

# Evaluationsmethoden zwischen Pragmatismus und Qualität – Ansätze zum Umgang mit dem Spannungsfeld

**Christian Rüefli** | *Bei der Realisierung von Evaluationen ist es wegen zeitlicher und finanziellen Restriktionen sowie Problemen der Datenverfügbarkeit oft kaum möglich, die aus methodologischer Sicht am besten geeigneten Verfahren einzusetzen, um die zur vollständigen und differenzierten Beantwortung der aufgeworfenen Evaluationsfragen notwendigen Informationen zu gewinnen. Evaluierende sind deshalb in methodischer Hinsicht oft zu einem pragmatischen Vorgehen gezwungen. Unter solchen Umständen methodische Qualität und hohe Nützlichkeit der Evaluation für den Auftraggeber sicherzustellen, kann jedoch eine Herausforderung darstellen. Der vorliegende Beitrag thematisiert dieses Spannungsfeld aus Sicht eines Auftraggebers und eines Evaluators und zeigt Wege auf, wie trotz beschränkter Ressourcen möglichst umfassende und qualitativ genügende Evaluationen durchgeführt werden können. Dabei kommt den SEVAL-Standards als Qualitätssicherungsinstrument eine gewisse Bedeutung zu.*

## Inhaltsübersicht

- 1 Ausgangslage: Evaluationsmethoden im Spannungsfeld zwischen Pragmatismus und Qualität
  - 1.1 Konkurrierende Ansprüche an Evaluationen
  - 1.2 Ziel und Aufbau des Beitrags
- 2 Die Evaluationsmethodik im Spannungsfeld verschiedener Ansprüche und Bedingungen
  - 2.1 Das Spannungsfeld aus der Sicht eines Auftraggebers und eines Evaluators
  - 2.2 Determinanten der Methodenwahl
  - 2.3 Risiken für die Qualität einer Evaluation
- 3 Ansätze zur Überwindung des Spannungsfelds
  - 3.1 Gestaltung des Evaluationsdesigns und der Methodik
  - 3.2 Gestaltung des Evaluationsprozesses
- 4 Ansätze zur Qualitätssicherung in der Evaluation
  - 4.1 Erfolgsbedingungen für pragmatische Vorgehensweisen
  - 4.2 Die Bedeutung der SEVAL-Standards als Qualitätssicherungsinstrument
- 5 Fazit

## **1 Ausgangslage: Evaluationsmethoden im Spannungsfeld zwischen Pragmatismus und Qualität**

### **1.1 Konkurrierende Ansprüche an Evaluationen**

In der Evaluation gelten dieselben Ansprüche an die Wissenschaftlichkeit wie an die Grundlagenforschung, insbesondere auch hinsichtlich der Verwendung wissenschaftlicher Methoden zur Erhebung und Auswertung von Daten (vgl. die Definition von Politikevaluation in Klöti 1997, 39). Diesem Umstand tragen auch mehrere SEVAL-Standards (vgl. Widmer et al. 2000) Rechnung. Sie verlangen den

Einsatz praktikabler Verfahren (Standard D1), die eine angemessene und systematische Analyse qualitativer und quantitativer Informationen nach den Regeln der jeweiligen Methodik ermöglichen (Standard G7). Dabei sollen Validität und Reliabilität gewährleistet sein (Standard G5), und es soll eine systematische Informationsüberprüfung erfolgen (Standard G6). Der adäquate und korrekte Einsatz wissenschaftlicher Methoden ist somit ein Qualitätselement von Evaluationen.

Ein anderes Qualitätselement stellt die Nützlichkeit für den Auftraggeber dar. Diesem Kriterium ist innerhalb der SEVAL-Standards eine eigene Gruppe<sup>2</sup> mit acht Standards gewidmet. Die Nützlichkeitsstandards stellen sicher, dass sich eine Evaluation an den Informationsbedürfnissen ihrer vorgesehenen Nutzerinnen und Nutzer ausrichtet. So sollen Evaluationen u.a. möglichst alle Stakeholder-Interessen einbeziehen (Standard N1). Die Berichterstattung soll aus der Sicht der Nutzerinnen und Nutzer «rechtzeitig» erfolgen (Standard N7), wobei ein vollständiger und klarer Bericht abzuliefern ist (Standard N6), der die Bewertungen transparent und nachvollziehbar darstellt (Standard N5).

Damit widerspiegeln die SEVAL-Standards die unterschiedlichen Ansprüche, die an Evaluationen gestellt werden. Widmer et al. (1996, 17 f.) haben dies am Beispiel des Spannungsfelds zwischen Politik und Wissenschaft illustriert. So erwartet die Politik als häufige Auftraggeberin und Adressatin von Evaluationen, dass Evaluationsergebnisse in nützlicher Form, termingerecht und kostengünstig bereitstehen, damit sie als eine von mehreren Entscheidungsgrundlagen herangezogen werden können. Demgegenüber stellt die Erwartung, dass Evaluationen die allgemeinen Gütekriterien wissenschaftlicher Forschung einhalten, eine wesentliche Legitimationsbasis von Evaluation als wissenschaftlicher Dienstleistung dar (Widmer 2001, 17). In der Praxis sind die beiden Erwartungen jedoch oft schwer gleichzeitig zu erfüllen, da finanzielle oder zeitliche Restriktionen es häufig verunmöglichen, Evaluationen gemäss dem Goldstandard wissenschaftlichen Arbeitens zu realisieren. Umgekehrt sind nach allen Regeln der Kunst erstellte, hochgradig wissenschaftliche Evaluationen für die Politik meist nur von beschränktem Nutzen, wenn sie nicht bedürfnisgerecht aufbereitet sind oder nicht die erwarteten eindeutigen Aussagen liefern.

## **1.2 Ziel und Aufbau des Beitrags**

Der vorliegende Beitrag geht der Frage nach, wie trotz beschränkter Ressourcen sichergestellt werden kann, dass eine Evaluation den inhaltlichen Ansprüchen des Auftraggebers gerecht wird (Kriterium der Nützlichkeit) und allgemeinen methodischen Standards genügt (Kriterium der Genauigkeit). Diese Frage betrifft im Wesentlichen zwei Ebenen: zum einen die Evaluationsmethodik, d.h. das Un-

tersuchungsdesign sowie die Erhebungs-, Auswertungs- und Interpretationstechnik (vgl. Klöti/Widmer 1997; Widmer/Binder 1997), zum anderen die Gestaltung des Evaluationsprozesses, d. h. der Interaktion zwischen Auftraggebenden, Evaluierenden und Evaluierten (vgl. Picciotto 2005; Widmer 2001).

Im folgenden Abschnitt wird das skizzierte Spannungsfeld ausgeleuchtet und aus den Perspektiven eines Auftraggebers und eines Evaluators illustriert. Zudem erfolgt eine kurze Diskussion möglicher qualitativer Risiken methodologischer Kompromisse. Anschliessend werden verschiedene praxisbezogene Lösungsansätze für den Umgang mit dem Spannungsfeld erörtert. Nach einer Darstellung von Erfolgsbedingungen für pragmatische Vorgehensweisen und einer Diskussion der Bedeutung der SEVAL-Standards für die Qualitätssicherung in der Evaluation wird ein kurzes Fazit gezogen.

## **2 Die Evaluationsmethodik im Spannungsfeld verschiedener Ansprüche und Bedingungen**

### **2.1 Das Spannungsfeld aus der Sicht eines Auftraggebers und eines Evaluators**

Die Workshopreferate von Kurt Bisang, Evaluationsverantwortlicher im Bundesamt für Energie, und Stefan Rieder von Interface, Institut für Politikstudien, illustrieren das einleitend skizzierte Spannungsfeld und zeigen auf, welche Ansprüche und Restriktionen die Gestaltung von Evaluationen beeinflussen können.

Zunächst ist der eingangs skizzierte Antagonismus zwischen Auftraggebenden und Beauftragtem zu relativieren. Auftraggeber sind im Lauf von Evaluationsprozessen mit denselben Problemen und Fragestellungen konfrontiert wie Evaluierende. Gegen aussen treten Evaluationsverantwortliche zwar als Auftraggeber auf. Intern sind sie jedoch gegenüber ihren übergeordneten Stellen für die Durchführung guter und nützlicher Evaluationen verantwortlich. Sie befinden sich somit ihrerseits in der Rolle des Beauftragten und sind ebenfalls dem skizzierten Spannungsfeld ausgesetzt.

Auftraggeberseitig sind die strategischen Ansprüche an eine Evaluation oft vielfältig, teilweise auch widersprüchlich. Ist eine Projektbegleitgruppe involviert, in der verschiedene Behörden oder Organisationen vertreten sind, kann sich dies mitunter zusätzlich akzentuieren. Die Ansprüche an eine Evaluation können sich im Verlauf eines Projekts zudem ändern. Um bei der Vergabe eines Evaluationsauftrags möglichst hohe Genauigkeit und Nützlichkeit sicherzustellen, gelten für Bisang gewisse nicht verhandelbare Mindestanforderungen an die Konzeption der Evaluation. So wird u.a. erwartet, dass der Evaluationsgegenstand nicht eindimensional betrachtet wird. Das Forschungsdesign soll sich an den Fragestellungen ausrichten, praktikabel und umsetzbar sein, dabei jedoch dem wissenschaftlichen Standard der vorgesehenen Methode möglichst nahe

kommen. Wo möglich sollen Vergleichsgruppendesigns eingesetzt werden. Auftragnehmende sind bei der Konzeption einer Evaluation gefordert, das verfügbare Budget sinnvoll einzusetzen, methodische Prioritäten zu setzen und ein möglichst gutes Verhältnis zwischen Ressourcenaufwand und methodischer Rigidität zu finden. Quantitative Erhebungen müssen statistisch belastbar sein, während bei qualitativen Erhebungen zu beachten ist, dass diese mit einem gewissen Aufwand verbunden sind. Weiter wird in Bezug auf die allgemeine Qualität von den Evaluierenden erwartet, dass sie ihre Empfehlungen und Ergebnisse verständlich herleiten und adressatengerecht formulieren sowie den Auftrag grundsätzlich professionell abwickeln.

Bei der Planung und Konzeption von Evaluationen bestehen auftraggeberseitig gewisse Restriktionen, insbesondere hinsichtlich der verfügbaren zeitlichen und monetären Ressourcen. Zum einen richtet sich das für eine Evaluation eingeplante Budget nach dem Evaluationsgegenstand. Es kann deshalb sehr unterschiedlich ausfallen. Dies hat Konsequenzen für die Wahl des Untersuchungsdesigns und der Erhebungsmethodik: In der Regel ist nicht alles, was methodisch wünschenswert wäre, wie quasi-experimentelle Designs, die Untersuchung zielgruppenspezifischer Teilsamples oder vertiefte Face-to-face-Interviews, auch finanzierbar. Zum anderen sehen sich Auftraggeber meistens auch mit zeitlichen Restriktionen konfrontiert. So ist die verfügbare Zeit zur Projektbegleitung und für Feedbackschlaufen oft knapp. Dasselbe gilt für die Auseinandersetzung mit den Empfehlungen und Ergebnissen einer Evaluation. Zeitknappheit setzt der Lernbereitschaft eines Auftraggebers oftmals Grenzen. Neben den Ressourcen ergeben sich auch aus dem auftraggeberseitigen Kontext einer Evaluation gewisse Restriktionen. Der Kontext, vor allem der politische, ist mitunter dynamisch, was sich einerseits auf die Projektabwicklung auswirken kann, aber auch für den Auftraggeber eine Herausforderung darstellt, wenn es gilt, den idealen Zeitpunkt für eine Evaluation und für die Präsentation von deren Ergebnissen zu finden.

Wie Rieder festhielt, sind an Evaluationen nicht nur Auftraggeber und Evaluierende beteiligt, sondern auch die Zielgruppen bzw. Betroffenen (Stakeholder). Durch die Erweiterung mit den Stakeholdern gewinnt das methodenbezogene Spannungsfeld an Komplexität. Neben dem hier karikierend überzeichneten Anspruch des Auftraggebers, von einer Evaluation unter Zeit- und Budgetdruck einfache, aber vollständige Antworten auf eine unrealistische Fragestellung zu erhalten, ergeben sich so zusätzliche methodische Herausforderungen an Evaluationen: So kommt es vor, dass die evaluierten Zielgruppen der Evaluation bzw. den Evaluierenden grundsätzlich ein gewisses Misstrauen entgegenbringen<sup>10</sup> oder bei der Preisgabe von Informationen und Daten zurückhaltend sind, nicht zuletzt, weil dies für sie mit einem gewissen Aufwand verbunden ist. Für Evalu-

ierende gilt damit, bei der Konzeption der Untersuchungsmethodik nicht nur dem Erkenntnisinteresse und den Rahmenbedingungen des Auftraggebers, sondern auch potenziellen Schwierigkeiten und Widerständen bei der Datenerhebung Rechnung zu tragen.

## 2.2 Determinanten der Methodenwahl

Angesichts dieser verschiedenen Ansprüche und Restriktionen kommt der Evaluationsmethodik aus verschiedenen Gründen eine zentrale Bedeutung zu:

- In einer Input-Throughput-Output-Optik besteht ein (empirisch-analytisch ausgerichteter) Evaluationsprozess darin, im Rahmen eines an die Fragestellungen angepassten Forschungsdesigns verschiedene Informationen bzw. Daten (Input) zu erheben und zu verarbeiten (Throughput), um so zu systematisch hergeleiteten, empirisch abgestützten Evaluationsergebnissen (Output) zu gelangen. Die Evaluationsergebnisse und ihre Aussagekraft sind somit stets davon abhängig, auf welche Weise (Erhebungsverfahren) und bei wem (z. B. Anzahl und Auswahl der Befragten, Datenquellen, Statistiken) die ihnen zugrundeliegenden Informationen erhoben wurden. Das Forschungsdesign und die Evaluationsmethodik sind somit von entscheidender Bedeutung für den Output der Evaluation – und damit auch für deren Nützlichkeit.
- Die Angemessenheit des Forschungsdesigns und der eingesetzten Methode(n) und die Einhaltung wissenschaftlicher Gütekriterien wie Reliabilität und Validität bei der Informationsauswertung sind Elemente der «handwerklichen» Qualität einer Evaluation (vgl. die entsprechenden Standards) und damit auch von Bedeutung für ihre weitere Nutzung durch den Auftraggeber. So hat z. B. Balthasar (2008, 265) festgestellt, dass die generelle Qualität einer Evaluation deren instrumentelle und konzeptionelle Verwendung fördert.
- Der zeitliche, personelle und damit auch finanzielle Aufwand für eine Evaluation hängt unter anderem vom Forschungsdesign und von der eingesetzten Methodik ab. Sind zur Beantwortung der Evaluationsfragen umfangreiche und arbeitsintensive Informationserhebungen nötig (z.B. weil keine Daten zur Verfügung stehen, weil zahlreiche verschiedene Zielgruppen befragt werden müssen, weil für einen Zeitvergleich eine Nullmessung nötig ist), ist dies mit entsprechenden Kosten verbunden.

Die Wahl des Forschungsdesigns und der Methode(n) ist somit ein zentraler Entscheid bei der Konzeption einer Evaluation. Wie dieser Entscheid im konkreten Fall ausfällt, ist von verschiedenen Faktoren abhängig, die teilweise miteinander in Zusammenhang stehen (vgl. auch Independent Evaluation Group 2006, 1f.):

- Aus dem Erkenntnisinteresse und den Fragestellungen der Evaluation leitet sich der zu deckende Informationsbedarf ab. Daraus und unter Berücksichtigung der Eigenschaften der Evaluationsgegenstände und allfälliger wichtiger Kontextfaktoren ergibt sich die Erhebungsmethodik, die bei der Evaluation idealerweise zur Anwendung kommt. In der Regel gilt, dass die Ansprüche an die Methodik bei zunehmender Komplexität der Untersuchungsgegenstände und Fragestellungen steigen.
- Das verfügbare Evaluationsbudget stellt eine zentrale Rahmenbedingung für die Methodenwahl dar: Grosse Budgets ermöglichen den Einsatz verschiedener Methoden und umfangreiche Erhebungen, kleine Budgets setzen der Anzahl und Komplexität der Methoden sowie der Erhebungs- und Analyse-schritte gewisse Grenzen. Vergleichbares gilt für den Zeitraum, der für die Realisierung der Evaluation vorgesehen ist.
- Die Methodenwahl richtet sich auch nach der Art, der Verfügbarkeit und Zugänglichkeit der zu erhebenden Daten und Informationen. Es kann vorkommen, dass sich gewisse Erhebungsmethoden aufdrängen, weil die zu erfassenden Informationen nur auf diese Weise gewonnen werden können oder weil bereits umfangreiche Datenbestände bestehen. Umgekehrt sind diverse Einschränkungen denkbar, welche die Anwendung der aus forschungstechnischer Sicht am besten geeigneten Methode nicht zulassen, wie z. B. fehlende Nullmessungen eines Ausgangszustands, schwer erreichbare Zielgruppen, lückenhafte oder wenig aussagekräftige Primärdaten.
- Die Methodenwahl hängt oft auch von den spezifischen Methodenkompetenzen und -präferenzen von Auftraggebern und Evaluierenden ab. So ist es für Evaluierende oft naheliegend, sich bei der Konzeption einer Evaluation auf bewährte Standardverfahren abzustützen, in denen sie eine gewisse Routine aufweisen. Umgekehrt kann es auch vorkommen, dass Forschende zu Innovationszwecken eine Evaluation dazu nutzen, um eine noch wenig etablierte bzw. für sie neue Methode zu testen. Auftraggebende ihrerseits haben tendenziell Interesse an der Anwendung allgemein etablierter und nachvollziehbarer Methoden, die möglichst klare und eindeutige Resultate liefern. Im Extremfall setzen sie sich mit Methodenfragen gar nicht auseinander. Besteht hinsichtlich Methodenkompetenz und -präferenz eine Asymmetrie zwischen Auftraggeber und Evaluierenden, kann dies den Evaluationsprozess belasten.<sup>4</sup> Das Vertrauen, das ein Auftraggeber einer evaluierenden Person und den von ihr eingesetzten Methoden entgegenbringt, ist deshalb von einer gewissen Bedeutung.

### 2.3 Risiken für die Qualität einer Evaluation

Die Independent Evaluation Group (2006, 22) der Weltbank skizziert vier Gruppen von «Gefahren» für die Qualität und Validität von (Wirkungs-)Evaluationen, die sich aus den genannten Einschränkungen der Wahl von Evaluationsdesign und -methode ergeben können (Tabelle 1). Zu dieser Übersicht ist anzumerken, dass die Autoren des zitierten Papiers von der Idealvorstellung einer Längsschnittuntersuchung mit einem Kontrollgruppendesign und Nullmessung ausgehen, die aus wissenschaftlicher Optik lange als «Goldstandard» für Wirkungsanalysen gehandelt wurden (vgl. auch Rossi et al. 2004, 237). Allerdings sind die Voraussetzungen für die Anwendung solcher Forschungsdesigns selten gegeben. Es kommen deshalb häufig auch andere Methoden zur Anwendung. In der schweizerischen Evaluationspraxis beispielsweise kommt qualitativen Verfahren eine ebenbürtige Bedeutung zu.<sup>5</sup> Entsprechend sind nicht alle der in Tabelle 1 aufgelisteten potenziellen Risiken von gleicher Relevanz. Die Tabelle ist darum als Illustration der möglichen Auswirkungen methodischer Restriktionen im spezifischen Fall von Kontrollgruppendesigns zu betrachten und nicht als allgemeingültige Auslegeordnung.

*Tabelle:*

#### **Potenzielle Gefahren für die Validität einer Evaluation bei zeit-, budget- oder datenbezogenen Restriktionen**

*Allgemeine Qualität des Designs und der Umsetzung einer Evaluation*

*(interne Validität)*

- Evaluationsplanung und Austausch zwischen Auftraggeber und Evaluierenden kommen zu kurz
- Wenig Augenmerk auf Entwicklung und Test von Erhebungsinstrumenten
- Zeit für begleitete Nachbetreuung der Umsetzung der Evaluationsergebnisse (Follow-up) fehlt
- Verzicht auf Einbezug von schwer erreichbaren Zielgruppen und schwer erhältlichen Informationen
- Begrenzte Anwendung von Mixed-Method-Ansätzen, dadurch Überprüfung der Konsistenz von Methodentriangulation erschwert
- Druck, Erhebungen durch billigere Arbeitskräfte durchzuführen; weniger Ressourcen für Schulung und Qualitätssicherung
- Angemessenheit von Sekundärdaten schwieriger abzuklären

*Stichprobenauswahl und statistische Analyse*

- Möglichkeiten zur Anwendung von Mixed-Method-Ansätzen eingeschränkt

- Ressourcen zur Erhöhung der Qualität der Fall- und Stichprobenauswahl fehlen
  - Verzerrungen bei Fall- und Stichprobenauswahl und Bildung von Kontrollgruppen schwieriger anzugehen
  - Schlechtere Qualität der Stichprobenziehung
  - Kleinere Stichprobengröße → Repräsentativität eingeschränkt
  - Druck, auf Erhebung von Baseline-Daten (Nullmessung) oder Untersuchung von Kontrollgruppen zu verzichten
  - Weniger differenzierte Analyse
- Theoretische Kohärenz und Angemessenheit des Kontrafaktuals*
- Weniger Zeit und Ressourcen zur Entwicklung eines Programmtheorie-Modells → Kernkonzepte und Indikatoren können weniger gut definiert werden, zentrale Hypothesen werden evtl. nicht erkannt oder falsch formuliert
  - Einsatz von Mixed-Method-Ansätzen und Methodentriangulation erschwert
  - Schwächere (kleinere oder suboptimal zusammengesetzte) Kontrollgruppen
  - Schwächere oder fehlende Nullmessungen
- Verallgemeinerbarkeit der Befunde und Empfehlungen (externe Validität)*
- Weniger Augenmerk auf Verzerrungen bei Stichproben- und Fallauswahl
  - Analyse und Berücksichtigung von Kontexteinflüssen erschwert

Quelle: Independent Evaluation Group (2006, 23)

### 3 Ansätze zur Überwindung des Spannungsfelds

Die vorangehenden Ausführungen zeigen, dass bei Evaluationen die Methodenauswahl von verschiedenen Faktoren abhängt, die sich aus den unterschiedlichen Funktionslogiken und Referenzsystemen von Auftraggebern und Evaluierenden sowie aus den exogen gegebenen Rahmenbedingungen ergeben. Sowohl Auftraggeber als auch Evaluierende stehen dabei im Spannungsfeld zwischen Qualität und Pragmatismus. Mit Qualität ist hier einerseits die handwerklich korrekte, fehlerfreie Anwendung der für die Umsetzung einer konkreten Evaluation adäquaten Methoden gemeint (Struktur- und Prozessqualität), andererseits die Gültigkeit, Klarheit, Aussagekraft und Nützlichkeit der Evaluationsergebnisse (Ergebnisqualität).<sup>6</sup> Pragmatismus demgegenüber meint hier ein – im schlechtesten Fall unreflektiertes – Vorgehen, das sich in erster Linie an den äusseren Umständen und den Präferenzen der an einer Evaluation Beteiligten orientiert und

nicht an wissenschaftlichen Grundsätzen. Um die Evaluation im Rahmen der gesetzten Vorgaben realisieren zu können, werden Abweichungen von «methodologischen Handwerksregeln» und die damit verbunden Risiken (vgl. Tabelle) in Kauf genommen. Pragmatismus und Qualität schliessen sich jedoch nicht gegenseitig aus; ein pragmatisches Vorgehen muss nicht zwangsläufig zu unbefriedigenden Ergebnissen führen. Die Kunst des Evaluierens besteht u. a. darin, das Optimum zwischen wissenschaftlicher bzw. methodischer Rigidität und nutzenorientiertem Praxisbezug zu finden (Widmer et al. 1996, 17). Die Konzeption einer Evaluation erfordert somit stets ein Abwägen zwischen verschiedenen Ansprüchen und die Suche nach dem optimalen Kompromiss. Entsprechend formulierten Rossi et al. (2004, 238) die folgende Grundregel: «[T]he evaluator should choose the strongest possible design from a methodological standpoint after having taken into account the potential importance of the results, the practicability and feasibility of each design, and the probability that the design chosen will produce useful and credible results.»

Bei der Herausforderung, das Design und die Methodik einer Evaluation an äussere Umstände anzupassen, stellt sich jeweils die zentrale Frage, wie sichergestellt werden kann, dass die Evaluation trotz aller Einschränkungen die in der Tabelle skizzierten (und weitere) Risiken minimiert und ausreichend valide, präzise und differenzierte Ergebnisse produziert (Independent Evaluation Group 2006, 4).

Im Folgenden sollen nun auf der Basis der beiden Workshopreferate und ausgewählter Literatur anwendungsorientierte Wege und Möglichkeiten diskutiert werden, wie bei Evaluationen trotz restriktionsbedingter Zugeständnisse an methodologische Idealvorstellungen eine möglichst hohe Qualität sichergestellt werden kann. Die diskutierten Ansätze können grundsätzlich auf zwei Ebenen verortet werden. Die erste Ebene betrifft das Evaluationsdesign und die Evaluationsmethodik, die zweite die Gestaltung des Evaluationsprozesses.

### **3.1 Gestaltung des Evaluationsdesigns und der Methodik**

In Bezug auf die Konzeption und Durchführung einer Evaluation stehen drei Ansätze im Vordergrund, die eine Begrenzung des Aufwands ermöglichen: die konzeptionelle, inhaltliche oder methodische Fokussierung der Evaluation, der Einsatz möglichst einfacher Methoden und Einschränkungen bei der Methoden-anwendung.

Ein erster Ansatz besteht in der Fokussierung der Untersuchung. In konzeptioneller Hinsicht kann dies beispielsweise durch den Verzicht auf eine systematische Wirkungsevaluation zugunsten einer Prozessevaluation erfolgen, allenfalls ergänzt durch Zufriedenheitsbefragungen der Zielgruppen. Eine

inhaltliche Fokussierung bedeutet eine sinnvolle Eingrenzung des Untersuchungsgegenstandes, eine methodische Fokussierung schliesslich bedeutet die optimale Gestaltung des Forschungsdesigns, was auch heisst, die Datenerhebungen auf das Notwendige zu reduzieren und möglichst effizient abzuwickeln.<sup>7</sup>

Zweitens bestehen diverse Möglichkeiten, die Evaluationsmethodik zu vereinfachen. So bietet es sich z. B. an, nur einzelne, gezielt ausgewählte Indikatoren zu erheben, auf Primärerhebungen zu verzichten und mit Sekundärdaten zu arbeiten, Nullmessungen zu rekonstruieren statt zu erheben, die Stichprobengrößen bei Befragungen einzuschränken usw. In seinem Referat skizzierte Rieder drei qualitative Evaluationsmethoden, die es trotz ausgeprägter Zeit- und Ressourcenknappheit ermöglichen, die notwendigen Informationen zu erheben: Das Expertenurteil sollte sich auf theoretisch hergeleitete Kriterien abstützen, mit diesen abgeglichen werden und im Evaluationsbericht explizit als subjektive Experteneinschätzung deklariert werden. Idealerweise grenzen die Expertinnen und Experten die Aussagekraft der Ergebnisse zudem selber ein. Nach einem ähnlichen Muster verlaufen Gruppengespräche unter Fachpersonen oder Stakeholdern: Ein gemeinsam definierter Kriteriensatz wird gemeinsam qualitativ angewandt; die so gewonnenen Ergebnisse und Schlussfolgerungen werden unter den Beteiligten ausgetauscht. Qualitative Fallstudien schliesslich erlauben die exemplarische Darstellung komplexer Zusammenhänge, sind einfach verständlich und bilden bei guter Fallauswahl sehr ergiebige Informationsquellen.

Drittens lassen sich über diverse weitere Massnahmen der Aufwand und damit die Kosten der Datenerhebung gering halten. Evaluierende können z. B. auf umfangreiche Pretests der Erhebungsinstrumente verzichten, die Anzahl Interviews auf ein vertretbares Minimum begrenzen, Interviews telefonisch statt face-to-face durchführen, Gruppen- statt Einzelgespräche abhalten oder persönliche qualitative Interviews durch standardisierte (Online-)Befragungen ersetzen. Ähnliche Möglichkeiten bestehen bei der Datenauswertung, z. B. durch Begrenzungen des Differenzierungsgrades von Analysen oder der Anzahl statistischer Qualitätstests.

Die soeben dargestellten Ansätze ermöglichen durchaus eine Begrenzung des Aufwandes. Allerdings sind verschiedene praxisbezogene Vorbehalte und Konzeptionen an die Wissenschaftlichkeit zu diskutieren.

Zum einen haben Evaluierende oftmals kaum die Möglichkeit, auf die Auftragsdefinition Einfluss zu nehmen und eine Fokussierung des Auftrags oder eine Begrenzung des Umfangs von Erhebungs- und Auswertungsarbeiten vorzunehmen. Dies gilt namentlich im Kontext der Auftragsforschung, in dem sich potenzielle Auftragnehmer in einer Wettbewerbssituation befinden. Sie werden sich entsprechend eher hüten, in einer Offerte eine Redimensionierung der ausge-

schriebenen Evaluation oder Anpassungen an der gewünschten Methodik vorzuschlagen, weil sie befürchten, das Mandat nicht zu erhalten, wenn sie die Erwartungen des Auftraggebers als unrealistisch taxieren und sich nicht an dessen Vorgaben ausrichten.<sup>8</sup>

Zum anderen sind zwar qualitative Vorgehensweisen in der Evaluationsforschung seit längerem als adäquate und qualitativ gleichwertige Alternative zu randomisierten Vergleichsgruppendesigns anerkannt (vgl. z. B. Flick, 2006; Pawson/Tilley, 1997; Widmer/Binder, 1997), allerdings haben auch sie sich an Gütekriterien wie Reliabilität und Validität zu orientieren. Deshalb weisen qualitative Evaluationen bzw. Vorgehensweisen, die hauptsächlich auf die Akkumulation und Wiedergabe von Expertenmeinungen abstellen, gewisse methodologische Angriffsflächen auf. Ein Problem kann sein, dass die so erhobenen Informationen keinen Mehrwert schaffen, weil die Evaluation auf diese Weise lediglich Wissen bündelt und Positionen aufzeigt, die unter den Akteuren des betreffenden Themenfeldes bereits allgemein bekannt sind. Auch die Gefahr, dass die Evaluationsergebnisse lediglich auf subjektiven und möglicherweise interessengeleiteten Einschätzungen von Themeninsidern beruhen, ist grösser, wenn nicht zusätzlich eine unabhängige und eigenständige, idealerweise mit zusätzlichen objektiven Daten gestützte Beurteilung erfolgt. Je nach Erkenntnisinteresse und Kontext des Auftrags kann ein einzig auf Expertenmeinungen basierendes Vorgehen durchaus sinnvoll und ertragreich sein. Allerdings ist dabei mit Blick auf die Validität dem Verfahren zur Auswahl der konsultierten Experten, der Perspektiventriangulation und der kritischen Auseinandersetzung mit den erhaltenen Einschätzungen erhöhte Aufmerksamkeit zu schenken. Zudem ist es in solchen Fällen angebracht, den spezifisch methodischen Charakter einer Studie sowie die methodischen Grenzen (z. B. Reichweite von Umfragen, selektive Auswahl von Interviewpartnerinnen und -partner) offenzulegen und im Evaluationsbericht anzusprechen, damit die Nutzenden die Ergebnisse entsprechend einordnen und interpretieren können (vgl. auch die SEVAL-Standards N6 «Vollständigkeit und Klarheit des Berichts» und G3 «Beschreibung von Zielen und Vorgehen»). Eine solche kritische Auseinandersetzung mit den Grenzen der Aussagekraft kann als allgemeines Qualitätsmerkmal einer Evaluation bezeichnet werden.

### **3.2 Gestaltung des Evaluationsprozesses**

Hinsichtlich der Gestaltung des Evaluationsprozesses bietet sich bei knappen Ressourcen an, diese möglichst auf die methodenrelevanten Arbeitsschritte zu konzentrieren und den administrativen Aufwand für die Evaluation möglichst gering zu halten. Dies kann bedeuten, dass nur wenige und eher knapp gehaltene Arbeitsdokumente (z. B. Zwischenberichte) erstellt werden oder dass sich die Eva-

luation auf die Datenerhebung und -analyse beschränkt und dem Auftraggeber lediglich Resultate liefert, ohne diese zu interpretieren.

Wie auch die Diskussionen in der Arbeitsgruppe Evaluationsstandards der SEVAL (vgl. hierzu Rieder, 2009) zeigen, ist die Phase der Projektplanung – sowohl seitens eines Auftraggebers (Auftragsdefinition, Ausarbeitung von Fragestellungen und Pflichtenheft) als auch seitens der Evaluierenden (Konzeption der Evaluation bei der Offertstellung) – von entscheidender Bedeutung für den weiteren Verlauf des Evaluationsprozesses und für die Qualität der Evaluation. Eine Auftragsanalyse und Machbarkeitsprüfung kann in diesem Stadium dazu beitragen, allfällige Diskrepanzen zwischen den Evaluationszielen und den verfügbaren Ressourcen, Fragen zur Datenverfügbarkeit, methodologische Stolpersteine und Kontextfaktoren wie z. B. den Zugang zu und die Kooperationsbereitschaft von Zielgruppen usw. anzusprechen und zu klären. Dies erfolgt idealerweise gemeinsam zwischen Auftraggeber und Evaluierenden. Aus Praxissicht ist allerdings festzustellen, dass einer derartigen prospektiven Auseinandersetzung mit der Evaluation in der Regel diverse Hindernisse entgegenstehen: Sie ist mit Aufwand verbunden, die Zeit dafür fehlt meistens, oder es ist unklar, ob die Klärung praktischer, methodenrelevanter Fragen, z. B. nach vorhandenen Datenbeständen oder zum Zielgruppenzugang, Sache des Auftraggebers oder der potenziellen Auftragnehmer ist bzw. ob sie bereits im Rahmen der Offertstellung oder erst nach Auftragsvergabe erfolgen soll. Ausserdem sind potenzielle Probleme zu Beginn des Evaluationsprozesses oft schwer absehbar und zeigen sich erst im Laufe der Arbeiten.

In der Workshopdiskussion wurde von einem Tagungsteilnehmer angemerkt, dass der Nutzen einer Evaluation dadurch erhöht werden kann, dass Evaluierende bereits bei der Evaluationsplanung Massnahmen vorsehen, um die Auftraggeber dazu zu bringen oder dabei zu unterstützen, die Evaluationsergebnisse zu reflektieren und deren Nutzung zu diskutieren. Dies bedeutet konkret eine Vermittlung der Ergebnisse über den eigentlichen Evaluationsbericht hinaus, im Sinne eines begleiteten Follow-ups.<sup>9</sup> Knappe Ressourcen oder fehlendes Interesse seitens des Auftraggebers stellen jedoch mögliche Hürden für derartige Aktivitäten dar.

Wie in der Workshopdiskussion auch erwähnt wurde, wird ein Evaluationsprozess dadurch vereinfacht, wenn bereits bei der Konzeption eines Programms oder Projekts unterstützende Vorkehren getroffen werden, z.B. indem eine Nullmessung des Ausgangszustands erfolgt und dafür gesorgt wird, dass die zur Evaluation notwendigen Daten routinemässig im Rahmen der Programm-/Projektumsetzung erhoben werden, beispielsweise über ein Monitoring oder über Selbstevaluationen. Im Idealfall tragen solche Massnahmen dazu bei, den Auf-

wand für Datenerhebungen gering zu halten. Dies setzt allerdings eine seriöse und kompetente Konzeption und Umsetzung der Monitoring- oder Selbstevaluationsmassnahmen voraus. Erfahrungsgemäss können Evaluierende nicht davon ausgehen, dass ihre Informationsbedürfnisse durch solche Vorkehren ausreichend und qualitativ befriedigend gedeckt sind, sodass bei einer Evaluation trotzdem gewisse eigene Nacherhebungen nötig sind.

#### **4 Ansätze zur Qualitätssicherung in der Evaluation**

In seinem Workshopreferat stellte Rieder neben verschiedenen methodischen Kompromissvorschlägen eine weitere Alternative zur Diskussion, wie sich Auftragnehmende bei Anfragen oder Ausschreibungen verhalten können, wenn Ressourcenknappheit oder andere Restriktionen es voraussichtlich verunmöglichen, einen Evaluationsauftrag so abzuwickeln, dass eine ausreichende Prozess- und Ergebnisqualität sichergestellt werden kann: den potenziellen Auftrag ablehnen. So vernünftig diese Option aus sachlichen Gesichtspunkten auch erscheint, ein derartiger Entscheid dürfte insbesondere für ein privatwirtschaftliches Unternehmen, das unter Akquisitions- und Konkurrenzdruck steht, unter Umständen nicht leicht zu fällen sein. Rieders Votum wirft aber eine – im Workshop nicht diskutierte – zentrale Frage auf: Gibt es (methodologische) Mindestanforderungen, die eine Evaluation einhalten muss, um als «qualitativ ausreichend» gelten zu können (Independent Evaluation Group 2006, 4)?

Auf diese Frage lässt sich keine allgemeingültige Antwort geben, nicht zuletzt, weil «Qualität» verschiedene Dimensionen aufweist und ein entsprechend schwer zu fassendes Kriterium darstellt (vgl. Widmer, 2001). Die vielschichtige Frage nach der Qualität einer Evaluation – bezüglich der Anforderungen an eine Evaluation, bezüglich der Gewährleistung dieser Anforderungen und bezüglich der Beurteilung der Qualität – ist jeweils im konkreten Anwendungsfall unter Berücksichtigung der spezifischen Umstände zu beantworten. Es bestehen allerdings diverse Hilfsmittel, die Evaluierende und Auftraggebende dabei unterstützen können.

##### **4.1 Erfolgsbedingungen für pragmatische Vorgehensweisen**

Kapitel 3 dieses Beitrags skizziert verschiedene Möglichkeiten zur Gestaltung des Evaluationsdesigns bzw. der Methodik, um trotz beschränkter zeitlicher und finanzieller Ressourcen qualitativ befriedigende Evaluationen mit ausreichend abgestützten Ergebnissen zu erarbeiten. Widmer et al. (1996) bezeichneten solche Evaluationen als «Kurzevaluationen»<sup>10</sup> und gingen der Frage nach, unter welchen Bedingungen «es angezeigt erscheint, als Alternative zu umfangreicheren Evaluationen die Durchführung von Kurzevaluationen in Erwägung zu ziehen. Sind

diese Bedingungen nicht erfüllt, erhöht sich die Gefahr, dass Kurzevaluationen ihre Zielsetzungen (Behandlung der evaluativen Fragestellung, methodisch ausreichend abgesicherte Ergebnisse) verfehlen.» (Widmer et al. 1996, 253ff.). In einer vergleichenden Untersuchung von 15 Evaluationsstudien identifizierten sie die folgenden Erfolgsbedingungen:

- Untersuchungsgegenstände ohne allzu grosse politische, soziale, ethische und finanzielle Relevanz
- bereits ausreichende Vorkenntnisse zu Untersuchungsgegenständen vorhanden
- eher geringe Komplexität der Untersuchungsgegenstände
- Der Untersuchungsgegenstand wurde im Hinblick auf eine Evaluation konzipiert und umgesetzt (z. B. Nullmessungen und Datenmonitoring vorhanden).
- hoher Zeitdruck
- hohe Akzeptanz der Evaluation
- präzise Evaluationsplanung
- formative statt summative Konzeption der Evaluation
- vor allem Prozessevaluation, weniger Wirkungsevaluation
- von der Evaluation werden keine Vertiefungen und keine Innovationen erwartet
- enge Projektbegleitung
- einfache Projektorganisation
- kompetente und flexible Evaluierende
- hohe Risikobereitschaft der Beteiligten und Betroffenen

#### **4.2 Die Bedeutung der SEVAL-Standards als Qualitätssicherungsinstrument**

Mit dem Ziel, einen Beitrag zur Professionalisierung der Evaluation in der Schweiz zu leisten, verabschiedete die SEVAL im Jahr 2000 ein Dokument mit 27 Evaluationsstandards (Widmer et al., 2000). Die Einhaltung der SEVAL-Standards soll die Glaubwürdigkeit, Qualität und Vertrauenswürdigkeit von Evaluationen erhöhen. Sie bilden insofern ein Qualitätssicherungsinstrument, das gewisse Mindestanforderungen an Evaluationsprozesse formuliert und so auch die Überprüfung der Qualität einer Evaluation ermöglicht. Die Standards setzen – formal nicht bindende – Spielregeln, die für alle an Evaluationsprozessen Beteiligten Geltung haben, d. h. für Evaluierende, für Auftraggeber und für allfällige weitere Beteiligte (Picciotto 2005, 40, 48; Widmer et al. 2000, 1).

Die SEVAL-Standards sind keine Methodenstandards, die sich primär auf die «handwerkliche» Umsetzung einer Evaluation beziehen, sondern befassen sich

mit Evaluationsprozessen im umfassenden Sinn (vgl. Widmer, 2005). Sie thematisieren alle Phasen im Projektzyklus einer Evaluation, von der Konzeption bis zur Verwendung der Ergebnisse, und gehen auch auf die Rahmenbedingungen einer Evaluation ein, namentlich auf das Verhältnis zwischen Auftraggebenden und Evaluierenden.

Um der Vielfalt und Diversität von Evaluationsprozessen Rechnung zu tragen, sind die SEVAL-Standards eher offen gehalten. Sie bieten deshalb keine allgemeingültigen Patentrezepte an, sondern müssen im Einzelfall jeweils an die spezifische Situation angepasst werden (Widmer et al. 2000, 1). Die Standards dienen damit allen Beteiligten als Orientierungshilfe bei der Konzeption, Planung und Umsetzung einer Evaluation sowie allenfalls bei der Metaevaluation, d. h. bei der Beurteilung der Qualität einer Evaluation. Insbesondere im Vorfeld einer Evaluation bieten sie eine Basis für eine gemeinsame Diskussion zwischen Auftraggebenden und Evaluierenden zur prospektiven Auseinandersetzung mit den verschiedenen Spannungsfeldern und methodischen Herausforderungen, die für sie im Laufe eines Evaluationsprozesses absehbar sind (vgl. Rieder 2009, 395 und Abschnitt 3.2). Die Erläuterungen zu den SEVAL-Standards greifen einige der in diesem Beitrag diskutierten Spannungsfelder auf und weisen auf die Notwendigkeit des Abwägens hin. Ein Beispiel dafür findet sich bei Standard D1 «Praktikable Verfahren» (Widmer et al. 2000, 6).

Die vorläufigen Ergebnisse der bisherigen Arbeiten der SEVAL-Arbeitsgruppe «Evaluationsstandards» (vgl. Rieder, 2009) deuten darauf hin, dass die SEVAL-Standards, sofern sie den an Evaluationsprozessen Beteiligten bekannt sind, eher implizit als explizit zur Anwendung kommen. Die Integration der Standards in institutionalisierte Qualitätssicherungssysteme fördert deren Beachtung, ist aber vermutlich nicht sehr verbreitet. Dies dürfte u. a. auch damit zusammenhängen, dass die explizite Auseinandersetzung mit den Standards im Laufe eines Evaluationsprozesses vor allem zeitlichen Zusatzaufwand mit sich bringt, der das meist ohnehin als knapp empfundene Evaluationsbudget zusätzlich belastet. In diesem Zusammenhang stellt sich die interessante Frage, inwiefern die SEVAL-Standards eine unterstützende Ressource darstellen und inwiefern sie zusätzliche Restriktionen und Kosten mit sich bringen. Diese diskussionswürdige Frage konnte im Rahmen des Workshops aus zeitlichen Gründen nicht mehr angepackt werden.

## 5 Fazit

Die Kunst des Evaluierens besteht gemäss Widmer et al. (1996, 17) auch darin, im Rahmen eines Optimierungsprozesses einen Kompromiss zu finden, um den konkurrierenden Ansprüchen von Wissenschaft und Politik<sup>11</sup> gerecht zu werden. Da-

mit ist die klassische Diskrepanz zwischen Theorie und Praxis angesprochen, die nicht nur für die wissenschaftliche Evaluation gilt, sondern sich auch auf die Auftragsforschung generell übertragen lässt. Evaluation sieht sich deshalb immer wieder mit der Herausforderung konfrontiert, die von einem Auftraggeber oder anderweitig gesetzten Rahmenbedingungen einzuhalten und dabei zugleich verschiedenen exogenen und endogenen Qualitätsansprüchen ausreichend gerecht zu werden.

Dieser Herausforderung ist durch eine den Umständen angepasste Ausgestaltung des Evaluationsdesigns, der Methodik und des Evaluationsprozesses zu begegnen. Die Kunst besteht darin, aus der breiten Palette der verfügbaren sozialwissenschaftlichen Erhebungs- und Analysemethoden die in einem konkreten Fall «richtige» Kombination qualitativer und quantitativer Verfahren zu finden, sodass im Rahmen der verschiedenen Vorgaben eine ausreichend gefestigte empirische Basis gelegt werden kann, um den Evaluationsauftrag erfüllen zu können. Dabei ist ein gewisser Pragmatismus hilfreich. Hier weisen qualitative Vorgehensweisen gegenüber quantitativen Methoden gewisse Vorteile und damit grosses Potenzial auf. Die Grenzen des Pragmatismus werden durch Mindestanforderungen an die methodische Qualität gesetzt, die sich allerdings nicht eindeutig und allgemeingültig festlegen lassen. Wichtige Kriterien sind sicher die Validität in ihren verschiedenen Ausprägungen, Transparenz und Nachvollziehbarkeit bei der Argumentation und der Herleitung der Evaluationsergebnisse sowie eine offene, (selbst)kritische Diskussion der Grenzen der Aussagekraft und der eingesetzten Methodik.

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

#### **Anmerkungen**

- 1 Der vorliegende Beitrag ist ein Produkt des Workshops «Evaluation im Spannungsfeld zwischen Pragmatismus und Qualität» an der SEVAL-Jahrestagung 2009. Er nimmt Bezug auf Elemente der Referate von Kurt Bisang (Bundesamt für Energie) und Stefan Rieder (Interface Institut für Politikstudien) sowie der Plenumsdiskussion im Rahmen dieses Workshops. Der Autor dankt den beiden Referenten sowie Christian Bolliger und Fritz Sager für die kritischen Anmerkungen und wertvollen ergänzenden Hinweise.
- 2 Die 27 SEVAL-Standards sind in die vier thematischen Gruppen Nützlichkeit (N1-N8), Durchführbarkeit (D1-D3), Korrektheit (K1-K6) und Genauigkeit (G1-G10) gegliedert.
- 3 Nicht auszuschliessen ist zudem, dass die evaluierten Zielgruppen beim Auftraggeber intervenieren, mit dem Ziel, auf die Evaluationsergebnisse Einfluss zu nehmen.
- 4 Als weiterer Faktor kommt hinzu, dass es eine Vielzahl von Erhebungs- und Analysemethoden mit jeweiligen Vor- und Nachteilen gibt und dass eine stetige Weiterentwicklung der Methoden stattfindet.
- 5 Für eine summarische Diskussion des Verhältnisses zwischen quantitativer und qualitativer Forschung vgl. z.B. Flick (2007, 39-55).
- 6 Für eine Diskussion des Begriffs und möglicher Dimensionen der Qualität von Evaluationen vgl. Widmer (2001).

- 7 In diesem Zusammenhang sei auf das von der Parlamentarischen Verwaltungskontrolle (PVK) entwickelte Geschäftsprüfungsaudit verwiesen (vgl. Janett 2008). Es handelt sich dabei um «ein Instrument mittlerer Reichweite, das im Spannungsfeld zwischen Untersuchungstiefe und -aufwand einen vernünftigen Mittelweg einschlägt» (ebd. 124). Dieses Instrument wurde erarbeitet, um das Kontrollinstrumentarium der Geschäftsprüfungskommissionen gezielt zu ergänzen. Das Geschäftsprüfungsaudit zielt darauf ab, die Managementstrukturen und -prozesse einer Verwaltungsstelle aus einer betriebswirtschaftlichen Perspektive zu bewerten. Es eignet sich aber nicht dafür, Policies, Programme oder Massnahmen zu evaluieren oder Kausalitäten aufzuzeigen. Es handelt sich somit um ein Instrument mit klar fokussiertem Einsatzzweck und kann als Beispiel für konzeptionelle Innovation betrachtet werden.
  - 8 Inwiefern diese Furcht berechtigt ist, ist selbstverständlich stets vom konkreten Setting, dem Konkretisierungsgrad der Vorgaben und vom Auftraggeber abhängig. Dieser kann eine kritische Auseinandersetzung mit dem Auftrag und offensichtliche Bemühungen um möglichst hohe Qualität durchaus auch positiv beurteilen.
  - 9 Wie Balthasar (2008) zeigt, ist das Interesse der Entscheidungsträger an einer Evaluation von herausragender Bedeutung dafür, ob eine Evaluation generell Verwendung findet oder nicht.
  - 10 Widmer et al. (1996) verzichten pragmatischerweise auf eine trennscharfe Abgrenzung zwischen Kurz- und Langevaluationen und treffen die Unterscheidung anhand des Medians der Kosten (100'000 CHF) und der Dauer (11 Monate) der 15 von ihnen untersuchten Evaluationen.
  - 11 Neben der Politik (oder anderen Auftraggebern) und der Wissenschaft richten auch die Öffentlichkeit als Adressatin von Evaluationen (Widmer 2001; Picciotto 2005), die evaluierten Personen bzw. Organisationen als Betroffene sowie andere datenliefernde Stellen Ansprüche an Evaluationen.
- Literatur**
- Balthasar, Andreas, 2008, «Der Einfluss unterschiedlicher Institutionalisierungsmodelle auf die Verwendung von Evaluationen: Erfahrungen aus der Schweiz», *Zeitschrift für Evaluation* 2/2008, S. 243-271.
- Flick, Uwe (Hrsg.), 2006, *Qualitative Evaluationsforschung*, Reinbek, Rowohlt.
- Flick, Uwe, 2007, *Qualitative Sozialforschung*, Reinbek, Rowohlt.
- Independent Evaluation Group, 2006, *Conducting Quality Impact Evaluations under Budget, Time and Data Constraints*, Washington D.C., IBRD/World Bank.
- Janett, Daniel, 2008, *Das Geschäftsprüfungsaudit – Entwicklung und Test eines neuen Instruments der parlamentarischen Aufsicht*, *LeGes* 2008/1, S. 123-137.
- Klöti, Ulrich, 1997, *Inhaltliche und methodische Anforderungen an wissenschaftliche Politikevaluationen*, in Bussmann, Werner/Klöti, Ulrich/Knoepfel, Peter (Hrsg.), *Einführung in die Politikevaluation*, Basel/Frankfurt a.M., Helbing & Lichtenhahn, S. 39-58.
- Klöti, Ulrich/Widmer, Thomas, 1997, *Untersuchungsdesigns*, in Bussmann, Werner/Klöti, Ulrich/Knoepfel, Peter (Hrsg.), *Einführung in die Politikevaluation*, Basel/Frankfurt a.M., Helbing & Lichtenhahn, S. 185-213.
- Pawson, Ray/Tilley, Nick, 1997, *Realistic Evaluation*, Thousand Oaks/London/New Delhi, Sage Publications.
- Picciotto, Robert, 2005, *The Value of Evaluation Standards: A Comparative Assessment*, *Journal of Multi-Disciplinary Evaluation* 3, S. 30-59.
- Rieder, Stefan, 2009, *Anwendung der SEVAL-Standards: Ergebnisse aus vier Fallstudien*, *LeGes* 2009/3, S. 387-397.
- Rossi, Peter/Lipsey, Mark W./Freeman, Howard E., 2004, *Evaluation – A Systematic Approach*. Seventh Edition. Thousand Oaks/London/New Delhi, Sage Publications.
- Widmer, Thomas/Rothmayr, Christine/Serdült, Uwe, 1996, *Kurz und gut? Qualität und Effizienz von Kurzevaluationen*. Zürich/Chur, Rüegger.
- Widmer, Thomas/Binder, Hans-Martin, 1997, *Forschungsmethoden*, in Bussmann, Werner/Klöti, Ulrich/Knoepfel, Peter (Hrsg.), *Einführung in die Politikevaluation*, Basel/Frankfurt a.M., Helbing & Lichtenhahn, S. 214-255.
- Widmer, Thomas/Landert, Charles/Bachmann, Nicole, 2000, *Evaluations-Standards der Schweizerischen Evaluationsgesellschaft (SEVAL-Standards)*, <http://www.seval.ch/de/standards/index.cfm>.
- Widmer, Thomas, *Qualitätssicherung in der Evaluation – Instrumente und Verfahren*, *LeGes* 2001/2, S. 11-39.
- Widmer, Thomas, 2005, *Leitfaden für Wirksamkeitsüberprüfungen beim Bund. Instrument zur Qualitätssicherung gestützt auf die Evaluationsstandards der Schweizerischen Evaluationsgesellschaft (SEVAL-Standards)*, Bern, Bundesamt für Justiz.

## **Résumé**

*Des restrictions temporelles et financières de même que le défaut de données font que souvent la procédure la plus adéquate du point de vue méthodologique ne peut guère être appliquée pour obtenir toutes les informations qui seraient nécessaires afin de répondre de manière complète et différenciée aux questions posées dans un projet d'évaluation. Les évaluateurs et évaluatrices sont alors contraints d'adopter une démarche pragmatique, ce qui constitue un défi s'ils veulent garantir à leurs mandants une évaluation qui leur soit utile et satisfasse aux critères de qualité méthodologique. La présente contribution traite de cette problématique des points de vue du mandant et de l'évaluateur et décrit quelques pistes qui doivent permettre de mener des évaluations aussi complètes que possible et de qualité suffisante malgré des ressources limitées. Dans ce contexte, les standards SEVAL servent d'instruments d'assurance-qualité.*