

Qualitätssicherung in der Evaluation – Instrumente und Verfahren

Thomas Widmer | *In zahlreichen Bereichen von Politik und Verwaltung sind Evaluationen üblich geworden. Nun ist es Zeit, sich mit Qualitätsstandards auseinander zu setzen. Die Schweizerische Evaluationsgesellschaft legt solche vor.*

1 Einleitung

1.1 Etablierung der Evaluationsforschung

Die Evaluation hat sich – zwar nicht vollständig, aber doch weitgehend – im letzten Jahrzehnt den ihr gebührenden Platz bei der Beurteilung staatlicher Aktivitäten in der Schweiz verschafft, besonders auf Bundesebene (Widmer/Rüegg/Neuenschwander 2001), aber auch zunehmend bei Kantonen und Gemeinden. Dazu beigetragen haben massgeblich die Arbeitsgruppe Gesetzesevaluation (AGEVAL; vgl. deren Schlussbericht aus dem Jahre 1991) sowie das Nationale Forschungsprogramm ‚Wirksamkeit staatlicher Massnahmen‘ (NFP 27; vgl. etwa Bussmann 1995 und Bussmann/Klöti/Knoepfel 1997). Diese Aktivitäten führten im Jahre 1996 zur Gründung der Schweizerischen Evaluationsgesellschaft (SEVAL), in der sich die mit Evaluationen befassten Kreise aus Politik und Verwaltung, Privatwirtschaft und Wissenschaft zusammengeschlossen haben.¹

Auch wenn (bisher) noch viele – durchaus gerechtfertigte – Postulate unerfüllt blieben (etwa ein Studiengang in Evaluation an einer schweizerischen Hochschule), hat sich die Auffassung weitgehend durchgesetzt, dass die Evaluation einen bedeutenden und unverzichtbaren Beitrag zur Gestaltung staatlicher Aktivitäten leisten kann. Doch die Erwartungen an die Evaluation sind hoch, teilweise wohl zu hoch; dies vor allem dort, wo von der Evaluation erwartet wird, dass sie flächendeckend, fundiert und ohne zeitlichen Verzug Auskunft über den Erfolg eines bestimmten Programms, einer Massnahme oder gar einer Politik geben könnte. Aus dieser Situation ergibt sich Handlungsbedarf in zweierlei Hinsicht.

1.2 Realistisches Verständnis

Erstens sind die Anstrengungen darauf auszurichten, die relevanten Akteure zu einem realistischen Verständnis des Leistungspotenzials der Evaluation hinzuführen. An manchen Orten ist eine angemessene Erwartungshal-

tung hinsichtlich der Möglichkeiten und Grenzen der Evaluation bereits fest etabliert. Doch immer wieder trifft man auf Konstellationen, in denen diesbezüglich vieles im Argen liegt. Dies lässt sich beispielsweise daran erkennen, dass es immer wieder zu Evaluationsausschreibungen kommt, die ein viel zu umfangreiches Pflichtenheft formulieren, das mit den vorgegebenen Mitteln nicht seriös bearbeitet werden kann. Häufig zu Befremden Anlass geben auch die zum Teil massiv an den Realitäten vorbeiweisenden Vorstellungen über die für die Ausführung einer Evaluation erforderlichen Zeiträume. Oder dann werden bereits in der Ausschreibung Vorgaben zur methodischen Umsetzung der Evaluation angeführt, die nach der Methodenlehre als absurd zu bezeichnen sind. Aber auch die Vorstellungen über die von einer Evaluation zu erwartenden Aussagen und Wirkungen weichen teilweise (sowohl in positiver wie negativer Richtung) massgeblich von den Möglichkeiten ab. Es wird also darum gehen, die Erwartungshaltungen der Beteiligten & Betroffenen („stakeholders“²⁾ dem Leistungspotenzial einer Evaluation anzunähern.

1.3 Qualitätssicherung

Zweitens liegt die Herausforderung darin, geeignete Massnahmen zu ergreifen, damit die Evaluation ihr Leistungspotenzial auch entwickeln kann. Die schweizerische Evaluationspraxis hat sich in den letzten Jahren nicht nur stark ausgeweitet, sie arbeitet heute auch deutlich professioneller als noch vor zehn oder fünfzehn Jahren (vgl. dazu die Befunde in Widmer 1996, Widmer/Rothmayr/Serdült 1996 und Widmer/Rüegg/Neuenschwander 2001). Trotz dieser unbestrittenen Fortschritte zeigt eine Betrachtung der aktuellen Praxis weiterhin bestehende Qualitätsmängel. Diese sind dadurch gekennzeichnet, dass Evaluationen ihre Möglichkeiten nicht ausschöpfen und damit auch durchaus angemessene Erwartungen der Beteiligten & Betroffenen nicht erfüllen können. Es sind also – im Sinne einer Konsolidierung der Evaluation – Anstrengungen erforderlich, die es erlauben, die Qualität von Evaluation in der Praxis zu gewährleisten.

1.4 Qualitätsbegriff

Nur wenn es gelingt, in diesen beiden Punkten das bisher Erreichte auch weiterhin zu gewährleisten bzw. in gewissen Fällen auch wesentliche Verbesserungen zu erzielen, lässt sich vermeiden, dass sich die Akteure, in ihren Erwartungen enttäuscht, von der Evaluation abwenden. Angesichts der Möglichkeiten, die sich durch einen sinnvollen Einsatz der Evaluation im Zusammenhang mit Staatstätigkeiten ergeben, wäre dies sehr bedauerlich.

Der vorliegende Artikel konzentriert sich auf den zweiten Punkt, also die Qualitätssicherung. Dies ist jedoch nicht in dem Sinne zu verstehen, dass diese bedeutender wäre als ein realistisches Verständnis der Möglichkeiten und Grenzen von Evaluationen oder die beiden Aspekte keine Querbezüge aufweisen würden.

Der Begriff *Qualität* wird heute überaus häufig verwendet, wobei auffällt, dass in der Regel eine griffige Spezifikation des Begriffs ausbleibt. Auch dort, wo man (in naiver Weise?) eine solche zu finden hofft, im Qualitätsmanagement, sucht man vergebens. Zwar stösst man ab und an auf eine Dimensionierung, etwa die neun Qualitätskriterien – unterteilt in ‚enablers‘ (Befähiger) und ‚results‘ (Ergebnisse) – im Modell der EFQM (European Foundation for Quality Management)³. Was aber eindeutig fehlt, ist eine einigermaßen befriedigende Antwort auf Fragen wie: Was unterscheidet einen guten von einem schlechten Schulunterricht? Was ist ein gutes Verkehrssystem oder gute Gesundheitspflege? Und was macht ein gutes Forschungsprojekt aus? Vor den für die Beantwortung dieser Fragen zu treffenden Werturteilen besteht ein grosser Respekt. Diese Fragen zeigen auch deutlich auf, dass es wenig Sinn machen kann, bereichsübergreifende Qualitätskriterien zu formulieren. Die Kriterien der Qualität sind dem jeweils zu bewertenden Qualitätsträger entsprechend zu spezifizieren.

In diesem Artikel wird der Versuch unternommen, die Qualitätsdebatte mit dem Ziel einer Definition der Qualität zu führen. Wie dargelegt, ist dabei eine Beschränkung auf den Qualitätsträger Evaluation sinnvoll.

1.5 Zum Aufbau des Artikels

Der Beitrag ist wie folgt strukturiert: Kapitel 2 wird der Versuch unternommen, die Evaluation zu charakterisieren, um auf eine bedeutende Grundlage der Debatte zur Evaluationsqualität zu verweisen. Das Kapitel 3 befasst sich mit den Erwartungen, die von verschiedener Seite an die Evaluation gerichtet werden, worin einer der Querbezüge zwischen den beiden eingangs diskutierten Handlungsfeldern zum Ausdruck kommt. Das Kapitel 4 gibt eine Einführung in die Referenzebenen zur Bewertung der Evaluationsqualität. Das Kapitel 5 stellt sodann ein Instrument der Qualitätssicherung in der Evaluation vor, die Evaluationsstandards der Schweizerischen Evaluationsgesellschaft (SEVAL-Standards). Kapitel 6 zeigt mögliche Verfahren der Qualitätssicherung in der Evaluationspraxis unter Anwendung der SEVAL-Standards auf. Das Schlusskapitel enthält eine Zusammenfassung der vorangegangenen Überlegungen sowie einen Ausblick.

2 Vielfalt der Evaluation

Der *Ansatz der Evaluation* zeichnet sich durch eine ausgesprochene Reichhaltigkeit auf verschiedenen Ebenen aus. Im Folgenden soll kurz diese Vielfalt resümiert werden, da sie eine eminent wichtige Rahmenbedingung der Debatte um die Evaluationsqualität und deren Sicherung darstellt.

2.1 Gegenstände der Evaluation

Die Unterschiedlichkeit potenzieller *Evaluationsgegenstände* – auch Evaluanda genannt – ist enorm. Es lässt sich so gut wie alles in irgendeiner Form evaluieren. Zur Spezifizierung eines bestimmten Evaluationstyps sind vielfältige Begriffe entstanden. An erster Stelle sei hier die prominente Programmevaluation angesprochen; hinzuweisen ist aber auch auf die Produktevaluation, die Personalevaluation, die Projektevaluation, die Reformevaluation, die Gesetzesevaluation, die Massnahmenevaluation, die Organisationsevaluation, die Politikevaluation oder auch die Evaluation einer Evaluation (sogenannte Meta-Evaluation; vgl. Kap. 6.2). Ebenso kann auch der Fokus der Evaluation erheblich variieren. So wird etwa unterschieden zwischen der umfassenden Globalevaluation, der spezifischeren Projektevaluation oder einer themenfokussierten Evaluation.

Aber auch bei der zu bewertenden Phase in einer *Wirkungskette* bestehen massgebliche Unterschiede. In der Sprache der Politikanalyse kann jede Phase des 'policy cycle' einer Evaluation unterzogen werden. So wird etwa die Politikformulierung oder die Politikimplementierung (Politikumsetzung; man spricht hier auch von Prozessevaluationen im Gegensatz zu Wirkungsevaluationen) ebenso evaluiert wie die drei Stufen der Wirkungsentfaltung, also der «output» (Leistungserbringung), der «outcome» (Wirkungen bei den direkten Adressaten) und der «impact» (gesellschaftliche Auswirkungen)⁴.

2.2 Verortung einer Evaluation

Wichtig bei der Spezifikation einer Evaluation ist die Frage nach deren Verortung, wobei die Orte der *Steuerung*, der *Durchführung* und der *Nutzung* zu unterscheiden sind (vgl. dazu und zum Folgenden Tabelle 1). Mit dem Ort der *Steuerung* einer Evaluation wird auf jene Stelle verwiesen, die über die Kompetenzen verfügt, die Funktion(en) einer Evaluation zu definieren und über die Wahl der Evaluationsthematik und -fragestellung zu entscheiden. Im Rahmen der Evaluationssteuerung werden auch der Ort der Durchführung und der Nutzung einer Evaluation – also die beiden anderen Verortungen – bestimmt. Mit dem Ort der *Durchführung* wird jene Instanz

angesprochen, die sich der operativen Ausführung der Evaluation annimmt, also die relevanten Informationen erhebt, erfasst und auswertet. Unter dem Ort der *Nutzung* einer Evaluation wird jener Raum verstanden, in dem die Evaluation ihre Wirkungen entfalten soll. Dies kann nicht nur nach Abschluss einer Evaluation in Form eines Berichtes der Fall sein, sondern auch während des Evaluationsprozesses oder sogar schon vor Evaluationsbeginn (präventive Wirkungen). Tabelle 1 zeigt diese drei Dimensionen mit den jeweils idealtypischen Ausprägungen. In der Praxis lassen sich Evaluationsstudien häufig nicht in der durch die Tabelle suggerierten Eindeutigkeit zuordnen; vielmehr sind Mischformen die Regel.

Dimension / Verortung	Ort der Steuerung einer Evaluation	Ort der Durchführung einer Evaluation	Ort der Nutzung einer Evaluation
Innerhalb der Institution / des Projektes	Selbstevaluation	Interne Evaluation	Formative Evaluation
Ausserhalb der Institution / des Projektes	Fremdevaluation	Externe Evaluation	Summative Evaluation

Tabelle 1: Dimensionen der Verortung einer Evaluation

Die Verortung innerhalb respektive ausserhalb des Projektes oder der Institution stellt im konkreten Anwendungsfall nicht immer eine trennscharfe Unterscheidung dar. Sie hängt sehr stark von der Definition der Systemgrenzen ab, die innen von aussen trennen. Dabei ist zu berücksichtigen, dass bei den Beteiligten & Betroffenen nicht immer Konsens über die Lokalisierung der Systemgrenzen besteht. Die Perspektive hängt auch eng mit der jeweiligen Position des einzelnen Akteurs zusammen.

Die Übersicht in Tabelle 1 will keineswegs darauf hindeuten, dass nur Kombinationen auf jeweils einer Zeile möglich wären. So ist es durchaus möglich, eine externe Fremdevaluation mit formativer Funktion zu realisieren. Sehr häufig bewegen sich aber konkrete Evaluationsstudien tendenziell eher auf der einen oder anderen Zeile. Weiter kann sich der Charakter einer Evaluation im Zeitverlauf auch verändern. Ursache derartiger Veränderungen sind oft externe Faktoren. Zum Ort der Steuerung und zum Ort der Durchführung ist zu bemerken, dass diese beiden Dimensionen in der Literatur häufig miteinander vermischt werden. So kennt meines Wissens die englische Evaluationsterminologie kein Äquivalent zum deutschen Begriff Fremdevaluation, dies im Gegensatz zu den anderen drei Typen («self evaluation», «internal» und «external evaluation»). Ich halte es für angemess-

sen, die in der deutschen Sprache mögliche Differenzierung zu verwenden, da dies zur begrifflichen Präzision beiträgt und eine wesentliche inhaltliche Unterscheidung ermöglicht.

Eine Zusatzbemerkung ist auch zur letzten Spalte der Tabelle nötig, in der die formative und die summative Evaluation aufgeführt sind. Diese beiden – funktional ausgerichteten – Begriffe sind unter der Nutzung (oder auch Verwendung) einer Evaluation subsumiert, weil sich die Funktion einer Evaluation sehr stark auf deren Nutzung bezieht. Die formative Evaluation (Verbesserungsevaluation) setzt sich zum Ziel, systeminternen Lernprozesse auszulösen und Ansatzpunkte zur Verbesserung aufzuzeigen. Die primären Adressaten einer formativen Evaluation sind demzufolge innerhalb des Systems zu finden. Die summative Evaluation (Bilanzevaluation) hingegen ist eher auf Adressaten ausserhalb der Systemgrenzen ausgerichtet. Sie soll Verantwortlichkeit gegenüber Aussenstehenden erzeugen, indem sie Entscheidungsgrundlagen zur Verfügung stellt oder zur Systemlegitimation beiträgt. Neben den beiden Funktionen der Verbesserung und der Verantwortlichkeit kann eine Evaluation grundsätzlich zwei weitere Funktionen wahrnehmen. Einmal können Erkenntnisse aus Evaluationen zur Wissenserweiterung beitragen. Zum Zweiten können Evaluationen auch mit strategischen Absichten eingesetzt und so auch instrumentalisiert werden. Die Instrumentalisierung ist jedoch nicht nur hier möglich, sondern ebenso auch im Rahmen der anderen drei genannten Funktionen.

2.3 Methodenpluralismus

Die Evaluation zeichnet sich auch hinsichtlich der zum Einsatz gelangenden methodischen Ansätze durch eine grosse Vielfalt aus. Während sie früher einen eher engen methodischen Zugang verfolgte, ist heute eine recht weitgehende methodische wie erkenntnistheoretische Offenheit festzustellen. Dies gilt sowohl auf der Ebene der epistemologischen Grundannahmen und der Untersuchungsdesigns (Klöti/Widmer 1997) als auch für die Methodik im engeren Sinne, also die Verfahren der Datenerhebung und der Datenauswertung (Widmer/Binder 1997).

Die unterschiedlichen methodischen Ansätze kennen jeweils ihre spezifischen Vorgehensweisen, Regeln und auch Qualitätskriterien. So folgt ein traditioneller Ansatz, der dem experimentellen Paradigma verbunden ist und eine quantitative Methodik einsetzt, anderen Regeln als ein konstruktivistischer Ansatz, der ein qualitatives Instrumentarium verwendet. Begriffe wie Objektivität, Validität und Reliabilität, geprägt durch ein klassisches Wissenschaftsverständnis, werden in alternativen Ansätzen anders inter-

pretiert. Dadurch wird natürlich die Aufgabe erschwert, über die Ansätze hinweg angemessene Qualitätskriterien zu formulieren.

2.4 Distanz und Umfang einer Evaluation

Die vier Funktionen der Evaluation (Verantwortlichkeit, Verbesserung, Wissensenerweiterung und Strategie; siehe Kap. 2.2) haben auch einen Bezug zur Nähe respektive zur Distanz zwischen Evaluation und Evaluandum. Die Evaluation kann, im Falle einer formativen, internen Selbstevaluation, weitgehend Bestandteil des Evaluandums sein. Aber auch wenn es sich um eine Fremdevaluation handelt, sind massgebliche Differenzen in der Distanz Evaluation – Evaluandum festzustellen. In Fällen, in der die Distanz sehr klein ist, geht die Evaluation fließend in Entwicklung und Beratung über (Organisationsberatung, Programmentwicklung usw.). Am anderen Ende des Spektrums, also bei Evaluationen, die eine sehr grosse Distanz zwischen Evaluandum und Evaluation aufweisen, verschwinden die Grenzen zur wissenschaftlichen Forschung oder je nachdem auch zur Kontrolle und Aufsicht fast gänzlich. Also ist auch bezüglich der Distanz einer Evaluation ein breites Spektrum an möglichen Ausrichtungen festzustellen. Dies ist in der Diskussion um die Evaluationsqualität angemessen zu berücksichtigen.

Hinsichtlich der Frage nach der Qualität einer Evaluation bildet auch der Umfang, also die zeitlichen, finanziellen, personellen, materiellen und rechtlichen Ressourcen, die einer Evaluation zur Verfügung stehen, einen zentralen Diskussionspunkt. Der Umfang von Evaluationen variiert in erheblichem Ausmass, wobei sich sehr kleine von sehr grossen Evaluationen schnell einmal um den Faktor hundert oder mehr unterscheiden. Unter dieser Voraussetzung überrascht es nicht, dass der Evaluationsumfang oft als ausschlaggebendes Kriterium hinsichtlich der Qualität einer Evaluation aufgefasst wird. Meines Erachtens spielt das Kriterium der Grösse für die Qualität zwar eine Rolle (vgl. dazu Widmer/Rothmayr/Serdült 1996 und Kap. 6.2), ich bin jedoch ebenso der Auffassung, dass hierfür auch die bisher genannten Unterschiede zu beachten sind. Weiter spielt für die Bewertung der Qualität einer Evaluation auch eine massgebliche Rolle, welche Ziele im Rahmen der Evaluation selbst gesetzt (vgl. Kap. 4) und welche Erwartungen an eine Evaluation gerichtet werden.

3. Erwartungen an eine Evaluation

3.1 Interaktionstheoretisches Modell

Die Evaluation, verstanden als *wissenschaftliche Dienstleistung*, wird mit unterschiedlichen Erwartungen verschiedener Beteiligter & Betroffener

konfrontiert. Der Zugang zu diesen Anforderungen soll nachfolgend mit einem interaktionstheoretischen Modell geschaffen werden.

Ausgangspunkt meiner Überlegungen bildet der Umstand, dass die *Evaluation als Prozess sozialer Interaktion* aufzufassen ist. Die Evaluatorin⁵ sieht sich in diesem Prozess mit einer Vielzahl verschiedener Interaktionen konfrontiert, die ihre Position als äusserst problematisch erscheinen lassen. Sie steht vor der Situation, von verschiedenen Seiten an sie herangetragenen, oft widersprüchlichen Ansprüchen genügen zu müssen. Zur Analyse derartiger Spannungsverhältnisse liegen bereits eine Vielzahl von theoretischen Überlegungen vor, die sich mit dem Verhältnis zwischen Wissenschaft (Theorie) und Politik (Praxis) beschäftigen (vgl. etwa Palumbo/Hallett 1993, Beck/Bonss 1989, Campbell 1984, Habermas/Luhmann 1971, Habermas 1968). Nachstehende Abbildung zeigt das hier vorgeschlagene, nicht an Personen, sondern an funktionalen Rollen orientierte interaktionstheoretische Modell:

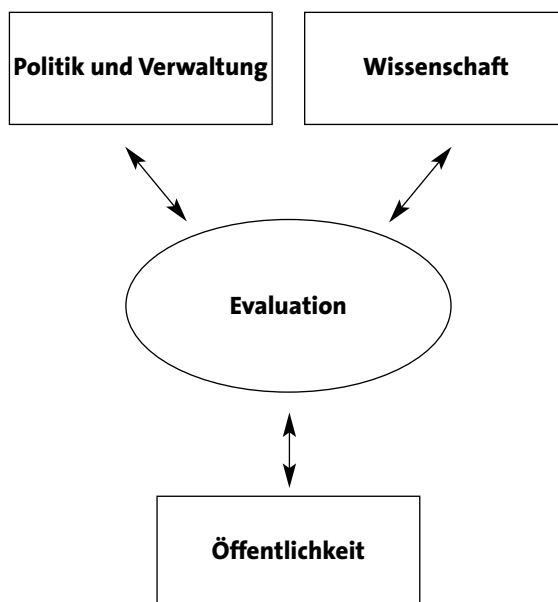


Abb. 1: Interaktionstheoretisches Modell

Der Kontext einer Evaluation kann – auf das Wesentliche beschränkt – durch drei verschiedene Akteure beschrieben werden, nämlich (1) Politik und Verwaltung, (2) Wissenschaft und (3) Öffentlichkeit. In diesem Modell werden die für eine demokratisch-verantwortlich (MacDonald 1993) ausge-

staltete Evaluation zentralen Interaktionsbeziehungen dargestellt. Diese Beziehungen, die je spezifisch strukturiert sind, sollen nachfolgend kurz charakterisiert werden.

3.2 Interaktion Politik/Verwaltung – Evaluation

Politik und Verwaltung sind zumeist in verschiedenen Rollen am Evaluationsprozess beteiligt. Sie sind häufig Auftraggeber einer Evaluation, aber oft auch wichtige Informationsquelle und zumeist gleichzeitig ihr primärer Adressat. Nicht zuletzt sind Politik/Verwaltung auch die intendierten Nutzer einer Evaluation. Im Gegensatz zu anderen hier zu diskutierenden Beziehungen ist das Verhältnis zwischen Politik/Verwaltung und Evaluatorsin zumeist vertraglich geregelt. Die Ansprüche, die seitens Politik/Verwaltung an die Evaluation gerichtet werden, lassen sich mit den folgenden Stichworten umschreiben: Nützlichkeit, Zeitgerechtigkeit, politische Opportunität, Anschlussfähigkeit, Sparsamkeit, Praktikabilität und Korrektheit.

Die Evaluatorsin andererseits hat ein ausgewiesenes Interesse, die Erwartungen des Auftraggebers auch zu erfüllen. Je nach Organisationsform ist die Evaluatorsin finanziell mehr oder weniger stark vom Auftraggeber abhängig. Sie kann zudem, sofern sie den Ansprüchen des Auftraggebers gerecht wird, darauf spekulieren, in Zukunft weitere Aufträge zu erhalten.

Das Verhältnis zwischen Politik/Verwaltung und Evaluation ist somit schwergewichtig ökonomisch und juristisch geprägt.

3.3 Interaktion Wissenschaft – Evaluation

Das Verhältnis von Wissenschaft und Evaluation hat ebenso seine spezifische Struktur. Die Wissenschaft profitiert von der Evaluation vor allem bezüglich zweier Aspekte. Die Evaluation ist in der Lage:

1. die gesamtgesellschaftliche Legitimation des Wissenschaftssystems zu unterstützen. Dies geschieht insbesondere dann, wenn sich die Evaluation als praxisrelevant erweist;
2. für wissenschaftliche Erkenntnisse wie Theorien und Methoden empirische Evidenz zu produzieren. Gleichzeitig kann die Evaluation die Fachdisziplinen auch mit zusätzlichem inhaltlichem Wissen versorgen (vgl. Kap. 2.2 und 2.4; Funktion der Wissenserweiterung).

Die Evaluation wiederum ist auf die Wissenschaft als Lieferantin wissenschaftlicher Wissensbeständen angewiesen. Die Evaluation kann nur mit Wissenschaft bestehen, da Letztere zu ihrer gesellschaftlichen Reputation beiträgt. Ohne die 'reine' Wissenschaft würde die Evaluation ihre Daseinsberechtigung als wissenschaftliche Dienstleistung verlieren.

Das Verhältnis zwischen Wissenschaft und Evaluation ist damit hauptsächlich durch den Austausch von Wissen und Reputation geprägt.

3.4 Interaktion Öffentlichkeit – Evaluation

Das Verhältnis zwischen Öffentlichkeit (im Sinne von Habermas als institutioneller Ordnung der Lebenswelt; 1981, II, 472) und Evaluation weist eine deutlich andere Qualität auf als die oben beschriebene Interaktion. Während sich die anderen beiden Interaktionen dadurch auszeichnen, dass eine direkte gegenseitige Abhängigkeit besteht, ist dies hier kaum der Fall. Die Evaluatorin ist damit vordergründig nicht auf Austausch mit der Öffentlichkeit angewiesen. Ihr bietet es sich deshalb geradezu an, auf Interaktionen mit der Öffentlichkeit gänzlich zu verzichten und sich primär in den für sie «lebensnotwendigen» Beziehungen mit Politik und Wissenschaft zu engagieren, was zur Vernachlässigung der Bedürfnisse der Öffentlichkeit führt. Die Evaluation liesse sich so im Dienste von Politik/Verwaltung und/oder Wissenschaft instrumentalisieren.

Ein demokratisch verantwortungsvoller Evaluationsprozess erfordert jedoch nicht nur, dass der resultierende Schlussbericht der Öffentlichkeit zugänglich gemacht wird; der Evaluationsprozess sollte von Beginn an die Anliegen der Beteiligten & Betroffenen berücksichtigen. Erforderlich ist die Aufnahme eines kommunikativen Diskurses, der – als Antipode zur ‘Entkoppelung von System und Lebenswelt’ – die Freiheit des Individuums steigert (in Anlehnung an Habermas 1981, II: 229–293). Dieser Anspruch ist sehr hoch gesteckt und in der Praxis wohl kaum je vollumfänglich einzulösen. Klaus von Beyme (1991, 269) dazu: «Der Sektor ‘Öffentlichkeit’ als Teil der Lebenswelt ist schwerlich je so kommunikativ gewesen, wie im Modell vorgesehen.»

Das Verhältnis zwischen Öffentlichkeit und Evaluation ist also primär ethisch-moralisch geprägt und zumindest latent gefährdet.

3.5 Zusammenfassung

Es lässt sich festhalten, dass an die Evaluation vielfältige Erwartungen gerichtet werden, die sich zudem oft gegenseitig konkurrenzieren. Weiter besteht die Gefahr, dass sich die Evaluation durch spezifische Anforderungen in unausgewogener Weise vereinnahmen lässt. Sich in diesem Spannungsfeld zu bewegen, stellt für alle an der Evaluation Beteiligten – also nicht nur die Evaluatorin, sondern ebenso den Auftraggeber und andere an der Evaluation beteiligte Kreise – hohe Anforderungen. Es lässt sich zudem schliessen, dass die Qualität einer Evaluation nicht eindimensional erfasst

werden kann, sondern dass dabei eine Vielzahl unterschiedlicher Aspekte zu berücksichtigen ist, weil eine eindimensionale Betrachtungsweise zu kurz greifen würde.

4 Referenzebenen der Evaluationsqualität

Die dargelegten Ansprüche an die Evaluation lassen sich – inspiriert durch die Vorgehensweise bei der historischen Quellenkritik – nach zwei verschiedenen Bewertungsgrundlagen differenzieren, nämlich nach den inneren und den äusseren Bewertungsgrundlagen. Diese Bewertungsgrundlagen können dazu eingesetzt werden, die Qualität einer Evaluation einzuschätzen. Die folgende Tabelle zeigt die jeweiligen Referenzebenen mit den dazugehörigen Bewertungsgrundlagen:

Ebene	Gegenstand	Innere Bewertungsgrundlagen	Äussere Bewertungsgrundlagen
Konstruktion erster Ordnung	Programm, Projekt, Massnahme etc.	Ziele des Programms, des Projektes etc.	gesamtgesellschaftliche Bedeutung des Programms
Konstruktion zweiter Ordnung	Evaluation	Ziele der Evaluation	Evaluationstheorie und Evaluationsmethodologie
Konstruktion dritter Ordnung	Meta-Evaluation	Ziele der Meta-Evaluation

Tabelle 2: Übersicht über die Referenzebenen

Die erste Spalte zeigt die Ebene des Gegenstandes, der in der zweiten Spalte aufgeführt ist. Die Evaluatorin kann ein Programm (oder einen anderen Evaluationsgegenstand; Konstruktion erster Ordnung) einerseits auf der Grundlage der Programmziele (innere Bewertungsgrundlage) und andererseits aufgrund der gesamtgesellschaftlichen Bedeutung des Programms (äussere Bewertungsgrundlage) bewerten. Analog lässt sich dies nun auch auf eine Evaluation als Konstruktion zweiter Ordnung beziehen. Die relevanten Bewertungsgrundlagen wären demzufolge erstens die Ziele der Evaluation (innere Bewertungsgrundlage) und zweitens die allgemeine Evaluationstheorie und Evaluationsmethodologie (äussere Bewertungsgrundlage). Wird eine Bewertung (oder Evaluation) einer Evaluation vorgenommen, spricht man auch von einer Meta-Evaluation (vgl. ausführlicher in Kap. 6.2). Selbstverständlich liesse sich dies auf der Ebene der Konstruktion dritter und weiterer Ordnungen weiterführen. Wir folgen hier aber der Auffassung von Michael Scriven, der zu Frage, wer die Meta-Evaluation evaluiert, folgendes festhält: «No infinite regress is generated because investi-

gation shows it usually doesn't pay after the first metalevel on most projects and after the second on any» (Scriven 1991, 230).

Von Bedeutung ist hier der Umstand, dass wir, wollen wir die Qualität einer Evaluation untersuchen, grundsätzlich zwei mögliche Bewertungsgrundlagen kennen. Während sich über die Ziele der Evaluation kaum Studien übergreifende Aussagen machen lassen, da diese fallspezifisch ausgehandelt werden, konzentrieren wir uns im Folgenden auf die äusseren Bewertungsgrundlagen einer Evaluation, also die Evaluationstheorie und Evaluationsmethodologie.

5 Standards der Evaluation

5.1 Entstehung der Standards

Im Zuge der Professionalisierung⁶ der Evaluation sind Ende der Siebzigerjahre in den USA verschiedene Initiativen ergriffen worden, um mittels eines Sets von Kriterien die vielfältigen Qualitätsdimensionen von Evaluationen zu erfassen. Diese Kriterien sollten uns in der Lage versetzen, die Qualität einer Evaluation hinsichtlich einer allgemeinen Evaluationstheorie und Evaluationsmethodologie – also im Hinblick auf die oben eingeführten äusseren Bewertungsgrundlagen – einzuschätzen. Die mehr oder weniger parallel verlaufenden Bemühungen, ein solches Kriterien-set zu formulieren, bezogen sich einerseits auf verschiedene Gegenstandsbereiche der Evaluation und andererseits auf unterschiedliche erkenntnistheoretische Grundprämissen der jeweiligen Urheber. Neben verschiedenen Kriterienrastern, die durch eine beschränkte erkenntnistheoretische Ausrichtung geprägt waren, publizierte das 'Joint Committee on Standards for Educational Evaluation' im Jahre 1981 die 'Standards for Evaluation of Educational Programs, Projects and Materials' (Joint Committee 1981), die hinsichtlich ihrer methodologischen Ausrichtung einen vergleichsweise offenen Ansatz verfolgten. In den frühen Achtzigerjahren wurde in der US-amerikanischen Evaluationsgemeinde die Frage, ob auch qualitative Methoden und nicht dem klassischen, (quasi-)experimentell-orientierten Evaluationsverständnis verbundene Ansätze zulässig seien, heftig und lange Zeit diskutiert. Durch ihre erkenntnistheoretische wie methodische Offenheit haben die erwähnten 'Standards' des 'Joint Committee' die spätere Entwicklung vorweggenommen, die mit sich brachte, dass heute die Daseinsberechtigung alternativer Zugangsweisen nicht (oder kaum) mehr in Frage gestellt wird. Die 'Standards' richten sich in erster Linie an Personen, die mit Evaluationen im Bereich von Bildung und Erziehung befasst sind. Auch die im genannten

‘Joint Committee’ vertretenen Organisationen stammen grösstenteils aus diesen Bereichen.

Diese ‘Standards’ haben sich in der amerikanischen Evaluationspraxis stark verbreitet, und zwar immer mehr auch in Themenfeldern, die weit über das ursprünglich beabsichtigte Anwendungsfeld der Bildung und Erziehung hinausreichten. Im Jahre 1994 präsentierte das ‘Joint Committee’ – welches inzwischen um Organisationen ergänzt worden war, die nicht ausschliesslich im Bildungsbereich tätig waren (vgl. die Übersicht in Widmer/Beywl 2000, 250) – eine überarbeitete Fassung der ‘Standards’, nun unter dem Titel: ‘The Program Evaluation Standards’ (Joint Committee 1994). Der Bezug zu Bildung und Erziehung wurde lediglich noch im Untertitel der Publikation erwähnt (‘How to Assess Evaluations of Educational Programs’). Der in dieser Titelformulierung vollzogene Wandel spiegelt (meines Erachtens etwas zögerlich) die in der Anwendung bereits etablierte Praxis, die ‘Standards’ vermehrt auch ausserhalb von Bildung und Erziehung einzusetzen.⁷

Im deutschen Sprachraum blieben die ‘Standards’ des ‘Joint Committee’ lange Zeit wenig beachtet (für eine frühe Ausnahme vgl. Beywl 1988, 113-123), auch wenn die Evaluation an und für sich in den Achtzigerjahren bereits einen kleineren Entwicklungsschub erlebte. Aber erst Mitte der Neunzigerjahre begannen sich die ‘Standards’ auch immer mehr im deutschsprachigen Raum zu verbreiten (Widmer 1996a-h, Widmer/Rothmayr/Serdült 1996). Um den Zugang der deutschsprachigen Evaluation zu den ‘Standards’ zu fördern, brachte man im Jahre 1999 eine deutsche Übersetzung auf den Markt (Joint Committee 1999). Inzwischen werden die ‘Standards’ auch im deutschsprachigen Raum recht häufig rezipiert, was sich etwa daran zeigt, dass die deutsche Übersetzung der ‘Standards’ bereits nach kurzer Zeit in zweiter Auflage erschienen ist (Joint Committee 2000; vgl. auch Beywl/Taut 2000).⁸

Im Rahmen ihrer Bemühungen um eine Professionalisierung der Evaluation in der Schweiz hat die Schweizerische Evaluationsgesellschaft (SEVAL) eine Arbeitsgruppe⁹ mit dem Auftrag betraut, Evaluationsstandards für die Schweiz zu entwickeln. Nachdem sich die Arbeitsgruppe vorerst intensiv mit verschiedensten bereits bestehenden Dokumenten (vgl. dazu auch die Hinweise in Beywl/Widmer 2000) auseinander gesetzt hatte, beschloss sie, ausgehend von den ‘Standards’ des ‘Joint Committee’, schweizerische Evaluationsstandards zu erarbeiten. Dabei folgte die Arbeitsgruppe der Grundstruktur der Standards, nahm jedoch an den einzelnen Standards gewisse Änderungen vor. Das resultierende Papier (Widmer/Landert/Bach-

mann 1999) wurde im Frühling 1999 von der Mitgliederversammlung der SEVAL mit grossem Interesse zur Kenntnis genommen. Nach einer weiteren Überarbeitung genehmigte die Mitgliederversammlung der SEVAL im Frühling 2001 die so entstandenen Evaluationsstandards als SEVAL-Standards (Widmer/Landert/Bachmann 2000). Damit hat die SEVAL als erste europäische Evaluationsgesellschaft eigene Evaluationsstandards etabliert und damit eine Vorreiterrolle übernommen (zu den Entwicklungen in anderen Ländern vgl. unten, Kap. 6.6).

5.2 Struktur und Inhalt der Standards

Die SEVAL-Standards basieren auf der Grundannahme, dass eine Evaluation gleichzeitig nützlich, durchführbar, korrekt und genau sein muss, um die Anforderungen zu erfüllen, die an sie gestellt werden. Eine gute Evaluation soll also die vier Eigenschaften Nützlichkeit, Durchführbarkeit, Korrektheit und Genauigkeit aufweisen. Um diese vier Konzepte fassbarer zu machen, gliedern sich die SEVAL-Standards in insgesamt 27 Einzelstandards, welche den vier Konzepten zugeordnet sind. Im Einzelnen lauten die Standards wie folgt (Widmer/Landert/Bachmann 2000):

N Nützlichkeit

Die Nützlichkeitsstandards stellen sicher, dass sich eine Evaluation an den Informationsbedürfnissen der vorgesehenen Evaluationsnutzerinnen und -nutzer ausrichtet.

N1 Ermittlung der Beteiligten & Betroffenen

Die an einer Evaluation beteiligten und die von ihr betroffenen Personen werden ermittelt, damit ihre Interessen und Bedürfnisse berücksichtigt werden können.

N2 Klärung der Evaluationsziele

Alle an einer Evaluation mitwirkenden Personen sind dafür besorgt, dass die Ziele der Evaluation für alle Beteiligten & Betroffenen geklärt sind.

N3 Glaubwürdigkeit

Wer Evaluationen durchführt, ist sowohl vertrauenswürdig als auch kompetent, damit die Evaluationsergebnisse ein Höchstmass an Glaubwürdigkeit und Akzeptanz erzielen.

N4 Umfang und Auswahl der Informationen

Der Umfang und die Auswahl der erhobenen Informationen ermöglichen die Behandlung sachdienlicher Fragen zum Evaluationsgegen-

stand und berücksichtigen gleichzeitig die Interessen und Bedürfnisse des Auftraggebers und anderer Beteiligter & Betroffener.

- N5 **Transparenz der Bewertung**
Die Perspektiven und Gedankengänge, auf denen die Interpretation der Ergebnisse beruht, werden so beschrieben, dass die Grundlagen der Werturteile klar ersichtlich sind.
- N6 **Vollständigkeit und Klarheit des Berichts**
Evaluationsberichte beschreiben den Evaluationsgegenstand einschliesslich seines Kontextes ebenso wie die Ziele, die Fragestellungen, die Verfahren und Befunde der Evaluation, damit die wesentlichen Informationen zur Verfügung stehen und leicht verstanden werden können.
- N7 **Rechtzeitigkeit der Berichterstattung**
Wichtige Zwischenergebnisse und Schlussberichte werden den vorgesehenen Nutzerinnen und Nutzern so zur Kenntnis gebracht, dass sie diese rechtzeitig verwenden können.
- N8 **Wirkung der Evaluation**
Planung, Durchführung und Darstellung einer Evaluation ermuntern die Beteiligten & Betroffenen dazu, dem Evaluationsprozess zu folgen und die Evaluation zu nutzen.

D Durchführbarkeit

Die Durchführbarkeitsstandards stellen sicher, dass eine Evaluation realistisch, gut durchdacht, diplomatisch und kostenbewusst ausgeführt wird.

- D1 **Praktikable Verfahren**
Evaluationsverfahren sind dazu geeignet, die benötigten Informationen zu beschaffen, ohne den Evaluationsgegenstand oder die Evaluation unnötig zu beeinträchtigen.
- D2 **Politische Tragfähigkeit**
Evaluationen werden unter Berücksichtigung der unterschiedlichen Positionen der verschiedenen Interessengruppen geplant und durchgeführt, um deren Kooperation zu erreichen und um mögliche Versuche irgendeiner dieser Gruppen zu vermeiden, die Evaluationsaktivitäten einzuschränken oder die Ergebnisse zu verzerren respektive zu missbrauchen.
- D3 **Kostenwirksamkeit**
Evaluationen bringen Informationen mit einem Wert hervor, der die eingesetzten Mittel rechtfertigt.

K Korrektheit

Die Korrektheitsstandards stellen sicher, dass eine Evaluation rechtlich und ethisch korrekt durchgeführt wird und dem Wohlergehen der Beteiligten & Betroffenen gebührend Aufmerksamkeit widmet.

K1 Formale Vereinbarungen

Die Pflichten der Vertragsparteien einer Evaluation (was, wie, von wem, wann getan werden soll) werden schriftlich festgehalten, damit die Parteien verpflichtet sind, alle Bedingungen dieser Vereinbarung zu erfüllen oder diese neu auszuhandeln.

K2 Schutz individueller Rechte

Evaluationen werden so geplant und durchgeführt, dass die Rechte und das Wohlergehen der Menschen respektiert und geschützt sind.

K3 Menschlich gestaltete Interaktion

Evaluationen sind so angelegt, dass die Kontakte zwischen den Beteiligten von gegenseitiger Achtung geprägt sind.

K4 Vollständige und faire Einschätzung

Evaluationen sind in der Überprüfung und in der Darstellung der Stärken und Schwächen des Evaluationsgegenstandes vollständig und fair, sodass die Stärken weiter ausgebaut und die Problemfelder behandelt werden können.

K5 Offenlegung der Ergebnisse

Die Vertragsparteien einer Evaluation stellen sicher, dass die Evaluationsergebnisse den betroffenen Personen zugänglich gemacht werden, ebenso wie all jenen, die einen ausgewiesenen Anspruch darauf haben.

K6 Deklaration von Interessenkonflikten

Interessenkonflikte werden offen und aufrichtig behandelt, damit sie die Evaluationsverfahren und -ergebnisse möglichst wenig beeinträchtigen.

G Genauigkeit

Die Genauigkeitsstandards stellen sicher, dass eine Evaluation gültige und verwendbare Informationen hervorbringt und vermittelt.

G1 Dokumentation des Evaluationsgegenstandes

Der Evaluationsgegenstand wird klar und genau beschrieben und dokumentiert, so dass er eindeutig identifiziert werden kann.

G2 Kontextanalyse

Die Einflüsse des Kontextes auf den Evaluationsgegenstand werden identifiziert.

- G3 Beschreibung von Zielen und Vorgehen
Ziele, Fragen und Vorgehen der Evaluation werden ausreichend genau dokumentiert und beschrieben, sodass sie identifiziert und beurteilt werden können.
- G4 Verlässliche Informationsquellen
Die in einer Evaluation genutzten Informationsquellen sind hinreichend genau beschrieben, damit die Angemessenheit der Informationen eingeschätzt werden kann.
- G5 Valide und reliable Informationen
Die Verfahren zur Gewinnung von Informationen werden so gewählt oder entwickelt und dann eingesetzt, dass Gültigkeit und Zuverlässigkeit der gewonnenen Interpretationen für den gegebenen Zweck sichergestellt sind.
- G6 Systematische Informationsüberprüfung
Die in einer Evaluation gesammelten, aufbereiteten und präsentierten Informationen werden systematisch auf Fehler überprüft.
- G7 Analyse qualitativer und quantitativer Informationen
Qualitative und quantitative Informationen einer Evaluation werden angemessen und systematisch analysiert, damit die gestellten Fragen durch die Evaluation effektiv beantwortet werden.
- G8 Begründete Schlussfolgerungen
Die in einer Evaluation gezogenen Folgerungen werden ausdrücklich begründet, damit die Beteiligten & Betroffenen diese nachvollziehen und einschätzen können.
- G9 Unparteiische Berichterstattung
Die Berichterstattung wird gegenüber Verzerrungen durch persönliche Gefühle und Vorlieben irgendeiner Evaluationspartei geschützt, so dass Evaluationsberichte die Ergebnisse fair wiedergeben.
- G10 Meta-Evaluation
Die Evaluation selbst wird anhand der vorliegenden oder anderer wichtiger Standards evaluiert, sodass die Durchführung entsprechend angeleitet werden kann und damit die Beteiligten und die Betroffenen bei Abschluss einer Evaluation deren Stärken und Schwächen einschätzen können.

Die SEVAL-Standards umfassen neben den obenaufgeführten Standards im engeren Sinne verschiedene weitere Informationen, wie eine allgemeine Einführung mit Angaben zu den Zielen, dem Anwendungsbereich, den Adressaten der Standards, Erläuterungen zu jedem einzelnen Standard und

verschiedene Begleitmaterialien (funktionale Übersicht, Vorgehen bei der Erarbeitung usw.)¹⁰.

Wie der obigen Darstellung der SEVAL-Standards entnommen werden kann, ist die Zahl der Einzelstandards in den vier Gruppen unterschiedlich. Zur Durchführbarkeit werden lediglich drei, zur Genauigkeit insgesamt zehn Standards aufgeführt. Dies ist jedoch keineswegs in dem Sinne zu verstehen, dass damit eine *Gewichtung* unter den vier Gruppen beabsichtigt ist. Vielmehr wird auf eine Gewichtung der Standardgruppen ebenso wie der Einzelstandards verzichtet, weil die Bedeutung eines Standards nur im konkreten Einzelfall festgelegt werden kann und deshalb eine generell Gültigkeit beanspruchende Gewichtung nicht angemessen wäre. Die Frage der Gewichtung ist besonders deshalb von Bedeutung, weil die SEVAL-Standards teilweise *konkurrierende Ansprüche* formulieren und so die Evaluatorin in der Praxis recht häufig mit der Frage konfrontiert ist, welche Standards gegenüber anderen Vorrang genießen sollten. Die fehlende innere Konsistenz der SEVAL-Standards ist aber keineswegs als Schwäche einzuschätzen, wird dadurch doch gerade das in Kap. 3) beschriebene Spannungsfeld abgebildet, in dem sich die Evaluation zu bewegen hat.

Die SEVAL-Standards formulieren *Maximalansprüche*. Sie sind keine Minimalstandards, die festlegen, was absolute Pflicht ist, sondern formulieren, was eine gute Evaluation anstreben sollte. In der Praxis wird es auch kaum je möglich sein, alle 27 Einzelstandards vollumfänglich zu erfüllen. Trotzdem sollten alle Beteiligten – nicht nur die Evaluatorin – anstreben, die SEVAL-Standards so weit wie möglich zu berücksichtigen.

Der Evaluation steht mit den SEVAL-Standards ein Instrument zur Verfügung, das die Qualitätsansprüche in der Evaluation auf präzise Weise und ausreichend differenziert umschreibt und damit den Ansprüchen einer Konzeptspezifikation ebenso wie der Vielseitigkeit der Evaluation gebührend Rechnung trägt.

6 Anwendung der Standards in der Praxis

6.1 Vorbemerkungen

Die Anwendungsmöglichkeiten der SEVAL-Standards in der Evaluationspraxis sind vielfältig. Im Folgenden soll der Versuch unternommen werden, Einblick in die verschiedenen Anwendungsformen zu geben und diese anhand von Beispielen zu konkretisieren. Selbstverständlich ist damit nicht eine vollständige Darstellung beabsichtigt.

Zuvor ist aber die Frage der generellen Anwendbarkeit der 'Standards' zu klären, präziser gefasst also auf die Frage einzugehen, ob die ursprünglich aus den USA stammenden und für die Bereiche Bildung und Erziehung entwickelten 'Standards' in anderen Themenfeldern und im deutschsprachigen Umfeld einsetzbar sind. Da diese Frage schon an anderer Stelle behandelt wurde (Widmer/Beywl 2000), kann ich mich hier auf eine knappe Darstellung beschränken. Aufgrund einer ausführlichen Argumentation wird dort im Grundsatz Folgendes festgestellt: «Zusammenfassend empfehlen wir aufgrund der bisherigen Erfahrungen die Anwendung der 'Standards' auf Evaluationen in Deutschland, Österreich und der Schweiz, und zwar auch dann, wenn Evaluationen ausserhalb des Bereiches Bildung und Erziehung durchgeführt werden» (Widmer/Beywl 2000, 257). Die Diskussion zeigt aber auch deutlich, dass die 'Standards' nicht unreflektiert übernommen werden sollten, sondern dass es angezeigt ist, die Angemessenheit der 'Standards' zu überprüfen und je nach Ergebnis gewisse Anpassungen vorzunehmen. Diese Anpassungsleistung bewegt sich jedoch gemäss dem bisherigen Erfahrungsstand in einem eng beschränkten Rahmen. Grundlegende Änderungen sind kaum zu erwarten. Im Rahmen der Arbeiten an den SEVAL-Standards wurden die für die allgemeine Verwendung in der Schweiz angemessen erscheinenden Anpassungen vorgenommen. Dies schliesst aber nicht aus, dass die SEVAL-Standards in der konkreten Anwendung auf die spezifische Situation angepasst werden.

6.2 Meta-Evaluation

Der Begriff der Meta-Evaluation, der bereits oben kurz angesprochen wurde (vgl. Kap. 4; vgl. dazu auch den Standard G10 'Meta-Evaluation'), bedarf einer etwas eingehenderen Erläuterung. Insbesondere ist auf die Differenzen zu anderen Ansätzen zu verweisen, die ebenfalls auf bereits bestehende Evaluationsstudien rekurrieren.

Neben der *Meta-Evaluation* sind dies einerseits die *Evaluationssynthese* und andererseits die *Meta-Analyse* (vgl. dazu und zum Folgenden Tab. 3). Diese beiden Instrumente sind auf die inhaltlichen Ergebnisse der zugrunde liegenden Evaluationsstudien ausgerichtet. Dies unterscheidet sie grundlegend von der Meta-Evaluation, die eine Bewertung einer oder mehrerer Evaluationen vornimmt. Für die beiden substantiell-orientierten Ansätze werden als Untersuchungsbasis Evaluationsstudien benötigt, die zur inhaltlich interessierenden Thematik Aussagen machen. Dabei ist der Fokus der Meta-Analyse, die quantitativ ausgerichtet ist, typischerweise deutlich enger als bei der Evaluationssynthese. Die Meta-Analyse benötigt zwingend

eine ausreichende Zahl bestehender Evaluationsstudien, die zu einer spezifischen Fragestellung quantitative Aussagen machen. Auch beim qualitativen Ansatz der Evaluationssynthese ist eine thematische Bündelung der Ausgangsstudien erforderlich, wobei diese aber bei weitem nicht so eng gefasst sein muss.

Instrument	Umschreibung
Evaluationssynthese	Inhaltliche Synthese verschiedener Evaluationsstudien (zumeist qualitativ) (Globalevaluation / Querschnittsanalyse)
Meta-Analyse	Quantitative Integration der Ergebnisse verschiedener Evaluationsstudien ('research synthesis')
Meta-Evaluation	Evaluation von Evaluation(en): Systematische Bewertung der Qualität einer oder mehrerer Evaluationsstudien

Tabelle 3: Übersicht über Evaluationsinstrumente dritter Ordnung

Die Meta-Evaluation hingegen hat eine deutlich andere Zielsetzung. Sie verfolgt die Absicht, in einer systematischen Weise den Wert und die Güte einer Evaluation zu bestimmen oder, mit anderen Worten, die Qualität einer Evaluation einzuschätzen. Um eine solche Bewertung vornehmen zu können, werden Bewertungskriterien benötigt. Wie in Kapitel 4 ausgeführt, sind dafür einerseits die Evaluationsziele zu berücksichtigen (innere Bewertungsgrundlagen) andererseits aber auch die Evaluationstheorie und die Evaluationsmethodologie (äussere Bewertungsgrundlagen). Zur präziseren Fassung der äusseren Bewertungsgrundlagen lassen sich Evaluationsstandards einsetzen.

Meta-Evaluationen lassen sich, wie eine Evaluation, in ganz unterschiedlicher Weise ausgestalten. Meta-Evaluationen können Selbst- oder Fremdevaluationen sein, sie können intern oder extern durchgeführt werden und formative oder summative Funktionen erfüllen (vgl. dazu Kap. 2.2, bes. Tab. 1).

Ein Beispiel für eine externe 'Fremd-Meta-Evaluation' mit summativem Charakter bildet eine Studie, die im Auftrag des Schweizerischen Nationalfonds (SNF) in der ersten Hälfte der 90er-Jahre erarbeitet wurde (Widmer 1996a-h). Darin werden zehn schweizerische Evaluationsstudien in detaillierten qualitativen Fallstudien *ex post* einer Bewertung unterzogen. Die Studie diente unter anderem dazu, die Qualität der schweizerischen Evaluationspraxis zu erfassen. Sie kommt zum Schluss, dass die untersuchten Evaluationen hinsichtlich Nützlichkeit, Durchführbarkeit und Korrektheit allgemein betrachtet eine hohe Qualität erreichen, dass hingegen bei der

Genauigkeit noch Verbesserungsmöglichkeiten bestehen. Eine weitere, ebenfalls vom SNF finanzierte Studie ging mit einer analogen Vorgehensweise der Frage nach, ob Evaluationen mit geringem Umfang und kurzer Untersuchungsdauer (so genannte Kurzevaluationen) gute Evaluationen sein können (Widmer/Rothmayr/Serdült 1996). Dabei wurden 15 Evaluationsstudien unterschiedlichen Umfangs im Hinblick auf ihre Qualität miteinander verglichen. Die Meta-Evaluation zeigt unter anderem, dass Kurzevaluationen unter bestimmten Bedingungen durchaus auch gute Evaluationen sein können, wobei die Risiken, dass sie in irgendeiner Form misslingen, deutlich höher sind, als dies bei umfangreicheren Evaluationsstudien der Fall ist.

Eine österreichische Untersuchung, die eine Kombination aus einer Evaluationssynthese und einer Meta-Evaluation darstellt, geht der Frage nach, wie Nachhaltigkeit in 43 Evaluationsstudien aus dem deutschsprachigen Raum evaluiert wurde (Kuffner 2000). Für die Bewertung der Qualität der Evaluationen setzte die Autorin die ‚Standards‘ des Joint Committee ein. Ein Beispiel für eine Meta-Evaluation, die sich mit einer einzelnen Evaluation befasst, bildet die Studie von Nideröst (2001). Er untersuchte (unter anderem mittels der SEVAL-Standards), weshalb die ausgewählte Evaluation als erfolgreich erscheint.

Im Rahmen einer Studie zur Evaluationspraxis beim Bund wurde anhand der SEVAL-Standards ebenfalls eine Meta-Evaluation durchgeführt. Dabei wurde die Qualität dreier ausgewählter Evaluationsstudien mit den SEVAL-Standards überprüft (Widmer/Rüegg/Neuenschwander 2001).

Zurzeit lässt das Bundesamt für Energie die Qualität seiner Evaluationen im Rahmen einer Meta-Evaluation anhand der SEVAL-Standards überprüfen. Die Resultate dieser Untersuchung stehen zwar zurzeit noch nicht zur Verfügung, sollten aber demnächst vorliegen.

6.3 Ratgeber für die Evaluationspraxis

Die SEVAL-Standards dienen jedoch nicht nur der Bewertung einer Evaluation, sondern sind primär darauf ausgerichtet, die Evaluatorin bei der Planung und Umsetzung eines Evaluationsprojektes zu unterstützen. In diesem beratenden und nicht kontrollierenden Sinne können die SEVAL-Standards auch als Leitlinie in der Evaluationspraxis dienen. Damit sind sie in der Lage, einen Gestaltungsbeitrag zuhanden der Evaluationspraxis zu leisten.

Seitens der Evaluationspraktiker (Evaluatorin, Auftraggeber, Adressaten) wird an den SEVAL-Standards zum Teil bemängelt, dass es schwierig sei,

sich in ihnen zurechtzufinden. Dies wird zumeist mit der hohen Zahl an Einzelstandards in Verbindung gebracht. Mittels der in den SEVAL-Standards enthaltenen funktionalen Übersicht (Widmer/Landert/Bachmann 2000, 14-16) lässt sich diese Schwierigkeit jedoch leicht überwinden. Die funktionale Übersicht ordnet nämlich jeweils den hauptsächlichen evaluativen Tätigkeiten ein bestimmtes Set von Einzelstandards zu, die für die spezifische Tätigkeit von besonders hoher Relevanz sind. Die funktionale Übersicht stellt das, die praktische Anwendung der SEVAL-Standards erleichternde Orientierungswissen bereit. Nachstehende Tabelle stellt diese Angaben in der Form einer Funktion-Standard-Matrix zusammen.

Standard Funktion	N 1	N 2	N 3	N 4	N 5	N 6	N 7	N 8	D 1	D 2	D 3	K 1	K 2	K 3	K 4	K 5	K 6	G 1	G 2	G 3	G 4	G 5	G 6	G 7	G 8	G 9	G 10
Entscheidung zur Evaluation																											
Definition des Problems																											
Planung der Evaluation																											
Informationserhebung																											
Informationsauswertung																											
Berichterstattung																											
Budget der Evaluation																											
Evaluationsertrag																											
Evaluationsmanagement																											
Persönliche Ausstattung																											

Tabelle 4: Funktionale Übersicht zu den SEVAL-Standards

Diese Matrix erlaubt es, die jeweils besonders relevanten Einzelstandards zu eruieren und daraus Checklisten zu generieren, die dem Bedürfnis nach Übersichtlichkeit ausreichend Rechnung tragen. Dieser vereinfachende Checklisten-Ansatz hat sich auch in äusserst kleinen, von Laien durchgeführten Selbstevaluationsprojekten als ein äusserst praktisches und nützliches Hilfsinstrument mit hoher Akzeptanz erwiesen. Die Checklisten erlauben es den Evaluierenden, die Selbstreflexion zur eigenen Tätigkeit gezielt zu systematisieren.

Der Checklisten-Ansatz hat aber auch einen bedeutenden Nachteil, der an dieser Stelle nicht unerwähnt bleiben sollte. In den 'Standards' wird ausdrücklich davor gewarnt, sich strikte auf die jeweils als relevant deklarierten Standards zu beschränken (Joint Committee 2000, 13). Man läuft nämlich so Gefahr, dass im konkreten Einzelfall fälschlicherweise durchaus relevante

Einzelstandards unbeachtet bleiben. In Anbetracht der Vorteile ist der Checklisten-Ansatz aber trotzdem zu empfehlen. Um die genannte Gefahr zu minimieren, empfiehlt sich eine regelmässige Überprüfung der Angemessenheit der eingesetzten Checklisten.

6.4 Qualitätssicherung

In der täglichen Evaluationspraxis zeigen sich immer wieder Qualitätsprobleme. Beteiligte & Betroffene vertreten nicht selten die Auffassung, dass eine bestimmte Evaluationsstudie qualitativ unbefriedigend ausgefallen sei. Auch wenn hinter derartigen Vermutungen oft ganz andere Motive stecken, ist es nicht von der Hand zu weisen, dass Evaluationen nicht immer den Anforderungen genügen. Ursachen für allfällige Qualitätsmängel im Evaluationsbereich sind unter anderem folgende:

- Der Markt der Evaluation hat in der Schweiz deutlich expandiert. Die Nachfrage nach dieser wissenschaftlichen Dienstleistung ist äusserst stark gestiegen. Das Angebot ist dieser Entwicklung zwar gefolgt, wobei aber nicht selten Evaluationsanbieter am Markt auftreten, denen die nötigen Grundlagen und Erfahrungen fehlen.
- Im Gegensatz zu anderen, bereits ex ante präzise zu umschreibenden Leistungen sind durch soziale Interaktionen geprägte Dienstleistungen – wie Bildung, soziale Arbeit, Beratung und auch Evaluation – nicht leicht zu fassen. Dies erschwert es dem Auftraggeber, sich im Markt zu orientieren.

Die SEVAL-Standards erlauben es dem Auftraggeber¹¹, die Evaluation fachlich fundiert zu bewerten. Die Sicherung eines hohen Qualitätsstandards ist besonders dort von grossem Interesse, wo häufig Evaluationen in Auftrag gegeben oder eigentliche Evaluationsprogramme durchgeführt werden.

Die SEVAL-Standards können aber auch dazu verwendet werden, einen Konflikt zwischen Auftraggeber und Evaluatorin zu bewältigen. So habe ich vor einiger Zeit eine Anfrage einer Organisation erhalten, ob er eine Evaluationsstudie im Hinblick auf ihre Qualität begutachten könne. Die Organisation war mit der Evaluation äusserst unzufrieden und erwog, der Evaluatorin das vereinbarte Honorar wegen Nichterfüllens des Vertrages vorzuenthalten. Mittels der Evaluationsstandards des Joint Committee begutachtete ich die Evaluation. Das abschliessende Gutachten wies auf verschiedene, zum Teil schwer wiegende Mängel der Evaluation hin. Da sich die 'Standards' aber nicht auf die Leistung der Evaluatorin beschränken, sondern die Evaluation als Ganzes betrachten, enthielt das Gutachten auch verschiede-

ne Hinweise auf Versäumnisse des Auftraggebers bei der Auftragsvergabe und der Projektbegleitung. Unter diesen Umständen hat sich die Organisation aufgrund des Gutachtens dazu entschlossen, das Honorar an die Evaluatorin auszuzahlen.

Die Motive zur Begutachtung einer Evaluationsstudie können aber auch ganz anders gelagert sein. So lässt etwa das schweizerische Bundesamt für Gesundheit regelmässig Evaluationsofferten sowie Zwischen- und Schlussberichte zu externen Evaluationen durch einen unabhängigen Sachverständigen begutachten. Dies geschieht mit dem Ziel, eine möglichst fundierte Offertenauswahl treffen respektive Qualitätsmängel in den Berichten erkennen zu können. Weiter wird natürlich damit auch die Absicht verfolgt, präventive Wirkungen zu erzielen. Auch in diesem Kontext haben sich die SEVAL-Standards, welche ich bei derartigen Begutachtungen regelmässig einsetzte¹², als äusserst hilfreich erwiesen.

Die Eidgenössische Finanzkontrolle (EFK), traditionell im Revisionsbereich tätig, führt vermehrt so genannte Wirtschaftlichkeitsprüfungen durch. Um in diesem Bereich die Qualität der Abklärungen zu sichern und Hinweise auf Weiterbildungsbedürfnisse der Mitarbeitenden zu erhalten, prüfen externe Berater die Qualität der intern durchgeführten Wirtschaftlichkeitsprüfungen. Die Berater setzen in diesem Rahmen ebenfalls die SEVAL-Standards ein.

Aber auch andere Bundesstellen wie die Direktion für Entwicklung und Zusammenarbeit (DEZA), der Schweizerische Wissenschafts- und Technologierat (SWTR), das Bundesamt für Justiz (BJ) oder das Bundesamt für Landwirtschaft (BLW) verwenden für die Qualitätssicherung bei Evaluationen, die sie durchführen oder durchführen lassen, Evaluationsstandards.

Ein letztes Beispiel für den Einsatz der SEVAL-Standards zur Qualitätssicherung bildet eine Meta-Evaluation zu drei externen Evaluationen von Reformprojekten im Rahmen der Verwaltungsreform wif! (Wirkungsorientierte Führung der Verwaltung des Kantons Zürich), die sich an den Prinzipien des «New Public Management» orientiert. Im Hinblick auf weitere, in Zukunft durchzuführende externe Evaluationen im selben Kontext wurde ich damit beauftragt, drei bereits durchgeführte respektive laufende Projektevaluationen anhand der SEVAL-Standards zu bewerten. Daraus wurden Empfehlungen für die zukünftige Durchführung von Evaluationen im Rahmen der Verwaltungsreform abgeleitet.

6.5 Didaktisches Hilfsmittel in Aus- und Weiterbildung

Im Rahmen von universitären Lehrveranstaltungen im Evaluationsbereich lassen sich die Evaluationsstandards als didaktisches Instrument einsetzen. Die SEVAL-Standards eignen sich dazu in idealer Weise. Als wertvoll erwiesen sich in diesem Zusammenhang auch die in der Buchfassung (Joint Committee 2000) enthaltenen ergänzenden Materialien, besonders auch die zu jedem Standard vorliegenden Anschauungsbeispiele. Eine andere Möglichkeit, die SEVAL-Standards im Unterricht zu verwenden, besteht in Fallstudien, bei denen die Studierenden die Aufgabe erhalten, eine bestehende Evaluationsstudie im Lichte der SEVAL-Standards kritisch zu hinterfragen. Auch bereits bestehende Meta-Evaluationen (siehe Kap. 6.2), in deren Rahmen die 'Standards' eingesetzt wurden, eignen sich hervorragend als Anschauungsmaterial für die Lehre. Im projektorientierten Studium dienen die SEVAL-Standards den Studierenden – ähnlich, wie dies oben (Kap. 6.3) für die Evaluationspraxis geschildert wurde – als Ratgeber für die Projektarbeit. Die Studierenden erhalten in diesem Rahmen jeweils projektphasenspezifische Aufträge zur Bewertung der eigenen Projektarbeit anhand der SEVAL-Standards.

Aber auch in der Weiterbildung haben sich die SEVAL-Standards als äusserst hilfreich erwiesen. Teilnehmende, die eine sozialwissenschaftliche Grundausbildung absolviert haben, jedoch über kein evaluationsspezifisches Knowhow verfügen, sind anhand der SEVAL-Standards in der Lage, die Spezifika der Evaluation in Relation zu den Sozialwissenschaften zu erkennen. Aber auch für Personen, die über keine einschlägigen Vorkenntnisse verfügen, jedoch Selbstevaluationen durchführen werden, erweisen sich die SEVAL-Standards als hilfreich. In diesem Kontext ist vor allem der oben geschilderte Checklisten-Ansatz auf grosses Interesse gestossen (vgl. Kap. 6.3).

6.6 Adaption durch Fachgesellschaften

Die 'American Evaluation Association' (AEA), weltweit die grösste Evaluationsvereinigung, setzt die 'Program Evaluation Standards' als Ergänzung zu den von der Gesellschaft selbst erarbeiteten 'Guiding Principles for Evaluators' ein (AEA 1995). Auch die 'Australasian Evaluation Society' (AES) und weitere Fachgesellschaften verwenden die 'Standards' des 'Joint Committee' (vgl. dazu Beywl/Widmer 2000).

Auf die Arbeiten an Evaluationsstandards im Rahmen der SEVAL wurde bereits in Kap. 5.1 hingewiesen. Die Deutsche Gesellschaft für Evaluation ist zurzeit ebenfalls daran, ein eigenes Set von Evaluationsstandards zu erar-

beiten. Sie orientiert sich dabei an den SEVAL-Standards (Beywl/Taut 2000: 367). Auch andere europäische Evaluationsgesellschaften haben es sich zu ihrem Ziel gemacht, Evaluationsstandards zu erarbeiten, so etwa die Société Française de l'Évaluation (siehe Perret/Barbier 2000), die UK Evaluation Society oder die neulich gegründete Österreichische Gesellschaft für Evaluation.

6.7 Standards in Rechtstexten und Handbüchern

Im Zuge der zunehmenden Verbreitung von Evaluationsstandards erscheinen diese auch immer häufiger in Texten mit rechtlicher Verbindlichkeit. Dabei stehen entsprechende Festlegungen in Ausschreibungen und Evaluationskontrakten im Vordergrund. Allmählich werden die Bezugnahmen auf Evaluationsstandards fester Vertragsbestandteil, was sich etwa darin zeigt, dass regelmässige Auftraggeber externer Evaluationen entsprechende Klauseln in ihre Mustertexte für Evaluationsverträge aufnehmen. Damit entwickeln sich die Evaluationsstandards – im Gegensatz zur Situation in den Achtzigern und der ersten Hälfte der Neunzigerjahre, als die Evaluationsstandards in der deutschsprachigen Evaluationspraxis nur selten überhaupt bekannt waren – sukzessive zu einem allgemein perzipierten und auch akzeptierten Referenzpunkt der Evaluationspraxis.

Dies zeigt sich auch darin, dass zunehmend in Handbüchern und Manualen diverser Projekte, Programme und Organisationen auf Evaluationsstandards Bezug genommen wird. Diese Dokumente haben zwar im Regelfall keinen Rechtscharakter, leisten aber einen bedeutenden Beitrag zur Verbreitung von Evaluationsstandards, besonders auch bei den entsprechenden Projektverantwortlichen, die häufig im Evaluationsbereich Auftraggeberfunktionen wahrnehmen. Beispiele dieser Art bilden das Handbuch zur bereits oben angesprochenen Verwaltungsreform wif! im Kanton Zürich (wif!-Stab 2000, Kapitel 3, 10–20) oder ein Expertenbericht zur Forschung und Entwicklung an den Pädagogischen Hochschulen (Kyburz-Grabner/Trachsler/Zutavern 2000, 20–22).

Die Europäische Kommission (DG Budget, Evaluation Network) hat ihre Anstrengungen zur Qualitätssicherung im Evaluationsbereich verstärkt. Bei der Erarbeitung von Leitlinien für eine gute Evaluationspraxis stützt sie sich ebenfalls auf die SEVAL-Standards ab.

Sukzessive scheint sich diese Entwicklung auch in verbindlichen rechtlichen Erlassen niederzuschlagen. Im Moment ist es zwar noch zu früh, dies als allgemeine Tendenz darzustellen. Aber es gibt Beispiele, die in diese Richtung deuten. So enthält das Evaluationsreglement der Universität

Zürich, das einen Bestandteil der Zürcher Gesetzessammlung bildet, die Vorschrift, dass die Evaluationsstelle die Einhaltung der SEVAL-Standards zu gewährleisten hat.

7 Schluss

Mit den SEVAL-Standards steht ein Instrument zur Verfügung, das die Qualitätsbewertung von Evaluationen ermöglicht. Mit ihren konzeptionellen und methodologischen Offenheit erlauben sie eine angemessene Bewertung der Qualität von Evaluationen. Besonders adäquat erscheinen die SEVAL-Standards, weil sie in der Lage sind, das vielfältige Spannungsfeld, in dem sich Evaluation bewegt, praxisnah abzubilden. Dank der Flexibilität der SEVAL-Standards eröffnet sich ein weites Feld von Anwendungsmöglichkeiten, sei dies in Meta-Evaluationen, als Ratgeber für die Evaluationspraxis, zur Qualitätssicherung, als didaktisches Hilfsmittel in Aus- und Weiterbildung, aber auch durch die in angepasster Form erfolgende Weiterverwendung durch Fachgesellschaften, in Rechtstexten oder Handbüchern.

Trotz dieser generell sehr positiven Einschätzung sind zum Schluss auch drei eher warnende Hinweise anzuführen, welche die Grenzen der SEVAL-Standards aufzeigen sollen:

- Die SEVAL-Standards formulieren Qualitätsansprüche, die an die Evaluation gestellt werden. Die Angemessenheit in der konkreten Anwendung ist aber sorgfältig zu beobachten, da möglicherweise Änderungen oder Anpassungen erforderlich sind. In jedem Fall repräsentieren die Evaluationsstandards nur die äusseren Bewertungsgrundlagen; die inneren Bewertungsgrundlagen können nur in Kenntnis der spezifischen Evaluationsstudie bestimmt werden.
- In letzter Zeit sind in verschiedenen Ländern Debatten über die Zertifizierung und Akkreditierung im Evaluationsbereich geführt worden. Im amerikanischen Bundesstaat Louisiana bestehen Zertifizierungskriterien für Evaluatoreninnen im Bildungsbereich (Louisiana State Department of Education 1996). In der 'Canadian Evaluation Society'/'Société canadienne d'évaluation' (CES/SCE¹³) und der 'American Evaluation Association' (AEA; Altschuld 1999a&b, Jones/Worthen 1999, Bickman 1999, Smith 1999 und Worthen 1999) laufen zurzeit Debatten zur Zertifizierung, wobei zum Teil heftige Widerstände angemeldet werden. In der Schweiz arbeitet zurzeit eine 'Groupe de réflexion meta-évaluation' (2000) daran, Standards zur Akkreditierung von Schulen im Bereich der Selbstevaluation zu formulieren. Ohne auf die Diskussion um Akkreditierung und Zertifizierung einzutreten, ist hier zu betonen, dass die SEVAL-Standards

(wie jene des ‚Joint Committee‘) dafür absolut ungeeignet sind. Erstens handelt es sich um Maximalstandards, die sich nicht dazu eignen, ein obligatorisches Minimum zu definieren. Zweitens ist der Bezugspunkt der SEVAL-Standards nicht die Evaluatorin, wie dies für die Zertifizierung nötig wäre, oder eine Institution, wie für die Akkreditierung erforderlich, sondern die Evaluation.

- Die zukünftige Entwicklung der Evaluation im Hinblick auf Theorien, Konzepte, Ansätze, Methoden und Verfahren lässt sich nicht voraussehen. Deshalb ist darauf zu achten, dass die SEVAL-Standards nicht dazu missbraucht werden, Weiterentwicklungen des Evaluationsansatzes zu behindern. Vielmehr sollten die SEVAL-Standards einer regelmässigen Überprüfung unterzogen und bei Bedarf angepasst werden.

Anmerkungen

- 1 Informationen zur SEVAL finden sich auf deren Homepage (<http://www.seval.ch/>).
- 2 Im vorliegenden Text wird die Wortkombination Beteiligte & Betroffene in diesem Sinne als Terminus technicus verwendet. Um dies zu verdeutlichen, werden die beiden Begriffe mit einem & verbunden.
- 3 Weitere Hinweise finden sich unter <http://www.efqm.org>.
- 4 Ich verwende hier die Begriffe ‚outcome‘ und ‚impact‘ nicht so, wie sie teilweise in der schweizerischen Literatur verwendet werden (so Knoepfel et al. 1997, 100–115), sondern halte mich an das angelsächsische Begriffsverständnis, das in der internationalen Literatur weit verbreitet ist (vgl. etwa Patton 1997, 193–194).
- 5 Ich verwende hier und im Folgenden für die evaluierende Person die weibliche Form, während für alle anderen Akteure die männliche Form steht. Selbstverständlich sind in jedem Fall Personen beider Geschlechter gemeint.
- 6 Zum Professionalisierungsprozess vgl. das Fünfstufenmodell von Harold Wilensky (1964); bezogen auf die Evaluation die Diskussion in Altschuld 1999a: 483–486 sowie die dort angegebene Literatur.
- 7 Vor kurzem hat der ehemalige Vorsitzende des Joint Committee an den ‚Standards‘ eine Bewertung von allgemeinen Evaluationsmodellen vorgenommen (Stufflebeam 2001).
- 8 Trotz der recht weiten Beachtung, welche diese Publikation in Fachkreisen gefunden hat, zeigt der im Jahre 2000 erschienene Beitrag von Schiffler und Hübner, dass noch Lücken bestehen.
- 9 In dieser Arbeitsgruppe haben mitgewirkt: Paul Ammann und Nicole Bachmann, Gesundheits- und Fürsorgedirektion des Kantons Bern; Renate Clémenton, Bundesamt für Justiz; Valérie Eggimann-Lanz und Armin Kühne, Eidgenössische Finanzkontrolle; Frohmut W. Gerheuser, Büro für Politikberatung und Sozialforschung; Charles Landert, Landert Farago Davatz & Partner; Marlène Läubli Loud, Office fédéral de la santé publique; Tobias Ledergerber, Organisationsamt des Kantons Bern; Urs Moser, Kompetenzzentrum Bildungsevaluation & Leistungsmessung, Universität Zürich; Urs Schmidt, e&e Entwicklung & Evaluation im Sozialwesen; Christian Simon, Schweizerischer Wissenschafts- und Technologierat; Heini Sommer, Ecoplan; Patrick Trees, Staatskanzlei des Kantons Bern; Thomas Widmer, Institut für Politikwissenschaft, Universität Zürich. Der Autor dankt den Mitgliedern der Arbeitsgruppe für ihre hilfreichen Beiträge und Anregungen.

- 10 Die vollständigen SEVAL-Standards finden sich auf der Homepage der SEVAL (<http://www.seval.ch/>). Dort steht auch eine Fassung in französischer Sprache zur Verfügung. Eine italienische und eine englische Fassung der SEVAL-Standards sind vorgesehen.
- 11 Ob die Evaluation intern oder extern ausgeführt wird, spielt hier keine Rolle. Auch bei internen Evaluationen existiert üblicherweise ein Auftraggeber, auch wenn die Evaluatoren in diesem Fall nicht in einem rechtlichen Auftragsverhältnis, sondern in einem Angestelltenverhältnis steht.
- 12 So wurden die SEVAL-Standards auch vom DLR, Projektträger des Deutschen Bundesministeriums für Bildung und Forschung, zur Bewertung von Evaluationsofferten im medizinischen Bereich eingesetzt.
- 13 Vgl. die entsprechenden Texte auf der Homepage der CES/SCE: <http://www.evaluation-canada.ca/devprof.html#certification>

Literatur

- Altschuld, James W., 1999a, The Certification of Evaluators: Highlights from a Report Submitted to the Board of Directors of The American Evaluation Association. *The American Journal of Evaluation* 20(3), S. 481–493.
- Altschuld, James W., 1999b, The Case for a Voluntary System for Credentialing Evaluators. *The American Journal of Evaluation* 20(3), S. 507–517.
- American Evaluation Association, Task Force on Guiding Principles for Evaluators, 1995, Guiding Principles for Evaluators. *New Directions for Program Evaluation* 66, S. 19–26.
- Beck, Ulrich/Bonss, Wolfgang, 1989, Verwissenschaftlichung ohne Aufklärung. Zum Strukturwandel von Sozialwissenschaft und Praxis, in: Dies. (Hgg.), *Weder Sozialtechnologie noch Aufklärung?* Frankfurt, Suhrkamp, S. 7–45.
- Beyme, Klaus von, 1991, Theorie der Politik im 20. Jahrhundert, Frankfurt, Suhrkamp.
- Beywl, Wolfgang, 1988, Zur Weiterentwicklung der Evaluationsmethodologie, Frankfurt, Lang.
- Beywl, Wolfgang/Taut, Sandy, 2000, Standards: Aktuelle Strategie zur Qualitätsentwicklung in der Evaluation, *Vierteljahresshefte zur Wirtschaftsforschung* 3/2000, S. 358–370.
- Beywl, Wolfgang/Widmer, Thomas, 2000, Die ‚Standards‘ im Vergleich mit weiteren Regelwerken zur Qualität fachlicher Leistungserstellung, in: Joint Committee on Standards for Educational Evaluation (Hrsg.), *Handbuch der Evaluationsstandards*, Opladen, Leske + Budrich, S. 259–295.

- Bickman, Leonard, 1999, AEA, Bold or Timid? *The American Journal of Evaluation* 20(3), S. 519–520.
- Campbell, Donald T., 1984, Can We Be Scientific in Applied Social Science? *Evaluation Studies Review Annual*, 9, S. 26–48.
- Cronbach, Lee J., 1982, Designing Evaluations of Educational and Social Programs, San Francisco, Jossey-Bass.
- Groupe de réflexion meta-évaluation, 2000, Mindeststandards für die Selbstevaluation von Schulen, 3. Entwurf vom 27.1.2000, (mimeo).
- Habermas, Jürgen, 1968, Technik und Wissenschaft als Ideologie. Frankfurt, Suhrkamp.
- Habermas, Jürgen, 1981, Theorie des kommunikativen Handelns, 2 Bde, Frankfurt, Suhrkamp.
- Habermas, Jürgen/Luhmann, Niklas, 1971, Theorie der Gesellschaft oder Sozialtechnologie – Was leistet die Systemforschung? Frankfurt, Suhrkamp.
- Joint Committee on Standards for Educational Evaluation, 1981, Standards for Evaluations of Educational Programs, Projects, and Materials, New York: McGraw-Hill.
- Joint Committee on Standards for Educational Evaluation, 1994, The Program Evaluation Standards, Newbury Park, Sage.
- Joint Committee on Standards for Educational Evaluation (Hrsg.), 1999, Handbuch der Evaluationsstandards, Opladen, Leske + Budrich.
- Joint Committee on Standards for Educational Evaluation (Hrsg.), 2000, Handbuch der Evaluationsstandards, 2. Auflage, Opladen, Leske + Budrich.
- Jones, Steven C./Worthen, Blaine R., 1999, AEA Members' Opinions Concerning Evaluator Certification. *The American Journal of Evaluation* 20(3), S. 495–506.
- Klöti, Ulrich/Widmer, Thomas, 1997, Untersuchungsdesigns, in: Bussmann, Werner/Klöti, Ulrich/Knoepfel, Peter (Hrsg.), *Einführung in die Politikevaluation*, Basel, Helbing & Lichtenhahn, S. 185–213.
- Knoepfel, Peter/Varone, Frédéric/Bussmann, Werner/Mader, Luzius, 1997, Evaluationsgegenstände und Evaluationskriterien, in: Bussmann, Werner/Klöti, Ulrich/Knoepfel, Peter (Hrsg.), *Einführung in die Politikevaluation*, Basel, Helbing & Lichtenhahn, S. 78–118.
- Kuffner, Astrid, 2000, Evaluation von Nachhaltigkeitsaspekten – Nachhaltige Evaluation?, Diplomarbeit an der Formal- und Naturwissenschaftlichen Fakultät der Universität Wien, Juni 2000, (mimeo).
- Kyburz-Graber, Regula/Trachsler, Ernst/ Zutz, Michael, 2000, Wissenschaftliche Zentren an den künftigen Pädagogischen Hochschulen der Region EDK-Ost. Aufgaben Vernetzung, Umsetzung, Schlussbericht der Expertengruppe ‚Forschung und Entwicklung‘ EDK-Ost, Januar 2000.
- Louisiana State Department of Education, 1996, Certification Criteria for Education Program Evaluators, Baton Rouge, Louisiana State Department of Education.
- MacDonald, Barry, 1993, A Political Classification of Evaluation Studies in Education, in: Hammersley, Martyn (ed.), *Social Research*, London, Sage, S. 105–108.
- Niederöst, Bruno, 2001, Eine erfolgreiche Politikevaluation, Die Evaluation der Eidgenössischen Volkszählung und die Gründe für deren Umsetzung. Lizentiatsarbeit am Institut für Politikwissenschaft, Universität Bern, März 2001, (mimeo).
- Palumbo, Dennis J./Hallett, Michael A., 1993, Conflict Versus Consensus Models in Policy Evaluation and Implementation, *Evaluation and Program Planning*, 16, S. 11–23.
- Perret, Bernard/Barbier, Jean-Claude, 2000, Ethical Guidelines, Process and Product Quality Standards, What For?, Paper presented at the European Evaluation Society Conference, Lausanne, October, 12th – 14th 2000. (mimeo)
- Schiffler, Angelina/Hübner, Sigrid, 2000, Allgemeine Standards für die Evaluationspraxis, in: Hager, Willi/Patry, Jean-Luc/Brezing, Hermann (Hrsg.), *Handbuch Evaluation psychologischer Interventionsmassnahmen*, Bern, Huber, S 141–152.
- Scriven, Michael, 1991, Evaluation Thesaurus, Fourth Edition. Newbury Park, Sage.
- Smith, M.F., 1999, Should AEA Begin a Process for Restricting Membership in the Profession of Evaluation, *The American Journal of Evaluation* 20(3), S. 521–531.
- Stufflebeam, Daniel L., 2001, Evaluation Models, *New Directions for Evaluation*, 89.
- Widmer, Thomas, 1996a, Meta-Evaluation, Kriterien zur Bewertung von Evaluationen, Bern, Haupt.
- Widmer, Thomas, 1996b-h, Fallstudien zur Meta-Evaluation, 7 Bde, Reihe ‚Schlussberichte‘ des Nationalen Forschungsprogramms ‚Wirksamkeit staatlicher Massnahmen‘ (NFP 27). Bern, Schweizerischer Nationalfonds.
- Widmer, Thomas/Beywl, Wolfgang, 2000, Die Übertragbarkeit der Evaluationsstandards auf unterschiedliche Anwendungsfelder, in: Joint Committee on Standards for Educational Evaluation (Hrsg.), *Handbuch der Evaluationsstandards*, Opladen, Leske + Budrich, S. 243–257.
- Widmer, Thomas/Binder, Hans-Martin, 1997, Forschungsmethoden, in: Bussmann, Werner/Klöti, Ulrich/Knoepfel, Peter (Hrsg.), *Einführung in die Politikevaluation*, Basel, Helbing & Lichtenhahn, S. 214–55.

- Widmer, Thomas/Landert, Charles/Bachmann, Nicole, 1999, Von der Schweizerischen Evaluationsgesellschaft (SEVAL) empfohlene Evaluationsstandards, Bern/Genève, SEVAL.
- Widmer, Thomas/Rothmayr, Christine/Serdült, Uwe, 1996, Kurz und gut? Qualität und Effizienz von Kurzevaluationen, Zürich, Rüegger.
- Widmer, Thomas/Rüegg, Erwin/Neuenschwander, Peter, 2001, Stand und Aussichten der Evaluation beim Bund (EvalBund), Bern, Schweizerische Bundeskanzlei.
- wifl-Stab (Hgg.), Wirkungsorientierte Führung der Verwaltung des Kantons Zürich, Projekt-handbuch, Aktualisierung 2000, Zürich: wifl-Stab.
- Wilensky, Harold L., 1964, The Professionalization of Everyone, *The American Journal of Sociology* 70(2), S. 137–58.
- Worthen, Blaine R., 1999, Critical Challenges Confronting Certification of Evaluators, *The American Journal of Evaluation* 20(3), S. 533 – 55.

Résumé

L'évaluation de l'activité étatique est devenue aujourd'hui un instrument de planification et de contrôle indispensable. Il faut se demander de quelle manière peut être assurée la qualité des travaux d'évaluation. Le présent article étudie la notion de qualité des évaluations et son importance. Il expose les instruments et les processus mis au point par la Société suisse d'évaluation, qui reposent sur le principe selon lequel une évaluation doit être utile, praticable, correcte et précise.

