

# Editorial

Dieses Heft der Zeitschrift für Evaluation kann mit einer höchst erfreulichen Nachricht beginnen: Nach doch beträchtlichen Unsicherheiten ist nunmehr das Überleben des Masterstudiengangs Evaluation am CEval gesichert. Das ist nicht nur eine gute Nachricht für Saarbrücken und alle, die sich für den Erhalt engagierten und engagieren, es ist auch eine gute Nachricht für die wissenschaftliche Evaluation und ihre Vertreter(innen). Ein solch komplexes und anspruchsvolles Paradigma wird ohne eine fundierte und spezialisierte Ausbildung beträchtliche Probleme auf dem Weg zu weiterer Professionalisierung haben. Gleichzeitig macht die gute Nachricht auch deutlich, wie unsicher der Weg zur Professionalisierung ist und wie nötig es ist, diesen Weg auszubauen, um im Bild zu bleiben: eine bessere Infrastruktur für Evaluation zu schaffen. Das wird nur mit verbreiteter Aus- und Weiterbildung gelingen.

Mittlerweile werden Feedbacks, Auditierung oder Monitoring gerne als Evaluation klassifiziert und sie schmücken sich mit dem Adjektiv „evaluativ“. Dies ist einerseits ein Ausweis des gewonnenen Renommées, gleichzeitig werden die Spezifika dessen, was als Evaluation – noch mehr: was als „gute“ Evaluation – gilt, aufgeweicht. Auch wenn diese Grenzen – in der Sache begründet – durchaus „offen“ sind, ist Evaluation ja nicht ein beliebiges Konzept. Wissenschaftsbasiert beurteilt Evaluation, so erinnert uns die nur scheinbar schlichte klassische Definition, den Wert und den Nutzen des evaluierten Gegenstandes. Wissenschaftlichkeit, Wert und Nutzen sind auch die drei Marker, die helfen, nicht nur die vier Hauptbeiträge des hier vorliegenden Heftes zu skizzieren.

Stefan Silvestrini und Reinhard Stockmann evaluieren im Auftrag der GIZ Evaluationen zur Berufsbildungsförderung und führen zudem Querschnittsauswertungen der Primärstudien durch. Mittels der Meta-Evaluation und der Queranalyse ist beabsichtigt, Erkenntnisse zu gewinnen, die über die Ergebnisse der einzelnen Evaluationen hinausgehen und zudem die Robustheit der Ergebnisse verbessern. Die Autoren verwendeten zur Analyse eine eigens für diese Studie entwickelte Methode, die auf der Methodologie der Grounded Theory basiert und, so der Befund, nicht nur große Datenmengen überblicken, sondern – anders als Systematic Reviews – auch ein sehr heterogenes Datenmaterial bearbeiten kann. Die Meta-Evaluationen der Autoren konnten Schwächen sowie Stärken der Evaluationen identifizieren, die mit einer gewissen Regelmäßigkeit auftreten, die Queranalysen konnten eine Vielzahl allgemeiner Erfolgs- und Misserfolgskriterien der evaluierten Programme herausarbeiten. Ihren doppelten Nutzen belegt die Studie, indem sie auf methodische Probleme hinweist wie eine gewisse „Lässigkeit“ von Evaluationen im Umgang mit dem Begriff ‚Multi-Methoden-Ansatz‘ oder den Mangel an Vergleiche erlaubenden Bewertungskriterien. Die „Querbefunde“ ermöglichen nicht nur wichtige methodische Hinweise

für zukünftige Evaluationen, sie geben auch deutliche Hinweise für die Effektivität und Strategie zukünftiger Programmentwicklung im Sektor Aus- und Weiterbildung. Silvestrini und Stockmann präsentieren eine Vielzahl von Faktoren, die zur Verbesserung der Programmplanung und strategischen Passung in die deutsche Entwicklungszusammenarbeit dienlich sind. Die Ergebnisse belegen die potenzielle Nützlichkeit guter Evaluationen und Meta-Evaluationen. Was der Auftraggeber jedoch daraus macht, die Autoren sprechen diesen Aspekt zum Schluss nur kurz, aber mit einer deutlichen Prise Pessimismus an, ist seine Sache.

Christoph Müller befasst sich in seinem Beitrag mit einem gleichermaßen praktischen wie methodologischen Problem. Was kann eine Wirkungsevaluation tun, wenn weder experimentelle Designs noch Pretest-Posttest-Designs angewendet werden können? Und damit ist ja eher der Normalfall als der Sonderfall beschrieben. In der Praxis ist es gängiger Ausweg, auf die Methode der retrospektiven Pretests (MRP) zurückzugreifen. Mittels dieses Verfahrens soll der Zustand widerspiegelt werden, in dem sich die Adressaten einer Intervention befunden hätten, wären sie nicht Teilnehmer gewesen. Der Autor beschreibt Vorgehen und Ergebnisse einer beeindruckenden Reihe von empirischen Studien, die MRP-Schätzungen mit den Wirkungsbefunden von experimentellen oder quasi-experimentellen vergleichen. Auch wenn, wie nicht anders zu erwarten, die Befunde dieser Forschungen kein völlig klares Bild zeichnen, so kann den MRP doch „unter dem Strich“ ein hohes Maß an Tauglichkeit attestiert werden. Müller befindet, dass die meisten Studien die interne Validität der Wirkungsmessung durch MRP bestätigen, bei der Anwendung jedoch bestimmte Bedingungen zu beachten sind. Und so ist eine Pointe der Diskussion, dass die scheinbare methodische Schlichtheit nicht dazu verführen sollte, dieses Verfahren unreflektiert einzusetzen. Im Fazit betont Müller, dass Evaluatoren die Anwendungsbedingungen und die Probleme der Methode kennen müssen, damit Fehlschlüsse vermieden werden. Dann allerdings kann die Wirkungsbewertung mittels retrospektiver Pretests die wissenschaftliche Güte der Wirkungsstudien sichern – auch wenn sie eher eine von der Praxis erzwungene Ausweichmethode ist.

Im dritten Hauptbeitrag nehmen Caroline Maria Tost und Heiner Rindermann ein Thema auf, das in früheren Heften schon mehrfach behandelt wurde: die Anwendung der Bibliometrie bei der Evaluation von Forschung. Tost und Rindermann vergleichen die Bewertung von Forschungsleistungen in den Fächern Erziehungswissenschaft und Psychologie, wie sie mit den bibliometrischen Datenbanken Scopus und Google Scholar Evaluation vorgenommen werden. Eingangs stellen sie die gewachsene und zunehmende Bedeutung von an Indikatoren gemessener Forschungsqualität dar und skizzieren einige der praktischen Konsequenzen, die sich für die Wissenschaft aus solchen Bewertungen ergeben. Es fehle, so die Verfasserin und der Verfasser, an Studien, die einen fairen Vergleich von fachbezogenen Leistungen, hier den Unterschieden zwischen pädagogischer und psychologischer Forschung, erlauben. Bibliometrisch ermittelte Fachdifferenzen werden unter anderem durch unterschiedliche historische Wurzeln der Fächer begründet: in der Erziehungswissenschaft die geisteswissenschaftliche, in der Psychologie die naturwissenschaftliche Tradition. Beim Vergleich der beiden Datenbanken wird jedoch auch deutlich, dass Differenzen durch die unterschiedliche Datengrundlage der beiden Datenban-

ken erklärt werden können. Scopus erfasst ausschließlich Zeitschriftenartikel, die für die Erziehungswissenschaft eine geringere Rolle spielen als für die Psychologie. Mit der Einbeziehung von Bewertungen anhand des Hirsch-Indexes wird zudem deutlich, dass eine höhere Produktivität psychologischer Forscher(innen) auch auf die größere Anzahl von Co-Autorenschaften zurückzuführen ist. Unerwähnt, aber auch bedenkenswert, dürfte die Tatsache sein, dass der Anteil lokal- oder regionalbezogener Forschung in der Erziehungswissenschaft deutlich größer sein dürfte als in der Psychologie, woraus unterschiedliche Publikationsinteressen und -chancen resultieren. Die Qualität von Datenbanken ist nicht nur durch die Beseitigung methodischer Mängel zu steigern. Die Beantwortung der Frage danach, wie Produktivität und Rezeptivität wissenschaftlicher Forschung vergleichbar gemacht werden können, legt die Forderung nahe, so Tost und Rindermann, spezifische Normwerte für Publikations- und Rezeptionszahlen zu entwickeln. Es bleibt angesichts des hier vorliegenden Versuchs, zur Verbesserung und Fairness des Systems der Leistungserfassung beizutragen, auch die Frage danach, ob das System nicht selbst Gegenstand einer Evaluation sein müsste: nicht Kritik „im System“, sondern Kritik „am System“.

Michael Brocks und Juliane Keitels Beitrag ist in gewisser Hinsicht grundsätzlicher Art: „Kritik am System“. Zwar wird die Notwendigkeit einer Evaluation akademischer Lehre nicht infrage gestellt, ganz im Gegenteil. Aber das dominante System der quantitativen Lehrbewertung wird massiv kritisiert. Die Alternative sei die „responsive Evaluation“ universitärer Lehrveranstaltungen. Da standardisierte Verfahren der Lehrkritik zu positiven wie negativen Sanktionen gegenüber den Lehrenden eingesetzt würden, müsste die Validität ihrer Ergebnisse gesichert sein. Allerdings setzen sie, dabei der quantitativ-hypothesenprüfenden Methodologie verpflichtet, ein bestimmtes und einseitig instruktionsorientiertes Modell guter Lehre voraus, wobei die zugrunde gelegten Standards zudem häufig nicht transparent dargelegt seien. Die aktive Rolle der Studierenden an Lernprozessen und -ergebnissen oder die Bedeutung der Rahmenbedingungen würden nicht oder nur unzureichend abgebildet. Daten ließen sich deshalb willkürlich interpretieren, der Entwicklung von Lehre könnten sie so nicht nutzen. Brock und Keitel legen alternativ ihr Verständnis von ‚guter Lehre‘ dar, das auf einem konstruktivistischen und partizipativen Lehr-Lernverständnis gründet, in das eine responsive Evaluation eingebettet ist. In ihrem Konzept einer partizipativen Lehre und Lehrevaluation könnten demnach die Bedürfnisse und Zielperspektiven der Lernenden ermittelt und die komplexen Interaktionsdynamiken und nicht antizipierten Folgen pädagogischen Handelns erfasst werden. Anders als standardisierte Verfahren zeichnen sich qualitative bzw. rekonstruktive Zugänge demnach dadurch aus, dass sie Lehre nicht nach vorher festgelegten, normativen Kriterien erfassen. Sie sind prozessbegleitend und damit integraler Teil der partizipativen Lehre. Sie dienen, so die Hauptthese des Beitrags, der Entwicklung von Lehre. Inwieweit die Kritik an standardisierten Verfahren, insbesondere die Beschreibung ihrer Limitierungen, stichhaltig ist, bleibt zu diskutieren. Dazu gehört auch zu fragen, was das Alternativmodell der responsiven Evaluation nicht zu leisten vermag und ob es nicht mindestens so stark normdurchsetzend ist wie standardisierte Verfahren.

Im Heftteil „Information & Service“ findet sich ein Praxisbericht von Susanne Johanna Väh und Oliver Rohde zur Evaluation internationaler Forschungsförderung. Spezifisch geht es um den Deutsch-Ägyptischen Forschungsfonds (GERF), der ein Instrument zur Förderung der bilateralen Zusammenarbeit in Wissenschaft und Forschung zwischen beiden Ländern ist. Die verantwortlichen Ministerien verfolgen das Ziel, Forschung und Industrie zu vernetzen. Bisher wurden insgesamt 52 deutsch-ägyptische Forschungsvorhaben aus verschiedenen Themenbereichen gefördert. Die Autorin und der Autor erläutern ihr Evaluationsdesign und die Schwierigkeiten einer Wirkungsanalyse, weil die nicht geförderten Projekte keine Kontrollgruppe zur Gruppe der geförderten sein kann, denn gerade die Differenzen zwischen beiden gaben ja den Ausschlag für die Auswahlentscheidung. Mittels Dokumentenanalysen und teilstandardisierten Befragungen ließ sich festhalten, dass das Förderprogramm in einer Vielzahl von Projekten die Weichen für ein Weiterführen der Kooperation und für Folgeaktivitäten stellte. Inwiefern es über lose Kooperationen hinaus gelingt, neue Projekte zu initiieren, sei allerdings fraglich. Bei der jüngsten Bekanntmachung GERF IV 2015 wurden Projekte zu einer sehr großen Breite an Themen eingereicht wie z.B. Nahrungsmittelproduktion und Lebensmittelsicherheit, Wasserressourcenmanagement, erneuerbare Energien oder Stadtplanung. Der Start der geförderten Projekte ist für diesen Sommer vorgesehen.

Richard Heidler, Anke Reinhardt und Uta Saß berichten über die Tagung „Evaluation der Förderung von Wissenschaft und Forschung“ (16. Juni 2016 in Bonn). Sie stellen eingangs fest, dass die in zwei Jahrzehnten systematisch betriebene Evaluation in Deutschland sich zu einem wichtigen Bestandteil der „Good Governance“ von Förderorganisationen entwickelt habe. Der Bericht verführt dazu, wenigstens anzudeuten, dass hier Themen und Fragen verhandelt wurden, die auch im Hauptteil dieser Ausgabe der Zeitschrift für Evaluation berührt werden. Lutz Bornmann z.B. fragte, ob und wie man gesellschaftlichen ‚Impact‘ von Forschung messen kann. Stefan Kuhlmann kritisierte simplifizierende Kennziffern zur Messung von Forschungsleistung wie Impact-Factor, Hirsch-Index oder Shanghai-Ranking sowie den opportunistischen Umgang von Forscher(inne)n mit solchen Indikatorensystemen. Als Folge eines verstärkten Evaluationshandelns lassen sich, so Widmer, verschiedene „pathologische Entwicklungen“ beobachten, die den eigentlichen Zielen der Evaluation entgegenstünden. Er macht Ermüdungserscheinungen, Resignation und adaptives Verhalten aus. Außerdem genügten einige Evaluationen selbst nicht den Qualitätsanforderungen. Nutzen und Anwendungspotenzial seien damit beschränkt. Die Tagung habe gezeigt, so die Berichterstatterinnen und der Berichterstatter, dass die Evaluation von Forschungsförderung heute ein professionell organisiertes und wissenschaftlich angebundenes Arbeitsfeld und Bestandteil guter Governance von Förderorganisationen darstellt.

Im Heftteil „Literaturüberblick und Rezensionen“ bespricht Petra Wagner den, wie sie feststellt, bunten und mit innovativen und zugleich brauchbaren Konzepten, Anregungen und Ideen für konkrete Evaluationsprojekte gefüllten Sammelband: „Evaluationspraxis. Professionalisierung – Ansätze – Methoden“, der von Susanne Giel, Katharina Klockgether und Susanne Mäder herausgegeben wurde. Michael Frais rezensiert den Erfahrungsbericht der „Schulinspektion in Hamburg“ (her-

ausgegeben von Marcus Pietsch, Barbara Scholand und Klaudia Schulte) und zeigt, dass die Verfahren der Inspektion zunehmend von der Evaluation und ihren Standards lernen. Schließlich schaut sich Maria Gutknecht-Gmeiner den Band „evaluiert“ von Lars Balzer und Wolfgang Beywl an, der als Planungsbuch für Evaluationen im Bildungsbereich konzipiert ist, eine, wie sie befindet, theoretisch fundierte, aber gleichzeitig praxisnahe Einführung in die Evaluation für Studierende sowie gleichermaßen für interessierte Praktiker(innen).

Die Rubrik „DeGEval ...Info“ ist von den Berichten der Frühjahrstagungen der Arbeitskreise geprägt. Jan Wessels, Sonja Kind und Leo Wangler berichten vom Frühjahrstreffen des Arbeitskreises Forschungs-, Technologie- und Innovationspolitik (FTI) in Berlin. Schwerpunkt des diesjährigen Treffens war das Thema „Projektträger und Evaluation“. Das Frühjahrstreffen sollte die veränderte Rolle von Projektträgern aus zwei möglichen Perspektiven beleuchten: ihre sich verändernde Rolle bei Evaluationen ihrer eigenen Fördermaßnahmen und als ‚Evaluatoren‘ fremder Maßnahmen.

Dörte Schott, Christine Nolte und Manuel Merz sprechen über „Die Nutzung von Wissensressourcen und von Zusammenarbeit für die Weiterentwicklung der Evaluationspraxis“. Ziel des Kolloquiums des AK Professionalisierung war es, in einen Austausch darüber zu kommen, welche frei oder leicht zugänglichen Wissensressourcen über Evaluation zur Weiterentwicklung der Evaluationspraxis genutzt und möglicherweise noch gewünscht werden und welche Potenziale im zwischenmenschlichen Austausch oder in der gezielten Zusammenarbeit liegen.

Evaluation und politische Gestaltungs- und Entscheidungsprozesse waren Tagungsthemen des AK Soziale Dienstleistungen. In Kooperation mit dem Forschungsinstitut für Gesellschaftliche Weiterentwicklung (FGW) behandelte der Arbeitskreis eines seiner Dauerthemen: Wie können die Nutzung von Evaluationsergebnissen durch die Adressatinnen und Adressaten, auch aus Politik und Verwaltung, sowie der Wissenstransfer zwischen Wissenschaft und Praxis in beiden Richtungen gefördert und intensiviert werden? Stefanie Reiter, Stefan Schmidt und Rainer Strobl verantworten den Bericht.

Unter dem Titel „Zukunft gestalten – Methoden der strategischen Vorausschau und Praxisbeispiele“ diskutierte der AK Evaluation in der Wirtschaft den Nutzen einer strategischen Vorausschau als Teil von Ex-ante-Evaluationsprozessen in Unternehmen, Clustern und Netzwerken. Sonja Kind, Roman Noetzel und Oliver Rohde berichten in ihrer Nachbetrachtung auch über die Praxisbeispiele, die zeigen, dass Methoden der Vorausschau unabhängig von Branchen oder Sektoren eingesetzt werden können.

Lucas Jacobs, Jessica Schendzielorz und Dorothea Giesen-Thole berichten von der Frühjahrstagung des AK „Methoden in der Evaluation“, der einlud, um die Möglichkeiten und Grenzen qualitativer Sozialforschung in der Evaluation zu erörtern. Auch auf die damit verbundenen Schwierigkeiten und den notwendigen Aufwand weisen sie dabei hin.

Das Forschungs- und Praxiskolloquium des Nachwuchsnetzwerks schloss sich an die Frühjahrstagung des AK Methoden an. Das Kolloquium wurde zum fünften Mal in Folge durchgeführt. Zielgruppe sind vorrangig Nachwuchswissenschaft-

lerinnen und -wissenschaftler sowie Berufseinsteigerinnen und -einsteiger im Bereich der Evaluation. Susanne Mäder, Miriam Burfeind, Annekatrin Gehre, Chiara Pierobon und Angela Ulrich berichten.

Last but not least schildern und reflektieren Michael Kalman und Anja Lietzmann „Aktuelle Formate der beruflichen Aus- und Weiterbildung – Evaluation in einem weiten Feld“. Die so betitelte Frühjahrstagung des Arbeitskreises Berufliche Bildung nahm neuere Entwicklungen zum Anlass, um Evaluationen relevanter aktueller Formate der beruflichen Aus- und Weiterbildung aufzugreifen. Eine Gewissheit, dass die Evaluationsergebnisse von den Stakeholdern genutzt werden, gibt es freilich nicht, konzedieren die Berichterstatterin und der Berichterstatter. Anwendung und Nutzung seien vom Auftragnehmenden auch nur bedingt beeinflussbar.

Zwei Aufrufe stehen am Ende des Heftes 2 aus 2016. Aus Anlass des 20. Jubiläums wird die DeGEval-Jahrestagung den Blick zurück mit einem Blick in die Zukunft verknüpfen. Ein gutes Stück ‚Selbstevaluation‘, wenn man so will. Die Jubiläumstagung der DeGEval – Gesellschaft für Evaluation e.V. wird vom 20. bis 22. September 2017 an der Johannes Gutenberg-Universität in Mainz stattfinden. Der „Call for Proposals“ ist mit diesem Heft der Zeitschrift für Evaluation offiziell.

Gleiches gilt auch für den DeGEval-Nachwuchspreis, der einmal jährlich für eine herausragende Arbeit im Bereich Evaluation im deutschsprachigen Raum an eine Nachwuchsevaluatorin/einen Nachwuchsevaluator oder eine Gruppe von Nachwuchsevaluatorinnen/Nachwuchsevaluatoren vergeben wird. Die Höhe des vom Waxmann Verlag mitfinanzierten Preisgeldes beträgt 1.000,- Euro. Alle Mitglieder der DeGEval und alle an Evaluation interessierten Akteure sind aufgefordert, Vorschläge einzureichen.

Schließlich bleibt mir, mich im Namen aller Herausgeberinnen und Herausgeber der Zeitschrift für die Beiträge bei den Autorinnen und Autoren zu bedanken. Den Leserinnen und Lesern wünschen wir Erkenntnisse und gerne auch Interesse an Gegenreden. Auf jeden Fall hoffen wir, dass sie die Vielfalt der Fragestellungen und der Felder, in denen sich Evaluation mehr und mehr etabliert, als interessante Herausforderung betrachten.

Wolfgang Böttcher

Für die Herausgeberinnen und Herausgeber