

# Anwendung der SEVAL-Standards: Ergebnisse aus vier Fallstudien

**Stefan Rieder**<sup>1</sup> | *Wie stelle ich die Qualität von Evaluationen sicher? Diese Frage stellen sich Auftragnehmer und Auftraggeber jeder Evaluation immer wieder neu. Die SEVAL-Standards sollen bei der Beantwortung der Frage eine Hilfestellung bieten. Doch wie bewähren sich diese Standards in der Praxis? Dieser Frage widmet sich die Arbeitsgruppe Evaluationsstandards der SEVAL. Als erstes Ergebnis ihrer Tätigkeit stellt die Arbeitsgruppe die Ergebnisse aus vier Fallstudien vor. Dabei zeigt sich unter anderem, dass die SEVAL-Standards primär implizit angewendet werden und die zur Verfügung stehenden Ressourcen dabei eine bedeutsame Rolle spielen.*

## Inhaltsübersicht

- 1 Einleitung
- 2 Bekanntheit und formale Anwendung der Standards
- 3 Inhaltliche Anwendung der Standards
- 4 Diskussion der Hypothesen
- 5 Folgerungen und Ausblick

### 1 Einleitung

Die Arbeitsgruppe Evaluationsstandards der SEVAL hat sich in ihrem Arbeitsprogramm 2008/2009<sup>2</sup> zum Ziel gesetzt, die expliziten und impliziten Anwendungen der SEVAL-Standards (Widmer et al. 2000) in der Praxis zu untersuchen. Ausgangspunkt bildete die Diskussion von Hypothesen zur Verwendung der Standards. Auf dieser Basis entschloss sich die Arbeitsgruppe, die Anwendung der Standards mittels vier Fallbeispielen zu untersuchen. Zu diesem Zweck wurden Untersuchungsfragen formuliert, welche die Erarbeitung und Auswertung der Fälle strukturiert haben. Hypothesen und Fragestellungen werden weiter hinten ausgeführt.

Die Auswahl der Fallbeispiele erfolgte pragmatisch mittels vier Kriterien: Erstens wurden Fälle gewählt, an denen mindestens ein Mitglied der Arbeitsgruppe beteiligt war. Dies stellte den Zugang zu den notwendigen Informationen sicher. Zweitens musste es sich um Evaluationen handeln, deren Ergebnisse seit mindestens einem Jahr publiziert waren. Damit ist eine gewisse Nachvollziehbarkeit der hier vorgestellten Analyse sichergestellt. Drittens sollten möglichst «typische» Evaluationen öffentlicher Politik analysiert werden. Untersuchungen, die sich auf einen Spezialbereich konzentrierten (z.B. Bewertung einer Organisation oder eines Instruments) oder bei denen ausschliesslich eine ganz spezifische Methode eingesetzt wurde (z.B. eine Netzwerkanalyse) wurden nicht berücksichtigt. Vier-

tens sollte jedes Fallbeispiel aus einem unterschiedlichen Politikbereich stammen.

Der vorliegende Werkstattbericht stellt die Ergebnisse aus den vier Fallbeispielen dar. Er basiert auf einer ausführlichen Beschreibung der Fälle, die durch jeweils zwei Mitglieder der Arbeitsgruppe erstellt wurden. Jeder Fall wurde an einer Sitzung der Arbeitsgruppe vorgestellt und mit den jeweiligen Auftraggebenden und Auftragnehmenden eingehend diskutiert. Schliesslich wurden die Fälle vergleichend ausgewertet, was Basis für die hier vorliegenden Ergebnisse bildete. Die vier Beispiele lassen sich wie folgt kurz beschreiben (die in Klammern angeführten Kürzel werden weiter hinten im Text zur Kennzeichnung der Fälle verwendet):

- *Evaluation der Schutzpolitik Trockenwiesen und -weiden (kurz TWW; Rieder et al. 2006)*: Die Schutzpolitik TWW hat zum Ziel, TWW-Flächen von nationaler Bedeutung nachhaltig zu sichern. Dazu können Bund und Kantone gestützt auf das Natur- und Heimatschutzgesetz NHG entweder Schutzverordnungen erlassen oder Finanzbeiträge entrichten, die an die Nutzungsaufgaben geknüpft sind. Die Evaluation untersuchte, wie diese Schutzpolitik bis 2007 umgesetzt wurde und welche Wirkungen erzielt werden konnten. Auftraggeber der Evaluation war das Bundesamt für Umwelt. Durchgeführt wurde die Studie von einem Konsortium von zwei privaten Büros.
- *Evaluation Kontrolle der Mehrwertsteuer (kurz MWST; EFK 2005)*: Die Eidgenössische Steuerverwaltung ist für die Kontrolle der Steuerpflichtigen im Bereich Mehrwertsteuer zuständig. Mittels Inspektionen und Revisionen wird die korrekte Abrechnung der Mehrwertsteuer durch Unternehmen geprüft. Die Eidgenössische Finanzkontrolle (EFK) untersuchte in ihrer Evaluation die Planung, die Durchführung und die Ergebnisse der Kontrollen durch die Eidgenössische Steuerverwaltung. Die EFK wählte den Evaluationsgegenstand selber aus, führte die Evaluation durch und stellte die Ergebnisse der Finanzdelegation der eidgenössischen Räte vor.
- *Evaluation Strategie Migration und Gesundheit (kurz M+G; Ledermann et al. 2006)*: Das Bundesamt für Gesundheit (BAG) verfügt über eine Strategie zur Verbesserung der Gesundheit von Personen mit Migrationshintergrund. Total wurden 2002-2006 16 Projekte zu Aus- und Weiterbildung, Information, Prävention, Gesundheitsförderung, Gesundheitsversorgung, Therapieangebote und Forschung umgesetzt. Damit sollten die Öffnung des Gesundheitswesens, die bessere Berücksichtigung migrationspezifischer Bedürfnisse und die Stärkung der Selbstverantwortung der Migrantinnen und Migranten gefördert werden. Die Evaluation wurde vom BAG in Auftrag gegeben und von

einer Arbeitsgemeinschaft mit vier Partnern, bestehend aus zwei privaten Büros sowie zwei Hochschulinstituten, umgesetzt.

- *Evaluation der Reglementierung des Arbeitsmarktes (kurz CEPP; CEPP 2008)*: Die paritätischen Kommissionen von Arbeitgebern und Arbeitnehmern sind für die Kontrolle der Einhaltung der Gesamtarbeitsverträge zuständig. Ein besonderer Fokus dieser Kontrollen liegt auf der Verhinderung von sogenannten «Dumping-Löhnen». Die Commission externe d'évaluation des politiques publiques (CEPP) des Kantons Genf prüfte in ihrer Evaluation die Umsetzung und die Wirkungen der Kontrollen des Arbeitsmarktes durch die Sozialpartner. Dies geschah sowohl bei Firmen, die den Gesamtarbeitsverträgen unterstellt sind, als auch bei Unternehmen ausserhalb der Gesamtarbeitsverträge. Die CEPP griff das Thema selber auf und führte die Untersuchung in Zusammenarbeit mit einem privaten Büro durch.

Die Auswahl der Fälle weist eine gewisse Verzerrung auf: Es handelt sich um Evaluationen, bei denen jeweils ein vergleichsweise hohes Budget zur Verfügung stand und die durchwegs von erfahrenen Auftraggebern wie Auftragnehmenden durchgeführt wurden. Wir werden bei den Schlussfolgerungen nochmals auf diesen Umstand zurückkommen.

In den folgenden Kapiteln werden die Ergebnisse der Fallstudien dargestellt. Zunächst wird darauf eingegangen, wie gut die Evaluationsstandards bekannt sind und inwieweit sie genutzt werden. Anschliessend wird detaillierter auf die einzelne inhaltliche Anwendung der Standards eingegangen. In Kapitel 4 werden die von der Arbeitsgruppe formulierten Hypothesen geprüft. Im letzten Kapitel finden sich einige vorläufige Schlussfolgerungen und ein Ausblick auf weitere Arbeiten der Arbeitsgruppe.

## **2 Bekanntheit und formale Anwendung der Standards**

Die Arbeitsgruppe hat als Ausgangspunkt der Analyse folgende Fragen gestellt:

- Sind die Standards bekannt und verständlich?
- Welche Qualitätssicherungssysteme (Standards, Instrumente) werden eingesetzt?
- Ist den Beteiligten bewusst, dass die Standards auf den konkreten Fall anzupassen beziehungsweise zu konkretisieren sind?
- Welche Bedeutung hatten die Standards bei der Konzeption und Durchführung der Evaluation? Werden sie explizit oder implizit angewendet?<sup>3</sup>
- Findet zwischen Auftraggebern und Auftragnehmenden eine Auseinandersetzung um die Standards statt?

Die Auswertung der vier Fälle erlaubt folgende Antworten auf die Fragen:

Die Standards sind in allen betrachteten Fällen sowohl bei Auftraggebenden als auch Auftragnehmenden bekannt. In drei von vier Fällen werden sie mehrheitlich implizit angewendet. Die Anwendung erfolgt schwergewichtig beim Start der Evaluation (allgemeine Referenz bei der Ausschreibung) und beim Abschluss der Untersuchung. In diesen beiden Phasen der Evaluation wurden Standards zum Teil systematisch (TWW), zum Teil punktuell in der Ausschreibung oder in der Metaevaluation (M+G) respektive in einem Debriefing (MWST) diskutiert. Im Falle der CEPP wurden die Standards nie explizit diskutiert. Allen Fällen ist gemeinsam, dass die Standards während der Durchführung der Evaluationen selten oder nie explizit angewendet wurden.

Eine explizite Selektion von Standards zu Beginn der Evaluation oder eine Konkretisierung von Standards fand mit einer Ausnahme (TWW) nicht statt. Im Nachhinein wäre dies vor allem im Fall M+G (Klärung der Verbindung von Selbstevaluationen mit der Drittevaluation) sowie im Fall der CEPP (politische Tragfähigkeit) wohl sehr nützlich gewesen. Im Fall M+G zeigten sich Schwierigkeiten bei der Verfügbarkeit der in Aussicht gestellten Informationen aus Selbstevaluationen (vgl. dazu auch Abschnitt 3.2). Im Fall CEPP wäre es in der Rückblende sinnvoll gewesen, über die Standards die politische Tragfähigkeit der Evaluation zu prüfen. Diese war offenbar nur teilweise gegeben, was mit ein Grund war, dass sich nach Abschluss der Arbeiten ein Konflikt um die Zuverlässigkeit der Ergebnisse entfaltete.

Von den betrachteten Auftraggebenden verfügen das Bundesamt für Gesundheit und die Eidgenössische Finanzkontrolle jeweils über ein eigenes System der Qualitätssicherung (QS). Grundsätzlich hat sich bei den Fallbeispielen gezeigt, dass keine Konkurrenz zwischen den internen QS-Systemen und den Standards besteht. Allerdings sind die Standards in unterschiedlichem Masse in die QS-Systeme eingeflossen: Beim BAG bildeten die Standards nach deren Publikation im Jahr 2000 explizit eine Grundlage bei der Überarbeitung der Qualitätsvorgaben. Bei dem von der EFK etablierten Verfahren sind die Standards nicht explizit eingeflossen.

### **3 Inhaltliche Anwendung der Standards**

In Bezug auf die inhaltliche Anwendung der Standards galt es, folgende Fragen zu prüfen:

- Wie werden die Standards im konkreten Fall in den verschiedenen Projektphasen angewandt? Welche Standards bereiten dabei Probleme und aus welchen Gründen?

- Welche Massnahmen treffen die Auftraggebenden und die Auftragnehmenden, um die Einhaltung der Standards sicherzustellen? Werden bestehende Hilfsmittel und Instrumente (vor allem SEVAL-Standards, Leitfaden für Wirkungsüberprüfungen beim Bund; Widmer 2005) genutzt?
- Könnte die Einhaltung der SEVAL-Standards durch eine explizite inhaltliche Anwendung verbessert werden?

### 3.1 Generelle inhaltliche Anwendung

Die Analyse der Fallbeispiele führte vier explizite Verwendungsarten der Standards zu Tage: Standards halfen *erstens* in einem Teil der Fälle mit, Vertrauen aufzubauen. Die Evaluationsstandards bieten eine anerkannte Referenzgrösse für Qualität und Vorgehensweise. Dies ist vor allem bei Auftragsverhältnissen von Bedeutung, bei denen sich Auftraggebende und Auftragnehmende noch nicht kennen. Standards dienten *zweitens* dazu, den Auftraggebenden die Auswahl von Offerten zu erleichtern. Sie halfen *drittens*, die sachgemässe Umsetzung der Evaluation zu prüfen. Schliesslich boten die Standards *viertens* in einzelnen Fallbeispielen Hilfestellung, um die Diskussion von heiklen Aspekten einer Evaluation zu erleichtern. Dazu gehört zum Beispiel die Publikation von Ergebnissen.

Bei den impliziten Nutzungsarten sind zunächst jene zwei Fälle zu erwähnen, bei denen interne Qualitätssicherungssysteme umgesetzt worden sind, die mit den Evaluationsstandards übereinstimmen. Dies gilt für die zwei Fälle MWST und M+G. Andere implizite Nutzungsarten wie beim Fallbeispiel CEPP (hier wurde bei der Machbarkeitsprüfung ein Teil der Standards angewendet) lassen sich nachträglich vermuten.

Trotz der beobachteten expliziten und impliziten Verwendungsarten muss festgehalten werden, dass die Standards die Planung, Durchführung und Verwendung der Evaluationsergebnisse in den vier Fällen keineswegs geprägt oder gar dominiert haben. Vielmehr entsteht der Eindruck, dass die genannten expliziten Nutzungsarten im Gesamtkontext der Evaluationen eine untergeordnete Rolle gespielt haben. Im vierten Fall (CEPP) sind die Standards gar nie explizit genutzt worden. Die implizite Nutzung lässt sich nur erahnen. Bei der impliziten Nutzung der Standards lässt sich zudem nicht eindeutig beurteilen, ob die beobachteten Verhaltensweisen nicht auch ohne die Standards eingetreten wären und die Nutzung der Standards damit überschätzt wird. In einem der Fälle (M+G) kamen Checklisten zum Einsatz, die explizit auf die SEVAL-Standards Bezug nehmen.

### 3.2 Spezifische inhaltliche Anwendung einzelner Standards

Für alle vier Fälle wurde einzeln geprüft, welche Standards explizit verwendet wurden und ob bei der Anwendung allenfalls Probleme auftraten. Wir gehen hier nicht auf alle geprüften Standards ein, sondern beschränken uns auf die Darstellung einiger exemplarischer Fälle. Wenden wir uns zunächst jener Gruppe von Standards zu, bei denen die Probleme in der Umsetzung am *grössten* ausgefallen sind:

- In zwei Fällen (TWW und CEPP) bildete die Verlässlichkeit und Validität der Informationen (Standards G4 und G5) einen Diskussionspunkt. Beim Fall TWW wurde vom Auftraggebenden bis zum Schluss ein Fragezeichen hinter die Zuverlässigkeit von Interviewergebnissen gesetzt. Weit zugespitzter sieht die Situation beim Fallbeispiel CEPP aus: Hier hat ein Teil der Stakeholder die Datenerhebung als falsch und unwissenschaftlich zurückgewiesen. Dies führte unmittelbar dazu, dass die politische Tragfähigkeit (Standard D2) der Ergebnisse nicht gegeben war. Von Seiten der Projektverantwortlichen wird allerdings festgehalten, dass die Anwendung von Standard D2 beim Start der Evaluation zu einer Blockade der Evaluation durch die Stakeholder hätte führen können, was kaum Ziel der Standards sein könne. Nicht ganz so ausgeprägt aber doch vorhanden waren Spannungen rund um Standard D2 (politische Tragfähigkeit) im Fallbeispiel TWW: Die Passivität und das geringe Interesse der Kantone haben der Verwendung der Ergebnisse der Evaluation enge Grenzen gesetzt.
- Im Falle MWST stellte die Einhaltung von Standard K5 (Offenlegung der Ergebnisse) ein Problem dar. Die Steuerverwaltung wehrte sich zunächst gegen die Publikation eines Teils der Ergebnisse. Mittels Gesprächen konnte erreicht werden, dass der Bericht dennoch ohne wesentliche Abstriche publiziert werden konnte. Allerdings gestaltete sich dieser Prozess sehr aufwendig.
- Im Fallbeispiel M+G lagen die Probleme primär bei der Deklaration von Interessenkonflikten (Standard K6) und der unparteiischen Berichterstattung (Standard G9): Offenbar haben Stakeholder versucht, auf die Ergebnisse des Berichts Einfluss zu nehmen. Gemeinsam mit dem Auftraggebenden konnten die Evaluatoren das Problem lösen. Möglicherweise hätte bei einer expliziten Diskussion der Interessenkonflikte bei Projektstart der Konflikt vermieden werden können.
- Ein zweites Problem im Rahmen des Fallbeispiels M+G bildete Standard D1 (praktikables Vorgehen). Zu Beginn der Evaluation waren den Evaluatoren Informationen aus Selbstevaluationen der zu untersuchenden Projekte in Aussicht gestellt worden. Es wurde aber unterlassen, die Durchführung der

Selbstevaluationen auch effektiv sicherzustellen. Deshalb fehlten die entsprechenden Daten teilweise. Im Rahmen von Standard D1 wäre zu Beginn der Evaluation intensiver zu prüfen gewesen, ob die avisierte Selbstevaluation auch tatsächlich praktikabel ist.

Neben den geschilderten Beispielen lässt sich eine zweite Gruppe von Standards benennen, deren Umsetzung sich als *teilweise problematisch* erwies. Der Standard Glaubwürdigkeit (Standard N3) wurde zwei Mal thematisiert: Im Fallbeispiel MWST galt die EFK für die betroffenen Stakeholder zunächst als thematisch zu unerfahren. Im Fallbeispiel CEPP brachte die Geschäftsprüfungskommission des Genfer Kantonsparlaments der Evaluation ein erhebliches Misstrauen entgegen. Die Gruppe der Standards unter dem Titel Nützlichkeit (N) ist im Kontext des Falles TWW zu thematisieren. Es stellte sich heraus, dass die Kantone die Ergebnisse in weit geringerem Masse nutzten, als dies anfänglich erwartet worden war. Das Interesse der Kantone an den Ergebnissen war bescheiden. Umfang und Auswahl der Informationen (Standard N4) waren im Beispiel MWST ein Thema: Die Perspektive der Unternehmen wurde explizit aus der Evaluation der Mehrwertsteuer ausgeklammert, was deren Aussagekraft schmälerte.

Abschliessend lässt sich festhalten, dass aus den vier untersuchten Fallbeispielen kein klarer Trend erkennbar ist: Die Mehrheit der Standards wird jeweils problemlos eingehalten. Einzelne Standards erweisen sich in der Umsetzung als teilweise, einige als sehr problematisch. Allerdings sind es immer unterschiedliche Standards, die Schwierigkeiten bereiten: In bestimmten Fällen sind Standards unproblematisch, in anderen wiederum schwierig zu erfüllen.

#### **4 Diskussion der Hypothesen**

Wie einleitend ausgeführt, hat die Arbeitsgruppe zu Beginn ihrer Tätigkeit sechs Hypothesen zur Nutzung von Standards formuliert (Hypothese 3 wird im Folgenden in drei Teile aufgeteilt). Auf Grund der Ergebnisse der vier Fallstudien werden die Hypothesen einer Überprüfung unterzogen. Daraus lassen sich erste Erkenntnisse gewinnen, die sich allerdings noch nicht verallgemeinern lassen. Vielmehr können die nachfolgenden Ergebnisse Ausgangspunkt für eine weiterführende Analyse bilden.

*H1 Auftraggebende und Auftragnehmende setzen sich heute nicht oder zu wenig intensiv und zu wenig bewusst mit den SEVAL-Standards auseinander. Die Standards werden in der Regel nicht explizit, sondern implizit über die Beachtung anderer Qualitätsstandard (z.B. wissenschaftliche Standards) angewendet.*

Dem ersten Teil der Hypothese kann in der Tendenz zugestimmt werden. Eine intensive Auseinandersetzung mit den Standards im Rahmen einer Evaluation bildet wohl eher die Ausnahme. Umgekehrt wäre es aber auch nicht richtig, von ei-

ner gänzlich fehlenden Anwendung zu sprechen. Die Anwendung ist momentan nicht intensiv. Zum Teil zutreffend ist der zweite Teil der Hypothese: In zwei der vier Fälle werden die Standards implizit über Qualitätssicherungsinstrumente berücksichtigt.

*H2 Die SEVAL-Standards werden von den Auftraggebenden unterschiedlich angewendet. Neben expliziten Bezügen (inkl. Prioritätensetzung) finden sich vor allem generelle Verweise auf die Standards. Auf Bundesebene besteht keine einheitliche Praxis.*

Dieser Hypothese kann insgesamt zugestimmt werden. Die Standards werden tatsächlich stark unterschiedlich angewendet. Die Anwendung reicht von der expliziten Einforderung einzelner Standards bei der Konzepterstellung über die implizite Anwendung in einem QS-System bis hin zur Nicht-Anwendung von Standards. Eine einheitliche Praxis ist nicht zu beobachten.

*H3a Die Auftragnehmenden setzen sich vor allem dann mit den SEVAL-Standards auseinander, wenn dies von den Auftraggebenden (z.B. in Pflichtenheften) verlangt wird.*

Dieser Teil der Hypothese kann als eher zutreffend gelten. Wenn der Auftraggebende die Standards explizit einfordert, findet eher eine Diskussion über die Standards statt.

*H3b Eine aktive und frühzeitige Auseinandersetzung mit den Standards findet bisher kaum statt.*

Diesem Teil von Hypothese 3 kann in der Tendenz ebenfalls zugestimmt werden. Nur in einem Fall wurde zu Beginn der Evaluation eine Priorisierung einzelner Standards vorgenommen. In der Regel ist dies unterblieben.

*H3c Die Standards haben sich auf Auftragnehmerseite noch nicht als Instrument zur Qualitätssicherung durchgesetzt.*

Diese Hypothese ist mehrheitlich zutreffend. In keinem der Fälle wurden Standards während der Umsetzung der Evaluation als Instrument einer systematischen Qualitätskontrolle eingesetzt.

*H4 Auftraggebende und Auftragnehmende passen die SEVAL-Standards dem konkreten Evaluationsprojekt ungenügend an. Damit bleibt in der Regel unklar, welche Standards wie erfüllt werden müssen und wie deren Erfüllung überprüft wird.*

Diese Hypothese erweist sich im Lichte der vier Fallstudien als insgesamt richtig. Nur in einem Fall wurden Standards priorisiert und angepasst.

*H5 Die SEVAL-Standards haben für Auftraggebende und Auftragnehmende im Evaluationsprozess eine unterschiedliche Bedeutung. Für Auftraggebende und Auftragnehmende sind im Evaluationsprozess jeweils unterschiedliche Standards besonders relevant.*



Die Ergebnisse zu H5 erlauben keine Beurteilung der Hypothese.

*H6 Insgesamt tragen die SEVAL-Standards heute in der Praxis noch kaum zur Qualitätssicherung und -förderung bei. Demzufolge ist der Beitrag der Standards zur Professionalisierung und zur Erhöhung der Glaubwürdigkeit, der Qualität und der Vertrauenswürdigkeit von Evaluationen bisher sehr begrenzt.*

Die Beurteilung von Hypothese 6 fällt ambivalent aus. Zwar kann gezeigt werden, dass ausgewählte Standards tatsächlich beachtet werden und Wirkungen zeigen. Allerdings sind diese Effekte nicht systematisch zu beobachten. Zudem gibt es Fälle, in denen die Wirksamkeit der Standards bei kritischer Betrachtungsweise gegen null tendiert.

## **5 Folgerungen und Ausblick**

Am Schluss des Beitrags sollen einige Schlussfolgerungen gezogen werden und ein Ausblick auf die weiteren Arbeiten der Arbeitsgruppe erfolgen. Zunächst zu den Schlussfolgerungen. Sechs scheinen uns von besonderer Bedeutung:

1. Die Evaluationsstandards der SEVAL sind den Auftragnehmenden wie auch den Auftraggebenden der untersuchten Fälle bekannt. Die Anwendung der Standards erfolgt primär implizit. Eine explizite Anwendung ist relativ selten.
2. Die Einhaltung von Standards ist (nicht nur aber wesentlich) von den verfügbaren Ressourcen für die Evaluation abhängig. Besonders die Standards der Gruppe G (Genauigkeit) sind davon betroffen. Wenn die Budgets knapp sind, sollte explizit über die Genauigkeitskriterien diskutiert werden. Allgemein formuliert: Die Ansprüche an die Reichweite und Genauigkeit der Ergebnisse müssen mit den vorhandenen Ressourcen übereinstimmen.
3. Eine Integration der Standards in ein internes QS-System kann als wirksamer Schritt zur Beachtung der Standards bezeichnet werden. Die Erfahrungen beim BAG und bei der EFK legen diesen Schluss nahe.
4. Eine Priorisierung der Standards zu Beginn der Evaluation ist auf jeden Fall lohnend. Damit steigen die Chancen, dass Konfliktpunkte früh erkannt und zumindest teilweise entschärft werden können. Eine Priorisierung der Standards zu Beginn der Evaluation stellt eine pragmatische Anwendung der Standards dar. Der entsprechenden Empfehlung in der Einleitung der SEVAL-Standards (Widmer et al. 2000: 1) kann uneingeschränkt zugestimmt werden.
5. Der Umgang mit dem Evaluationsstandard D2 (politische Tragfähigkeit) ist heikel. Einerseits kann der Einbezug der Stakeholder einen wichtigen Beitrag zur Nutzung der Evaluationsergebnisse leisten. Wenn die Einhaltung von Standard D2 zu Beginn der Evaluation zu einem Aushandlungsprozess mit Betrof-

fenen führt, besteht andererseits aber die Gefahr, dass bei Konflikten Evaluationen stark verzögert, eventuell inhaltlich verwässert oder gar blockiert werden. Die Ambivalenz dieses Standards hat sich insbesondere im Fallbeispiel CEPP gezeigt.

6. Offenbar wird am Anfang und am Schluss einer Evaluation öfter auf die Standards Bezug genommen. Während der Durchführung der Evaluation treten die Standards hingegen stark in den Hintergrund. Offen bleibt, ob eine intensivere explizite Verwendung der Standards in der Umsetzungsphase anzustreben ist. Die Erfahrungen aus den Fallbeispielen lassen kaum Schlüsse zu.

Eine Bemerkung ist an dieser Stelle anzubringen: In den vier untersuchten Fällen sind durchwegs Auftraggebende und Evaluatoreninnen/Evaluatoren aktiv geworden, die über eine vergleichsweise grosse Evaluationserfahrung verfügen. Ferner handelt es sich bei allen vier Fällen um Arbeiten mit einem vergleichsweise grossen Finanzvolumen, welches deutlich über 100'000 Franken lag. Diese Beispiele dürften somit kaum repräsentativ sein für die Mehrheit der Evaluationen, die in der Schweiz durchgeführt werden. Vielmehr dürfte es wohl zutreffen, dass oft mit weit weniger Mitteln und mit weniger Erfahrung evaluiert werden muss. Somit ist nicht von der Hand zu weisen, dass auf Grund der ermittelten Ergebnisse die Bedeutung der Standards in der Evaluationspraxis eher überschätzt denn unterschätzt wird.

Die Arbeitsgruppe Evaluationsstandards der SEVAL wird aufgrund der Ergebnisse ihre weiteren Aktivitäten planen und sich verstärkt jenen Aspekten zuwenden, welche nicht oder nur unzureichend beantwortet werden konnten. Dazu gehören die Fragen, wie die Anwendung der SEVAL-Standards erhöht werden könnte, welche Hilfsmittel sich dafür anbieten und welchen Beitrag die SEVAL leisten könnte. Die Arbeitsgruppe wird daher versuchen, die empirische Basis für die Beurteilung der Standards mit geeigneten Methoden zu vergrössern und ein entsprechendes Arbeitsprogramm für die nächsten zwei Jahre festlegen. Damit soll ein Beitrag zur Nutzung der Evaluationsstandards in Ausbildung und Praxis geleistet werden.

*Stefan Rieder, Dr. rer. pol., Interface Politikstudien Forschung Beratung, Luzern,  
E-Mail: rieder@interface-politikstudien.ch*

## Anmerkungen

- 1 Der Beitrag wurde von Stefan Rieder unter der Mitarbeit der Arbeitsgruppe Evaluationsstandards der SEVAL verfasst. Der Arbeitsgruppe Evaluationsstandards der SEVAL gehören gegenwärtig folgende Personen an: Stephan Hammer, INFRAS (Leiter der Arbeitsgruppe), Lars Balzer, Eidgenössisches Hochschulinstitut für Berufsbildung (EHB), Christoph Bättig, Parlamentarische Verwaltungskontrolle (PVK), Kurt Bisang, Bundesamt für Energie (BFE), Joan Ribi (selbstständige Beraterin), Stefan Rieder, Interface, Christian Rüefli, Büro Vatter, Emmanuel Sangra, Eidgenössische Finanzkontrolle (EFK).
- 2 [http://www.seval.ch/documents/Arbeitsgruppen/konzept\\_ag\\_standards\\_050608.pdf](http://www.seval.ch/documents/Arbeitsgruppen/konzept_ag_standards_050608.pdf)
- 3 Eine *explizite* Anwendung liegt dann vor, wenn Auftraggebende oder Auftragnehmende einen oder mehrere Standards benennen und festhalten, ob und wie sie einzuhalten sind. Als *implizite* Anwendung werden Fälle bezeichnet, in denen Standards eingehalten werden, ohne dass explizit auf sie Bezug genommen wird.

## Literatur

- Commission externe d'évaluation des politiques publiques CEPP, 2008, Evaluation de la politique de réglementation du marché du travail. Partie I, Contrôles effectués par les commissions paritaires, Genève, CEPP.
- Eidgenössische Finanzkontrolle EFK, 2005, Kontrolle der Mehrwertsteuer. Evaluation der Strategie, der Umsetzung und der Ergebnisse der Kontrollen bei den Steuerpflichtigen, Bern, EFK.
- Ledermann, Simone et al., 2006, Evaluation der Strategie «Migration und Gesundheit» 2002–2006, Bern, Bundesamt für Gesundheit.
- Rieder, Stefan et al., 2006, Erfolgskontrolle des Schutzes von Trockenwiesen und -weiden, Bericht der Nullmessung, Luzern und Bern: Interface/BUWAL.
- Widmer, Thomas/Landert, Charles/Bachmann, Nicole, 2000, Evaluations-Standards der Schweizerischen Evaluationsgesellschaft (SEVAL-Standards), <http://www.seval.ch/de/standards/index.cfm>.
- Widmer, Thomas, 2005, Leitfaden für Wirksamkeitsüberprüfungen beim Bund, Bern: Bundesamt für Justiz.

## Résumé

*Le groupe de travail Evaluation des standards a étudié sur la base de quatre cas si les Standards SEVAL ont fait leur preuve dans la pratique. Il a en particulier examiné si ces standards étaient connus et compréhensibles et de quelle manière ils ont été appliqués. Il ressort de l'étude qu'ils sont connus et qu'ils sont mis en pratique, souvent de manière implicite. Le respect des standards dépend pour une part non négligeable du budget à disposition pour l'évaluation: plus celui-ci est élevé, plus il est aisé de se conformer aux standards du groupe P (Précision). Les Standards SEVAL déploient leurs effets aussi du fait que différents mandants les ont déjà intégrés dans leur système d'assurance-qualité. Il est important dans la pratique de discuter et de hiérarchiser les standards avant de mener l'évaluation. Le standard de l'acceptabilité politique de l'évaluation s'est révélé particulièrement délicat à respecter.*