

## Editorial

Evaluationen oder der Evaluation verwandte Verfahren haben in der jüngeren Vergangenheit in Bildung und Wissenschaft eine hohe Bedeutung erfahren. Organisation, Prozesse und Ergebnisse von Unterricht und Forschung werden kritischen Analysen unterzogen, die in der Regel von Teams durchgeführt werden, in denen sich auch – manchmal ausschließlich – Fachkollegen bzw. -kolleginnen der zu bewertenden Einrichtungen und Tätigkeiten befinden.

*Silke Gülker, Dagmar Simon und Marc Torke* thematisieren im ersten Originalbeitrag dieser Ausgabe der *Zeitschrift für Evaluation* diese kollegiale bzw. Peer-Evaluation im Bereich der Wissenschaft. Sie sei durch eine strukturelle und aus ihrer Sicht unaufhebbare Spannung gekennzeichnet. Obwohl als Mechanismus der externen Kontrolle intendiert, soll die Wissenschaftsevaluation eben nicht nur kontrollieren, sondern sie soll auch Rückmeldungen über Möglichkeiten zur Verbesserung der Forschung und der Organisation liefern. Die Autorinnen und der Autor analysieren ihre Ausgangsthese vor dem Hintergrund dreier verschiedener Verfahrenstypen: dem Evaluationsverfahren der deutschen Leibniz-Gemeinschaft, dem des niederländischen Standard Evaluation Protocol und dem der britischen Research Assessment Exercise. Damit sind Verfahren ausgewählt, die maximal kontrastierend sind. Im Fazit wird deutlich, dass beide Orientierungsmuster, nämlich die kollegiale Rückmeldung und der Auftrag der externen Kontrolle in allen analysierten Evaluationsverfahren in permanenter Spannung zueinander standen. Ein zentraler Befund des Beitrags ist es, dass trotz des Auftrags der externen Kontrolle alle Beteiligten – Evaluierte und Evaluierende – ein kollegiales Feedback erwarten. Den Selbstauskünften zufolge geht es sogar eher darum, Hilfen und Anregungen zu geben, als das jeweilige Forschungsfeld zu kontrollieren. Diesem Ansinnen jedoch sind durch die Verfahrensregeln mehr oder weniger enge Grenzen gesetzt.

Der Aufsatz von *Maria Gutknecht-Gmeiner* befasst sich mit einem ähnlichen Thema – aber unter einer anderen Perspektive. Auch sie verweist auf die Tatsache, dass bei der Bewertung von Schulqualität und im Bereich der Hochschulakkreditierung davon ausgegangen wird, dass Fachkolleginnen und -kollegen als Sachverständige die Arbeit in den Bildungseinrichtungen angemessen beurteilen können. Sie macht in ihrem Beitrag deutlich, dass leider bisher zu wenig Wert auf die Eva-

luationskompetenz der Experten bzw. Expertinnen im sogenannten Peer-Review gelegt wurde. Sie nimmt Bezug auf einen der wenigen Versuche, dieses zu ändern: Im Rahmen einer europäischen Initiative zur Etablierung des Peer-Review als Evaluierungsverfahren in der beruflichen Bildung beschreibt sie ein Curriculum zum Erwerb von Wissen und Können im Feld der Evaluation. Die Autorin kritisiert die häufig vertretene Annahme, dass Peers quasi automatisch in der Lage seien, Organisation und Unterricht in Bildungseinrichtungen kompetent zu bewerten. Sie geht davon aus, dass ein Peer-Review-Verfahren Gütekriterien unterworfen sein muss, die sich u.a. auch aus den Standards der Evaluation ableiten. Ausführlich beschreibt Frau Gutknecht-Gmeiner Kompetenzprofile für Peers und ein Curriculum für die Peer-Schulungen. Das hier vorgestellte Curriculum wurde von der Autorin selbst umgesetzt und evaluiert. Ihr Fazit lautet, dass bei Peer-Reviews auf Kenntnisse und Fähigkeiten im Bereich der Evaluation nicht verzichtet werden kann. Gleichzeitig jedoch befindet sie auf Basis ihres Schulungsprogramms, dass eine auf die jeweiligen Anforderungen fokussierte Schulung ausreichend sein dürfte.

Die zwei folgenden Originalbeiträge lassen sich auch unter einer gemeinsamen Perspektive behandeln. Beiden geht es um die Frage der Evidenz für Wirkungen von Programmen, beide zeichnen sich auch durch methodische Reflexion aus.

*Georg Stadtmann, Christian Pierdzioch und Philipp Decker* bearbeiten eine in der Lehr-Lernforschung zentrale Frage, nämlich die nach den Wirkungen der Teilnahme an unterrichtlichen Arrangements. Zu einer Vorlesung im Fach Angewandte Wirtschaftstheorie wurde studienbegleitend ein webbasiertes Tutorium eingesetzt. Effektvariable war der Erfolg in der Klausur. Dabei ging es besonders um die Behaltensleistung der Studierenden. Das experimentelle Design der Studie, das freilich nicht alle lernerfolgsrelevanten Faktoren erfassen konnte, führt zu der ersten vorsichtigen Schlussfolgerung, dass die Nutzung zusätzlicher E-Learning-Angebote einen positiven Einfluss hat.

Der Beitrag von *Christoph Emanuel Müller und Ingo Konradt* befasst sich mit einer methodologisch und methodisch höchst aktuellen Fragestellung, nämlich der Messung kausaler Interventionseffekte. Der Goldstandard der Wirkungsmessung von sozialen Programmen, Projekten oder z.B. pädagogischen Interventionen, das kontrollierte und randomisierte Experiment, ist in der Evaluationspraxis nicht leicht umsetzbar. Neben einer Vielzahl anderer Gründe erweisen sich häufig zu geringe Größe der Stichprobe und Unklarheit darüber, ob sie tatsächlich ein repräsentatives Abbild der Grundgesamtheit darstellt, als Hindernisse. Dies jedoch führt dazu, dass die in experimentellen Wirkungsevaluationen in der Regel angewandten Signifikanztests nicht geeignet sind, Zufall von Systematik zu unterscheiden. Müller und Konradt stellen als Alternative den Randomization-Test – auch als Permutations-Test bekannt – vor. Im Rahmen dieses Tests müssen keine Annahmen über die Grundgesamtheit getroffen werden. Er verbessert die interne Validität der Wirkungsanalyse. Damit können zwar keine statistischen Aussagen über die Grundgesamtheit getroffen werden, die externe Validität einer Intervention ist damit also nicht überprüfbar, im Feld der Evaluation ist, so die Autoren, dies jedoch nicht unbedingt ein Nachteil. Die Autoren demonstrieren das Verfahren am Beispiel einer Wirkungsevaluation, die sich auf ein Internetportal zum Klimaschutz bezieht. Auf

eine besonders relevante Feststellung der Autoren ist hinzuweisen. Sie beklagen mit Recht die häufig unreflektierte Nutzung klassischer Signifikanztests in der experimentellen Wirkungsevaluation und mahnen an, dass alternative, nicht parametrische Verfahren Gegenstand der Methodenausbildung und -weiterbildung sein müssten.

Über die vier Originalbeiträge hinaus bietet die vorliegende Ausgabe im Informations- und Serviceteil sowie in der Rubrik DeGEval ...Info zahlreiche weitere lesenswerte Beiträge.

In der Reihe „Evaluationsinstitute stellen sich vor“ schreiben ihr Geschäftsführer, *Helmut Asche*, und Senior Evaluator *Helge Roxin* über das deutsche Evaluierungsinstitut der Entwicklungszusammenarbeit (DEval), das im Frühjahr 2012 seine Arbeit aufgenommen hat. Alleinige Gesellschafterin der gemeinnützigen Gesellschaft mit beschränkter Haftung ist die Bundesrepublik Deutschland, vertreten durch das Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung. In ihrem Beitrag thematisieren die Autoren unter anderem die Frage, ob die große Nähe zur Arbeit des Entwicklungsministeriums die wissenschaftliche Souveränität des Instituts gefährden könnte. Die Autoren sind im Hinblick auf die Unabhängigkeit positiv gestimmt, machen aber deutlich, dass es auf die Praxis ankommen werde, in der sich erweisen wird, inwieweit wissenschaftliche Distanz gesichert wird. Als Leitgedanken des DEval wird die Erhöhung der entwicklungspolitischen Wirksamkeit, ihrer Wirtschaftlichkeit und Nachhaltigkeit ausführlich erläutert. Mit der Absicht, Evaluationsergebnisse für evidenzbasierte Entscheidungen zu nutzen, wird das DEval möglicherweise dazu beitragen, nicht nur das Feld der Entwicklungspolitik im Sinne systematischer Evaluation positiv zu beeinflussen. Helmut Asche, der Teilnehmer des Podiums der letzten DeGEval-Jahrestagung war, hat, wenn ich es richtig sehe, durchaus das Interesse, die Idee der Evaluation auch in anderen Politikbereichen zu stärken.

*Anke Reinhardt* beschreibt das Evaluationsvorgehen der Deutschen Forschungsgemeinschaft (DFG). Sie zeigt, wie sehr die DFG daran interessiert ist, ihre eigene Förderpolitik bewerten zu lassen. Sie beschreibt dabei auch die Förderungsarten der DFG, der größten Förderorganisation für Grundlagenforschung in Europa. Evaluationsstudien, so die Autorin, sind ein wichtiges Instrument der Selbstkontrolle und Qualitätssicherung. Um Evaluationsmüdigkeit zu verhindern, führt die DFG Evaluation immer anlassbezogen durch. Vorgehen und Methoden der Evaluation werden beschrieben. Die von der DFG benutzten Standards profitieren, so die Autorin, von den Evaluationsstandards der DeGEval. In ihrem Beitrag spricht Frau Reinhardt auch eine interessante Entwicklung an: Evaluation und wissenschaftliche Forschung nähern sich ihrer Meinung nach immer stärker an. Die DFG will die Forschungsevaluation auch in internationaler Perspektive weiterentwickeln und ihre Professionalisierung vorantreiben.

„Nicht das Werkzeug soll am Anfang stehen, sondern das Nachdenken über die Fragen, die beantwortet werden sollen – und mit welchen Mitteln sie beantwortet werden können“, so eine Stellungnahme zum Stellenwert von Wirkungsevidenz, die sich im Tagungsbericht von *Michaela Raab* nachlesen lässt. Frau Raab beschreibt die spannenden und abwägenden Debatten, die bei der Konferenz „Politics of Evi-

dence' in Brighton im April dieses Jahres geführt wurden. Mehr als 100 internationale Expertinnen und Experten waren der Einladung von *Big Push Forward*, einer Initiative von Fachleuten für Monitoring und Evaluierung und der Entwicklungszusammenarbeit, gefolgt. Der Wunsch von Programminitiatoren nach demonstrierbarer Wirksamkeit von Maßnahmen und eine transparente Rechenschaftslegung seien nachvollziehbar, die Gefahr einer Einengung des Methodenrepertoires als Folge eines schlichten Impact-Denkens sei aber sehr ernst zu nehmen. Veränderungsprozesse seien so zum Beispiel nicht zu verbessern. *Big Push Forward*, so die Berichtserstatterin, „plädiert für offenere Ansätze, die auf die unterschiedlichen Bedarfe und Kapazitäten zugeschnitten werden können“.

Abgerundet wird der allgemeine Serviceteil durch insgesamt vier Rezensionen zu aktuellen Buchpublikationen im Feld der Evaluation.

In der Rubrik DeGEval ...Info befindet sich der Call for Papers für die Jahrestagung 2014. Diese wird in Kooperation mit der schweizerischen Evaluationsgesellschaft SEVAL in Zürich stattfinden und das Thema ‚Professionalisierung‘ haben. Wie üblich müssen Beiträge bis Ende März auf der Kongress-Homepage angemeldet werden. Hingewiesen sei auch auf die Ausschreibung des Nachwuchspreises 2014 der DeGEval, der im Rahmen der Jahrestagung verliehen werden wird.

*Iris Fischl, Marianne Kulicke und Jan Wessels* berichten vom Frühjahrestreffen des Arbeitskreises Forschungs-, Technologie- und Innovationspolitik, das sich einem aktuellen Thema widmete. Während in der Vergangenheit Evaluationen zumeist nur einzelne Programme im Blick hatten, beziehen sich aus guten Gründen immer mehr Evaluationen auf ein definiertes Portfolio von Maßnahmen. Allerdings gelte dieser Befund lediglich für Österreich, während in Deutschland die Portfolioevaluation noch selten sei.

Komplexität und Evaluation von Programmen der Beruflichen Bildung war das Thema der Frühjahrstagung des Arbeitskreises Berufliche Bildung. *Michael Kalman* und *Jürgen Götzhaber* berichten über diese Veranstaltung und referieren zentrale Inhalte der Vorträge. Es ging dort u.a. um die Frage des Herangehens und der Strategien, wie mit Komplexität umzugehen sei. Dabei, so die Autoren, wurde gefragt, wie eine adäquate Reduktion von Komplexität aussehen oder transparent gemacht werden könne. Dass Komplexitätsreduktion ihre Grenzen hat, wurde hier auch deutlich.

*Maria Gutknecht-Gmeiner, Brigitte Wotha und Angela Wroblewski* diskutieren die Frage, wie Evaluierungen durch Einbeziehung der Geschlechterperspektive ihre Qualität sichern bzw. verbessern können. Ihr Beitrag basiert auf einer aktuellen Entwicklung innerhalb der DeGEval: Mit Beschluss der Jahrestagung 2012 bemüht sich eine kleinere Ad-hoc-Gruppe um Vorschläge zur Revision der Standards der Evaluation. Mit dem hier vorliegenden Beitrag bringt sich der Arbeitskreis Gender Mainstreaming konstruktiv in diese Debatte ein.

Im vorliegenden Heft der *Zeitschrift für Evaluation* befassen sich, wie wir gesehen haben, verschiedene Beiträge mit dem Thema Evidenz, das auch Gegenstand des letzten Positionspapiers der DeGEval war. Der AK Schulen der DeGEval hat sich nun als Ergebnis seiner Frühjahrstagung 2013 entschlossen, ein gegenstandsbezogenes Positionspapier zum Thema vorzulegen. Sein Titel „Schulqualität zwischen

Evidenzbasierung und Effekthascherei“ verweist auf den von der Schulforschung häufig belegten Befund, dass sich zwischen den aufwendig erzeugten Belegen zur Schul- und Unterrichtsqualität und -wirksamkeit einerseits und den tatsächlichen Verbesserungen andererseits große Distanzen ausmachen lassen.

*Michael Fraiss* und *Angela Stumpp-Marx*, Sprecher und Sprecherin des AK, formulieren scharf: „Es hat sich in den vergangenen Jahren an vielen Beispielen gezeigt, dass Evaluationen im schulischen Feld zum Teil mit Nutzlosigkeit, Wirkungslosigkeit und Effekthascherei in Verbindung gebracht werden“. Das Positionspapier fordert die Verstärkung partizipativer Ansätze der Evaluationen, damit deren Nutzen und Wirksamkeit erhöht werden. Auch wenn eine Top-down-Steuerung im Schulsystem nicht grundsätzlich abgelehnt wird, so sei doch „eine prozessorientierte Vorgehensweise auf Ebene der einzelnen Schuleinheit notwendig“. Es ist zu wünschen, dass das Positionspapier zu Diskussionen anregt, insbesondere auf Ebene der schulpolitisch Verantwortlichen.

Ich danke allen, die zum Gelingen dieser Ausgabe der *Zeitschrift für Evaluation* beigetragen haben, und wünsche viel Freude bei der Lektüre.

*Wolfgang Böttcher*

Für die Herausgeberinnen und Herausgeber