

# Evaluation

Fritz SAGER und Markus HINTERLEITNER

erschienen in SCHUBERT, Klaus, und Nils BANDELOW (Hg.). *Lehrbuch der Politikfeldanalyse*. München/Wien: De Gruyter Oldenbourg (3. Auflage), 2014, S. 437-462.

## 1. Begriff - Was ist Evaluation?

Politikevaluation ist die Bewertung öffentlicher Politik. Auf die Vielschichtigkeit des Begriffs Evaluation weist bereits sein französischer etymologischer Ursprung hin: „évaluer“ bedeutet ‚bewerten‘, ‚begutachten‘, ‚abschätzen‘, oder ‚taxieren‘. Für eine zweckdienliche Begriffsbestimmung kann Evaluation im Policy Cycle (siehe *Jann/Wegrich* in diesem Band) eingeordnet werden. Evaluation in diesem ursprünglichen und engen Verständnis schließt sich an Genese und Vollzug einer Politik an und debattiert deren Wirkungen. Ob man nun sämtliche Aktivitäten nach erfolgtem Vollzug bis zum Entscheid über Weiterführung, Neuausrichtung, oder gar Termination einer öffentlichen Politik unter dem Terminus Evaluation zusammenfasst bzw. nur die empirisch- wissenschaftliche Wirkungsanalyse als Evaluation begreift, zeigt weiteren begrifflichen Klärungsbedarf an. Dass mit dieser Einordnung in den Policy Cycle der Begriff Evaluation noch nicht annähernd und befriedigend bestimmt ist, zeigt nicht zuletzt die Tatsache, dass zwischen vorausschauenden (ex ante), begleitenden und nachträglichen (ex post) Evaluationen differenziert werden kann.

Eine in Stein gegossene Definition des Begriffs Evaluation scheint im Licht einer sich rasant verändernden und ausbreitenden Tätigkeit bzw. Disziplin als schwierig, wenn nicht gar nachteilig. Vielmehr gilt es, sich der Evaluation über Schlüsselbegriffe oder Komponenten zu nähern, die dieser Begriff umfasst. So sprechen bspw. *Widmer* und *deRocchi* (2012: 11-13) bewusst von einer Arbeitsdefinition, die auf für das Begriffsverständnis unerlässliche Elemente fokussiert. Auch der vorliegende Beitrag orientiert sich an dieser Praxis, indem eine *funktions- bzw. zweckgeleitete Begriffsbestimmung* vorgeschlagen wird, die der Frage nachgeht: Was bezweckt eine Evaluation und welche begrifflichen Komponenten ergeben sich aus der Funktion von Evaluation? Drei zusammenhängende Komponenten ragen hierbei heraus: *Bewertung/Beurteilung*, *Wissenschaftlichkeit* und *Dienstleistungscharakter*.

Die *Bewertung* und *Beurteilung* durch eine Evaluation weist zunächst auf den Evaluationsgegenstand bzw. das Evaluationsobjekt hin. Gegenstand einer Evaluation ist in der Regel eine öffentliche Politik, die einzelne Massnahmen genauso umfassen kann wie komplexe und weitreichende politische Programme. An dieser Stelle sei bereits erwähnt, dass nicht nur substanzielle Politiken, sondern auch institutionelle Politiken als Evaluationsgegenstand fungieren können (vgl. Abschnitt 4). Ausserdem ist zu betonen, dass die Bewertung einer öffentlichen Politik im Rahmen einer Evaluation nicht politisch motiviert, sondern möglichst wissenschaftlich und transparent zu erfolgen hat und sich an den Zielen ausrichtet, die die öffentliche Politik zu erreichen versucht

Mit einer wissenschaftlichen Evaluation nutzt die Bewertung das Renommee der Wissenschaft und muss sich daher auch an den dort üblichen Kriterien orientieren. Damit ist die *Wissenschaftlichkeit* einer Evaluation von unerlässlicher Bedeutung. Wissenschaftlichkeit bedeutet, dass die Evaluation intersubjektiv nachvollziehbar sowie reproduzierbar ist und nach allgemein anerkannten Standards systematisch und wissenschaftlich vorgegangen wird. Nur dies sichert ein Mindestmaß an Qualität, das auch im Hinblick auf die noch zu thematisierende (partei-)politische Instrumentalisierung von Evaluationen und Evaluationsergebnissen von verstärkter Bedeutung ist.

Schliesslich gilt es aber auch zu berücksichtigen, dass Evaluationen im politischen Prozess nicht aus eigenem Antrieb erfolgen bzw. einem Selbstzweck gehorchen, sondern von externen Akteuren in Auftrag gegeben werden. Daraus folgt, dass Evaluationen meist auch einen gewissen *Dienstleistungscharakter* aufweisen. Die Auftraggeber erwarten von der Evaluation einen gewissen Nutzen, wie die Beantwortung von Fragen, die Aufdeckung von Funktionsstörungen oder konkrete Verbesserungsvorschläge. Diese Ansprüche an die EvaluatorInnen müssen sowohl bei der Konzeption und der Durchführung der Evaluation, als auch bei der Kommunikation der Ergebnisse berücksichtigt werden.

Aus den drei Schlüsselkomponenten Bewertung/Beurteilung, Wissenschaftlichkeit und Dienstleistungscharakter geht zudem eindeutig hervor, was eine Evaluation *nicht* sein kann: Sie grenzt sich ab von der *Grundlagenforschung*, welche stärker auf Wissensgenerierung, Generalisierbarkeit der Ergebnisse und Theorieentwicklung, als auf Nutzengenerierung für den Auftraggeber und Situationsbezug ausgerichtet ist (Ritz 2003: 35, vgl. Tabelle 1).

Tabelle 1: Abgrenzung von Evaluation und Grundlagenforschung

	<b>Grundlagenforschung</b>	<b>Evaluation als angewandte Forschung</b>
Ziel und Zweck	Wissensgenerierung	Nutzung und Nützlichkeit
Fragestellung	Forschungsdebatte, Theorie	Auftraggeber und Anspruchsgruppen
Ergebnisse	Generalisierbarkeit, Theorieentwicklung	Situationsbezug, Falltreue
Bewertung	Forschungsdebatte, Theorie	Programmziele, normative Kriterien, Auftraggeber
Kausalität	Kein Unterschied: abhängige, unabhängige und intervenierende Variablen	
Methoden	Kein Unterschied: gesamtes Methodenspektrum	
Kontext	Prinzip der Unabhängigkeit	Politischer Druck, Zeit- und Dienstleistungsdruck

nach Ritz 2003:35

Auch unterscheidet sich die Evaluation von anderen Instrumenten der Wirkungsprüfung, wie dem wirtschaftlichen *Controlling* (permanentes und führungsorientiertes Überwachungs- und Steuerungssystem) und dem *Monitoring* (permanentes Überwachungssystem ohne Bewertungscharakter), wie aus Tabelle 2 hervorgeht.

Tabelle 2: Unterscheidung von Evaluation, Controlling und Monitoring

	<b>Evaluation</b>	<b>Controlling</b>	<b>Monitoring</b>
Definition	Evaluationen sind <i>gezielte und zeitlich begrenzte</i> Untersuchungen, die insbesondere das Verständnis und die Erklärung nicht nur der Wirkungen, sondern auch der Wirksamkeit von staatlichen Massnahmen zum Ziel haben.	Controlling ist eine unterstützende Führungstätigkeit und umfasst ein <i>permanentes System</i> für die Beobachtung und Beurteilung des gesamten Planungs- und Steuerungsprozesses in einem bestimmten Aufgabenbereich.	Monitoring ist die <i>routinemässige, permanente</i> und systematische Sammlung von vergleichbaren Daten.
Ziel und Zweck	Spezifische Fragen im Hinblick auf die Wirkungen von Massnahmen beantworten. Zweck kontextabhängig: Rechenschaft, Lernen usw.	Durch Informationen zur effektiven und effizienten Steuerung von Prozessen beitragen.	Veränderungen und/oder Trends bei der Umsetzung, beim Verhalten der Zielgruppen oder bei den Wirkungen feststellen.
Bewertung	Ja	Ja	Nein
Verhältnis zu anderen Instrumenten	Evaluation kann Monitoring-Daten und Resultate des Controlling berücksichtigen	Controlling kann Monitoring-Daten einbeziehen	Evaluation und Controlling können Hinweise für geeignete Monitoring-Indikatoren liefern

Interdepartementaler Kontaktgruppe «Wirkungsprüfungen» (IDEKOWI) 2004

Das bereits gesagte eröffnet einen recht klaren Blick auf die im Folgenden zu thematisierenden Etappen. Aus den oben genannten elementaren Komponenten der Evaluation ergibt sich eine natürliche Gliederung eines Beitrags über Evaluation: Eine dem Kriterium der Wissenschaftlichkeit gerecht werdende Evaluation stützt sich zunächst auf *theoretische Grundlagen*. Dementsprechend wird im folgenden Abschnitt ein Überblick zu Evaluationstheorien und –ansätzen gegeben. Da eine Evaluation in ihrem Kern eine Wirkungsanalyse ist, wird anschliessend ein theoretisch fundiertes *Wirkungsmodell* dargestellt, das die Wirkungsweise einer öffentlichen Politik darstellt, die Politik in ihre Bestandteile zerlegt, und Kriterien für die Bewertung letzterer bereitstellt. Im nächsten Abschnitt steht dann der Gegenstand der Evaluation nochmals im Vordergrund und es wird darauf eingegangen, inwiefern sich der Ablauf und das Vorgehen ändern, wenn nicht eine „klassische“ substantielle Politik, sondern eine *institutionelle Politik* evaluiert wird. Die letzten drei Abschnitte stehen dann im Zeichen der *Kontexteinbettung* von Evaluation. Die Verwendungsarten von Evaluation und deren Rolle und Bedeutung im politischen Prozess werden thematisiert. Ausserdem werden mit den Stichwörtern Evidence-based Policymaking (EBP) und Regulatory Impact Assessments (RIAs) zwei eng mit der Evaluation verbundene Themenbereiche diskutiert, deren

Kenntnis dazu dient, die Bedeutung von Evaluation für eine moderne Politikfeldanalyse erfassen zu können.

Die Gliederung unseres Beitrags streicht eine Tatsache hervor, die für das Verständnis von Evaluation von unerlässlicher Bedeutung ist: Eine Evaluation steht nicht frei im Raum, sondern ist auf zweifache Art *kontextsensitiv*. Einerseits muss bei der Evaluation der Kontext der öffentlichen Politik analysiert werden, um die Wirkungen selbiger überhaupt erst eindeutig erfassen und beurteilen zu können. Andererseits kann die Bedeutung von Evaluation im Rahmen der Politikfeldanalyse nur verstanden werden, wenn man den politischen und institutionellen Kontext thematisiert, in dem sie durchgeführt wird. Diesen beiden „kontextuellen Berührungspunkten“ wird im Folgenden durchgehend Rechnung getragen.

## 2. Entwicklung und Theoriebildung

Auch wenn die Beschäftigung mit Evaluation als Forschungsgegenstand in den Sozialwissenschaften als relativ junge Disziplin gilt, existiert gleichwohl eine lebendige Theoriedebatte. Allerdings wäre es übertrieben, von einer einheitlichen und konsolidierten Evaluationstheorie zu sprechen. Vielmehr ist die Systematisierung der theoretischen Beiträge, wie bspw. von *Guba* und *Lincoln* (1989) versucht, im Bereich der Evaluationsforschung schwierig. Gründe hierfür sind vor allem in der starken Heterogenität evaluationstheoretischer Vorstellungen, der begrifflichen Vielfalt und der primär amerikanischen Literatur zu sehen (*Ritz* 2003: 57). Dessen ungeachtet kann eine gewisse Systematisierung gelingen, wenn man den jeweiligen politischen und institutionellen Kontext, vor dem sich die Politikevaluation in den letzten Jahrzehnten entwickelt hat, betrachtet (vgl. *Derlien* 1997). Im Rahmen dieses Beitrags muss eine Beschränkung auf vereinzelt Beiträge zur Entwicklung und Theorie von Evaluation erfolgen. Es werden Autorinnen und Autoren herausgegriffen, die das zeitgenössische Verständnis als auch die Durchführung von Evaluation massgeblich beeinflusst haben.

Mit den USA als Vorreiter, entwickelte sich mit Beginn der Nachkriegszeit in Ländern der westlichen Hemisphäre eine Evaluationspraxis, die durch volle Staatskassen, Reformparteien und eine generelle Affinität zu den Sozialwissenschaften gekennzeichnet war. Das Interesse an der Funktions- und Wirkungsweise politischer Programme zur Zeit der „Experimenting Society“ (*Campbell* 1969) resultierte für die Evaluation in der Aufgabe, bestehende und neue Massnahmen und Programme effektiver zu machen (*Derlien* 1997: 7). Vor diesem Hintergrund wurde Evaluationsforschung als Teil der Grundlagenforschung verstanden und zum Zweck der Erkenntnisgewinnung durchgeführt. Im Zentrum stand die Entwicklung wissenschaftlicher Methoden, um Kausalzusammenhänge mit angemessenen Kriterien bestimmen zu können (*Ritz* 2003: 59-60). Auch wenn sich das theoretische Verständnis von Politikevaluationen seitdem beträchtlich weiterentwickelt hat, stammen aus dieser ersten Periode der Evaluationsforschung doch Erkenntnisse, die bis heute Gültigkeit haben. So betont *Suchman* (1967) die Unterscheidung von Programmfehler (heute Umsetzungsfehler), bei dem eine Aktivität falsch durchgeführt bzw. umgesetzt wird, und Theoriefehler, bei dem die an sich richtig ausgeführte Aktivität nicht zu den Programmzielen beiträgt. *Campbell* (1969) hat, basierend auf der realistischen Einsicht, dass der Bewertungsvorgang nie vollkommen ohne subjektive Werturteile auskommt, auf der transparenten Trennung von Fakten und Werturteilen in der Evaluation bestanden. In Anbetracht häufiger politischer Instrumentalisierungen von Evaluationen verliert diese Unterscheidung keineswegs an Aktualität.

Ende der 1970er Jahre erfuhr der politische und sozioökonomische Kontext in den meisten westlichen Ländern tiefgreifende Veränderungen, die sich auch auf das Verständnis von und die formulierten Ansprüche an die Evaluationstätigkeit auswirkten. Durch den „Neoliberal Turn“ (vgl. *Harvey* 2005) bestärkte konservative Regierungen (*Thatcherism, Reaganomics*) sahen in der Evaluation ein Mittel zur Durchsetzung effizienterer Programmgestaltung, um knapper werdende Ressourcen zu schonen. Auch wenn die Effektivitätssteigerung weiter ein Anliegen ist, ging es nun doch primär um die Kostenfrage. Hierbei ist das

Evaluationsinstrumentarium behilflich, indem es aufzeigt, wo politische Programme gekürzt werden können. Allgemein erfolgte eine Hinwendung zur Outputorientierung staatlichen Handelns. Von diesen Veränderungen blieben Evaluationstheorien nicht unberührt: Fortan wurde Evaluation zusehends als angewandte und nutzenorientierte Forschung konzipiert (z. B. Weiss 1974, Wholey 1983). Als Vorreiterin fungierte Carol H. Weiss, welche „die bisher zu wenig berücksichtigten Schwierigkeiten und Interessenkonflikte im Rahmen der politischen und organisatorischen Kontexte des Evaluationsobjekts“ (Ritz 2003: 66) erkannte. Zentral ist dabei die Einsicht, dass Politik und Wissenschaft unterschiedlichen Logiken folgen. Will die Evaluation gehört werden (vgl. hierzu Abschnitt 5), muss sie in gewissem Maß die Sprache der Politik sprechen. Diese Tatsache schwingt noch heute in einem Evaluationsverständnis mit, das auf Anwendungsorientierung und Dienstleistungscharakter ausgerichtet ist.

Seit Mitte der 1990er Jahre betonen Wissenschaftler wie Patton (1997) noch stärker die Anwendungsorientierung von Evaluationen, um den tatsächlichen Nutzen in Form von Programmverbesserungen zu forcieren. Hierfür ist es Patton zufolge unerlässlich, Evaluationen so zu gestalten, dass sie von Entscheidungsträgern direkt genutzt werden können. Dieses Verständnis von Nutzung und Nützlichkeit einer Evaluation wird unter den Schlagwörtern *Utilization*, *Usability*, und *Influence* zusammengefasst. In diesem Zusammenhang steht auch der Ansatz der „Realistic Evaluation“ nach Pawson und Tilley (1997). Die Kontextbedingtheit von objektiv erkennbaren Phänomenen impliziert, dass politische Programme und Instrumente je nach Kontext unterschiedliche Wirkungen zeitigen. Wirkungen dürfen nicht isoliert betrachtet, sondern müssen in den jeweiligen programmspezifischen Kontext eingebettet werden, um zu realistischen Aussagen zu gelangen.

Aus diesen Meilensteinen der Evaluationstheorieentwicklung resultiert ein aktuelles Bild von Evaluation, das einerseits wissenschaftsbasiert und theoriegeleitet ist, aber auch die Anwendungsorientierung und die Kontextabhängigkeit nicht vernachlässigt. Dieses Bild bestimmt den im nächsten Abschnitt dargelegten Ansatz der theoriegeleiteten Evaluation und das hierauf aufbauende Wirkungsmodell zur Analyse öffentlicher Politik.

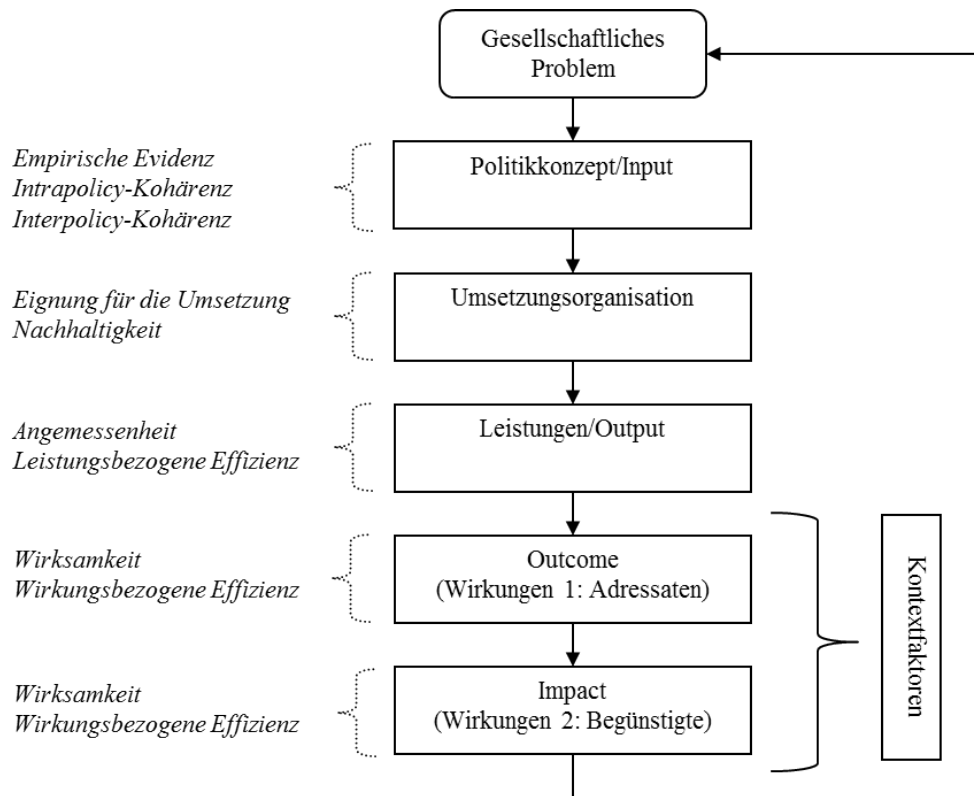
### **3. Evaluationsansatz, Evaluationsgegenstände und Wirkungsmodell**

Um nun den einleitend erläuterten Schlüsselkomponenten eines modernen Evaluationsverständnisses zu entsprechen und gleichzeitig die bisher aufgezeigten theoretischen Schwerpunkte zu integrieren, gilt es, ein wissenschaftlich fundiertes, theoriegeleitetes und kontextsensitives Evaluationsmodell zu erarbeiten, das der Evaluation konzeptuell zu Grunde liegt und in unterschiedlichen Politikfeldern zur Analyse öffentlicher Politik eingesetzt werden kann. Hierfür wird der theoriegeleitete Evaluationsansatz nach Rossi (Rossi et al. 1999) um den kontextsensitiven realistischen Evaluationsansatz nach Pawson und Tilley (1997) ergänzt.

Die *theoriegeleitete Evaluation* geht zunächst von der impliziten Programmtheorie der Akteure aus: „Every program embodies a conception of the structure, functions, and procedures appropriate to attain its goals. This conception constitutes the ‚logic‘ or plan of the program, which we have called program theory. The program theory explains why the program does what it does and provides the rationale for expecting that doing things that way will achieve the desired results“ (Rossi et al. 1999: 156). Hier setzt die Evaluation an, indem sie diese Programmtheorie einer wissenschaftlichen Untersuchung zugänglich macht. Zu diesem Zweck werden Bewertungskriterien, Indikatoren und Erhebungsmethoden benötigt. Dabei gilt es zu beachten, dass die der Evaluation zugrunde gelegte Programmtheorie, genau wie das Programm bzw. die öffentliche Politik selbst, disziplinübergreifend angelegt ist. Konkret bedeutet dies zumeist, dass eine Evaluation eine politikwissenschaftliche Perspektive verfolgt, aber gleichzeitig politikbereichsspezifische Besonderheiten nicht vernachlässigt werden. Auf diese Weise stellt der theoriegeleitete Evaluationsansatz sicher, dass Lernprozesse ausgelöst und soziale Interventionen optimiert werden können.

Der Evaluationsansatz der theoriegeleiteten Evaluation fokussiert auf die in der Programmtheorie implizierten Kausalketten, welche die Teile eines Programms mit den angestrebten Programmzielen verbinden. Die oben bereits thematisierten Kontextfaktoren, die den Erfolg oder Misserfolg eines Programms massgeblich beeinflussen, werden von diesem Ansatz allerdings eher vernachlässigt. Um diese Schwäche zu beheben, können Teile des theoriegeleiteten Evaluationsansatzes mit Ideen des Ansatzes der „Realistic Evaluation“ nach Pawson und Tilley (1997) kombiniert werden. Dieser Ansatz ist in seiner zugrunde liegenden Forschungslogik zwischen Positivismus und Konstruktivismus angesiedelt, indem er die Kontextbedingtheit von objektiv erkennbaren Phänomenen betont. Übertragen auf die Evaluation einer öffentlichen Politik bedeutet dies, dass letztere je nach Kontext andere Wirkungen zeitigt. Eine Evaluation kann also nur dann tatsächliche Wirkungen analysieren, wenn Wechselbeziehungen zwischen öffentlicher Politik und Kontext miteinbezogen werden. Um diese Wechselbeziehungen abzubilden, werden sogenannte *Context-Mechanism-Outcome*-Konfigurationen (kurz: CMO-Konfigurationen) modelliert. Bestimmte, von einer öffentlichen Politik ausgelöste Mechanismen (M) werden in spezifischen lokalen, historischen, soziokulturellen und institutionellen Umfeldern (C) wirksam und führen zu entsprechenden Outcomes (O). Aufgabe der Evaluation ist es nun zu untersuchen, welche CMO-Konfigurationen jeweils vorherrschen und wann, wie und wo eine öffentliche Politik wirkt (vgl. Befani et al. 2007, Sager/Andereggen 2012).

Die Elemente einer öffentlichen Politik, die von einer Politikevaluation untersucht und abgebildet werden, können mittels unterschiedlicher theoretischer Modelle strukturiert werden. Hierfür empfiehlt sich eine Typisierung nach dem Policy Cycle (vgl. Knoepfel et al. 2001; Jann/Wegrich in diesem Band). Dieser Heuristik liegt die Vorstellung zugrunde, dass eine öffentliche Politik verschiedene Stufen durchläuft. Generell können *drei Phasen* unterschieden werden: Politikformulierung, Implementierung respektive Umsetzung und Wirkungsentfaltung. In der Phase der *Politikformulierung* wird zunächst das Problem definiert und sodann ein Programm konzipiert, welches das Problem adressiert und möglichst beseitigt. Für die *Implementierung* werden Umsetzungsstrukturen konstituiert und Umsetzungsprozesse definiert, damit die tatsächliche Leistung erbracht werden kann. Mit diesen Leistungen soll eine *Wirkung*, genauer eine Verhaltensänderung bei den Adressaten des Programms hervorgerufen werden. Die Verhaltensänderungen der Adressaten tragen schliesslich in gewissem Grad zur Lösung des in der Phase der Politikformulierung definierten Problems bei. Am Ende dieses Zyklus stellt sich zumeist die Frage, wie das existierende Programm verbessert bzw. an veränderte kontextuelle Bedingungen angepasst werden kann, womit eine neue Phase der Politikformulierung beginnt und der Zyklus von neuem durchlaufen wird. In jeder Phase des Policy Cycle entstehen „Zwischenprodukte“, die Gegenstände einer Evaluation sein können und mittels Evaluationskriterien untersucht werden. Diese Gegenstände mit den zugehörigen Evaluationskriterien werden im Folgenden dargestellt. Abbildung 1 stellt das Wirkungsmodell grafisch dar. Aufgrund der linearen Abfolge der Evaluationsgegenstände ist von einem linearen Wirkungsmodell die Rede.

**Abb. 1: Lineares Wirkungsmodell von Politikevaluation**

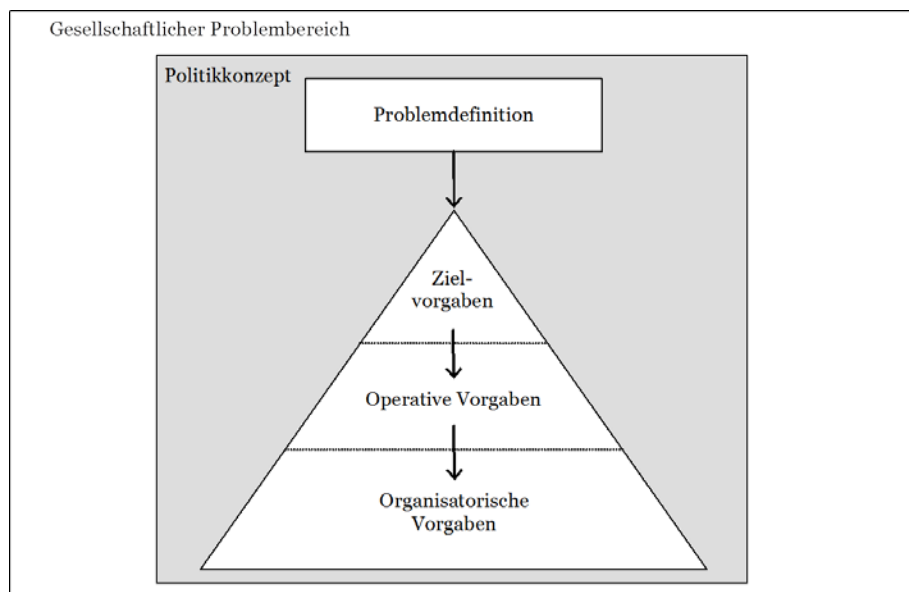
Quelle: eigene Darstellung nach Knoepfel/Bussmann 1997

Ursprung einer Policy ist das als gesellschaftlich relevant anerkannte Problem, zu dessen Lösung in der Politikformulierungsphase ein *Politikkonzept* entwickelt und beschlossen wird. Während das Problem Auslöser einer öffentlichen Politik ist, stellt es aber keinen Evaluationsgegenstand dar, da es in einer Politikevaluation keiner Bewertung unterzogen wird. Vielmehr geht es um die Bewertung der Lösung. Der erste Evaluationsgegenstand ist deshalb das *Politikkonzept*. Das *Politikkonzept* oder der *Input* bezeichnet die Gesamtheit der rechtlichen Bestimmungen und der diese konkretisierenden Vorgaben, die zu einer Politik formuliert wurden, also die Politik, wie sie auf dem Papier steht, noch ungeachtet ihrer tatsächlichen Umsetzung. Im Idealfall umfasst das *Politikkonzept* vier aufeinander abgestimmte Komponenten (vgl. Abbildung 2). Die *Problemdefinition* enthält Annahmen über die Art des Problems, dessen Ursachen, Intensität und Verbreitung. Hierfür werden im *Politikkonzept* Kausalhypothesen formuliert, welche aufzeigen, wie das in einem bestimmten gesellschaftlichen Bereich ausgemachte Problem durch eine öffentliche Politik adressiert werden soll. Unnötig zu erwähnen, dass die *Problemdefinition* eine hohe politische Komponente enthält. Die *Zielvorgaben* enthalten nun konkrete Vorgaben zur Zustandsänderung im Problemfeld. Es können *substanzielle*, die Wirkungen des Programms betreffende Vorgaben, und *operative*, die Umsetzungsstruktur regelnde Vorgaben unterschieden werden. Ausserdem können entsprechend dem Zeithorizont und dem Konkretisierungsgrad Zielvorgaben gegliedert und analysiert werden. Um die Vision, die daraus abgeleiteten Haupt- und Unterziele sowie beabsichtigte Nebenwirkungen systematisieren zu können, empfiehlt sich die Abbildung in einem Zielsystem (vgl. *Ledermann/Sager 2009*). Damit die gemachten Zielvorgaben erreicht werden können, benötigt ein politisches Programm *operative Vorgaben*, welche Policy-Instrumente definieren, durchzuführende Aktivitäten beschreiben und Interventionsbereiche abstecken. Je nach Umfang des *Politikkonzepts* enthalten die operativen Vorgaben unterschiedliche Konkretisierungsstufen, Ansatzpunkte und modellierte Wechselwirkungen zwischen

Aktivitäten. Schliesslich enthält ein vollständiges Politikkonzept *organisatorische Vorgaben*, welche Zuständigkeiten definieren und benötigte Ressourcenausstattungen festlegen. Unterschieden werden strukturelle und prozedurale Vorgaben. Abbildung 2 fasst die dargelegten Komponenten des Politikkonzepts zusammen und stellt schön deren hierarchische Beziehung dar. Um nun den Evaluationsgegenstand *Politikkonzept* zu evaluieren, können folgende Evaluationskriterien herangezogen werden:

- *Empirische Evidenz*: Dieses Kriterium beschreibt die wissenschaftliche Abstützung des Politikkonzepts. Werden bei der Problemdefinition und den Interventionshypothesen auf wissenschaftliche Erkenntnisse zurückgegriffen, die das Politikkonzept in bestehender Form rechtfertigen bzw. unterstützen? Ist es somit realistisch und enthält überhaupt die reale Möglichkeit, dass die öffentliche Politik „funktioniert“?
- Zudem wird die *logische Kohärenz* sowohl des Politikkonzepts selbst (*Intrapolicy-Kohärenz*) als auch in Bezug auf andere politische Massnahmen im jeweiligen gesellschaftlichen Problembereich (*Interpolicy-Kohärenz*) untersucht. Ist das Politikkonzept in sich stimmig und widerspruchsfrei, werden Synergien genutzt und befindet es sich in Einklang mit anderen Politiken?

**Abb. 2: Das Politikkonzept und seine Bestandteile**



Quelle: Ledermann/Sager 2009: 10.

Die *Umsetzungsorganisation* oder das Behördenarrangement bezeichnen die Gesamtheit der *tatsächlich* existierenden Umsetzungsstrukturen und Umsetzungsprozesse einer öffentlichen Politik. Diese umfassen somit die gesamte Organisation, die den Ablauf und die Zusammenarbeit und Koordination zwischen den involvierten Akteuren bei der Umsetzung einer Politik regelt. Oft wird zwischen der Umsetzungsorganisation auf der Programmebene der öffentlichen Politik und den Projektebenen einzelner enthaltener Projekte oder Massnahmen unterschieden. Für die Evaluation der Umsetzung empfehlen sich folgende Kriterien:

- *Eignung für die Umsetzung*: Die Umsetzungsorganisation (Behördenarrangement) wird auf ihre Eignung für den Vollzug analysiert. Sind die Aufgaben zweckmässig verteilt bzw. eignen sich die Vollzugsträger prinzipiell für den Vollzug? Besitzen sie



das entsprechende Know-How, einschlägige Erfahrung und genügend zeitliche und materielle Ressourcen?

- *Nachhaltigkeit*: Mit diesem Kriterium wird überprüft, ob die Umsetzungsorganisation institutionell verankert ist und eine Finanzierungssicherheit in Bezug auf die nötige Aufgabenerfüllung besteht, um die zukünftige Aufgabenerbringung entsprechend den Vorgaben aus dem Politikkonzept zu sichern.

Der *Output* beschreibt die Gesamtheit der Endprodukte des politisch-administrativen Umsetzungsprozesses, d.h. die Angebote und Leistungen der Verwaltung und weiterer in die Umsetzung der Politik involvierter Akteure. Er stellt die direkte Verbindung zwischen Umsetzungsakteuren und Adressaten dar. Oft werden allerdings sog. Zwischenoutputs produziert, die von Akteuren „weiterverarbeitet“ werden und erst dann die Adressaten erreichen. Selbstverständlich kann der Output einer öffentlichen Politik unterschiedlichste Formen annehmen. Die folgenden Kriterien dienen der Evaluation dieses Evaluationsgegenstands:

- Das Kriterium der *Angemessenheit* untersucht, ob die Leistung in Art, Umfang und Qualität den Vorgaben aus dem Politikkonzept entspricht. Es ist zu betonen, dass mit diesem Kriterium noch nicht die Wirkung der öffentlichen Politik evaluiert wird. So können zwar die Adressaten mit dem Output zufrieden sein, allerdings trifft dies noch keine Aussage in Bezug auf die Wirkung und den Problemlösungsbeitrag der Politik.
- Die *leistungsbezogene Effizienz* fragt schliesslich, ob das Verhältnis von Ressourceneinsatz und Leistungserbringung angemessen ist. Auch hier werden noch keine die Wirksamkeit der Politik betreffende Aussagen getroffen.

Der *Outcome* umschliesst die Verhaltensänderung bei den Zielgruppen, wobei letztere unterteilt werden können in Adressaten und Endbegünstigte. Adressaten sind Personen oder Institutionen, deren Verhalten durch die Politik verändert werden soll, sodass bei den Endbegünstigten ein Nutzen generiert wird. Endbegünstigte sind Personen, Institutionen oder andere Subjekte, die von der öffentlichen Politik profitieren sollen. Folglich beschreibt der Outcome die Gesamtheit der durch die Politik ausgelösten Verhaltensänderungen, die bei den Adressaten beobachtet werden können. Besonderes Augenmerk gilt hierbei etwaigen nichtbeabsichtigten Verhaltensänderungen, die gegebenenfalls die Wirkung der Politik negativ beeinflussen können oder gar vollends zu Nichte machen. Um den Outcome zu evaluieren, werden folgende Kriterien herangezogen:

- Die *Wirksamkeit des Outcome* bestimmt den Grad der Übereinstimmung zwischen beabsichtigter und tatsächlich gezeigter Verhaltensänderung. Zeigen also die Adressaten eine Verhaltensänderung, die so im Politikkonzept angestrebt wurde?
- Mit der *wirkungsbezogenen Effizienz des Outcome* wird das Verhältnis des Ressourceneinsatzes und dadurch erzielten Verhaltensänderungen beurteilt. Hier stellt sich die konkrete Frage, ob die Verhaltensänderung auch ressourcenschonender hätte erreicht werden können.

Der *Impact* als letzter Evaluationsgegenstand umfasst die Gesamtheit der Veränderungen im gesellschaftlichen Problembereich, die sich ursächlich auf die Bestandteile der öffentlichen Politik zurückführen lassen. Zentral ist die Beantwortung der Frage, ob sich die Situation der Endbegünstigten tatsächlich entsprechend den im Politikkonzept formulierten Zielen verändert hat. Auch rücken etwaige indirekte Auswirkungen der öffentlichen Politik (auf die Gesamtgesellschaft und/oder auf die Volkswirtschaft) in den Fokus. Für die Bewertung des Impact bieten sich folgende Evaluationskriterien an, die den letztendlichen Erfolg der Politik beurteilen:

- Die *Wirksamkeit des Impact* misst den Grad der Übereinstimmung zwischen beabsichtigten und tatsächlich eingetretenen Wirkungen in Bezug auf den Problembeitrag.
- Die *wirkungsbezogene Effizienz des Impact* bewertet schliesslich das Verhältnis von eingesetzten Ressourcen und eingetretenen Wirkungen der Politik.

Wie aus den dargestellten Evaluationsgegenständen des Wirkungsmodells ersichtlich wird, können diese bis zu einem gewissen Grad einzeln bzw. separat evaluiert werden. Diese Tatsache entspricht auch der Realität der Evaluationspraxis. Die meisten Evaluationen setzen bei der Bewertung und Beurteilung gewisse Schwerpunkte, sei dies aus zeitlichen, ressourcentechnischen oder praktischen Gründen. Allerdings liegt auf der Hand, dass diesen Schwerpunktsetzungen strenge Grenzen gesetzt sind, will man zu realistischen Erkenntnissen in Bezug auf Funktionieren und Wirken einer öffentlichen Politik gelangen. Das zeigt schon das zuletzt dargestellte Evaluationskriterium der *wirkungsbezogenen Effizienz des Impact*: Will man das Verhältnis von eingesetzten Ressourcen und eingetretenen Wirkungen der Politik beurteilen, müssen zuvor die Eignung der Umsetzungsevaluation und die Wirkungen der Politik separat untersucht worden sein. Viele Beurteilungen und Bewertungen in einer Evaluation bauen also aufeinander auf. Generell hält die Verbindung von Konzept- und Umsetzungsevaluation in einer sog. *Globalevaluation* gewichtige Vorteile bereit: So erlaubt eine Globalevaluation die genaue Zurückführung von Wirkungsdefiziten auf *Policy Failure* (auch Theoriefehler) und *Implementation Failure* (auch Umsetzungsfehler) sowie die Aufdeckung von Wechselwirkungen zwischen diesen beiden Fehlerquellen, und stellt eine notwendige Bedingung für die Möglichkeit dar, zielführende Lehren für die Weiterentwicklung und Verbesserung der öffentlichen Politik ableiten zu können (Ledermann/Sager 2009).

#### **4. Evaluation institutioneller Politik**

Das im vorherigen Abschnitt dargestellte Wirkungsmodell stellt in obiger Form natürlich ein verallgemeinertes, idealtypisches Modell dar, das je nach Evaluationsobjekt und spezifischem Politikkontext angepasst werden muss (vgl. Sager/Rüefli 2005). Es versteht sich gerade nicht als starre Schablone, die für jede erdenkliche öffentliche Politik in dieser Form verwendet werden kann. Trotzdem sind die Änderungen, die sich für die Evaluation ergeben, dann besonders signifikant, wenn der Gegenstand der Evaluation keine substanzielle Politik, sondern eine *institutionelle Politik* darstellt. Deshalb wird in diesem Abschnitt die Evaluation institutioneller Politik eigens erläutert und auf wichtige Unterschiede bei dieser speziellen Form der Evaluation eingegangen.

Zunächst muss hierfür die institutionelle Politik von der substanziellen, nachgerade „klassischen“ Politik abgegrenzt werden. Letztere versucht durch inhaltliche Vorgaben und unter Verwendung unterschiedlicher Policy-Instrumente das anvisierte Adressatenverhalten direkt zu verändern und somit das gesellschaftliche Problem zu lösen. Institutionelle Politik geht einen Umweg, indem sie das Institutionengefüge verändert und damit die Machtverteilung zwischen Akteuren und Prozessen beeinflusst, sodass durch die veränderte institutionelle Politik in einem zweiten Schritt auch Policy-Ziele erreicht werden können. Weiterhin müssen zwei Arten von institutioneller Politik unterschieden werden. Die erste Gattung institutioneller Politik umfasst *politische Institutionen* bzw. die Gestaltung derselben. Politische Institutionen sind anerkannte Regelsysteme politischer Willensbildung und Willensumsetzung (March/Olson 1989) wie Wahl- und Regierungssysteme sowie die gliedstaatliche Kompetenzverteilung in föderalistischen Staaten. Sie beziehen sich auf die Metaebene des politischen Systems und kennzeichnen dessen Inputseite (Sager 2009). Die zweite Gattung beschreibt *Institutionenpolitik*, die auf der tiefer gelegenen Meso-Ebene angesiedelt ist und die Outputseite des politischen Systems charakterisiert, indem sie die Voraussetzungen schafft, damit Sachpolitik ablaufen und wirken kann (Sager 2009). Für die Evaluation besonders

relevante Beispiele sind Verwaltungsreformen und Reorganisationen im öffentlichen Sektor sowie Vollzugsstrukturen öffentlicher Politik (vgl. Kuhlmann 2009; 2008).

Speziell die *Institutionenpolitik* stellt EvaluatorenInnen vor Herausforderungen, die sich zuvorderst auf die Anwendbarkeit des oben dargestellten Wirkungsmodells erstrecken. Konkret wird bei der Evaluation von Institutionenpolitik das Verwaltungshandeln nicht mehr als Teil der Umsetzungsorganisation einer substantziellen Politik evaluiert, sondern wird selbst primärer Gegenstand des Evaluationsinteresses. Die untersuchten Wirkungsketten verlaufen gänzlich innerhalb des Behördenarrangements und enden beim Output des Verwaltungshandelns (Sager 2009). In der Folge sind auch die oben dargestellten Wirkungsvariablen gänzlich *innerhalb des Behördenarrangements* angesiedelt. Dies stellt EvaluatorenInnen oft vor Probleme, da die untersuchten Wirkungsketten im Vergleich zu substantziellen Politiken meist komplexer sind und oft einen hohen Abstraktionsgrad erfordern (Widmer 2002; Kuhlmann 2009). Zudem ist mit einem derartigen Wirkungsmodell noch nichts über die Wirkung der Institutionenpolitik ausserhalb des Behördenarrangements gesagt. Die Beschränkung auf die Output-Ebene ist problematisch, da so noch nichts über den gesellschaftlichen Problemlösungsbeitrag der Institutionenpolitik bekannt ist. Verstärkt wird diese analytische Unzulänglichkeit noch dadurch, dass sich die eigentlichen Wirkungen von Institutionenpolitik häufig erst langfristig einstellen. In der Praxis behilft man sich deshalb oft mit der Verwendung von outputorientierten Zielgrößen, was allerdings Leistungserbringung mit Wirkungserzielung fälschlicherweise gleichsetzt. Weitere Herausforderungen für ein Wirkungsmodell zur Evaluation von Institutionenpolitik ergeben sich aus dem oft *nicht vorhanden Politikkonzept*, wodurch die Umsetzung einer Politik nicht mehr als realisiertes Politikkonzept begriffen werden kann und die Unterscheidung in Policy Failure und Implementation Failure erschwert wird. Schliesslich stellt sich speziell im Fall einer Reformevaluation das Problem des *Moving Target*, nach dem die Umsetzungsorganisation zur Durchführung der Reform selbst Gegenstand der Veränderung ist. Dies birgt die Gefahr zirkulärer Wirkungsmodelle.

Zusammenfassend stellt die Evaluation institutioneller Politik große Herausforderungen an das Untersuchungsdesign und die Überprüfung der Wirksamkeit. Um diese Schwierigkeiten bei der Evaluation und daraus resultierende Qualitätsprobleme (Widmer 2002) in den Griff zu bekommen, empfiehlt sich eine stärkere Nutzensausrichtung der Evaluation, die in Anbetracht der zahlreichen Wechselwirkungen von Institutionenpolitik speziell eine stärkere Kontextsensitivität bei der Analyse impliziert (Sager 2009).

Bei der Evaluation von *politischen Institutionen* können generell zwei Arten von Evaluationen unterschieden werden. *Soll-Ist-Vergleiche* gehen davon aus, dass politische Institutionen gewisse Ziele auf der Politics- und Policy-Ebene verfolgen. Bei der Evaluation wird nun untersucht, inwieweit die politische Institution ihre Ziele erreicht und wo Funktionsprobleme zu finden sind. Ein Beispiel ist die Untersuchung eines föderalistischen Systems auf seine tatsächlichen Vorteile und Nachteile in den Bereichen Bürgerbeteiligung, ökonomische Performanz, Verteilungsgerechtigkeit, etc. (vgl. Vatter/Sager 1996). Im Unterschied zu Soll-Ist-Vergleichen versuchen *wirkungsorientierte Evaluationen* sämtliche Wirkungen einer politischen Institution zu eruieren. Im Allgemeinen findet die Evaluation von politischen Institutionen zumeist als Grundlagenforschung statt, auch weil die Reformfähigkeit selbiger aufgrund selbsterhaltender Mechanismen und Veto-Punkten (siehe dazu den Beitrag von Wenzelburger/Zohlnhöfer in diesem Band) stark beschränkt ist. Evaluationen haben deswegen in erster Linie die Aufgabe, die Wechselwirkungen von politischen Institutionen und der Gesellschaft reflektierend zu begleiten und so im besten Fall die Wahrnehmung von politischen Institutionen nachhaltig zu verändern.

## 5. Nutzung von Evaluationen

Nachdem die vorangegangenen Abschnitte die Entstehung der Evaluation als Disziplin, ihre Bestandteile sowie ihr Funktionieren erläutert haben, stellt sich im Folgenden die Frage nach

der eingangs erwähnten Kontexteinbettung von Evaluation. Welche Bedeutung haben Evaluationen für den politischen Prozess, für die wissenschaftliche Politikfeldanalyse und für modernes Regierungs- und Verwaltungshandeln? Um diese Fragen zu beantworten und die Politikevaluation zu situieren, müssen die unterschiedlichen Verbindungen der Evaluation zum politisch-administrativen Komplex eines staatlichen Systems herausgearbeitet werden. Die Politikevaluation stellt eine wissenschaftlich abgestützte Form der Politikberatung dar, deren Erkenntnisse in unterschiedlicher Form im politischen Prozess nutzbar gemacht werden können. In Tabelle 3 sind die drei Dimensionen der Valorisierung zusammengefasst (vgl. *Stockbauer 2000; Sager/Ledermann 2008*).

<i>Dimension</i>	<i>Erläuterung</i>
Nutzung (Verwendung, Verwertung)	Wertfreier, <i>objektiv beobachtbarer Gebrauch</i> von Evaluationen
Nützlichkeit (Verwendbarkeit, Verwertbarkeit)	Beschreibt das <i>Nutzungspotenzial</i> von Evaluationen
Nutzen	Wertmäßige, <i>subjektive Beurteilung</i> der Qualität und Nützlichkeit von Evaluationen

Tabelle 3: Drei Dimensionen von Nutzbarmachung von Evaluationen

Wichtigste Dimension ist dabei die tatsächliche Nutzung von Evaluationen im politischen Prozess. Die Untersuchung von Bedingungen für und Arten von Nutzung hat im Fall von Evaluationen lange Tradition, rechtfertigt doch letztendlich die (positive) Beeinflussung des Politikprozesses durch die Evaluation die Durchführung selbiger. So wurde in den 1960er Jahren die Verwendungsforschung durch die Einsicht initiiert, dass die durch die Evaluation erarbeiteten wissenschaftlichen Ergebnisse und Empfehlungen meist nicht umgesetzt wurden. Erklärungsgründe hierfür wurden in erster Linie im jeweiligen politischen Kontext und in der Rolle der an der Evaluation beteiligten Stakeholder, die häufig auch die Nutzer der Evaluationsergebnisse darstellen, ausgemacht. Auch wurde schnell offensichtlich, dass zwischen unterschiedlichen Verwendungsarten differenziert werden musste. So wurden Empfehlungen zwar selten direkt umgesetzt, allerdings kam es durch Evaluationsergebnisse häufig zu einem Umdenken der Stakeholder. *Carol H. Weiss* hat diese indirekte Form der Nutzung als Enlightenment- bzw. Aufklärungsnutzung bezeichnet. Diese Ausrichtung der Verwendungsforschung führte in den 1970er und 1980er Jahren zu zahlreichen, zumeist quantitativ ausgerichteten Studien, welche versuchten, die Determinanten der Nutzung von Evaluationen zu bestimmen. Allerdings kristallisierte sich kein eindeutiges Kompendium an die Nutzung bedingender Faktoren heraus, sodass in den 1990er Jahren eine gewisse Ernüchterung in der Verwendungsforschung einkehrte und manche ForscherInnen folgerten, dass die EvaluatorInnen den Einfluss der von ihnen durchgeführten Evaluation aufgrund ihrer Kontextabhängigkeit schlussendlich nicht beeinflussen könnten (vgl. *Weiss 1988a; 1988b*). *Patton (1988)* widersprach dieser Schlussfolgerung und setzte ihr die bereits im Theorieteil dieses Beitrags diskutierte verstärkte Anwendungsorientierung von Evaluationen entgegen, um die Nutzung von Evaluationsergebnissen zu forcieren.

In der Verwendungsforschung hat sich mittlerweile eine Differenzierung von Nutzungsarten herausgebildet, die hilft, die Analyse des Einflusses spezifischer Kontextsituationen auf die Nutzung genauer zu untersuchen. Generell wird zwischen vier Nutzungsarten unterschieden (vgl. *Johnson 1998*):

- *Instrumentelle Nutzung*: Die Ergebnisse und Empfehlungen einer Evaluation lösen konkrete Veränderungen aus. Aufgrund der Evaluation handeln die NutzerInnen entsprechend.
- *Konzeptuelle Nutzung*: Die Ergebnisse und Empfehlungen einer Evaluation lösen kognitive Veränderungen bei den NutzerInnen aus (Denken, Wahrnehmung).

- *Prozedurale Nutzung*: Durch die Teilnahme am Evaluationsprozess kommt es bei NutzerInnen zu konkreten Handlungsänderungen oder kognitiven Veränderungen.
- *Symbolische Nutzung*: Diese auch als legitimatorische, taktische oder persuasive Nutzung bezeichnete Form der Verwendung stellt keine wirkliche Veränderung da. Evaluationsergebnisse und -empfehlungen werden lediglich instrumentalisiert, um eine bereits bestehende Haltung zu festigen bzw. zu begründen.

Diese vier Formen der Nutzung dienen in dieser oder leicht abgeänderter Form bis heute den meisten Studien zur Verwendungsforschung als Analyseraster (z.B. *Balthasar*, 2006; *Weiss et al.*, 2005). Allerdings konnten, trotz zahlreicher Studien, nur wenige generelle Erkenntnisse zur Verwendung von Evaluationen erarbeitet werden. So gilt es als wahrscheinlich, dass es aufgrund von Evaluationen dann kaum zu Veränderungen kommt, wenn Konflikte zwischen Stakeholdern existieren. In solchen Fällen dient die Evaluation meist nur zur Unterstützung der bereits bestehenden Position, es kommt also nur zu einer symbolischen Nutzung. Folglich wirken sich der Konfliktgrad und der Problemdruck im Evaluationsumfeld negativ auf die Verwendung aus (*Ledermann* 2012; *Valovirta* 2002). Eine zweite Erkenntnis betrifft den positiven Einfluss der Einbeziehung der NutzerInnen auf die Verwendung von Evaluationsergebnissen und -empfehlungen. Ein partizipativ ausgestalteter Evaluationsprozess erhöht die Nachvollziehbarkeit der Ergebnisse bei den NutzerInnen und kann dadurch die instrumentelle und konzeptuelle Nutzung steigern (*Ledermann* 2012).

Trotz ihrer ungebrochenen Popularität in der Verwendungsforschung wurden diese vier Formen der Nutzung auch kritisiert (*Mark/Henry* 2004). So fehlen klare Indikatoren für die einzelnen Nutzungsarten, was eine empirische Untersuchung der tatsächlichen Verwendung erschwert. Zudem ist die Kategorisierung nicht frei von Überschneidungen: Die Quelle der Veränderung ist im Fall der prozeduralen Nutzung die Teilnahme am Evaluationsprozess, im Fall der instrumentellen und konzeptuellen Nutzung sind es die Ergebnisse und Empfehlungen von Evaluationen. Zudem beschreiben die vier Kategorien unterschiedliche Typen von Veränderungen, die teils aufeinander aufbauen: Einer instrumentellen Nutzung geht häufig eine konzeptuelle Nutzung, oder besser, Verinnerlichung von Evaluationserkenntnissen und –empfehlungen voraus.

Diese Abgrenzungs- und Operationalisierungsprobleme haben schliesslich dazu geführt, dass der Begriff ‚Verwendung‘ in Zusammenhang mit Evaluationen generell in Frage gestellt wurde (*Kirkhart* 2000). Anstelle von Verwendung sollte der Begriff *Einfluss* (Influence) treten, um diffuse, zeitlich schwer abgrenzbare und wenig zielgerichtete Wirkungen von Evaluationen ebenfalls erfassen zu können. Darauf aufbauend schlugen *Mark* und *Henry* (2004) vor, die einzelnen Prozesse, von allgemeinen, kognitiven und motivationalen Prozessen bis zu Verhaltensprozessen, die durch Evaluationen beeinflusst werden, verstärkt zu untersuchen. Zudem lenken sie die Aufmerksamkeit auf unterschiedliche Kontextbedingungen, die den Einfluss von Evaluationen jeweils behindern oder fördern können.

Ob nun von Nutzung oder Einfluss im Zusammenhang mit Evaluationen gesprochen wird, ist es doch unzweifelhaft, dass professionell durchgeführte und qualitativ hochwertige Evaluationen die *potenzielle Nutzung* bzw. den potenziellen Einfluss im politischen Prozess zu steigern vermögen. Zu diesem Zweck wurden in jüngerer Zeit vielfach *Evaluationsstandards* festgelegt, die gegenstandsunabhängige und bereichsübergreifende Anforderungen an Evaluationen formulieren und nicht nur von EvaluatorInnen, sondern von sämtlichen Stakeholdern berücksichtigt werden müssen. Beispielsweise haben die deutsche und die schweizerische Evaluationsgesellschaft Standards publiziert, die sich in vier Gruppen untergliedern (*DeGEval* 2002; *SEVAL* 2000):

- Die Einhaltung von *Nützlichkeitsstandards* soll dafür sorgen, dass eine Evaluation die Informationsbedürfnisse designierter NutzerInnen möglichst befriedigt sowie eindeutige und klar festgelegte Evaluationsziele verfolgt.

- *Durchführbarkeitsstandards* sollen gewährleisten, dass eine Evaluation realistisch, durchdacht und diplomatisch, aber auch kostenbewusst konzipiert und durchgeführt wird.
- *Fairness-Standards* bzw. *Korrektheitsstandards* sollen den fairen und respektvollen Umgang mit sämtlichen Beteiligten während der Evaluation sowie eine rechtlich und ethisch korrekte Durchführung ermöglichen.
- Die *Genauigkeitsstandards* sollen schließlich sicherstellen, dass eine Evaluation gültige und verwendbare Informationen produziert und vermittelt.

Diese sowie vergleichbare Standards erfreuen sich steigender Akzeptanz in der Praxis, allerdings sind sie auch kontinuierlichem Veränderungsdruck ausgesetzt, nicht zuletzt weil die Umsetzung und Operationalisierung der Standards weiterhin schwierig ist sowie teilweise Überschneidungen und sogar Widersprüche existieren (vgl. *Sager/Ledermann 2009*).

## 6. Evidenzbasierte Politik

Um die gestiegene Bedeutung von Evaluationen verstehen zu können, muss die Politikevaluation im Kontext des in jüngerer Zeit v.a. im deutschsprachigen Raum immer bekannter werdenden *Evidence-Based Policy-Making* (EBP) betrachtet werden (*Frey/Ledermann 2010*). Dieser Begriff orientiert sich an den etablierten Prinzipien der Evidence-Based Medicine und geht auf eine durch die Labour-Regierung unter Tony Blair in Großbritannien initiierte Reformbestrebung zurück, die unter dem Motto „What matters is what works“ anstelle ideologiegetriebener Politik eine informierte, wissenschaftsgestützte Art der Politikformulierung setzen wollte. Politikformulierung sollte einem rationalen, bewährten Erkenntnissen folgenden Machbarkeitsdiktat unterworfen werden (*Frey 2012, Frey/Ledermann 2010, Ledermann 2013*).

Allerdings ist Evidenz im Kontext des EBP kein klar definierter Begriff. So herrscht geteilte Meinung, ob hiermit nur wissenschaftliche Erkenntnisse, oder auch gesicherte Informationen anderer Art gemeint sind. Eine enge Definition von Evidenz hat jedenfalls den Nachteil, dass viele Formen politikrelevanten Wissens vernachlässigt werden. Auch das britische *Cabinet Office* hat *evidence* recht breit definiert:

“The raw ingredient of evidence is information. Good quality policy making depends on high-quality information, derived from a variety of sources - expert knowledge; existing domestic and international research; existing statistics; stakeholder consultation; evaluation of previous policies; new research, if appropriate; or secondary sources, including the internet.” (*Cabinet Office 1999: 33*).

Anstelle einer unkontrollierten begrifflichen Ausweitung, die Gefahr läuft, Kausalzusammenhänge zu vernachlässigen, empfiehlt sich für Analysezwecke eine klare Abgrenzung von Evidenz zu anderen Formen politikrelevanten Wissens (*Ledermann 2013*). *Evidenz* kann als forschungsbasierte Information verstanden werden, die mittels eines systematischen Verfahrens generiert und in objektiv nachvollziehbarer Form festgehalten worden ist (*Sager/Ledermann 2008*). Evidenz ist objektgebunden und hat statischen Charakter. Dadurch unterscheidet sie sich von einer anderen Form politikrelevanten Wissens, der *Expertise*. Letztere ist durch ihre Personengebundenheit weitaus weniger statisch, dafür aber situativ anwendbar (*Dreyfus/Dreyfus 2005*). Zudem umfasst sie im Gegensatz zur Evidenz auch implizites Wissen (*Nutley et al. 2003*). Eine solche Abgrenzung, in der Definition des *Cabinet Office* im Übrigen noch nicht klar ersichtlich, hat den Vorteil, genau analysieren zu können, welche Formen politikrelevanten Wissens von welchen Akteuren ins System eingespeist und von welchen Akteuren unter welchen Kontextbedingungen genutzt werden. EBP wird so einer wissenschaftlichen Untersuchung zugänglich gemacht.

Im Rahmen des EBP kommt der Evaluation eine besondere Rolle zu: In vielen Situationen reicht die Zeit nicht aus, um durch eine ausführliche (und zeitaufwändige) Evaluation eine informierte Entscheidungssituation herbeiführen zu können. Die durch die Evaluation

produzierte Evidenz kommt also oft zu spät, um in die Politikformulierung integriert werden zu können. Stattdessen ist die Evaluation im Zuge der *systematischen Zweitauswertung von existierendem Wissen* von großer Bedeutung. Diese *Metaanalysen von Evaluationen* werden im Rahmen des EBP immer wichtiger (vgl. bspw. Sager 2005, Sager 2007). Dabei stellt sich die Frage, wie die in unterschiedlichen Evaluationen gewonnen Erkenntnisse aggregiert und für eine spezifische Entscheidungssituation aufbereitet werden können. Eine erste Möglichkeit besteht in *quantitativ ausgerichteten Metaanalysen*, eine zweite in sog. *Narrative Reviews*, welche die qualitative Zusammenführung existierender Einzelfallstudien beschreiben. Beide Möglichkeiten haben ihre Vor- und Nachteile, die zu einem Trade-Off zwingen. *Quantitative Metaanalysen* liefern generalisiertes Wissen, leiden aber an mangelnder Kontextberücksichtigung. *Narrative Reviews* hingegen sind kontextsensitiv, die Generalisierbarkeit der Erkenntnisse ist dafür beschränkt. Einen Mittelweg bietet die auf dem realistischen Evaluationsansatz basierende Realist Synthesis nach Pawson (2002). Dabei werden anlässlich von sog. *Systematic Reviews* die in Abschnitt 3 erläuterten CMO-Konfigurationen in einzelnen Studien ermittelt, um die Wechselwirkungen von unterschiedlichen Mechanismen und Kontexteinflüssen sowie deren Auswirkungen auf die Outcomes studienübergreifend zu ermitteln. Dieses Vorgehen ermöglicht im Idealfall eine kontextsensitive Generalisierung von Evaluationsergebnissen.

Die im Zusammenhang der EBP-Bewegung gestiegene Bedeutung von Evaluationen wird noch verstärkt durch die Tatsache, dass in vielen Fällen auch seitens der Politik eine *verstärkte Nachfrage nach Evidenz* besteht. Diese Nachfragewirkung führt Solesbury (2002: 91) zufolge zu einer „utilitaristischen Wende“, die das Umdenken in der Wissenschaft hin zu stärkerer Anwendungsorientierung beschreibt.

Allerdings muss auch über die *Grenzen des EBP* und damit über die Grenzen des Einflusses von Evaluationen auf die Politikformulierung gesprochen werden. Die Komplexität letzterer steht nämlich der Idealform von Wissensspeisung, die von EBP gewissermaßen impliziert wird, entgegen. Ideal wäre eine rationale Abwägung zwischen Handlungsalternativen, von denen jeweils sämtliche Konsequenzen im Voraus bekannt sind. Dies ist aber meist nicht der Fall. Die Gründe hierfür liegen auf der Hand: Die Komplexität des politischen Prozesses schmälert die Aussage und Prognosekraft von wissenschaftlichen Erkenntnissen (Sanderson 2006), sodass viele Konsequenzen und „nicht-intendierte Nebenwirkungen“ unberücksichtigt bleiben. Zudem besteht nicht immer vollkommene Klarheit über das Problem, das gelöst werden muss (Parsons 2004: 50), was im Umkehrschluss bedeutet dass auch unklar bleibt, welche Formen öffentlicher Politik überhaupt einer evidenzbasierten Überprüfung unterzogen werden sollten, um das Problem zu adressieren (Ledermann 2013). Auch können aufgrund von Zeit- und Ressourcenbeschränkungen häufig schlicht nicht alle Maßnahmen ausgiebig überprüft werden. Schließlich muss Evidenz auch aufbereitet werden, um Kenntnis zu erlangen (Tenbensel 2004). Aufgrund dieser Beschränkungen sprechen manche Autoren auch von „evidence-informed“ oder „evidence-inspired“ und schwächen so die Rolle von Evidenz für die Politikformulierung ab (z. B. Nutley et al. 2003). Auch soll nicht unerwähnt bleiben, dass EBP manchen Formen der Nutzung von politikrelevantem Wissen in gewisser Weise diametral entgegensteht: EBP impliziert nämlich, dass bei den politischen Akteuren der *Wille zur Verwendung* von Evidenz stets vorhanden ist. Die im vorangegangenen Abschnitt thematisierte symbolische Nutzung unterstellt aber eine legitimatorische, taktische oder persuasive Nutzung von Evidenz. In solchen Fällen wird der zweifellos vorhandene Spielraum beim Rückgriff auf Evidenz genutzt.

Zusammenfassend gilt, dass dem Ansatz des EBP komplexitäts- und kontextbedingte Grenzen gesetzt sind. In diesem Zusammenhang ist das im nächsten Abschnitt erläuterte institutionalisierte Regulatory Impact Assessment (RIA) zu verorten.

## 7. Regulatory Impact Assessment (RIA)

Um die Berücksichtigung von Evidenz bei der Politikformulierung – trotz der im vorherigen Abschnitt erläuterten Grenzen des EBP-Ansatzes – zu fördern oder unter bestimmten Umständen gar zu erzwingen, existiert die Idee, die *Evidenzeinspeisung in die Politikformulierung zu institutionalisieren*. Das in vielen Staaten verankerte *Regulatory Impact Assessment* (RIA) ist ein Produkt dieser Bestrebungen. RIA ist ein ex-ante Instrument, dem die Annahme zugrunde liegt, dass die Folgen konkreter politischer Maßnahmen modelliert und damit zu einem guten Teil vorausgesagt werden können, indem auf bestehendes Wissen über die Wirksamkeit und Problemlösungskapazität unterschiedlicher Interventionen zurückgegriffen wird. In diesem Sinne basieren RIAs auf dem oben erläuterten EBP, reichen in ihrem Anspruch aber noch über letzteres hinaus.

Die Institutionalisierung des RIA-Ansatzes ist vielen Ländern ein Anliegen und wird durch die OECD standardisiert und gefördert (OECD 2008; 2012). Die Bertelsmann Stiftung misst im Rahmen des Sustainable Governance Indicators (SGI) Projekts systematisch die Nutzung von RIA in den OECD-Staaten, die letzten Daten liegen zum Jahr 2011 vor ([www.sgi-network.org](http://www.sgi-network.org)). Dem RIA-Ansatz kommt in der so genannten „Better-Regulation-Debatte“, die durch die zahlreichen während der Finanzkrise aufgedeckten Regulierungsfehler neu befeuert wurde, eine Schlüsselrolle zu (Wegrich 2010). Chronisch finanzschwache Staaten haben ein besonderes Interesse daran, staatliche Eingriffe auf Erfolg und Rentabilität zu überprüfen (vgl. bspw. Lodge/Wegrich 2010).

Um dabei behilflich zu sein, geht der RIA-Ansatz zunächst von der Grundeinsicht aus, dass staatliche Eingriffe weitreichende, oft schwierig zu identifizierende Konsequenzen zeitigen: „they affect many different groups in society and the effects may be of many different types. Many of the effects are ‘hidden’, or at least are difficult to identify when a regulation is being considered” (OECD 2008: 3). Um diese Konsequenzen bestimmen zu können, werden in einem ersten Schritt sämtliche verfügbaren und entscheidungsrelevanten Informationen gesammelt und ausgewertet. Kommt man zum Schluss, dass basierend auf diesen Informationen ein staatlicher Eingriff zielführend ist und das jeweilige gesellschaftliche Problem adressiert werden kann, wird die betrachtete staatliche Maßnahme auf ihre vermeintlichen Kosten untersucht. In einem letzten Schritt wird das ermittelte Kosten-Nutzen-Verhältnis der staatlichen Maßnahme mit anderen potentiell in Frage kommenden Policies verglichen. Dieser idealtypisch zu durchlaufende Prozess der Politikanalyse soll den politischen Prozess informieren und, den Zielen des EBP entsprechend, auf eine rationale, systematische und evidenzbasierte Grundlage stellen.

Werden RIAs bei der Politikformulierung vorgeschrieben, ist ersichtlich, wie Staaten versuchen, EBP zu institutionalisieren. Ein Beispiel hierfür ist die in der Schweiz gesetzlich verankerte Regulierungsfolgenabschätzung, die vorschreibt, dass sämtliche neu verabschiedeten Gesetzesbeschlüsse sowie Gesetzesänderungen einem RIA unterzogen werden müssen. Dabei werden zur Diskussion stehende Vorschläge, bevor im Parlament über ihre Umsetzung abgestimmt wird, einem Test unterzogen, der die Notwendigkeit und Möglichkeit staatlichen Handelns eruiert, Auswirkungen auf einzelne gesellschaftliche Gruppen und die Gesamtwirtschaft ermittelt sowie alternative Regelungen und die Zweckmäßigkeit im Vollzug untersucht.

Diese institutionalisierte Form wissenschaftlicher Politikformulierung klingt auf dem Papier natürlich einleuchtend, allerdings muss auch hier thematisiert werden, wie RIAs den tatsächlichen politischen Prozess beeinflussen bzw. in diesen integriert werden können. Um die komplexen Zusammenhänge und unklaren Grenzen zwischen der Folgeabschätzung und dem politischen Prozess untersuchen zu können, kann auf die in Abschnitt 5 erläuterten vier Arten der Nutzung politikrelevanten Wissens zurückgegriffen werden. “RIA is a particularly fascinating case for the analysis of the role of knowledge in policy-making because it has quasi-scientific ambitions, but also takes place at the heart of government” (Hertin et al. 2009: 413). Untersuchungen zeigen, dass für die Nutzung von Wissen im Politikprozess nicht nur die



spezifische Ausgestaltung des RIA-Prozesses, sondern auch die Konstellationen relevanter, in die Prozesse involvierter Akteure und der Kontext des jeweiligen politischen Prozesses sowie institutionelle Gegebenheiten des politischen Systems entscheidende Faktoren sind (Rissi/Sager 2013; Sager/Rissi 2011). So scheint es zwar prinzipiell möglich, dass RIAs instrumentell und konzeptionell genutzt werden und dadurch den Prozess der Politikformulierung verbessern. Allerdings handelt es sich auch hier um Instrumente, die überwiegend symbolisch genutzt werden und somit nicht ungeachtet ihres spezifischen politischen Kontexts erfolgreich eingeführt werden können.

## 8. Fazit

Die Politikevaluation entstand aus dem Bedarf nach Wissen über die Wirksamkeit von öffentlicher Politik heraus, sowohl um diese zu rechtfertigen als auch um sie zu verbessern. Der hier nachgezeichnete Weg von der Evaluation als Grundlagenforschung zur nutzungsorientierten Bereitstellung von Wissen zeigt, dass die Nachfrage nach Evidenz in nächster Zeit nicht abbrechen wird. Ob damit der Schluss gezogen werden kann, dass der Evaluation eine goldene Zukunft bevorsteht, ist aber fraglich. *Evidence-based Policymaking*, wissenschaftliche Politikformulierung, sowie *Regulatory Impact Assessment* haben nicht nur eine stärkere Nachfrage nach Wissen erzeugt, sondern auch eine „utilitaristische Wende“ in der Wissenschaft hin zu Anwendungsorientierung und Einbezug der NutzerInnen mitverursacht. Gerade im Falle der RIA stellt sich dabei auch die Frage, inwiefern Empirie, die bei der klassischen Politikevaluation im Zentrum steht, überhaupt noch eine Rolle spielt und nicht von reinen Modellschätzungen abgelöst wird. Insgesamt bringt die Ära des „Evidence-Based-Everything“ (Pawson et al. 2011: 519) zahlreiche neue Herausforderungen. Wie insbesondere die Abschnitte 3 und 4 gezeigt haben, braucht es seitens der EvaluatorInnen ein großes Maß an Sachverstand und Expertise, um in komplexen Politikbereichen qualitativ hochwertige und v.a. aussagekräftige Evaluationen durchführen zu können. Betrachtet man die in Abschnitt 2 dargelegte Theoriedebatte im Zeitraffer, wird klar, dass hierbei ein Spagat zwischen Wissenschaftlichkeit, Theoriefundierung und Qualität, sowie Anspruchsorientierung und Berücksichtigung des „evaluativen Kontexts“ vollzogen werden muss. Die verstärkte Bedeutung von Evaluationen im politischen Prozess formuliert neue Ansprüche an Wissenschaft und Praxis und wird diese besondere Disziplin der Politikfeldanalyse noch lange beschäftigen.

## Literatur

- Balthasar, Andreas*, 2006: The Effects of the Institutional Design on the Utilization of Evaluation, in: *Evaluation* 12 (3), 353-371.
- Befani, Barbara/Ledermann, Simone/Sager, Fritz*, 2007, Realistic Evaluation and QCA: Conceptual Parallels and an Empirical Application, in: *Evaluation: The International Journal of Theory, Research and Practice* 13 (2): 25-46.
- Cabinet Office*, 1999: Professional Policy Making for the Twenty First Century, London.
- Campbell, Donald T.*, 1969: Reform as Experiments, in: *American Psychologist* 24 (4), 409-429.
- DeGEval*, 2002: Standards für Evaluation, Köln: Deutsche Gesellschaft für Evaluation (DeGEval).
- Derlien, Hans-Ulrich*, 1997: Die Entwicklung von Evaluationen im internationalen Kontext, in: *Werner Bussmann/Ulrich Klöti/Peter Knoepfel* (Hrsg.), Einführung in die Politikevaluation, Basel/Frankfurt a. M., 4-12.
- Dreyfus, Hubert L./Dreyfus, Stuart E.*, 2005: Peripheral Vision: Expertise in Real World Contexts, in: *Organization Studies*, 26 (5), 779-792.
- Frey, Kathrin*, 2012: Evidenzbasierte Politikformulierung in der Schweiz: Gesetzesrevisionen im Vergleich. Baden-Baden.
- Frey, Kathrin/Ledermann, Simone*, 2010: Evidence-Based Policy: A Concept in Geographical and Substantive Expansion, in: *German Policy Studies* 6(2), 1-15.
- Guba, Egon G./Lincoln, Yvonna S.*, 1989: Fourth Generation Evaluation, Newbury Park/London/New Delhi.
- Harvey, David*, 2005: A Brief History of Neoliberalism, Oxford.
- Hertin Julia/Jacob Klaus/Pesch Udo/Pacchi Carolina*, 2009: The Production and Use of Knowledge in Regulatory Impact Assessment – An Empirical Analysis, in: *Forest Policy and Economics* 11, 413-421.
- Interdepartementale Kontaktgruppe «Wirkungsprüfungen» (IDEKOWI)*, 2004: Wirksamkeit von Bundesmassnahmen. Vorschläge zur Umsetzung von Artikel 170 der Bundesverfassung bei Bundesrat und Bundesverwaltung. Bericht der Interdepartementalen Kontaktgruppe "Wirkungsprüfungen" an die Generalsekretärenkonferenz der Schweizerischen Eidgenossenschaft, Bern.
- Johnson, R. Burke*, 1998: Toward a Theoretical Model of Evaluation Utilization, in: *Evaluation and Program Planning* 21 (1), 93-110.
- Kirkhart, Karen E.*, 2000: Reconceptualizing Evaluation Use: An Integrated Theory of Influence, in: *Valerie J. Caracelli/Hallie Preskill* (Hrsg.), *The Expanding Scope of Evaluation Use*, San Francisco, CA, 5-24.
- Knoepfel, Peter/Bussmann, Werner*, 1997: Die öffentliche Politik als Evaluationsobjekt, in: *Werner Bussmann/Ulrich Klöti/Peter Knoepfel* (Hrsg.), Einführung in die Politikevaluation, Basel/Frankfurt a. M., 58-77.
- Knoepfel, Peter/Larrue, Corinne/Varone, Frédéric*, 2001: Analyse et pilotage des politiques publiques. Genève/Bâle/Munich: Helbing & Lichtenhahn, 128-41.
- Kuhlmann, Sabine*, 2008: Politik- und Verwaltungsreform in Kontinentaleuropa: Subnationaler Institutionenwandel im deutsch-französischen Vergleich. Habilitationsschrift an der Universität Potsdam, Baden-Baden.
- Kuhlmann, Sabine*, 2009: Die Evaluation von Institutionenpolitik in Deutschland: Verwaltungsmodernisierung und Wirkungsanalyse im föderalen System, in: *Thomas Widmer/Wolfgang Beywl/Carlo Fabian* (Hrsg.), *Evaluation. Ein systematisches Handbuch*, Wiesbaden, 371-380.
- Ledermann, Simone*, 2012: Exploring the Necessary Conditions for Evaluation Use in Program Change, in: *American Journal of Evaluation* 33 (2), 159-178.

- Ledermann, Simone*, 2013: Evidence-Based Policy - Die Rolle der Bundesverwaltung. Dissertation, Bern: Kompetenzzentrum für Public Management.
- Ledermann, Simone/Sager, Fritz*, 2009: Problem erkannt, aber nicht gebannt. Der Nutzen einer Verknüpfung von Konzept- und Umsetzungsevaluation am Beispiel der schweizerischen Strategie ‚Migration und Gesundheit‘, in: *Zeitschrift für Evaluation* 1/2009, 7-25.
- Lodge, Martin/Wegrich, Kai*, 2010: Letter to the Editor of *Public Administration Review* in Response to a Recent Symposium on Financial Regulatory Reform, in: *Public Administration Review* 70 (2), 336-41.
- Mark, Melvin M./Henry, Gary T.*, 2004: The Mechanisms and Outcomes of Evaluation Influence, in: *Evaluation* 10 (1), 35-57.
- Nutley, Sandra/Walter, Isabel/Davies, Huw T. O.*, 2003: From Knowing to Doing: A Framework for Understanding the Evidence-into-Practice Agenda, in: *Evaluation*, 9 (2), 125-148.
- OECD*, 2008: Introductory Handbook for Undertaking Regulatory Impact Analysis (RIA) online: <http://www.oecd.org/gov/regulatory-policy/ria.htm> (23.07.2013).
- OECD*, 2012: Recommendation of the Council on Regulatory Policy and Governance, online: <http://www.oecd.org/gov/regulatory-policy/49990817.pdf> (23.07.2013).
- Parsons, Wayne*, 2004: Not Just Steering But Weaving: Relevant Knowledge and the Craft of Building Policy Capacity and Coherence in: *Australian Journal of Public Administration*, 63 (March), 43-57.
- Patton, Michael Q.*, 1988: The Evaluator's Responsibility for Utilization, in: *Evaluation Practice* 9 (2), 5-24.
- Patton, Michael Q.*, 1997: *Utilization-Focused Evaluation: The New Century Text*, Third Edition, Thousand Oaks/London/New Delhi.
- Pawson, Ray*, 2002: Evidence-based Policy: The Promise of 'Realist Synthesis', in: *Evaluation* 8 (3), 340-358.
- Pawson, Ray/Tilley, Nick*, 1997: *Realistic Evaluation*, London/New Delhi.
- Pawson, Ray/Wong, Geoff/Owen, Lesley*, 2011: Known knowns, known unknowns, unknown unknowns: The Predicament of Evidence-based Policy, in: *American Journal of Evaluation* 32 (4), 518-546.
- Rissi, Christof/Sager, Fritz*, 2013: Types of Knowledge Utilization of Regulatory Impact Assessment (RIA). Evidence from Swiss Policy-making, in: *Regulation & Governance* 7(3), 348-364.
- Ritz, Adrian*, 2003: *Evaluation von New Public Management – Grundlagen und empirische Ergebnisse der Bewertung von Verwaltungsreformen in der schweizerischen Bundesverwaltung*, Bern/Stuttgart/Wien.
- Rossi, Peter H./Freeman, Howard E./Lipsey, Mark W.*, 1999: *Evaluation. A Systematic Approach*, Sixth Edition, London/New Delhi.
- Sager, Fritz*, 2005: Die Wirksamkeit verkehrspolitischer Massnahmen: Eine Meta-Analyse, in *Thomas Bieger/Christian Laesser/ Rico Maggi* (Hrsg.), *Schweizerische Verkehrswirtschaft – Jahrbuch 2004/2005*, St. Gallen, 213-227.
- Sager, Fritz*, 2007: Making Transport Policy Work: Polity, Policy, Politics, and Systematic Review, in *Policy & Politics* 35(2), 269-288.
- Sager, Fritz*, 2009: Die Evaluation institutioneller Politiken in der Schweiz., in: *Thomas Widmer/Wolfgang Beywl/Carlo Fabian* (Hrsg.), *Evaluation. Ein systematisches Handbuch*, Wiesbaden, 361-370.
- Sager, Fritz/Rüefli, Christian*, 2005: Die Evaluation öffentlicher Politiken mit föderalistischen Vollzugsarrangements. Eine konzeptionelle Erweiterung des Stufenmodells und eine praktische Anwendung, in: *Swiss Political Science Review* 11 (2), 101-129.
- Sager, Fritz/Ledermann, Simone*, 2008: Valorisierung von Politikberatung, in *Stefan Bröchler/Rainer Schützeichel* (Hrsg.), *Politikberatung*, Stuttgart, 310-325.

- Sager, Fritz/Rissi, Christof*, 2011: The Limited Scope of Policy Appraisal in the Context of Referendum Democracy – the Case of Regulatory Impact Assessment in Switzerland, in: *Evaluation: The International Journal of Theory, Research and Practice* 17 (2), 151-164.
- Sager, Fritz/Andereggen, Céline*, 2012: Dealing with Complex Causality in Realist Synthesis: The Promise of Qualitative Comparative Analysis (QCA), in: *American Journal of Evaluation* 33 (1), 60-78.
- Sanderson, Ian*, 2006: Complexity, 'Practical Rationality' and Evidence-Based Policy Making, in: *Policy and Politics*, 34 (1), 115–132.
- SEVAL*, 2000: Evaluations-Standards der Schweizerischen Evaluationsgesellschaft (SEVAL-Standards), Schweizerischen Evaluationsgesellschaft (SEVAL).
- Solesbury, William*, 2002: The Ascendancy of Evidence, in: *Planning Theory & Practice* 3 (1), 90-96.
- Stockbauer, Uta*, 2000: Was macht Evaluationen nützlich? Grundlagen und empirische Untersuchungen zum Thema Verwertung und Verwertbarkeit von Evaluationen, Dissertation, Salzburg: Universität Salzburg.
- Suchman, Edward A.*, 1967: *Evaluative Research: Principles and Practice in Public Service & Social Action Programs*, New York.
- Tenbenseel, Tim*, 2004: Does more evidence lead to better policy?, in: *Policy Studies*, 25 (3), 189–207.
- Wegrich, Kai*, 2010: Governing 'Better Regulation' in Europe: The Logic, Limits of and Prospects for a 'middle-aged' Reform Policy, in: CESifo DICE Report 1/2010.
- Weiss, Carol H.*, 1974: *Evaluierungsforschung: Methoden zur Einschätzung von sozialen Reformprogrammen*, Opladen.
- Weiss, Carol H.*, 1988a: Evaluation for Decisions: Is Anybody There? Does Anybody Care?, in: *Evaluation Practice* 9 (1), 5-19.
- Weiss, Carol H.*, 1988b: If Program Decisions Hinged Only on Information: A Response to Patton in: *Evaluation Practice* 9 (3), 15-28.
- Weiss, Carol H./Murphy-Graham, Erin/Birkeland, Sarah*, 2005: An Alternate Route to Policy Influence: How Evaluations Affect D.A.R.E, in: *American Journal of Evaluation* 26 (1), 12-30.
- Wholey, Joseph S.*, 1983: *Evaluation and Effective Public Management*, Boston.
- Widmer, Thomas*, 2002: Staatsreformen und Evaluation: Konzeptionelle Grundlagen und Praxis bei den Schweizer Kantonen, in: *Zeitschrift für Evaluation* 1/2002, 101-114.
- Widmer, Thomas/deRocchi, Thomas*, 2012: *Evaluation – Grundlagen, Ansätze und Anwendungen*, Zürich/Chur.
- Valovirta, Ville*, 2002: Evaluation Utilization as Argumentation, in: *Evaluation* 8 (1), 60-80.
- Vatter, Adrian/Sager, Fritz*, 1996: Föderalismusreform am Beispiel des Ständemehrs, in: *Swiss Political Science Review* 2 (2), 165-200.