

Editorial

Im ersten Heft des Jahrgangs 2017 können wir Ihnen, liebe Leserinnen und Leser, eine breite Palette interessanter Beiträge und Informationen anbieten. Die Originalarbeiten reichen vom konkreten Analyseschema für die Evaluation politisch-demokratischer Prozesse über die Evaluationsergebnisse von Maßnahmen zur Verbesserung von Arbeitsmarktchancen und von Auswahlverfahren zur Studienzulassung bis hin zur Entwicklung und Evaluation eines Wirkmodells für ein Jugend-Sportprogramm und zu einer breiten Literaturreview über Methode und Anwendungen der „Realistic Evaluation“. Bezüglich der Herkunft der Beiträge kann eine erfreuliche Internationalität konstatiert werden, denn von den fünf Originalarbeiten kommen zwei aus Deutschland, zwei aus der Schweiz und eine aus Österreich. Auch sind von den Autorinnen und Autoren erfreulich viele unterschiedliche Institutionen repräsentiert, nämlich vier Universitäten, ein Institut der Bundesagentur für Arbeit, zwei Fachhochschulen, eine Sporthochschule sowie eine GmbH für Datenanalyse. Neben den fünf Originalbeiträgen finden sich im Informations- und Serviceteil des Heftes mehrere Praxisberichte, gefolgt von Rezensionen aktueller Publikationen zur Evaluation. Über Aktuelles aus der Gesellschaft für Evaluation e.V berichtet die Rubrik *DeGeval ...Info* und abschließend gilt es den Gutachterinnen und Gutachtern der *ZfEv* im Jahr 2015 und 2016 zu danken, deren Namen nach den Originalbeiträgen aufgelistet sind.

Das Ziel des ersten Originalbeitrags „Entwicklung eines Analyseschemas für die Evaluation der Strategischen Umweltprüfung in Deutschland“ von *Anke Rehausen* und *Lena Franziska Burchartz* (Technische Universität Berlin) besteht darin, einen bisher fehlenden methodischen Ansatz für eine sektorübergreifende Fallstudienanalyse zur Evaluation der strategischen Umweltprüfung (SUP) in Deutschland vorzustellen und zu diskutieren.

Ausgehend von verschiedenen Konzepten der SUP-Effektivität (procedural/substantive/transactive) wird zur Verbindung von Ergebnissen der internationalen Literaturanalyse mit Expert(inn)enmeinungen/Interviews eine reduzierte Delphi-Methode („one-stop Delphi“) angewendet, bei der neun international ausgewiesene SUP-Expert(inn)en gebeten wurden, die vorgeschlagenen Kriterien und deren Indikatoren sowie die Beschreibung der idealtypischen SUP-Praxis frei zu kommentieren. Darauf aufbauend wurden für die Indikatoren Qualitätsstufen formuliert, die zusammengenommen eine Einschätzung der Qualität der jeweiligen Kriterien ermöglichen. Hierbei wird aber nicht das Ziel verfolgt, durch Aggregation eine Gesamteinschätzung über die SUP zu erhalten; vielmehr soll ein differenziertes Bild dazu dienen, Schwächen und Stärken der SUP zu identifizieren.

Das detailliert entwickelte Analyseschema ist in Form einer Tabelle dargelegt, die eine Reihe von Kriterien enthält (rechtzeitige Integration der SUP in den Planungsprozess; Abschichtung; Substanz und Qualität des Scopings; Qualität der Alternativenwahl und Alternativenprüfung; Qualität der Öffentlichkeitsbeteiligung; Konzeption der Überwachung der Umweltauswirkungen). Die jeweils zu beurteilende hohe/(mittlere)/geringe Berücksichtigung der den Kriterien zugeordneten Indikatoren liefert die differenzierten qualitativen Bewertungen der einzelnen Kriterien, die sich zum qualitativen Gesamtbild zusammenfügen lassen.

Die Praktikabilität und Validität des methodischen Ansatzes und des darauf aufbauenden Analyseschemas wurde anhand von zwei aktuellen Fallstudien (SUP zum Regionalplan zur Windenergieerzeugung; SUP zur Bedarfsanalyse der Stromübertragungsnetze) getestet. Hierbei mussten manche der auf Basis von Dokumentenanalysen und in einem Pre- und Posttestverfahren unter Einbeziehung des one-stop Delphis erarbeiteten und inhaltsanalytisch ausgewerteten Kriterien und Indikatoren modifiziert werden, bis eindeutige Formulierungen der Qualitätsstufen festlagen und die Ergebnisse in die jeweiligen Schritte des Planungsprozesses integriert werden konnten.

In der Diskussion wird insbesondere die Nützlichkeit des Rechnerprogramms MaxQDA 11 hervorgehoben, welches sich zur inhaltsanalytischen Gewinnung geeigneter Indikatoren aus den Dokumentenanalysen sowie aus den Interview-Audiodateien sehr bewähren konnte. Die Ergebnisse zeigen, dass sich das erarbeitete Analyseschema vor allem zur Evaluation der „procedural effectiveness“ als praktikabel erwies und für eine sektorenübergreifende Evaluation der Qualität von SUP-Maßnahmen herangezogen werden kann.

Der zweite Originalbeitrag von *Juliane Achatz, Brigitte Schels, Joachim Wolff* (Institut für Arbeitsmarkt- und Berufsforschung Nürnberg und Friedrich-Alexander-Universität Erlangen-Nürnberg) befasst sich mit dem Thema „Wie wirken Zusatzjobs und Trainingsmaßnahmen auf die Arbeitsmarktchancen von jungen Erwachsenen in unterschiedlichen familiären Situationen?“ Im Unterschied zu Studien über die allgemeine Wirksamkeit von arbeitsmarktpolitischen Maßnahmen verschiedener Jobcenter wird hier bei jungen Arbeitslosen (15- bis 30-jährige Frauen und Männer mit Bezug von Arbeitslosengeld II) der Fokus der Untersuchung auf eine Differenzierung nach den familiären Gegebenheiten gerichtet und insbesondere darauf, ob der/die Betroffene in einer „Bedarfsgemeinschaft“ mit einem/-r Partner/-in und ggf. auch mit mindestens einem Kind lebt.

In einer Ex-post-Evaluation, die auf einem sehr großen Datenbestand der Bundesagentur für Arbeit basiert, wird die Wirksamkeit verschiedener Maßnahmen bei Bedarfsgemeinschaften und bei Einzelpersonen untersucht. Um eine Kontrollgruppe mit ähnlichen Merkmalen, aber ohne Teilnahme an den Maßnahmen zu erzielen, werden Matching-Verfahren eingesetzt. Hierbei werden jedem Maßnahmeteilnehmenden ein oder auch mehrere Nichtteilnehmende zugeordnet, die den Teilnehmenden hinsichtlich der relevanten Determinanten so ähnlich wie möglich sind. Neben einem exakten Matching hinsichtlich ausgewählter Merkmale kommt auch das statistische ‚Propensity Score Matching‘-Verfahren zur Anwendung.

Hinsichtlich der Arbeitsmarktchancen zeigt sich im Vergleich der verschiedenen Gruppen, dass Mütter und insbesondere Mütter in Paar-Bedarfsgemeinschaften besonders von Fördermaßnahmen profitieren, vor allem durch Annahme von Zusatzjobs („Ein-Euro-Jobs“); bei schulischen und insbesondere bei betrieblichen Trainingsmaßnahmen ist der höchste Netto-Effekt bei jungen Vätern mit Partnerin zu verzeichnen. Die dahinterliegenden Wirkmechanismen werden ausführlich diskutiert. Die Studie zeigt, dass die Beachtung der familiären Situation eine genauere Prognostik ermöglicht und somit einen Beitrag zu einem möglichst effektiven Einsatz von passgenauen arbeitsmarktpolitischen Maßnahmen leisten kann.

Thema des dritten Originalbeitrages „Evaluierung des Aufnahmeverfahrens eines Fernstudiengangs anhand des Studienerfolges: Vorhersage des Studienerfolges durch studienbezogene Lerntests, Englischtests, objektive Leistungsmotivations-tests, Intelligenztests und Eignungsinterviews“ von *Ingrid Wahl* und *Christa Walenta* (Ferdinand Porsche Fern-FH Wiener Neustadt, Österreich) ist ein Auswahlverfahren für Studierende im Bachelor-(Fern-)Studiengang „Betriebswirtschaft und Wirtschaftspsychologie“ einer österreichischen Fachhochschule. Überprüft werden soll, welche der im Auswahlverfahren eingesetzten Verfahren den Studienerfolg am besten vorhersagen können.

Als Auswahlverfahren diene eine bis zu dreistündige computeradministrierte Testbatterie, bestehend aus einem Intelligenztest, einem Englischtest, einem objektiven Leistungsmotivationstest und einem studienbezogenen Lerntest; als Ergänzung wurde ein 20-minütiges Eignungsinterview anhand eines standardisierten Interviewleitfadens durchgeführt. Zur Vorbereitung auf den studienbezogenen Lerntest wurde eine 105-seitige schriftliche Lernunterlage (fachliche Informationen, Auszüge aus den Studienmaterialien der Anfangssemester) mindestens drei Wochen vor der Computertestung verschickt; der Lernerfolg wurde mit 31 Multiple-Choice-Fragen mit vier Antwortmöglichkeiten und einer bis drei Richtigitworten erfasst, ergänzt durch sieben Entscheidungsfragen bezüglich der Richtigkeit von gegebenen Aussagen. Untersucht wurden 267 Studierende (89 männlich, 169 weiblich), die für den Studiengang auf Basis des Auswahlverfahrens zugelassen worden waren. Als Maße des Studienerfolges dienten der Notendurchschnitt im ersten und zweiten Studiensemester sowie auch dichotome Angaben bezüglich Wiederholungsprüfungen (ja/nein), Semester erfolgreich abgeschlossen (ja/nein) oder Studienabbruch (ja/nein).

Die mit multivariaten Varianzanalysen abgesicherten Ergebnisse zeigen, dass das Auswahlverfahren als Ganzes signifikante Zusammenhänge mit dem Studienerfolg aufweist. Zur Identifikation jenes Teiles des Auswahlverfahrens, welches für sich genommen die höchste prognostische Validität aufweist, wurden hierarchische Regressionsanalysen durchgeführt, die eindeutig den studienspezifischen Lerntest an die Spitze stellten ($R^2=.16$). Von den anderen Teiltests liefert lediglich der Intelligenztest (und im Detail das numerisch-induktive Denken) einen Zuwachs auf $R^2=.18$, wobei alle Teiltests untereinander korrelierten.

Die Ergebnisse werden dahingehend interpretiert, dass durch die studiennahe Vorbereitung der schriftlichen Lernunterlagen und die nachfolgende Prüfung im studienbezogenen Lerntest für die Studienbewerber die Anforderung des selbstregulierten Lernens erfahrbar wird. Da das selbstregulierte Lernen im Fernstudium eine

zentrale Rolle spielt, weisen Personen, die in der Studieneingangsprüfung diesbezüglich gut abschneiden, wohl auch später entsprechend bessere Studienleistungen auf. Eine Gesamtbetrachtung der Ergebnisse gibt Anlass zur Überarbeitung der verschiedenen Teile des Auswahlverfahrens.

Die vierte Originalarbeit mit dem Titel „Entwicklung eines Wirkmodells für die Evaluation des Programms Jugend und Sport“ von *Alain Dössegger, David Weibel, Klaus M. Frei, Bartholomäus Wissmath* und *Jan Hense* (Eidgenössische Hochschule für Sport Magglingen, Schweiz, Universitäten Bern und Gießen sowie w hoch 2 GmbH Bern) befasst sich zunächst mit einem Überblick zur Rolle von Wirkmodellen in der gegenwärtigen Evaluationsliteratur. Anstelle des aus dem Englischen kommenden Begriffs „Programmtheorie“ präferieren die Autoren den weniger anspruchsvollen Begriff „Wirkmodell“, unter dem sie eine Zusammenstellung der möglichst kausalen Annahmen über die spezifische Funktions- und Wirkungsweise eines Programms verstehen. Während weniger differenzierte Modelle die genaue Wirkungsweise von Programmen häufig nur als Input-/Output-Modelle mit dazwischenliegender „Black Box“ thematisieren, sollen Wirkmodelle durch die Einbeziehung von Mediator(inn)en und Moderator(inn)en eine differenzierte Aufdeckung von nicht direkt beobachtbaren Wirkmechanismen ermöglichen und als Ergebnis aufzeigen, wie und auf welchem Weg die Programmaktivitäten ihre beabsichtigten Wirkungen entfalten.

Nach einer Erörterung empfohlener Vorgehensweisen zur Füllung der Black Box und der Grundgedanken theoriebasierter Evaluationsansätze (Literaturrecherche/-review, Programmarchetypen, Analytische Verfahren, Stakeholder-Interviews und -Workshops, Analyse von Programmdokumenten, Fill-Out-Approach) werden drei kritische und bisher nicht angemessen beantwortete Fragen aufgeworfen: (1) Machen Wirkmodelle eine Evaluation blind für nicht intendierte Wirkungen? (2) Sollen Wirkmodelle eher die Sichtweise von Stakeholdern abbilden oder sich an ‚echten‘ wissenschaftlichen Theorien und Befunden orientieren? (3) Ist der zusätzliche Aufwand für die Entwicklung eines Wirkmodells gerechtfertigt? Im Zuge der konkreten Entwicklung eines Wirkmodells für das Programm Jugend und Sport mit Fokus auf Kindersport werden diese drei Fragen aus evaluationstheoretischer Sicht diskutiert. Demnach sind (1) Wirkmodelle eher geeignet, auch nicht intendierte Nebenwirkungen aufzuzeigen, die im konkreten Programm aber nicht auftraten. (2) Die in der Literatur häufig genannten Widersprüche zwischen der fachwissenschaftlichen Perspektive und den Stakeholdern waren im konkreten Fall nicht vorzufinden; vielmehr war die implizite Wirktheorie der Programmbeteiligten weitgehend kompatibel mit der Forschungslage. (3) Der Entwicklungsaufwand lässt sich vor allem dadurch rechtfertigen, dass ein Wirkmodell die Aufdeckung unerwünschter Nebenwirkungen erleichtert sowie etwaige Divergenzen der Sichtweisen von Stakeholdern erkennbar macht. Das Vorliegen eines Wirkmodells fördert somit auch das Steuerungswissen zur Optimierung des Programms.

Die fünfte Originalarbeit mit dem Titel „Wirkungsevaluationen in der Sozialen Arbeit mittels Realistic Evaluation: empirische Anwendungen und methodische Herausforderungen. Eine systematische Literaturreview“ von *Sigrid Haunberger* und *Edgar Baumgartner* (Fachhochschule für soziale Arbeit Olten, Nordwestschweiz)

beginnt mit einer Differenzierung des Evaluationskontinuums von der rein effektorientierten Black-Box-Evaluation über die Grey-Box bis hin zum Wirkungszusammenhänge detailliert darstellenden White-Box-Ansatz und betont dabei die Bedeutung der „Realistic Evaluation“ (RE), welche bei einer differenzierten Betrachtung theoriebasierter Ansätze das Potenzial zu einer „White-Box-Evaluation“ aufweist.

Das Ziel einer RE ist in der Erklärung von sozialen Regelmäßigkeiten und Mustern für Theorien mittlerer Reichweite zu sehen, die zwischen Universal- und Mikrotheorien anzusiedeln sind und keinen Anspruch auf Generalisierungen aufweisen, sich aber auch nicht auf Einzelfallaussagen beschränken. Die RE geht über theoriebasierte Evaluationsansätze insoweit hinaus, als sie zentrale theoretische Komponenten benennt und die Kontextabhängigkeit von Effekten besonders hervorhebt, indem sie Kontexte, Mechanismen und Outcomes zu KMO-Konfigurationen verbindet, die darüber informieren, wie die Outcomes einer Intervention von den zugrunde liegenden Mechanismen im jeweils spezifischen Kontext abhängen.

Eine sorgfältig dokumentierte Literaturreview soll Auskunft darüber geben, wie die theoretischen Grundlagen der RE in empirischen Anwendungen umgesetzt werden und wie dabei methodisch vorgegangen wird. Auch sollen mittels RE Hinweise auf Stärken und Schwächen der Evaluationspraxis sowie Empfehlungen für die Evaluationspraxis in der Sozialen Arbeit gewonnen werden.

Die Literaturreview wurde einerseits in quantitativer und andererseits in qualitativer Richtung ausgewertet. In quantifizierter Form wurde mittels Interraterreliabilität eine insgesamt als zufriedenstellend bewertete Übereinstimmung der beiden Reviewer (Autoren) bei der Erstellung eines nominalskalierten Kategorienrasters festgestellt. Aufbauend auf der anschließenden narrativ-qualitativen Betrachtung finden sich als Ergebnis der systematischen Review Antworten auf die Fragen

- warum die RE angewendet wird,
- in welchen Anwendungsbereichen die RE Verwendung findet,
- welche Forschungsmethoden zur Umsetzung der RE angewendet werden
- welche ungelösten methodischen Anwendungsprobleme für die RE existieren sowie
- welches die zentralen methodischen Implikationen sind und wo es weiteren methodischen Klärungsbedarf gibt.

Alle Fragen/Antworten werden ausführlich diskutiert, wobei darauf hingewiesen wird, dass RE für sich alleine genommen „nicht als Modell für Evaluationsdesigns mit konkreten Vorgaben“ angesehen werden kann; vielmehr handle es sich „um ein heuristisches Rahmenwerk, mit dem Mechanismen und Effekte im Sinne einer White-Box-Evaluation herausgearbeitet und kontextabhängig dargestellt werden können“ wobei „das Auffinden von KMO-Konfigurationen mehr eine interpretative als eine logisch-deduktive Aufgabe zu sein“ scheint. Empfehlungen für die Evaluationspraxis mittels RE in der Sozialen Arbeit werden abgeleitet.

Über die fünf Originalbeiträge hinaus bieten die Rubriken *Information & Service* sowie *DeGEval ...Info* auch in diesem Heft eine Vielzahl aktueller Beiträge.

Drei Praxisberichte beschäftigen sich mit unterschiedlichen Aspekten der Evaluation. *Benoit Simon* und *Wolfgang Meyer* untersuchen, welche Rolle das Thema Evaluation in den ersten freiwilligen nationalen Berichten zur Implementie-

rung der Sustainable Development Goals (SDG) spielt. *Kirsten Vorwerk, Christoph Hartmann, Felix Gaisbauer* und *Stefanie Krapp* berichten über die vom Deutschen Evaluierungsinstitut der Entwicklungszusammenarbeit (DEval) durchgeführte Evaluierung eines umfangreichen Programms zur Förderung von Entwicklungspartnerschaften mit der Wirtschaft. *Heidemarie Kelleter* stellt die Ergebnisse eines Projekts zur evidenzbasierten Steuerung von Einrichtungen der Langzeitpflege vor. Hinzu kommen drei Rezensionen aktueller Buchveröffentlichungen zur Evaluation, die von *Antonia Baumeister, Franziska Heinze, Frank König, Stefanie Reiter* sowie *Hellmut Wollmann* verfasst wurden.

Auch die *DeGEval ...Info* enthält wieder eine große Vielfalt interessanter Informationen. Eröffnet wird diese Rubrik von einem Bericht zur 19. Jahrestagung der DeGEval – Gesellschaft für Evaluation e.V. 2016 in Salzburg. Weiter geht es mit der Laudatio zu dem auf dieser Tagung verliehenen Nachwuchspreis sowie einer Kurzfassung des mit dem Nachwuchspreis ausgezeichneten Journal-Artikels; es folgen das Positionspapier des Vorstandes der DeGEval zum Tagungsthema „Zwischen Nutzung, Einfluss und Nachhaltigkeit – Wie wirken Evaluationen in unterschiedlichen Systemen?“ sowie ein Bericht über eine Session des AK Methoden auf der Jahrestagung 2016, in deren Rahmen eine politikfeldübergreifende Diskussion zu methodischen Standards von Evaluationen stattfand. Abgerundet wird die DeGEval ...Info durch einen Beitrag aus dem AK Gender Mainstreaming zum Thema „Genderkompetenz und Genderexpertise – Anforderungen an Kompetenzprofile von Evaluator(inn)en“ sowie durch die Vorstellung des niederländischen „Evaluatorennetzwerks“.

Darüber hinaus enthält dieses Heft die Liste der Gutachterinnen und Gutachter der Originalbeiträge aus den Jahren 2015 und 2016, denen für ihren Einsatz für die Qualität und den wissenschaftlichen Standard der Zeitschrift für Evaluation im Namen aller Herausgeberinnen und Herausgeber herzlich gedankt sei.

Explizit hingewiesen sei auch auf die Ankündigung der 20. Jahrestagung der DeGEval zum Thema „Evaluation (in) der Zukunft“, die vom 20. bis 22. September 2017 in Mainz stattfindet, sowie die Ausschreibung für den Nachwuchspreis der DeGEval, der in diesem Rahmen verliehen wird.

Schließlich ist auch noch von einem erfreulichen Neuzugang im Herausgebergremium zu berichten: Prof. Dr. Philipp Mayring, Universität Klagenfurt, wird während seiner Amtszeit als Vorsitzender der DeGEval im Herausgebergremium als Mitglied kooptiert. Wir freuen uns auf die Zusammenarbeit.

Viel Vergnügen beim Lesen und viele interessante Anregungen wünscht Ihnen

Helfried Moosbrugger

Für die Herausgeberinnen und Herausgeber