

Evaluationsmethoden im Spannungsfeld von Ansprüchen, Ressourcen und Kompetenzen – Einführung

Christian Rüefli | Die SEVAL-Jahrestagung 2009 widmete sich dem Thema «Evaluationsmethoden». Drei Impulsreferate und vier Workshops thematisierten diverse Spannungsfelder, denen Evaluationen ausgesetzt sind, deren Bedeutung für die Evaluationsmethodik und Möglichkeiten, um mit diesen Spannungsfeldern umzugehen. Politische und konzeptionelle Fragen oder exogene Rahmenbedingungen wie Ressourcenknappheit, Zeitdruck und Komplexität prägen Evaluationsprozesse meistens stärker als erkenntnistheoretische Methodenfragen und führen oft zu einem pragmatischen Vorgehen. Deshalb sind fundierte Methodenkenntnisse und ein gemeinsamer Austausch zwischen Evaluierenden und Auftraggebenden wichtig, damit Evaluationen einen hohen Nutzen bringen.

Inhaltsübersicht

- 1 Einleitung
- 2 Evaluationsmethoden zwischen Theorie und Praxis – ein Brückenschlag
 - 2.1 Die wissenschaftstheoretische Perspektive
 - 2.2 Die pragmatische Nutzerperspektive
 - 2.3 Wirkungsevaluation zwischen Theorie und Praxis
- 3 Tagungsfazit
 - 3.1 Methoden im Spannungsfeld unterschiedlicher Ansprüche
 - 3.2 Pragmatismus im Umgang mit den Spannungsfeldern
- 4 Ausblick

1 Einleitung

Für die Evaluation gelten vergleichbare Ansprüche an die Wissenschaftlichkeit wie für die empirische Sozialforschung. Deshalb ist der adäquate Umgang mit Methoden ein zentrales Qualitätsmerkmal von Evaluationen. Methoden sind jedoch nicht aus einem Selbstzweck heraus einzusetzen, sondern stets an den konkreten Fall und Kontext anzupassen. Dies bringt zahlreiche Herausforderungen mit sich:

- Es gibt eine Vielzahl von Erhebungs- und Analysemethoden mit jeweiligen Vor- und Nachteilen, und es findet eine stetige Weiterentwicklung der Methoden statt. Dies macht es nicht einfach, auf dem Gebiet à jour zu bleiben.
- Im Kontext von Auftragsstudien besteht oft ein Spannungsfeld zwischen den Anforderungen an eine angemessene Methodik und begrenzten zeitlichen sowie finanziellen Ressourcen, welche den methodischen Möglichkeiten Grenzen setzen.

- Bei konkreten Evaluationsvorhaben kann es vorkommen, dass die Daten- und Informationslage die Anwendung der am besten geeigneten Methode nicht zulässt, weil die dazu notwendigen Informationen nicht erhoben wurden, nicht zugänglich sind oder nur unter unverhältnismässig grossem Aufwand erhoben werden können.
- Evaluationsprozesse können unter Umständen belastet werden, wenn wegen unterschiedlich ausgeprägten Methodenkompetenzen und –präferenzen zwischen Auftraggebenden und Evaluierenden eine ungenügende Vertrauens- und Gesprächsbasis besteht.

Sowohl Evaluatorinnen und Evaluatoren als auch ihre Auftraggeberinnen und Auftraggeber müssen sich somit immer wieder mit der Frage auseinandersetzen, was eine «gute» Evaluationsmethodik ist und benötigen dafür entsprechend breite Methodenkenntnisse. Konsequenterweise haben die Mitglieder der SEVAL in einer im Sommer 2008 durchgeführten Mitgliederbefragung unter anderem ein Bedürfnis nach einer verstärkten Auseinandersetzung mit methodischen Fragen der Evaluation und entsprechenden Ausbildungsbedarf geäussert.²

Mit dem Ziel, die Verständigung über die Bedeutung von Methoden in der Evaluation zwischen Evaluierenden, Auftraggebern sowie anderen interessierten und betroffenen Kreisen zu verbessern und die Diskussion über die Grenzen und Möglichkeiten von Methoden in der Evaluation anzuregen, widmete die SEVAL ihre Jahrestagung 2009 dem Thema «Evaluationsmethoden».³ Die von rund 130 Personen besuchte Tagung fand am 11. September 2009 in den Räumlichkeiten des Eidgenössischen Personalamts in Bern statt. Drei Inputreferate und vier Workshops zu gezielt ausgewählten Fragestellungen boten den Tagungsteilnehmenden die Möglichkeit zur Auseinandersetzung mit allgemeinen Fragen und spezifischen Problemen in Zusammenhang mit der Anwendung von Methoden in Evaluationen. Um dem ausgeprägten Interesse der Schweizerischen Evaluationsgemeinschaft an diesem Tagungsthema Rechnung zu tragen, werden die Inhalte der Referate und Workshops in der vorliegenden LeGes-Ausgabe einem breiteren Publikum zugänglich gemacht. Damit soll nicht zuletzt auch eine Diskussion über die Bedeutung von Methoden in der Evaluationspraxis lanciert werden.⁴

2 Evaluationsmethoden zwischen Theorie und Praxis – ein Brückenschlag

Der erste Teil der Tagung war von drei Impulsreferaten geprägt, die sich aus verschiedenen, eher übergeordneten Blickwinkeln mit Evaluationsmethoden auseinandersetzten. Ziel der ersten beiden Referate war es, einer theoretisch ausgerichteten Auslegeordnung der Bedeutung von Methoden in Evaluationen die eher pragmatischen Interessen einer Nutzerin von Evaluationen gegenüberzustellen,

die in erster Linie darin bestehen, verlässliche und aussagekräftige Evaluationsergebnisse zu erhalten. Das dritte Referat stellte die Verbindung zwischen Theorie und Praxis her und zeigte auf, wie Evaluationen konzipiert werden können, die unter Beachtung wissenschaftlicher und methodologischer Grundsätze einen hohen Nutzen für die interessierten Kreise abwerfen. Die Ausführungen werden im Folgenden zusammengefasst.

2.1 Die wissenschaftstheoretische Perspektive

Prof. Max Bergman vom Institut für Soziologie der Universität Basel befasste sich aus einer theoretischen Sicht mit den Ansprüchen an und den Grenzen von Methoden im Kontext von Evaluationen. Entlang der verschiedenen Elemente und Schritte von Evaluationsprozessen (Kontext, Evaluandum und Methode, Daten, Analyse und Interpretation) wies er auf die zahlreichen methodologischen Anforderungen hin, die es bei der Konzeption und Umsetzung einer Evaluation zu beachten gilt. Er nahm eine entsprechende Auslegeordnung von Fragen, Überlegungen und Abwägungen vor, mit denen sich Evaluierende im Zusammenhang mit der Anwendung von Methoden konfrontiert sehen (vgl. seinen nachfolgenden Beitrag in dieser Ausgabe). Seine Ausführungen mündeten im Fazit, dass Evaluation in verschiedener Hinsicht stets mit der Notwendigkeit zur Komplexitätsreduktion verbunden ist (vgl. hierzu Ackermann/Pfyl in dieser Ausgabe) und insbesondere im Auftragskontext auch Verhandlungen nötig macht. Methodenbezogene Entscheidungen sind damit nicht nur abhängig von methodologischen Gesichtspunkten, sondern auch von der Konstellation der Beziehungen zwischen den in einer Evaluation involvierten Parteien, von deren Ressourcen und Erwartungen. Evaluierende stehen deshalb auch vor kommunikativen Herausforderungen. Abgesehen von methodologischen Gütekriterien bemisst sich die Qualität einer Evaluation auch daran, ob es den Evaluierenden gelingt, die Beziehungen zwischen den Evaluationszielen, den theoretischen Grundlagen, den erhobenen Daten, der Analyse, den Interpretationen und daraus abgeleiteten Empfehlungen kohärent darzulegen. In diesem Sinne ist Evaluation auch Argumentation und erfordert die Verbindung von wissenschaftlichen und policy-orientierten Kommunikationsformen.

2.2 Die pragmatische Nutzerperspektive

Diesen grundsätzlichen methodologisch-theoretischen Ausführungen stellte Martine Brunschwig Graf, amtierende Nationalrätin und ehemalige Regierungsrätin des Kantons Genf, eine bewusst damit kontrastierende Sichtweise entgegen. Sie bettete die Evaluationstätigkeit in einen grösseren staatstheoretischen und politischen Kontext ein und erläuterte ihre Erwartungshaltung als politisch verantwortliche Auftraggeberin und Nutzerin von Evaluationen. Für Brunschwig

Graf soll und kann Evaluation zur Verbesserung der Regierungsführung und der Gesetzgebung beitragen, indem sie hilft, die Verteilung der strategischen, operativen und kontrollierenden Rolle zwischen Parlament und Regierung zu klären und die Qualität, Umsetzung und Auswirkungen von Gesetzen zu beurteilen. Evaluation soll der Politik Entscheidungsgrundlagen zur Verfügung stellen, steht dabei aber vor dem Problem, dass ihre Antworten – der gesellschaftlichen Realität entsprechend – häufig komplex und widersprüchlich ausfallen, während die Politik einer eigenen Logik folgt und klare und eindeutige Aussagen bevorzugt.

Im politischen Kontext ist es eine Funktion der Evaluation, zur Effizienz und Transparenz staatlichen Handelns beizutragen. Damit trägt sie eine politische und gesellschaftliche Verantwortung. Sie ist nicht nur eine Entscheidungsgrundlage, sondern kann auch zur Rechenschaftslegung gegenüber der Öffentlichkeit eingesetzt werden. Die Publikation von Evaluationsergebnissen stellt in diesem Zusammenhang ein Kontroll- bzw. Sanktionsinstrument dar, weil sich die interessierte Öffentlichkeit so eine Meinung über die evaluierten staatlichen Massnahmen – und damit auch über die Projekt- oder Programmverantwortlichen oder ihre politischen Vorgesetzten – bilden kann, was sich unter Umständen auf das Abstimmungs- und Wahlverhalten der Bevölkerung auswirkt.

Aus diesen verschiedenen Funktionen leitet Brunschwig Graf konkrete Anforderungen an Evaluation ab. So erachtet sie es als unabdingbar, dass Evaluatoreninnen und Evaluatoren den vollständigen Kontext berücksichtigen, das Umfeld des Auftrags, des Untersuchungsgegenstandes und des Feldes verstehen und dem Umstand Rechnung tragen, dass die Realität nicht immer der theoretischen Modelle und Annahmen folgt, z. B. bei der Evaluation eines Pilotprojekts. Wenn ein solches Projekt nicht wie geplant verläuft und Anpassungen erfährt oder seine Ziele nicht erreicht, sollte das nicht per se negativ beurteilt, sondern vor dem Hintergrund des Kontexts (Unsicherheit bei Projektbeginn, Lerneffekte usw.) konstruktiv gewürdigt werden. Brunschwig Graf mahnte davor, sich bei Evaluationen dogmatisch an der «reinen Methodenlehre» zu orientieren und plädierte für einen kontextsensiblen Pragmatismus. Als Nutzerin von Evaluationen ist für sie nicht von vordringlichem Interesse, welche Methode konkret zur Anwendung kommt. Zentral ist für sie hingegen, dass Evaluatoreninnen und Evaluatoren in Bezug auf die Methodik transparent sind, damit ihre Befunde nachvollzogen werden können. Transparenz dient so der Vertrauensbildung. Zugleich nahm Brunschwig Graf auch die Auftraggebenden und die Politik in die Pflicht und plädierte dafür, Evaluationsergebnisse zu nutzen und daraus zu lernen sowie diese Nutzung zu verfolgen und Rechenschaft darüber abzulegen. Auch dies diene der Vertrauensbildung. Um dieser Idealvorstellung einer lernorientierten Evaluationskultur näher zu kommen sei allerdings nach wie vor ein Lernprozess nötig.

2.3 Wirkungsevaluation zwischen Theorie und Praxis

Frans L. Leeuw, Soziologieprofessor am Netherlands Justice Research Institute Den Haag und der Universität Maastricht, hatte die Aufgabe, den Brückenschlag zwischen der Wissenschaft mit ihren methodologischen Grundsätzen und Gütekriterien und der Politik, die nicht an methodischen Fragen, sondern an möglichst zuverlässigen und nachvollziehbaren Aussagen über die Wirkung ihres Handelns interessiert ist, zu bewerkstelligen. In seinem Referat «Impact evaluations and social reality» zeigte er auf, welche methodologischen Schlüsselaspekte bei der Durchführung von Wirkungsevaluationen zu beachten sind, damit diese ihr Potenzial entfalten und sowohl wissenschaftlich fundiertes als auch handlungsrelevantes Wissen über die Wirksamkeit von Projekten, Programmen oder öffentlichen Politiken generieren können.

Leeuw begann seine Ausführungen mit der Feststellung, dass es sich bei der allerersten jemals durchgeführten und schriftlich dokumentierten Evaluation um eine Wirkungsevaluation handelte, nämlich um James Linds Abhandlung über die Ursachen und Behandlungsmöglichkeiten der Seefahrerkrankheit Skorbut aus dem Jahr 1753. Nachdem insbesondere im angelsächsischen Raum Wirkungsstudien mit experimentellen Längsschnittdesigns bis in die 1960er Jahre hinein als Goldstandard der Evaluationsforschung gegolten hatten, rückte die Wirkungsevaluation zwischen Mitte der 1970er und Ende der 1990er Jahre zunehmend in den Hintergrund. Wie die Entwicklung in Staaten wie Grossbritannien, den Vereinigten Staaten, Schweden, Dänemark, den Niederlanden oder Deutschland zeigt, gewannen andere Evaluationsformen wie Prozessevaluationen, *performance audits*, *management assessments*, Machbarkeitsstudien, Inspektionen und Akkreditierungen sowohl im öffentlichen als auch im privaten Sektor stark an Popularität und wurden zu dominierenden Managementinstrumenten, welche die öffentliche Verwaltung und die Politikumsetzung unterstützen sollten. Mit der Zeit wurden jedoch diejenigen Stimmen lauter, die an diesen Evaluationsformen kritisieren, dass sie sich weitgehend darauf beschränkten, Prozesse und Outputs zu beschreiben, ihre Übereinstimmung mit Vorgaben und ihre Effizienz zu überprüfen, jedoch keine Aussagen über die Wirksamkeit der untersuchten Projekte, Programme oder Politiken zuließen. Sowohl die Politik als auch zivilgesellschaftliche Kreise verlangten zunehmend nach Evidenz über die Ergebnisse, Nettowirkungen und allfälligen Nebenwirkungen staatlicher Interventionen. Es besteht somit ein *evaluation gap*, eine «Evaluationslücke»: Trotz zahlreicher Evaluationen ist das Wissen über die eigentlichen Ergebnisse von Politiken und Programmen äusserst gering.⁵

Angesichts dieses *evaluation gaps* rückt der Ansatz der Wirkungsevaluation (*impact evaluation*) erneut ins Blickfeld des Interesses. Obwohl dieser Ansatz in

der Evaluationsgemeinde in der Vergangenheit oft kritisch betrachtet wurde, gab Leeuw ein engagiertes Plädoyer zu dessen Gunsten ab. Er rief dazu auf, ideologische und paradigmatische Gräben zu überwinden und sich auf die «Mutter aller Evaluationen» zurückzubedenken, um die Evaluationslücke zu schliessen und sich mit den Auswirkungen von öffentlichen Politiken, Programmen, Projekten, Massnahmen, Organisationen, Verfahren etc. auf das menschliche und gesellschaftliche Verhalten zu befassen und so die wissenschaftliche Evidenz zu generieren, nach der Entscheidungsträger zunehmend fragen.

Einen jüngst publizierten Weltbank-Leitfaden (Leeuw/Vaessen 2009) zusammenfassend, vertiefte Leeuw in der Folge sechs methodologische Schlüsselaspekte, die es bei der Durchführung von Wirkungsevaluationen zu beachten gilt:

1. *Identifikation der Art und Reichweite der Intervention*

Die Reichweite einer Wirkungsevaluation ergibt sich anhand von zwei Fragen: Die Wirkung von was (Intervention) und die Wirkung auf was? Bei Interventionen kann es sich um klar isolierbare, einzelne Massnahmen mit einem klar messbaren Ziel handeln, aber auch um komplexe Programme mit zahlreichen Ebenen und Handlungsfeldern. Während erstere im Idealfall im Rahmen von Experimental- und Kontrollgruppendesigns einer kausalen Wirkungsanalyse unterzogen werden können, muss bei komplexen Programmen ein systematisches Wirkungsmodell erstellt werden (program mapping; logic model), welches die Priorisierung und Analyse einzelner Programmkomponenten ermöglicht. Bei der Analyse von Wirkungen sind die verschiedenen Ebenen, auf denen sich diese entfalten können, zu berücksichtigen. Hier können beispielsweise die direkten Adressaten einer Intervention und die weiteren Begünstigten unterschieden werden.

2. *Einigung über den zu beurteilenden Evaluationsgegenstand*

Wirkungsevaluationen haben auch danach zu fragen, für wen die Wirkungen anfallen. Damit soll festgelegt werden, bei wem welche Indikatoren zu messen sind, um Aussagen über die Wirkungen einer Intervention machen zu können. Dies beinhaltet eine Analyse und Abwägung der verschiedenen involvierten Stakeholder-Interessen. Weiter ist in Betracht zu ziehen, dass Interventionen auch un-intendierte Effekte zeigen können, dass sich ihre Wirkungen allenfalls erst langfristig einstellen und dass deshalb auch die Nachhaltigkeit einer Intervention in der Regel schwierig zu beurteilen ist.

3. *Sorgfältige Rekonstruktion der Theorien zur Verbindung von Interventionen und Ergebnissen*

Interventionen liegen zumeist explizite oder implizite Annahmen (so genannte Programmtheorien) über ihre Wirkungsweise zugrunde. Diese gilt es zu rekon-

struieren und zu artikulieren. In einem zweiten Schritt sind diese Annahmen zu überprüfen – entweder anhand der Konstruktion von so genannten causal stories oder über den formalen Test der Kausalannahmen mittels angemessener Methoden.

4. *Diskussion des Zuordnungsproblems*

Die beobachteten Ergebnisse unterliegen zumeist vielfachen Einflussfaktoren und sind deshalb nur teilweise kausal auf die untersuchte Intervention zurückzuführen. Der eigentliche Mehrwert von Wirkungsevaluationen besteht deshalb darin, dass sie den Beitrag einer konkreten Intervention an die vorgefundenen Ergebnisse identifizieren, das heisst das so genannte Zuordnungsproblem (*attribution problem*) erörtern. Dies erfolgt in der Regel durch einen Vergleich einer gegebenen Situation *mit* erfolgter Intervention mit der Situation, wie sie sich *ohne* Intervention ergeben hätte. Für solche Wirkungsevaluationen gibt es keine allgemeingültige «beste» Methode. Diese gilt es je nach Reichweite, Ziel und Ausgestaltung der Massnahme, aber auch je nach Datenverfügbarkeit fallweise immer wieder neu zu bestimmen. Quantitative Techniken wie randomisierte Kontrollgruppendesigns, quasi-experimentelle Verfahren oder Regressionsanalysen sind bei geschicktem Einsatz in der Regel gut geeignet, um das Zuordnungsproblem anzugehen. Sie sind aber methodisch anspruchsvoll, und die Voraussetzungen für ihre Anwendung sind oft nicht gegeben. Vor allem in komplexen Kontexten sind quantitative Techniken oft nicht angemessen. Qualitative Verfahren sind meist weniger effektiv in Bezug auf das Zuordnungsproblem, sind aber besser geeignet, um in Wirkungsevaluationen Informationen über die Reichweite, die Ziele und die Programmtheorie zu gewinnen und zusätzliche Daten und Evidenz zu generieren. Hier sind insbesondere partizipative Ansätze zentral. Diese bauen auf dem Grundsatz auf, dass Stakeholders in einzelne oder alle Phasen einer Evaluation einbezogen werden sollten, wobei es verschiedene Formen mit einem unterschiedlichen Ausmass des Einbezugs gibt. Mögliche Probleme qualitativer Ansätze sind der hohe Aufwand und die unter Umständen eingeschränkte Validität der gewonnenen Informationen. Die Vor- und Nachteile der verschiedenen Methoden sind wie erwähnt fallweise abzuwägen.

5. *Verwendung eines Mixed-Methods-Ansatzes*

Weil jede Methode ihre spezifischen Vor- und Nachteile aufweist und es bei Wirkungsanalysen keinen allgemeingültigen «Goldstandard» gibt, mit dem die Validität der kausalen Annahmen, die Validität der Operationalisierungen, die Verallgemeinerbarkeit und die statistische Validität gleichzeitig maximiert werden können, empfiehlt es sich, verschiedene Methoden zu kombinieren, das heisst In-

formationen, die mit verschiedenen Verfahren gewonnen wurden, zu triangulieren. So lässt sich zudem die Datenbasis, auf denen eine Wirkungsevaluation beruht, verbreitern.

6. *Aufbau auf bestehendem Wissen über die Wirkung von Interventionen*

Die Aussagekraft von Wirkungsevaluationen kann gestärkt und vertieft werden, indem neben den eigenen erhobenen Daten bereits vorhandenes Wissen einbezogen wird. Dies stärkt zum einen die äussere Validität, indem die untersuchte Intervention mit ähnlichen Interventionen in anderen Kontexten verglichen wird, und trägt zum anderen zur Verfeinerung der Interventionshypothesen bei. Hier können Methoden wie die Evaluations-synthese (*systematic review*), die Meta-Analyse, *narrative review* oder *realist synthesis* zur Anwendung gelangen.

Vertiefungsworkshops: Das Nachmittagsprogramm der SEVAL-Tagung umfasste vier Workshops, in denen praxisrelevante Fragen in Zusammenhang mit Evaluationsmethoden erörtert wurden. Dabei ging es nicht darum, konkrete Evaluationsmethoden vorzustellen und deren Vor- und Nachteile zu erörtern, sondern es sollten verschiedene Spannungsfelder, denen Evaluationen ausgesetzt sind, und die sich daraus ergebenden Konsequenzen für die Evaluationsmethodik beleuchtet und diskutiert werden. Die Workshopthemen waren zudem so ausgewählt, dass sie für Evaluierende, für Auftraggebende und für Forschende gleichermaßen von Interesse sind. Die Themen, Grundfragen und Diskussionen der vier Workshops sind jeweils in einem Beitrag der vorliegenden LeGes-Ausgabe dokumentiert. Die nachfolgende kurze Zusammenfassung gibt einen ersten Überblick über die Workshops.

Workshop 1: *Qualität evaluieren und entwickeln an Hochschulen – externe Ansprüche und interne Realitäten*» stand unter der Leitung von Désirée Donzallaz, Leiterin der Dienststelle Evaluation und Qualitätsmanagement der Universität Freiburg. Ausgangspunkt stellen die methodischen Herausforderungen dar, mit denen Hochschulen konfrontiert sind, wenn sie Qualität für externe Rechenschaftslegung und interne Entwicklungsarbeit evaluieren bzw. evaluieren sollen. Anhand von zwei Praxisbeispielen, vorgestellt von Michel Jaccard, directeur des affaires institutionnelles, EPFL und Eva Tov, Beauftragte für Qualitätsmanagement der Hochschule für Soziale Arbeit, Fachhochschule Nordwestschweiz, befassten sich die Teilnehmenden mit der Wahl und Entwicklung der Evaluationsmethodik an Hochschulen, insbesondere auch mit der Frage nach den entscheidungsrelevanten Kriterien. Der Workshop zeigte unter anderem eine Diskrepanz zwischen den künftigen gesetzlichen Anforderungen an die Qualitäts-

sicherung im Hochschulbereich und den bestehenden konzeptionellen und methodischen Voraussetzungen auf. Diese Fragen haben im vergangenen Jahr auch die Gründung einer entsprechenden SEVAL-Arbeitsgruppe ausgelöst.

Workshop 2: *Die Rolle von Kennzahlen für die Steuerung von politischen Programmen: Zwischen Theorie und Praxis* wurde von Katia Horber-Papazian, Professorin am Idheap in Lausanne, geleitet und widmete sich den methodologischen Herausforderungen, die sich aus einer indikatorenbasierten Steuerung politischer Programme ergeben. Die beiden Inputreferate von Françoise Gianadda, ehemalige Vorsteherin der Dienststelle für Bevölkerung und Migration des Kantons Wallis, und Jean-Claude Rochat, ehemaliger Kantonsrat und Präsident der Finanzkommission des Kantons Waadt, zeigten auf, dass sich in Zusammenhang mit einer Steuerung anhand von Kennzahlen zwar durchaus methodische und technische Fragen stellen, dass diese jedoch in den Hintergrund rücken, solange die Voraussetzungen für die Anwendung von Indikatoren noch nicht gegeben sind. So bestehen in Politik und Verwaltung einerseits insbesondere Akzeptanzprobleme und – auch politisch motivierte – Vorbehalte gegenüber diesem Instrument, andererseits fehlen vielenorts noch die Kompetenzen, um mit Indikatoren adäquat umgehen zu können.

Workshop 3: *Evaluation im Spannungsfeld zwischen Pragmatismus und Qualität* widmete sich der Frage, wie trotz beschränkter Ressourcen möglichst umfassende und qualitativ hochstehende Evaluationen durchgeführt werden können, und fokussierte dabei auf mögliche methodische Ansätze. Unter der Moderation von Christian Rüefli vom Büro Vatter, Politikforschung & -beratung, illustrierten ein Auftraggeber (Kurt Bisang, Bundesamt für Energie) und ein Auftragnehmer (Stefan Rieder, Interface Institut für Politikstudien) jeweils aus ihrer Sicht, mit welchen Ansprüchen und Restriktionen eine Evaluation konfrontiert sein kann und wie sich diese auf die Evaluationsmethodik auswirken können. Die Diskussion brachte zahlreiche Vorschläge für konzeptionelle und methodologische Anpassungen an Evaluationen hervor, die einen Ausgleich zwischen Pragmatismus und Qualität ermöglichen. Beidseits wurde eine vertiefte Auftragsanalyse als wichtiger erster Schritt einer Evaluation identifiziert, um deren Machbarkeit abzuklären und methodische Stolpersteine zu identifizieren und so eine Grundlage für eine gemeinsame Verständigung über die Grenzen und Möglichkeiten der Evaluation zu schaffen. Eine interessante Diskrepanz besteht darin, dass auf der einen Seite der Einbezug und die Synthese von Expertenmeinungen als gangbarer Weg gesehen wird, um bei Zeit- und Ressourcenknappheit eine Evaluation vorzunehmen, während auf der anderen Seite den Evaluierenden gelegentlich feh-

lender Mut zur eigenen kritischen Beurteilung attestiert wird. Ausgehend von dieser Diskrepanz wäre der weiterführenden Frage nach der (methodischen und empirischen) Grundlage von Evaluationsurteilen nachzugehen.

Workshop 4: *Umgang der Evaluation mit Komplexität – ein Beispiel aus der Gesundheitsförderung*, wurde von Roland Pfyl von der WOV-Geschäftsstelle im Finanzdepartement des Kantons Schwyz geleitet. Ausgangspunkt bildete das Spannungsfeld zwischen der Notwendigkeit, gesellschaftliche Komplexität differenziert zu erfassen und zu analysieren, um dem Evaluationsgegenstand gerecht werden zu können, und der Notwendigkeit, bei der Evaluation die Komplexität inhaltlich und methodisch sinnvoll zu reduzieren, um den Ressourcenaufwand in Grenzen zu halten und möglichst klare und eindeutige Aussagen generieren zu können. Anhand eines konkreten Praxisbeispiels nahmen sich die beiden Referenten (Günter Ackermann, Leiter Wirkungsmanagement bei Gesundheitsförderung Schweiz, und Robert Schmuki, Geschäftsführer Midnight Projekte Schweiz) der Frage der Evaluation komplexer Projekte aus theoretischer und praktischer Sicht an. Dabei stellten sich wiederum die Offenlegung der Grenzen einer Evaluation und die Machbarkeitsdiskussion als zentrale Elemente des Evaluationsprozesses heraus. Eine enge Zusammenarbeit zwischen Evaluierenden und Evaluierten, z. B. durch partizipative Evaluationsansätze, Methodenvielfalt und eine stärkere Berücksichtigung qualitativer Methoden sowie ein gezielter Einsatz von Methoden zur Strukturierung und Reduktion von Komplexität (z. B. von Wirkungsmodellen) stellen aus Sicht der Referenten gewinnbringende Ansätze für den Umgang mit Komplexität in Evaluationen dar.

3 Tagungsfazit

Wie bereits erwähnt standen an der SEVAL-Tagung 2009 nicht konkrete Evaluationsmethoden im Mittelpunkt des Interesses, sondern die Auseinandersetzung mit der Bedeutung von Methoden für Evaluationsprozesse und für die verschiedenen daran beteiligten Akteure, namentlich Auftraggebende und Evaluierende, aber auch für Betroffene bzw. Evaluierte und für weitere Nutzerkreise.

3.1 Methoden im Spannungsfeld unterschiedlicher Ansprüche

Die Präsentationen und Diskussionen dienten der Illustration verschiedener Spannungsfelder, die sich im Zusammenhang mit der Wahl und dem Einsatz von Evaluationsmethoden jeweils bemerkbar machen. Ein zentrales solches Spannungsfeld besteht zwischen der Politik als häufige Auftraggeberin und Nutzerin von Evaluationen und der Forschung, deren Logik und Grundsätze sich Evaluie-

rende verpflichtet fühlen (sollten) (vgl. Widmer et al. 1996, 17 ff.). Dieses Spannungsfeld kann anhand der unterschiedlichen Qualitätsbegriffe illustriert werden, die sich aus den Referaten von Max Bergman und Martine Brunschwig Graf ableiten lassen: Im einen Fall bemisst sich die Qualität einer Evaluation an der methodologisch korrekten und bewussten Wahl und Anwendung der Methoden, im anderen Fall an der kontextsensiblen, verantwortungsbewussten Konzeption der Evaluation und an der Transparenz über die eingesetzten Methoden. Diese Auffassungen schliessen sich jedoch nicht gegenseitig aus, sondern unterscheiden sich dadurch, dass sie den Fokus auf andere Aspekte einer Evaluation legen. Das Desinteresse der Politik an der Methodik bedeutet nicht, dass methodische Fragen vernachlässigt werden können, sondern ist so zu deuten, dass die Politik bei den Evaluierenden Methodenkompetenz und Qualitätsbewusstsein voraussetzt und darauf vertraut.

Das Beispiel der unterschiedlichen Qualitätsauffassungen verweist auf eine wesentliche Erkenntnis der SEVAL-Tagung und der vier Workshops: Methodologische Fragen stellen sich vermutlich in jeder Evaluation in der einen oder anderen Form, ihnen sind jedoch zumeist andere Probleme vorgelagert. Dabei kann es sich um Probleme politischer Art handeln – gesetzliche Auflagen sind wegen konzeptioneller und methodischer Unsicherheiten (fehlende Definitionen, Indikatoren und Messinstrumente) schwer umsetzbar; fehlendes Vertrauen und die Befürchtung eines Einfluss- und Kontrollverlusts stehen einer verstärkten indikatorengestützten politischen Steuerung entgegen –, oder um exogene Rahmenbedingungen wie Ressourcenknappheit und Komplexität des Kontexts, welche ihrerseits ein Spannungsfeld bilden und bei der Wahl der Evaluationsmethode ebenso wichtige Kriterien darstellen wie rein erkenntnistheoretische Gesichtspunkte. Es ist deshalb zu vermuten, dass die Evaluatorinnen und Evaluatoren gegenüber den Auftraggebern zumeist über die methodologische Definitionshoheit verfügen, dass aber in der Evaluationspraxis vorgelagerte Fragen, wie sie soeben beispielhaft skizziert wurden, die fachliche Diskussion über Methoden dominieren und damit die kritische Auseinandersetzung mit ihnen tendenziell zu kurz kommt. Mit der Definitionsmacht geht jedoch ebenso die Verantwortung der Evaluierenden einher, angemessen und seriös mit Methoden umzugehen und das Vertrauen, das ihnen seitens der Auftraggebern entgegengebracht wird, zu rechtfertigen. Die Auftraggeberseite verfügt ihrerseits über die Definitionsmacht bei der Verwendung von Evaluationsergebnissen (vgl. Balthasar 2007) und steht entsprechend in der Verantwortung, diese adäquat zu interpretieren, die methodisch gegebenen Grenzen der Machbarkeit und der Aussagekraft anzuerkennen und eine Instrumentalisierung von Resultaten zu vermeiden.

Die Problematik der Komplexitätsreduktion bildet ein anderes Kernthema der verschiedenen Präsentationen und Diskussionen. Die Konzeption einer Evaluation, der Einsatz von Methoden, aber auch die Vermittlung von Evaluationsergebnissen sind stets mit der Notwendigkeit zur Komplexitätsreduktion verbunden, die auf geschickte Art und Weise vollzogen und nachvollziehbar vermittelt werden muss. Dabei besteht die Herausforderung darin, dass methodisch fundierte Komplexitätsreduktion zumeist einen gewissen Aufwand mit sich bringt, was die Ressourcenproblematik aufwirft, und stets mit Einschränkungen der Aussagekraft verbunden ist, was mitunter auch Konsequenzen für die Qualität und den Nutzen einer Evaluation haben kann. Auch in dieser Hinsicht sind Evaluation und der Umgang mit ihren Ergebnissen mit der oben angesprochenen Verantwortung verbunden.

3.2 Pragmatismus im Umgang mit den Spannungsfeldern

Evaluation erfordert immer wieder das Fällen von Entscheidungen. Die Wahl des Forschungsdesigns und der Erhebungs- und Analysemethoden, aber auch die Wahl der bei einer Evaluation zu berücksichtigenden Datenquellen, der zu befragenden Personen oder der zu konsultierenden Dokumente und Literatur stellen zentrale solche Entscheidungen dar. Dabei wären die Vor- und Nachteile verschiedener Optionen gegeneinander abzuwägen und ihre möglichen Konsequenzen auf die Evaluation und ihre Ergebnisse zu berücksichtigen. In der realen Evaluationspraxis ist allerdings in der Regel ein gewisser Pragmatismus nötig, da derartige Entscheidungssituationen meistens von unvollständiger Information, Zeitdruck und Ressourcenknappheit geprägt sind.

Gegen einen gesunden Pragmatismus ist nichts einzuwenden. In den verschiedenen Tagungsreferaten wurde aber auch auf die damit verbundenen Gefahren hingewiesen. So ist Evaluation immer mit Verantwortung gegenüber den Auftraggebenden, aber auch gegenüber der Öffentlichkeit verbunden, die von allen Beteiligten wahrgenommen werden muss. Deshalb ist ein informierter, verantwortungsbewusster Pragmatismus wünschenswert, der die Einhaltung der Evaluationsstandards der SEVAL (Widmer et al. 2000) gewährleistet.

4 Ausblick

Evaluationen und insbesondere die Wahl und Anwendung von Evaluationsmethoden erfolgen stets in einem Spannungsfeld unterschiedlicher und oftmals divergierender Ansprüche. Um die damit verbundenen Herausforderungen besser meistern zu können, sind deshalb fundierte Methodenkenntnisse nötig, und zwar sowohl in der Tiefe als auch in der Breite: Grössere Vertrautheit mit spezifischen Evaluationsmethoden erleichtert es, ihre Stärken und Schwächen abzuschätzen

und mögliche Grenzen und Risiken zu erkennen. Routine bei der Anwendung hält zudem den damit verbundenen Aufwand in Grenzen. Eine vielfältige Palette verfügbarer Methoden bringt grössere Auswahlmöglichkeiten und erhöht die Wahrscheinlichkeit, eine auf den jeweiligen Kontext und die Ressourcensituation angepasste Methode zu finden.

Wie eingangs erwähnt, orten zahlreiche SEVAL-Mitglieder Ausbildungsbedarf in Bezug auf methodische Fragen. Dieser betrifft nicht nur Evaluierende, sondern auch Auftraggebende: Es ist zu vermeiden, dass bezüglich der Methodenkompetenz eine zu grosse Diskrepanz zwischen den beiden Seiten besteht, weil gemeinsames Fachwissen auch eine Vertrauensbasis bildet und die Zusammenarbeit erleichtert. Auftraggebende sollten in der Lage sein, die methodische Qualität einer Evaluation beurteilen zu können, aber mit Rückgriff auf ihre Methodenkompetenz auch gegenüber Evaluierenden Verständnis für allfällige methodische Probleme (Datenqualität, Repräsentativität, Validität, Konflikt mit verfügbaren Ressourcen usw.) aufbringen und gegebenenfalls die Grenzen der Machbarkeit anerkennen können. Evaluierende ihrerseits sollten grundsätzlich über möglichst hohe Methodenkompetenz verfügen, um das Vertrauen der Auftraggeber in sie zu rechtfertigen. Dies beinhaltet u. a., dass sie die oben erwähnte Definitionsmacht nicht überstrapazieren und einen adäquaten Umgang mit Evaluationsmethoden pflegen, das heisst die allgemeinen Gütekriterien (Repräsentativität, Validität, Reliabilität usw.) beachten, aber auch der Versuchung widerstehen, hochkomplexe, nur für Experten nachvollziehbare Methoden einzusetzen oder die analytischen Möglichkeiten einer bestimmten Methode exzessiv auszureizen, ohne damit einen Mehrwert für den Auftraggeber zu schaffen.

Die SEVAL will sich diesem Ausbildungsbedarf annehmen und so einen Beitrag zur weiteren Professionalisierung der Evaluationstätigkeit in der Schweiz leisten. Um die bereits bestehenden zahlreichen spezifischen Kursangeboten zur methodologischen Weiterbildung, die zum Teil von der SEVAL mitorganisiert werden⁶, zu ergänzen, findet deshalb im Vorfeld der SEVAL-Tagung 2010 erstmals ein Methodenatelier statt, an dem Evaluationen mit innovativen oder interessanten methodologischen Aspekten vorgestellt und diskutiert werden.⁷

Angesichts der skizzierten Herausforderungen ist aber auch die gemeinsame Auseinandersetzung über die Grenzen und Möglichkeiten von Evaluation und von Evaluationsmethoden wichtig, damit sich alle an einer Evaluation Beteiligten diesen bewusst sind, nicht überhöhten Erwartungen erliegen und einen verantwortungsvollen Umgang mit Evaluationen pflegen können. Zur Illustration dieses Punkts sei hier die Erläuterung zu Standard D1 (Praktikable Verfahren) der SEVAL-Standards zitiert:

Bei der Planung und Durchführung der Evaluation darf nicht nur danach gestrebt werden, die aus der Sicht der Wissenschaft geeignetsten Verfahren zu verwenden. Genauso wichtig ist es, darauf zu achten, dass die gewählten Methoden und Verfahren möglichst praktisch anwendbar sind und dass dadurch weder der Evaluationsgegenstand noch die von der Erhebung betroffenen Personen unnötig belastet werden. Im Rahmen von Evaluationsprojekten können die aus wissenschaftlicher Sicht aussagekräftigsten Methoden oft nicht verwendet werden, da sie zu aufwendig, zeitraubend oder im entsprechenden Kontext ethisch nicht akzeptabel sind. Wichtig ist hier, die Vor- und Nachteile sowie die Aussagekraft der gewählten Verfahren bei der Planung des Evaluationsprojektes transparent zu machen und mit den Auftraggeberinnen und Auftraggebern und den Adressatengruppen zu diskutieren. (Widmer et al. 2000, 6).

Dieser Standard verweist auf die Bedeutung der gemeinsamen Abwägung bei der Entscheidungsfindung über die sinnvolle Evaluationsmethode. Die bereits mehrfach angesprochene Verantwortung, die die bei der Planung, Umsetzung und Verwendung einer Evaluation beteiligten Akteure tragen, ist somit nicht nur, aber auch eine gemeinsame Verantwortung. Die SEVAL-Tagung 2009 und die daraus resultierenden Artikel in der vorliegenden LeGes-Ausgabe sollen einen Beitrag an die gemeinsame Diskussion leisten und die Reflexion über die damit verbundenen Fragen anregen.

*Christian Rüefli, lic. rer. soc., Büro Vatter, Politikforschung & -beratung, Bern,
E-Mail: ruefli@buerovatter.ch*

Anmerkungen

- 1 Ich danke Max Bergman, Eveline Hügli und Fritz Sager für ihr kritisches Gegenlesen und ihre Hinweise und Bemerkungen.
- 2 Der Auswertungsbericht dieser Mitgliederbefragung ist auf der Website der SEVAL publiziert (www.seval.ch/rapport/rapport_enquete.pdf).
- 3 Die Tagung wurde von einer Arbeitsgruppe vorbereitet, der folgende Personen angehörten: Gabriella Bardin Arigoni (Vorsitz), Max Bergman, Désirée Donzallaz, Roland Pfyl, Christian Rüefli, Fritz Sager und Patrick Zadory (Sekretariat).
- 4 Die Aktualität der Themenwahl zeigt sich auch daran, dass die Deutsche Evaluationsgesellschaft DeGEval ihre Jahrestagung 2010 (13.-15. September 2010 in Luxemburg) ebenfalls dem Thema «Evaluation und Methoden» widmet (www.degeval.de/index.php?class=Calimero_Article&id=21287).
- 5 Vaessen und Leeuw (2010) illustrieren diesen *evaluation gap* anhand konkreter Beispiele und verweisen auf die wachsende Distanz zwischen sozialer (und ökonomischer) Forschung einerseits und der vorwiegend als Managementinstrument eingesetzten Evaluationspraxis.
- 6 Weiterführende Informationen hierzu finden sich auf der SEVAL-Website (www.seval.ch/de/veranstaltungen/ausbildung.cfm).

- 7 Die SEVAL-Jahrestagung 2010 zum Thema «Studien und Evaluationen auf internationaler Ebene: Erfahrungen aus der Schweiz» findet am 10. September in Bern statt; das Methodenatelier tags zuvor, am 9. September.

Literatur

- Balthasar, Andreas, 2007, Institutionelle Verankerung und Verwendung von Evaluationen. Praxis und Verwendung von Evaluationen in der schweizerischen Bundesverwaltung, Zürich/Chur, Rüegger.
- Leeuw, Frans L./Vaessen, Jos (Hrsg.), 2009, Impact Evaluations and Development. NONIE Guidance on Impact Evaluation, Washington: World Bank Independent Evaluation Group.
- Vaessen, Jos/Leeuw, Frans L. (Hrsg.), 2010, Mind the Gap: Perspectives on Policy Evaluation and the Social Sciences, New Brunswick, Transaction Publishers.
- Widmer, Thomas/Rothmayr, Christine/Serdült, Uwe, 1996, Kurz und gut? Qualität und Effizienz von Kurzevaluationen. Zürich/Chur, Rüegger.
- Widmer, Thomas/Landert, Charles/Bachmann, Nicole, 2000, Evaluations-Standards der Schweizerischen Evaluationsgesellschaft (SEVAL-Standards), <http://www.seval.ch/de/standards/index.cfm>.