zu reflektieren. In allen drei Politikfeldern wurde deutlich, dass Evaluationen mit Anforderungen sowie Fragen der Realisierbarkeit und des Nutzens von *Modellierungen bzw. Messungen von Wirkungen* konfrontiert sind. Diese Anforderungen werden an Evaluierende insbesondere von extern, bspw. von Auftraggebern im Zuge veränderter Formen der (daten- bzw. evidenzbasierten) Programmsteuerung im Feld der Sozialen Dienstleistungen, herangetragen. In den Bereichen Gesundheitswesen und Soziale Dienstleistungen bearbeiten Evaluierende vor allem Fragen der Wirkungszusammenhänge von Intervention und Outcome, die es zunächst qualitativ (z.B. über die Programmtheoriemodellierung oder die Rekonstruktion von theories of change) aufzuklären gelte (Wirkungsmodellierung). *Rainer Strobl* ergänzte mit Blick auf komplexe (Mehrebenen-)Programme, dass es hier darauf ankäme, insbesondere auch Ziele auf Programmebene möglichst konkret zu formulieren, um darauf aufbauend Aspekte der Zielerreichung durch die Projekte überprüfen zu können.

Gerade schwer messbare Wirkungsziele auf Outcome- und Impactebene würden Evaluierende jedoch vor besondere (methodische) Herausforderungen stellen bzw. Verfahren der Wirkungsmessung ließen sich oft gar nicht umsetzen. *Christa Peinhaupt* regte hier – mit Blick auf ein Spezifikum des Gesundheitswesens – an, die Vielzahl an vorhandenen Sekundärdaten (z.B. Sozialversicherungsdaten, Abrechnungsdaten der Krankenkassen) zu nutzen, um anspruchsvolle Kontrollgruppendesigns umzusetzen. Im Feld der Sozialen Dienstleistungen bestehe im Hinblick auf Fragen nach Wirkungen vor allem auch ein verstärktes Interesse daran, erfolg-

11 Dies schien zunächst widersprüchlich, da gerade Hochschulen bzw. Universitäten als grundsätzlich innovativ bzw. innovationsorientiert gelten.

reiche Strategien der Veränderung bspw. von Einstellungen oder Verhaltensweisen nach einer Intervention bzw. Maßnahme zu identifizieren. Für den Hochschulbereich schlug *Thorsten Braun* (Universität Stuttgart) vor, zum Nachweis von Wirkungen beispielsweise Prüfungen kompetenzorientiert zu gestalten und sie als Instrumente der Outcome-Messung zu nutzen. Gerade die Empirische Bildungsforschung stelle seiner Auffassung nach eine gute Referenz dar, wie kompetenzorientierte Tests, die probabilistische Testtheorie oder die Item-Response-Theorie für die Erfassung von Outputs und Outcomes nutzbar gemacht werden können. Seiner Ansicht nach ließen sich derartige Verfahren auch im Feld der Sozialen Dienstleistungen, bspw. in der politischen Bildung, einsetzen. Als Restriktion für entsprechende Wirkungsmessungen benannte er die oftmals gering(er)en Fallzahlen im Feld der Sozialen Dienstleistungen gegenüber der Evaluation von Lehre und Studium im Hochschulbereich. Auf die Unmöglichkeit der Wirkungsmessung mit Randomized Control Trials (RCT) in komplexen sozialen Programmen wies *Thomas Widmer* (Universität Zürich) hin unter Rekurs auf entsprechende Unternehmungen in den USA in den 1960er und 1970er Jahren. *Alexandra Caspari* schlug vor, anstelle von RCT quasi-experimentelle Designs in kleinen, ausgewählten Settings zu nutzen, um die Plausibilität von Wirkungsannahmen zu stützen. Beispielsweise könnten ausgewählte, einfach messbare Teiloutcomes mit Mixed-Methods-Verfahren gemessen bzw. nachgewiesen werden und darüber wiederum ein Beitrag zur Plausibilität der Gesamt-Outcomes geleistet werden. Aufseiten von *Thomas Widmer, Philipp Pohlenz und Rainer Strobl* wurde dies kritisch gesehen: Vor allem simple Wirkungsketten ließen sich mit quasi-experimentellen Designs bzw. RCT erfassen, komplexe Zusammenhänge in sozialen Programmen jedoch selten adäquat modellieren. Für die Bereiche Soziale Dienstleistungen und Hochschulen wurde darüber hinaus konstatiert, dass Koproduktionsprozesse den sinnvollen Einsatz (quasi-)experimenteller Designs erschwerten. *Thomas Widmer* merkte zudem an, dass die Fokussierung auf einfach messbare Teiloutcomes möglicherweise den Blick für die eigentlich relevanten Zusammenhänge und Wirkungen in sozialen Programmen verstelle.

Politikfeldübergreifend ebenso kontrovers diskutiert wurden *methodische Standards in der Arbeit mit qualitativen Daten* und hier insbesondere die Frage, ob Gespräche, Interviews oder Gruppendiskussionen nicht nur aufgezeichnet, sondern auch (voll-)transkribiert werden sollen. Dieser Punkt wurde – anders als von den gesetzten Diskussionsteilnehmenden – im erweiterten Kreis der Diskutierenden als Mindeststandard hinterfragt. Einerseits wurde eine (Voll-)Transkription vor dem Hintergrund von Aufwand und Kosten als nicht zwingend notwendig argumentiert, andererseits wurden Wissenschaftlichkeitsstandards als Argumente ins Feld geführt. Die Frage der Kosten wurde in Teilen dahingehend relativiert, dass es mittlerweile Spracherkennungssoftware gibt, die zumindest für Standardsprache eine mögliche Alternative zur Transkription bietet. *Susanne Mäder* (Univation) erläuterte, dass qualitativ hochwertige Evaluation und die Arbeit ohne Transkription aus ihrer Perspektive keinen Widerspruch darstellen würden. Sie umriss, wie über eine ausführliche Dokumentation der Daten mit integrierten Interpretationsschritten hinreichende Qualität erzeugt werden könne, insbesondere wenn eine inhaltsbezogene Auswertung im Vordergrund stehe. Vor dem Hintergrund dieser Vorgehensweise reflektierte

sie die Frage nach einer (Voll-)Transkription im Hinblick auf methodische bzw. gegenstandsbezogene Angemessenheit und die Passfähigkeit von Auswertungsmethode und -interesse. Während *Thomas Widmer* weniger die Transkription als vielmehr die ‚Black Box‘ der Interpretation von qualitativen Daten problematisierte, ergänzte *Christa Peinhaupt* die zusätzliche Herausforderung, dass Auftraggebende einerseits qualitative, differenzierte Aussagen nachfragen, andererseits aber die Ergebnisdarstellung in Tabellen oder Grafiken einfordern und die Länge qualitativer Evaluationsberichte als unpraktikabel bzw. wenig nutzbar bewerten. Diesbezüglich sieht sie es als notwendig an, Überzeugungsarbeit in Richtung Auftraggebende zu leisten.

Wohin bewegen sich Ansätze und Methoden in der Evaluation? – Fazit und Ausblick

Aus den Diskussionen der Jahre 2016 und 2017 ging hervor, dass die sechs betrachteten Politikfelder von einer mitunter großen Heterogenität von Subbereichen gekennzeichnet sind, somit typische Ansätze und auch der Grad der Standardisierung von Methoden nicht nur zwischen den Politikfeldern, sondern auch innerhalb dieser stark variieren können. Dies hängt sowohl mit den jeweiligen Gegenständen und den damit korrespondierenden angemessenen Methoden sowie Ansätzen als auch teilweise mit historischen Entwicklungslinien in den jeweiligen Feldern (bspw. Tradition der Selbstevaluation im Bereich der Sozialen Dienstleistungen, Qualitätsdiskurs durch den Einzug von New Public Management in die Hochschulen im Bereich der Hochschulevaluation) und weiteren Kontextbedingungen zusammen (siehe bspw. Unterschiede im Bereich der Forschungs-, Technologie- und Innovationspolitik in Deutschland und Österreich). Entsprechend muss die Frage nach typischen und innovativen Methoden immer auch im Kontext dieser Subbereiche gestellt und diskutiert werden. Dieser Befund der vorhandenen Heterogenität trug auch dazu bei, dass die in der Session 2017 gestellte abschließende Frage zur pauschalen Einordnung der methodischen Qualität von Evaluationen in den Politikfeldern von den Diskutierenden weitgehend zurückgewiesen wurde. Zugleich formulierten sie für die von ihnen repräsentierten Bereiche Soziale Dienstleistungen, Gesundheitswesen und Hochschule mittlere bis erhöhte Handlungsbedarfe im Hinblick auf die methodische Qualität von Evaluationen. Neben politikfeldspezifischen Besonderheiten (bspw. der 2016 konstatierten geringen Akzeptanz der Evaluation von Projekten im Bereich von Kunst und Kultur), die sich auf die gewählten Vorgehensweisen und Verfahren auswirken¹², zeichneten sich in den Diskussionen auch Kontextbedingungen ab, die für mehrere Politikfelder methodische Herausforderungen aufwerfen (bspw. der Aspekt der Koproduktionsprozesse für den Bereich der Sozialen Dienstleistungen und der Hochschulen). Darüber hinaus wurde aus den Austauschformaten der Jahre 2016 und 2017 ersichtlich, dass Methodenfragen bzw. entsprechende Min-

12 *Vera Hennefeld* aus dem AK *Kultur und Kulturpolitik* wies in der Diskussion 2016 in diesem Zusammenhang darauf hin, dass in der Konsequenz oftmals viel Aufklärungsarbeit über die Ziele von Evaluation – welche explizit nicht eine Bewertung von Kultur an sich intendiere – notwendig sei, um die Akzeptanz in diesem Feld zu erhöhen.

deststandards nicht nur mit Blick auf die Gegenstände von Evaluationen zu diskutieren sind, sondern auch maßgeblich von zur Verfügung gestellten finanziellen und zeitlichen Ressourcen abhängen. Hierbei scheinen – so der Eindruck aus der Diskussionsrunde 2017 – einige Politikfelder, z.B. das Feld der Sozialen Dienstleistungen, tendenziell stärkeren Restriktionen zu unterliegen als andere. Es zeigten sich über die Politikfelder hinweg auch Unterschiede in den Auswirkungen von Ressourcenfragen bzw. im Umgang mit diesen. Während in der Diskussion der Felder Kultur und Kulturpolitik, Forschungs-, Technologie- und Innovationspolitik sowie Entwicklungspolitik 2016 diese Aspekte beispielsweise mit Blick auf die Anforderungen einer transparenten Darstellung von Methoden und auf den Prozess der Erkenntnisgenerierung thematisiert wurden, standen 2017 vor allem Fragen der (Voll-) Transkription von Datenmaterial und der Realisierung aufwendiger Designs (z.B. Kontrollgruppendesigns) im Fokus. In nahezu allen Bereichen spielen Fragen der angemessenen Erfassung und Überprüfung von Wirkungen eine (zusehends) wichtige Rolle, wengleich die Thematik nicht in allen Bereichen gleichermaßen intensiv diskutiert wird. So scheint der Rückgriff auf Plausibilitätsüberlegungen und die Offenlegung von Grenzen der Erbringung von Wirkungsnachweisen im Bereich der Kultur und Kulturpolitik weit verbreitet zu sein, wohingegen methodische Fragen der Wirkungsmessung in den anderen betrachteten Politikfeldern stärker im Fokus stehen und kontroverser diskutiert werden. Experimentelle Designs können jedoch kaum realisiert werden und auch quasi-experimentelle Designs stellen in keinem der betrachteten Felder einen Standard dar. Vielmehr werden in der Regel, in Abhängigkeit vom Gegenstand und von zur Verfügung stehenden Ressourcen, alternative Wege gesucht, um sich Wirkungsfragen zu widmen. Insgesamt wurde deutlich, dass hierzu nicht nur zwischen den Politikfeldern, sondern auch innerhalb von ihnen, durchaus unterschiedliche Auffassungen dazu bestehen, was Wirkungsmodellierung bzw. -messung leisten soll und wie dies methodisch gut zu bearbeiten ist. Methodische Innovationen speisen sich – wie in den auf der Session 2016 diskutierten Politikfeldern – vorrangig aus den eigenen Politikbereichen und den darin zu bearbeitenden Fragen. Gerade für die Evaluation von Lehre und Studium und für weitere pädagogische Felder können aber auch nahestehende Wissenschaftsbereiche Anregungen bieten. Darüber hinaus wurden in allen drei, 2017 vertretenen Politikfeldern Innovationspotenziale durch zunehmende Digitalisierung, neue Kommunikationsformen wie Social Media und Formen der massenhaften Datengewinnung bzw. -nutzung (Big Data) gesehen, die eher auf gesamtgesellschaftlichen Trends beruhen. Hier sind weitere Entwicklungen zu erwarten, welche auch methodische Fragen aufwerfen und daher in weiteren Sessions thematisiert werden sollten.

Literatur

- Altenburg, Thomas (2017): Zwischen Schema F und Innovation. Eine politikfeldübergreifende Diskussion zu methodischen Standards. In: Zeitschrift für Evaluation, 16 (1), S. 210-217.
- European Association for Quality Assurance in Higher Education (2015): Standards and Guidelines for Quality Assurance in the European Higher Education Area (ESG). Brussels, Belgium. Verfügbar unter: <http://www.enqa.eu/index.php/home/esg/> [25.11.2017].

- Halves, Edith/Lück-Filsinger, Marianne/Schmidt, Stefan (2014): Evaluation in der Sozialen Arbeit. Entwicklungen und Herausforderungen. In: Böttcher, Wolfgang/Kerlen, Christiane/Maats, Peter/Schwab, Oliver/Sheik, Sonja (DeGEval-Vorstand) (Hg.): Evaluation in Deutschland und Österreich, Stand und Entwicklungsperspektiven in den Arbeitsfeldern der DeGEval – Gesellschaft für Evaluation. Münster: Waxmann, S. 133-138.
- Harris-Huermann, Susan/Mittauer, Lukas/Pohlenz, Philipp (Hg.) (2015): Heterogenität der Studierenden. Herausforderungen für die Qualitätsentwicklung in Studium und Lehre, neuer Fokus für Evaluation? Bielefeld: Universitätsverlag Webler.
- Mitterauer, Lukas/Pohlenz, Philipp/Harris-Huermann, Susan (2017): Aktuelle Trends der Evaluation an Hochschulen. In: Zeitschrift für Evaluation, 16 (2), S. 275-276.
- Reiter, Stefanie/Schmidt, Stefan/Strobl, Rainer/Astleithner, Florentina/Froncek, Benjamin/Stepanek, Peter (2015): Methodische Herausforderungen der Wirkungsanalyse bei knappen Ressourcen. Frühjahrstagung 2015 des AK Soziale Dienstleistungen. In: Zeitschrift für Evaluation, 14 (2), S. 319-327.
- Reiter, Stefanie/Strobl, Rainer/Buchheit, Frank (2017): Kurzbericht des AK Soziale Dienstleistungen in der DeGEval über Entwicklungen in diesem Feld und die Rolle der DeGEval. In: Zeitschrift für Evaluation, 16 (2), S. 293-296.
- Wirtz, Angela/Cappellaro, Marcus/Spiel, Georg (2017): Der Arbeitskreis Gesundheit in der DeGEval: ein Rückblick und eine Standortbestimmung. In: Zeitschrift für Evaluation, 16 (2), S. 270-273.

Nachwuchs stärken – Das Forschungs- und Praxiskolloquium des Nachwuchsnetzwerks der DeGEval 2017

Michael Bigos,¹ Laurina Bleier,² Barbara Flatters,³ Susanne Mäder,⁴ Katharina Schurz⁵

Am 06. April 2017 war es wieder soweit. Das Nachwuchsnetzwerk der DeGEval richtete zum sechsten Mal erfolgreich das Forschungs- und Praxiskolloquium für Einsteiger und Nachwuchs im Bereich Evaluation aus. Dieses innovative Format wurde ins Leben gerufen, um für die über verschiedenste Disziplinen und Wissenschaftsbereiche verstreuten Nachwuchswissenschaftlerinnen und -wissenschaftler eine geeignete Unterstützung zu schaffen. Das ist insbesondere interessant, da der interdisziplinäre Charakter von Evaluation oftmals für den einzelnen Evaluierenden bedeutet, dass kein ausgereiftes Unterstützungssystem für evaluationsfachliche Themen in den Heiminstitutionen etabliert ist. Diese Lücke zu schließen oder zumindest einen Anlaufpunkt für Peer-to-Peer-Unterstützung zu ermöglichen, ist daher der Teil der Anstrengungen des Nachwuchsnetzwerkes.

Wie in den vorangegangenen Jahren auch, wurden auch in diesem Jahr neben Qualifikationsarbeiten mit Evaluationsfokus ebenfalls Praxisprojekte der Nachwuchsevaluatorinnen und -evaluatorenen im Programm berücksichtigt, um das Format auch für Berufseinsteigerinnen und Berufseinsteiger mit praktischen Fragen zu öffnen. Das Forschungs- und Praxiskolloquium ermöglicht daher einerseits bereits eine frühzeitige Vernetzung und Rückhalt in der Fachgemeinde der Evaluation und andererseits eine niedrigschwellige und offen-konstruktive Form der Unterstützung für alle Interessierten. In 2017 konnte das Kolloquium dankenswerterweise an die Frühjahrstagung der Arbeitskreise Professionalisierung sowie Kultur und Kulturpolitik angeschlossen werden. Die Vortragenden berichteten aus drei Praxisprojekten und einer Masterarbeit. Die Anwendungsfelder waren breit gestreut über Kultur, Hochschule, Schule und Verwaltung. Mit ca. 15 Teilnehmenden stieß die Veranstaltung auch in 2017 erneut auf Resonanz.

1 Johannes Gutenberg-Universität Mainz

2 Hochschule für Musik Franz Liszt Weimar / Duisburger Philharmoniker

3 Universität zu Köln

4 Univation – Institut für Evaluation – Dr. Beywl & Associates GmbH

5 Hochschule Osnabrück

Nachfolgend werden die einzelnen Vorträge inhaltlich kurz angerissen und zentrale Diskussionslinien des Kolloquiums beschrieben. Ergänzt wird dieser Veranstaltungsbericht durch die Eindrücke der Vortragenden, um den konkreten Nutzen für die Berufseinsteigerinnen und Berufseinsteiger sowie die Nachwuchswissenschaftlerinnen und -wissenschaftler darzustellen.

Im Anschluss an die thematische Einführung und die Vorstellung des Nachwuchsnetzwerks folgte eine Kennenlernrunde, die den Teilnehmenden erlaubte, persönlich und fachlich ins Gespräch zu kommen sowie erste Kontakte zu knüpfen. Den konkreten Einstieg in das Kolloquium bot dann der Vortrag von Michael Bigos (Johannes Gutenberg-Universität Mainz) zur Implementierung von Evaluationskompetenz als schulpraktisches Thema im Lehramtsstudium. Das konzipierte Praxisprojekt ‚Schulentwicklung und Evaluation – Schulpraktische Kompetenzen im Lehramt‘ verfolgt dabei zwei sich ergänzende Zielrichtungen. Einerseits wird für eine Projektschule eine formativ-partizipative Evaluation des vorliegenden Schulentwicklungskonzepts konzipiert und umgesetzt und andererseits werden Lehramtsstudierende und beteiligte Lehrkräfte in ihren Evaluationskompetenzen weitergebildet. Die tatsächliche Evaluation der Schulentwicklung wird dabei vom Zentrum für Lehrerbildung der Johannes Gutenberg-Universität in Mainz durchgeführt, während parallel in einem Seminar die Lehramtsstudierenden die grundlegenden Fähigkeiten zur Konzeption, Durchführung und Auswertung einer formativen Evaluation am konkreten Beispiel erlernen sollen. Sowohl die Lehrkräfte als auch die Lehramtsstudierenden sollen dabei im partizipativen Design ihre jeweiligen Stärken einbringen und so am eigenen Leib die Relevanz und Potenziale von Evaluationen im Schulbereich erleben. Die Diskussion im Kolloquium fokussierte vor allem auf die Verzahnung von grundlegenden Seminareinheiten zur Evaluation mit den praktischen Evaluationsabläufen an der Projektschule. Diskutiert wurde unter anderem, wie die Summe der Inhalte und Praxisphasen in die Zeitläufe eines Semesters integriert werden können oder wie beispielsweise Themen aus den Empfehlungen für die Aus- und Weiterbildung in der Evaluation (DeGEval 2008) unter der Berücksichtigung von Vorwissen aus dem Lehramtsstudium in die Seminargestaltung einbezogen werden können. Ein weiterer Diskussionsstrang beschäftigte sich mit Vorschlägen der Teilnehmenden zur partizipativen Gestaltung der Auswertung der Ergebnisse zwischen Schule, Studierenden und Universität.

Kommentar von Michael Bigos: „Durch das Forschungs- und Praxiskolloquium hatte ich die Chance, in einer frühen Phase mit anderen Interessierten über wichtige Kernfragen meines Projekts zu diskutieren. So haben mir die Erfahrungen und Rückmeldungen der anderen Teilnehmer sehr dabei geholfen, mein Konzept zur Vermittlung von Evaluationskompetenzen für Lehramtsstudierende organisatorisch und inhaltlich sinnvoll anzupassen. Ein konkreter Ansatz war beispielsweise die Frage nach der Vernetzung der grundständigen methodischen Ausbildung der Studierenden mit den Methodenkompetenzen für die Evaluation. Diesen Austausch habe ich als sehr wertschätzend und fruchtbar erlebt.“

Laurina Bleier (Hochschule für Musik FRANZ LISZT Weimar) stellte ausgehend von ihren Erfahrungen im Rahmen eines Qualifizierungsprojekts mit On-Going-Evaluation ein mögliches Evaluationsdesign im Bereich von Kultur und Wis-

senschaft zur kontinuierlichen Überprüfung und Weiterentwicklung des Projektes ‚Safar – Musik aus Afghanistan‘ vor. In der 2016 entstandenen Masterarbeit wurde das Drittmittelprojekt ‚Safar‘ umfassend evaluiert, um die Anforderungen an Projektevaluation/-evaluatore:n im Kulturbereich zu untersuchen. Die Evaluationsergebnisse verdeutlichten, dass die Durchführung von Evaluationen bedeutsam für die Projektentwicklung ist und daher auch im folgenden Projektverlauf Anwendung finden sollen. Jedoch zeigte sich im Projektalltag, dass aufgrund fehlender Kapazitäten (zeitliche, personelle, finanzielle) Evaluation leicht in den Hintergrund geraten kann. Das Praxiskolloquium wurde genutzt, um sich über Methoden und Herangehensweisen für eine Weiterführung der Evaluationstätigkeit sowie die Implementation einer ‚Evaluationskultur‘ auseinanderzusetzen und auszutauschen.

Diskutiert wurde insbesondere, welche Möglichkeiten zur Beteiligung von und Kommunikation mit Stakeholdergruppen im Evaluationsprozess bestehen. Vorgesprochen wurde hier bspw. die Einrichtung einer begleitenden Steuerungsgruppe, die sich auch via Skype kontinuierlich austauschen kann. Zudem wurde erörtert, wie langfristige, gesellschaftliche Ziele und wie die höhere Bekanntheit afghanischer Musik ermittelt werden kann. Hierzu wurde eingebracht, dass die intendierten Wirkungen in Bezug auf die erwünschte Netzwerkbildung zunächst konzeptionell beschrieben werden sollten. Als eine mögliche Strategie zur Ermittlung der Impacts wurde vorgeschlagen, die Bekanntheit und Bedeutung afghanischer Musik in bestehenden Musikorganisationen und -festivals zu erfragen.

Kommentar von Laurina Bleier: „Beim Forschungs- und Praxiskolloquium konnte ich mich mit Akteuren aus den verschiedensten Disziplinen austauschen. Damit eröffneten sich mir neue Perspektiven auf und Herangehensweisen an projektbezogene Evaluationen, die es mir ermöglichen, meinen Blickwinkel auf die Evaluation des Projekts zu erweitern. Im Kulturbereich bzw. in diesem Projekt, treffen vielerlei Interessen und Ziele der handelnden Personen aufeinander, so dass diese in einer Evaluation besonderer Berücksichtigung bedürfen. In der offenen Diskussion wurde deutlich, dass für eine On-Going-Evaluation die Kommunikation mit und unter den Projektbeteiligten unerlässlich ist. Ich richte nun daher meinen Blick auf mögliche Kommunikationsformen, wie z.B. eine digitale Plattform, um sämtliche Meinungen, Anmerkungen etc. zentral bündeln und Evaluationsinstrumente ansetzen zu können.“

Als Praxisprojekt wurden von Katharina Schurz erste Ideen für eine kompetenzorientierte Evaluation von fachübergreifenden Förderangeboten der Hochschule Osnabrück vorgestellt. Mithilfe des vorgestellten Ansatzes sollen fachübergreifende Angebote, die durch das LearningCenter der Hochschule Osnabrück (finanziert durch den Qualitätspakt Lehre) selbst konzipiert und durchgeführt werden, intern evaluiert werden. Zielgruppen der Angebote sind Lehrende und Studierende der Hochschule in verschiedenen Qualifikationsphasen. Die Evaluation soll Aufschluss über den Kompetenzerwerb der Angebote geben und damit auch Hinweise, ob eine Anpassung oder Weiterentwicklung notwendig ist. Da sich das Vorhaben noch in der Entwicklung befindet, wurden erste Ideen zur Umsetzung im Rahmen des Kolloquiums vorgestellt. So ist angedacht auf Basis bereits gut geprüfter Selbstbeurteilungsskalen einen Online-Fragebogen zu entwickeln, um die Kompetenzen zu ermitteln.

Unklarheit herrschte noch, ob eine Messung des Erwerbs generischer Kompetenzen durch fachübergreifende Angebote überhaupt möglich und sinnvoll ist. Im Plenum wurde daher diskutiert, ob nicht auch qualitative Verfahren – auch aufgrund der Komplexität des Ansatzes und der diversen Zielgruppen – als Evaluationsmethoden in Frage kommen. Dazu wurden von den anwesenden Kolleginnen und Kollegen einige Beispiele z.B. Planspiele und Gruppendiskussionen gesammelt. Zudem wurde nach Möglichkeiten gefragt, wie Studierende zur Teilnahme an der Befragung angeregt werden könnten. Hier wurde als Anreiz genannt, den Studierenden die Ergebnisse der Befragung zur Verfügung zu stellen, damit sie ihren Kompetenzerwerb eigenständig verfolgen und reflektieren können. Dies passt gut zu der Idee, das Vorhaben in Form eines längsschnittlichen Paneldesigns zu konzipieren.

Kommentar von Katharina Schurz: „Das Forschungs- und Praxiskolloquium war für mich die ideale Möglichkeit unsere ersten Überlegungen zu unserem Vorhaben in einer offenen aber vertrauensvollen Atmosphäre vorzustellen. Die anschließende Diskussion empfand ich als sehr wertschätzend und kollegial. Die verschiedenen Impulse des Fachpublikums aus ganz unterschiedlichen Evaluationsbereichen haben mir Denkanstöße zum Design aber auch eine Bekräftigung für bereits bestehende Ideen gegeben. So fand ich den Vorschlag, qualitative Methoden wie z.B. Gruppendiskussionen einzubinden, um beispielsweise die Selbsteinschätzung einer Fremdeinschätzung gegenüber zu stellen, eine gut umsetzbare Ergänzung für unser Vorhaben.“

Außerdem präsentierte Barbara Flatters (Universität zu Köln) das Vorgehen und den aktuellen Sachstand ihrer Masterarbeit zu „Relevanz von Evaluation in deutschen Kommunen am Beispiel der Stadt Köln“. Die Bedeutung der Evaluation wird in dem Projekt anhand von Kriterien aus dem ‚Evaluation Culture Maturity‘-Index und verschiedenen Instrumenten zur Nützlichkeit von Evaluation abgeleitet. Ziel ist es, die Relevanz von Evaluierung für Kommunen zur Erfüllung und Steuerung ihrer Aufgaben und Projekte zu ermitteln. Die bisherigen Erfahrungen und Daten aus dem Feld wurden ebenso wie erste Ergebnisse der Studie vorgestellt. Die Diskussion beschäftigte sich insbesondere mit der Interpretation und Gewichtung der Ergebnisse aber auch mit den besonderen Erfahrungen und Schwierigkeiten im Feldzugang, da die Rückmeldungen der Befragten teils spontan über Telefoninterviews, teils per Mail und teils per Face-to-Face-Interview entstanden sind. Auch wurde diskutiert, inwiefern in der kommunalen Verwaltung beispielsweise Evaluationsmaßnahmen auch als solche bezeichnet bzw. inwieweit der Begriff ‚Evaluation‘ auch mit weiteren Konzepten assoziiert wird und daher möglicherweise in der Erhebung zu Missverständnissen führen kann.

Kommentar von Barbara Flatters: „Ich habe gerne die Chance genutzt, meine Masterarbeit schon vor deren Abschluss im Kolloquium zu präsentieren. Dies war sehr hilfreich, da meine Arbeit hier nicht beurteilt wurde und ich ehrlich konkrete Schwierigkeiten benennen und die übrigen Teilnehmenden um Rat bitten konnte. Diese kommen aus verschiedenen Disziplinen der Sozialwissenschaften, kennen sich aber gut mit Evaluierung aus. So wurde aus diesen verschiedenen Perspektiven die Stimmigkeit meiner bisherigen Ausarbeitung betrachtet, was mich darin bestärkt hat, mich auf einem guten Weg zu befinden. Besonders hilfreich für das wei-

tere Vorgehen waren zum einen die Hinweise auf Forschungsprojekte zu ähnlichen Themen wie dem meinen, die aber teilweise nicht publiziert sind sondern über die Vermittlung von Kontakten einsehbar sind. Zum anderen war die Diskussion um die Problematik der Verwendung des Begriffs ‚Evaluierung‘ in der Verwaltung und der Abgrenzung zu allgemeiner Berichterstattung sehr ertragreich.“

In 2018 ist ein weiteres Forschungs- und Praxiskolloquium an der FOM Hochschule Essen geplant. Weitere Hinweise zum Nachwuchsnetzwerk und anderen Formaten (z.B. LinkedIn-Gruppe „Nachwuchsnetzwerk der Gesellschaft für Evaluation (DeGEval)“) und dem Mailverteiler des Netzwerks finden sich unter <http://www.degeval.de/arbeitskreise/nachwuchsnetzwerk/>.

Literatur

DeGEval (2008): Empfehlungen für die Aus- und Weiterbildung in der Evaluation. Anforderungsprofile an Evaluatorinnen und Evaluatoren. Mainz: DeGEval – Gesellschaft für Evaluation (2. Aufl.).

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 Nachwuchsevaluatorin/einen Nachwuchsevaluator oder eine Gruppe von Nachwuchsevaluatorinnen/Nachwuchsevaluatoren vergeben (bei Gruppenarbeiten muss die Gesamtgruppe aus Nachwuchsevaluatorinnen/Nachwuchsevaluatoren 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, Masterarbeiten) als auch beispielhafte Evaluationsberichte.

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

Die DeGEval dankt dem Waxmann Verlag, der durch eine Spende das bisherige Preisgeld verdoppelt hat. Eine zusätzliche Ehrung erfolgt über eine Meldung in der Zeitschrift für Evaluation (ZfEv) und 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 der Preisträgerin/dem Preisträger bzw. den Preisträgern/Preisträgerinnen vereinbart.

Jeder Bewerbung müssen zwei einseitige schriftliche Befürwortungen renommierter Evaluatorinnen/Evaluatoren beiliegen. Alternativ kann eine der beiden Befürwortungen auch von einer Auftraggeberin/einem Auftraggeber erstellt werden. Zum Zeitpunkt der Preisverleihung soll die Preisträgerin/der Preisträger bzw. die Preisträger/Preisträgerinnen das 38. Lebensjahr noch nicht vollendet haben und keine Lebenszeitprofessur oder vergleichbare Position innehaben.

Einreichungen werden **bis zum 1. Juli** des jeweiligen Jahres erbeten. Sie sollen – neben der vorgeschlagenen Arbeit – ein Schriftenverzeichnis und Curriculum Vitae der vorgeschlagenen Person/Personen 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.

Die Auswahl der Preisträgerin/des Preisträgers bzw. der Preisträger/Preisträgerinnen 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).

**„Wirkungsorientierung und Evaluation“
21. Jahrestagung 2018 der DeGEval –
Gesellschaft für Evaluation e.V.
vom 12. bis 14. September 2018
in Dresden**

Die Forderung nach mehr Wirkungsorientierung hat in den vergangenen Jahren in verschiedenen Kontexten wachsende Bedeutung erlangt. Kern dieser Forderung ist, politische, administrative und organisationale Steuerungs- und Entscheidungsprozesse stärker als bisher an der Überprüfung der erzielten Wirkungen von Maßnahmen und Aktivitäten auszurichten.

In Österreich etwa ist die wirkungsorientierte Verwaltungsführung explizit im Bundesverfassungs- und Bundeshaushaltsgesetz verankert, während entsprechende Prinzipien in Deutschland Bestandteil des neuen Steuerungsmodells sind. Vorbild sind hier wie in vielen anderen Politikfeldern auch Managementansätze, die einer allgemeinen Produktionslogik folgen. Auch im Non-Profit-Bereich gewinnt das Thema Wirkungsorientierung an Bedeutung, da zunehmend eine Legitimation der gemeinnützigen Arbeit über den ‚Social Impact‘ von Maßnahmen und Aktivitäten eingefordert wird.

Die Verschiebung von der Input-Steuerung, die vorwiegend über die Planung und Allokation von Ressourcen funktioniert, hin zu einer stärkeren Output- bzw. Outcome-Steuerung hat besondere Bedeutung für die Evaluation, da die Wirkungsüberprüfung traditionell eines der wichtigsten Elemente bei der systematischen Bewertung von Maßnahmen in Evaluationen darstellt. Der Nachweis von Wirkungen in Evaluationen ist allerdings aus verschiedenen Gründen problembehaftet, da ‚Wirkung‘ begrifflich folgenreiche Implikationen beinhaltet.

Denn zu jeder Wirkung gehört eine Ursache, wodurch die Frage nach Wirkungen immer mit der Frage nach den sie auslösenden Bedingungen und damit mit der Frage der Kausalität verbunden ist. Klassisch kausalprüfende Evaluationsdesigns sind unter Praxisbedingungen jedoch oft nicht realisierbar oder stoßen an ethische Grenzen. Auch wird die Reduktion der untersuchten Wirkungszusammenhänge auf eine einfache unilaterale Ursache-Wirkung-Beziehung selten der sozialen Komplexität im Feld gerecht.

Daher ist auch die ausschließliche Beantwortung der Frage, *ob* eine einzelne Maßnahme (kausal) wirkt, in vielen Evaluationskontexten zu informationsarm, um hilfreich zu sein. Wichtiger vor allem für Verbesserungs- und Generalisierungszwecke ist oft die Frage, *wie* Maßnahmen wirken. Damit stellt sich zusätzlich die Frage nach der

Wirkungsweise von Evaluationsgegenständen, die seit Längerem im Kontext theoriebasierter Evaluationsansätze oder der Stichworte ‚Wirkungsmodelle‘ und ‚Wirkungsmechanismen‘ diskutiert wird („what works for whom in what circumstances?“; Pawson, 2003).

Die 21. Jahrestagung der Gesellschaft für Evaluation will daher ein Forum zur Diskussion der verschiedenen wechselseitigen Beziehungen von Wirkungsorientierung und Evaluation bereitstellen. Eingeladen sind insbesondere Beiträge zu den folgenden Aspekten:

- Wie verhalten sich wirkungsorientierte Steuerung und Evaluation zueinander? Wie verhält sich Evaluation zu Konzepten wie etwa Wirkungsanalysen oder Impact Assessments, die teils alternativ zu ‚Evaluation‘ verwendet werden?
- Was meinen wir genau mit ‚Wirkung‘ und ‚Wirkungsorientierung‘? Gibt es ein unterschiedliches Verständnis von ‚Wirkung‘ in verschiedenen Handlungsfeldern und Fachtraditionen? Welche Konsequenz hat das für die Evaluation?
- Sollte sich Evaluation mehr auf den Kausalnachweis von Wirkungen konzentrieren oder eher auf die Aufhellung von Wirkungsmechanismen („what works for whom in what circumstances?“)? Kann sie beides gemeinsam leisten?
- Was sind Möglichkeiten und Grenzen kausalprüfender Evaluationsdesigns? Welche alternativen Zugänge zur Aufhellung der Wirkungsfrage in Evaluationen stehen zur Verfügung? Können auch Designs jenseits des ‚gold standards‘ (randomisierte Kontrollgruppendesigns) Wirkungsbehauptungen empirisch untermauern?
- Welchen Beitrag können Programmtheorien, Wirkungsmodelle und Wirkungsmechanismen als Grundlage der Analyse von Wirkungszusammenhängen in Evaluationen leisten?

Nähere Informationen zur Tagung finden Sie unter: <http://www.degeval.de/veranstaltungen/jahrestagungen/dresden-2018>.

Kontakt:

DeGEval – Gesellschaft für Evaluation e.V.
Johannes Gutenberg-Universität Mainz
c/o Zentrum für Qualitätssicherung
und -entwicklung (ZQ)
Colonel-Kleinmann-Weg 2, SBII, 04-543
D-55099 Mainz

Tel.: +49-(0)6131- 39 2 68 69
Fax: +49-(0)6131- 39 2 68 68
E-Mail: info@degeval.de
<http://www.degeval.de>