

Simone Ledermann

**„Evidence-Based Policy“:  
Die Rolle der Bundesverwaltung**

*Inauguraldissertation zur Erlangung der  
Würde eines Doctor Administrationis Rei  
Publicae der Wirtschafts- und  
Sozialwissenschaftlichen Fakultät der  
Universität Bern*

Die Fakultät hat diese Arbeit am  
23. Mai 2013 auf Antrag der Gutachter  
Prof. Dr. Fritz Sager, Prof. Dr. Frédéric  
Varone und Prof. Dr. Adrian Vatter als  
Dissertation angenommen, ohne damit  
zu den darin ausgesprochenen  
Auffassungen Stellung nehmen zu  
wollen.

KPM-Schriftenreihe Nr. 51

Die Schriftenreihe des Kompetenzzentrums für Public Management der Universität Bern wird herausgegeben von:

Prof. Dr. Andreas Lienhard  
Prof. Dr. Adrian Ritz  
Prof. Dr. Fritz Sager  
Prof. Dr. Reto Steiner

Die inhaltliche Verantwortung für jeden Band der KPM-Schriftenreihe liegt beim Autor resp. bei den Autoren.

Simone Ledermann

# **„Evidence-Based Policy“: Die Rolle der Bundesverwaltung**

KPM-Verlag  
Bern

*Dr. admin. publ. Simone Ledermann* ist Projektleiterin bei der Parlamentarischen Verwaltungskontrolle, dem Kompetenzzentrum der schweizerischen Bundesversammlung für Evaluationen.

Die Deutsche Bibliothek – CIP-Einheitsaufnahme

*„Evidence-Based Policy“: Die Rolle der Bundesverwaltung*

Ledermann Simone  
Bern 2014  
ISBN 978-3-906798-50-9

Alle Rechte vorbehalten  
Copyright © 2014 by KPM-Verlag Bern  
Jede Art der Vervielfältigung ohne Genehmigung des Verlags ist unzulässig.

[www.kpm.unibe.ch](http://www.kpm.unibe.ch)

---

## INHALTSVERZEICHNIS

<b>1</b>	<b>EINFÜHRUNG.....</b>	<b>1</b>
1.1	Evidence-Based Policy in der Schweiz.....	2
1.2	Fokus und Fragestellungen der Dissertation .....	4
1.3	Fazit.....	9
1.4	Bibliographie.....	11
<b>2</b>	<b>EVIDENZ UND EXPERTISE IM VORPARLAMENTARISCHEN PROZESS: DIE ROLLE VON VERWALTUNG UND EXTERNEN EXPERTEN.....</b>	<b>17</b>
2.1	Typologie politikrelevanten Wissens.....	20
2.2	Hypothesen .....	25
2.3	Empirisches Vorgehen.....	30
2.4	Resultate .....	35
2.5	Diskussion .....	44
2.6	Schlussfolgerungen .....	52
2.7	Bibliographie.....	56
2.8	Anhang .....	65
<b>3</b>	<b>EXPLORING THE NECESSARY CONDITIONS FOR EVALUATION USE IN PROGRAM CHANGE .....</b>	<b>67</b>
3.1	Conceptualization of Evaluation Use .....	69
3.2	A Context-Mechanism-Actor Model of Evaluation Use .....	71
3.3	Method.....	76
3.4	Findings of the Comparative Case Analysis .....	85
3.5	Discussion .....	91
3.6	Limitations of the Study .....	95
3.7	Conclusions .....	97
3.8	Bibliography.....	99
<b>4</b>	<b>PROBLEM ERKANNT, ABER NICHT GEBANNT – DER NUTZEN EINER VERKNÜPFUNG VON KONZEPT- UND UMSETZUNGSEVALUATION AM BEISPIEL DER STRATEGIE „MIGRATION UND GESUNDHEIT“ (MIT PROF. DR. FRITZ SAGER) .....</b>	<b>105</b>
4.1	Programmkonzept und Umsetzung als Evaluationsgegenstände.....	107
4.2	Empirisches Vorgehen und Bewertungskriterien .....	110
4.3	Strategiekonzept.....	111
4.4	Bewertung des Konzepts.....	115
4.5	Umsetzung der Strategie .....	120
4.6	Schlussfolgerungen .....	126
4.7	Bibliographie.....	129



## DANK

Ich danke allen, die in irgendeiner Weise zum Zustandekommen dieser Dissertation beigetragen haben. Ein besonderer Dank geht an Prof. Dr. Fritz Sager, Kompetenzzentrum für Public Management der Universität Bern, für die immer sehr prompte Unterstützung und aufmunternden Worte, dank welcher das Vorhaben doch noch zu einem Abschluss gekommen ist. Herzlich danken möchte ich meinen weiteren Gutachtern, Prof. Dr. Frédéric Varone und Prof. Dr. Adrian Vatter, die meinen Dissertationswunsch von Anfang an unterstützt haben. Prof. Dr. Pascal Sciarini gebührt ein grosses „Merci“ für seine Unterstützung während meiner Zeit an der Universität Genf. Bedanken möchte ich mich auch bei Prof. Dr. Andreas Balthasar von Interface und Dr. Jean-Daniel Delley, ehemals Titularprofessor an der Universität Genf, für die Gelegenheit, an den Forschungsprojekten mitzuwirken, aus welchen die Fachartikel für die Dissertation hervorgegangen sind. Meinen Kolleginnen und Kollegen in den verschiedenen Projekten danke ich für die angenehme Zusammenarbeit und den Austausch von Freud und Leid. Vielen Dank auch an das Team bei der Parlamentarischen Verwaltungskontrolle, das für meine zuweilen beschränkte Verfügbarkeit Verständnis zeigte. Meinen Freundinnen, Freunden und Verwandten werde ich nie vergessen, dass sie unerschütterlich an mich geglaubt haben. Der grösste Dank schliesslich gebührt Nik Streit, der mich mit einer Engelsgeduld durch die Höhen und Tiefen des Dissertationsprojekts begleitet und mir die Zeit dafür ermöglicht hat.

---

## 1 EINFÜHRUNG

Die vorliegende Dissertation besteht aus drei verwaltungswissenschaftlichen Fachartikeln zur Rolle der Bundesverwaltung in Bezug auf „Evidence-Based Policy“. „Evidence-Based Policy“ oder kurz EBP ist eine Reformbestrebung, die aus England stammt und darauf abzielt, Politik besser auf vorhandenes Wissen abzustützen. Tony Blairs Labour-Regierung förderte EBP ab 1999 als Teil einer Modernisierungsstrategie in Richtung einer ergebnisorientierten staatlichen Steuerung (Cabinet Office 1999a; Solesbury 2001). „What matters is what works“ lautete der EBP-Slogan, womit gemeint ist, dass Politik nicht nach ideologischen Überlegungen ausgestaltet werden soll, sondern danach, was erwiesenermassen funktioniert (Nutley et al. 2007: 125).

Ausgangspunkt für EBP war das Konzept der evidenzbasierten Medizin. Diese propagiert ein Vorgehen, bei dem die wirksamsten Interventionen mittels experimenteller klinischer Studien, so genannter „Randomized Controlled Trials“ (RCTs), eruiert werden (Hansen & Rieper 2010; Pawson et al. 2011). EBP stellt analog hierzu das Postulat auf, dass die vorhandene Evidenz zur Wirksamkeit von politischen Massnahmen bei der Politikformulierung berücksichtigt werden soll. Grundsätzlich ist diese Forderung nichts Neues, verfolgt die Politikanalyse doch seit ihrer Entstehung in den 1950er-Jahren das Ziel, wissenschaftliche Erkenntnisse in die Politik einzubringen (Parsons 2004). Unter dem Label EBP erhielten entsprechende Bestrebungen jedoch einen neuen Impetus (Sanderson 2004; Solesbury 2001; Van Dooren 2008). Gleichzeitig ist auch die Forschung über das Verhältnis von Wissen und Politik wieder erstarkt.

Von England aus hat EBP in verschiedenen Ländern Verbreitung gefunden (Head 2010a; Nutley et al. 2010), beispielsweise in Kanada (Howlett 2009), Dänemark (Hansen & Rieper 2010), den USA (Heinrich 2007), Südafrika (Strydom et al. 2010) oder in der EU (Radaelli & Meuwese 2009). Zudem hat sich EBP von der Gesundheitspolitik, in der sie zuerst Fuss fasste, auf andere Politikbereiche wie die Bildungs-, Sozial- oder Infrastrukturpolitik ausgedehnt (Frey & Ledermann 2010).



## 1.1 Evidence-Based Policy in der Schweiz

Die Schweiz gehört bei EBP-Reformen nicht zu den Vorreitern (Frey 2010); der Begriff des EBP taucht bisher hierzulande kaum auf. Dennoch sind auch in der Schweiz zumindest auf nationaler Ebene durchaus Bestrebungen hin zu einem auf gesicherten Erkenntnissen basierendem staatlichen Handeln zu erkennen (Frey & Widmer 2011). Beispielsweise ist die Evaluation seit den 1990er Jahren stark ausgebaut worden. Nicht nur ist die Zahl von Evaluationen, welche die Umsetzung und Wirkung von Politiken und Programmen untersuchen, gestiegen, sondern die Evaluationsfunktion ist beim Bund im internationalen Vergleich auch stark institutionell verankert worden (Balthasar 2007; Beywl & Widmer 2009; Bussmann 2008; Varone 2007; Widmer & Beywl 2009; Widmer & Neuenschwander 2004). Seit dem Jahr 2000 wird zudem bei sämtlichen Gesetzesvorlagen eine Regulierungsfolgenabschätzung durchgeführt. Insbesondere bei den seit 2006 durchgeführten vertieften Regulierungsfolgenabschätzungen werden die zu erwartenden Wirkungen dabei genau analysiert (Sager & Rissi 2011).

Wie stark Evidenz auf Bundesebene in die Politikgestaltung einfließt, bleibt jedoch kontrovers. Einerseits gibt es eher pessimistische Einschätzungen: Laut Klöti und Schneider (1989) werden wissenschaftliche Studien wenig verwendet. Auch gemäss Spinatsch (2006) sowie Widmer und Neuenschwander (2004: 404) fließen Evaluationsergebnisse nur selten in Gesetzgebungsprozesse ein und werden, wenn überhaupt, vor allem zur Legitimation genutzt. Bei der prospektiven Regulierungsfolgenabschätzung ist die Qualität der Prüfungen als mangelhaft und ihre Relevanz bei der Entscheidungsfindung als gering bezeichnet worden (OECD 2006; PVK 2005). Das Instrument der vertieften Regulierungsfolgenabschätzung ist zwar besser und wird auch genutzt (Rissi & Sager 2012), doch blieb seine Anwendung bisher auf sehr wenige Gesetzesvorhaben beschränkt. Als wichtigster Hemmfaktor für eine evidenzbasierte Politikgestaltung wird in der Schweizer Literatur die Konsens- und Kompromissorientierung genannt, die sich aus den direkt-demokratischen Instrumenten heraus entwickelt hat (Bussmann 1997, 2008; Klöti & Schneider 1989; Sager & Rissi 2011; Sager 2007a; Spinatsch 2002).

---

Am pointiertesten formulieren es Widmer und Neuenschwander (2004: 391):

„To avoid failures in subsequent plebiscites, administration, government and parliament strongly emphasize the need to ensure broad support prior to suggesting any new policy. Therefore policy making in Switzerland is oriented toward gathering support more than toward gathering evidence.“

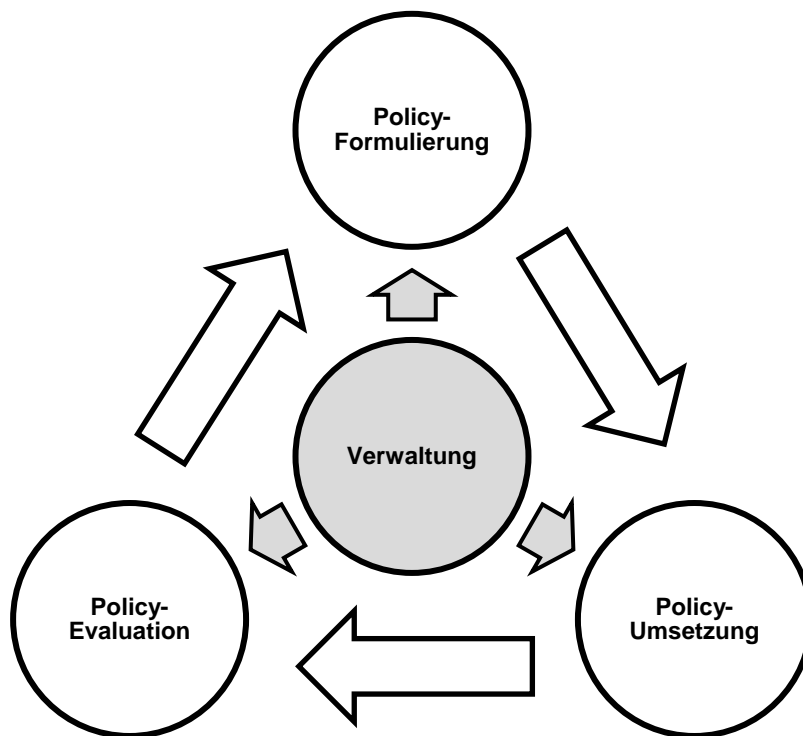
In anderen Studien wurde durchaus ein Einfluss von Evidenz festgestellt. Laut Balthasar (2007) werden knapp zwei Drittel der Evaluationen des Bundes intensiv verwendet und wird knapp die Hälfte der ans Parlament gerichteten Evaluationen von diesem tatsächlich genutzt. Auch Niederöst (2002) und Frey (2010) dokumentieren Fälle, welche einer idealtypisch evidenzbasierten Politikgestaltung sehr nahe kommen. In der Literatur werden dabei die vielfältigen Konsultationsmechanismen im vorparlamentarischen Prozess (Vernehmlassungen, Expertenkommissionen, Anhörungen, Ämterkonsultationen und Mitberichtsverfahren), die aus dem systembedingtem Kompromisserfordernis heraus entstanden sind, als wichtige Zugangskanäle für Wissen im Gesetzgebungsprozess genannt (Bussmann 2008; Germann 1985; Sciarini 2006). Zudem sei aufgrund der Mehrparteienregierung und der wechselnden Mehrheiten im Parlament häufig ein Interesse an kritischen Informationen vorhanden (Bussmann 2008). Schliesslich erzeuge der ausgeprägte Vollzugsföderalismus auf nationaler Ebene einen Informationsbedarf über die Politikumsetzung in den Kantonen, was der Evidenzgenerierung ebenfalls förderlich sei (Widmer & Frey 2006).

Laut in der jüngsten Untersuchung zu EBP in der Schweiz von Frey (2012) kann Evidenz innerhalb ein und derselben Gesetzesrevision bei einer Massnahme berücksichtigt und bei einer anderen ignoriert werden. Ausschlaggebend seien dabei eher fallspezifische als institutionelle Faktoren. Insgesamt zeichnet die bisherige Forschung jedenfalls kein klares Bild zur Bedeutung und den Bedingungen einer evidenzbasierten Politik in der Schweiz.

## 1.2 Fokus und Fragestellungen der Dissertation

Um zu einer Klärung der Möglichkeiten von EBP in der Schweiz beizutragen, fokussiert die Dissertation auf die Rolle der Bundesverwaltung, da diese gemäss bestehenden Studien für die Realisierung einer evidenzbasierten Politik in der Schweiz entscheidend ist (Frey 2012; Sager 2007b). Sie ist an allen drei Phasen des Politikzyklus beteiligt (vgl. Abbildung 1-1).

**Abbildung 1-1:** Zentrale Rolle der Verwaltung im Policy-Zyklus



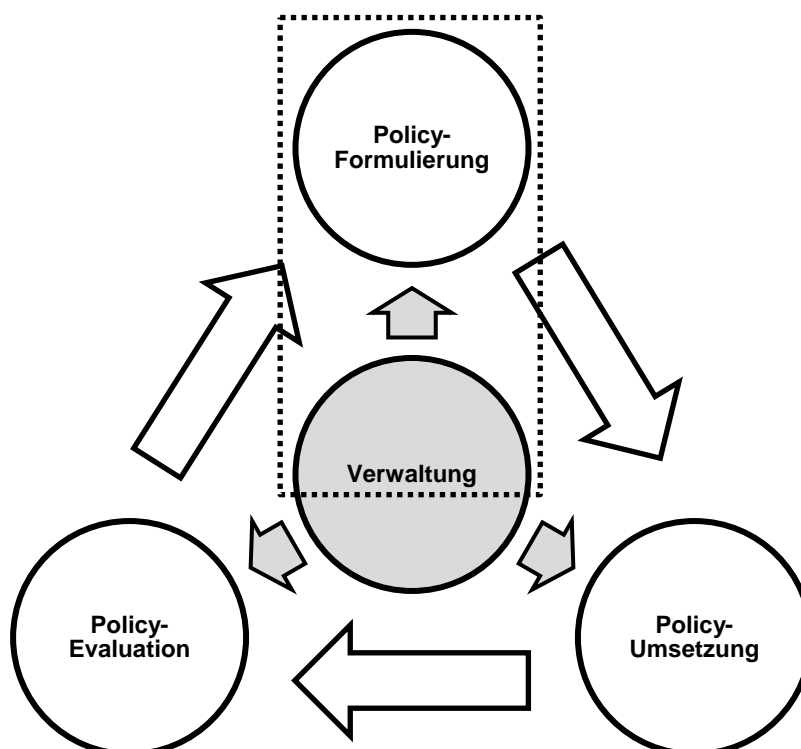
Erstens ist die Verwaltung auf Bundesebene prägend bei der Politikformulierung. Auch wenn bei Gesetzen der Entscheid beim Parlament liegt, so geschieht die Vorbereitung des Gesetzestexts im Wesentlichen dennoch durch die Verwaltung (Lüthi 2009). Bei der Ausführung der Gesetze ist es ebenfalls die Verwaltung, welche die massgeblichen Bestimmungen, Programme und Projekte formuliert. Sie ist es auch, welche die Policy umsetzt bzw. den Vollzug durch externe Akteure beaufsichtigt. Schliesslich initiiert sie die Evaluation, bereitet die Ergebnisse auf und speist sie die Policy-Revision ein. Inwiefern die Bundesverwaltung auf

den verschiedenen Stufen des Politikzyklus einen Beitrag zu einer evidenzbasierten Politik liefert, wird in der Dissertation näher analysiert.

Dabei schliesst die Dissertation an die drei zentralen Fragestellungen der wissenschaftlichen Debatte um evidenzbasierte Politik an (Sager & Pawson 2008): Was ist überhaupt Evidenz und wie wird sie generiert? Wann, wie und durch wen gelangt Evidenz in die Politik? Ist evidenzbasierte Politik letztlich auch wirksamer? Wie aus den nachfolgenden kurzen Beschreibungen der einzelnen Beiträge hervorgeht, steht jeweils eine dieser Fragen im Zentrum.

Der Artikel „Evidenz und Expertise im vorparlamentarischen Prozess“ setzt am Anfang des Policy-Zyklus bei der Politikformulierung an (vgl. Abbildung 1-2) und nimmt die Frage, was überhaupt als Evidenz zählt, als Ausgangspunkt. Der Artikel erscheint in überarbeiteter Form in der *Swiss Political Science Review*.

**Abbildung 1-2:** Verortung des Artikels „Evidenz und Expertise im vorparlamentarischen Prozess“



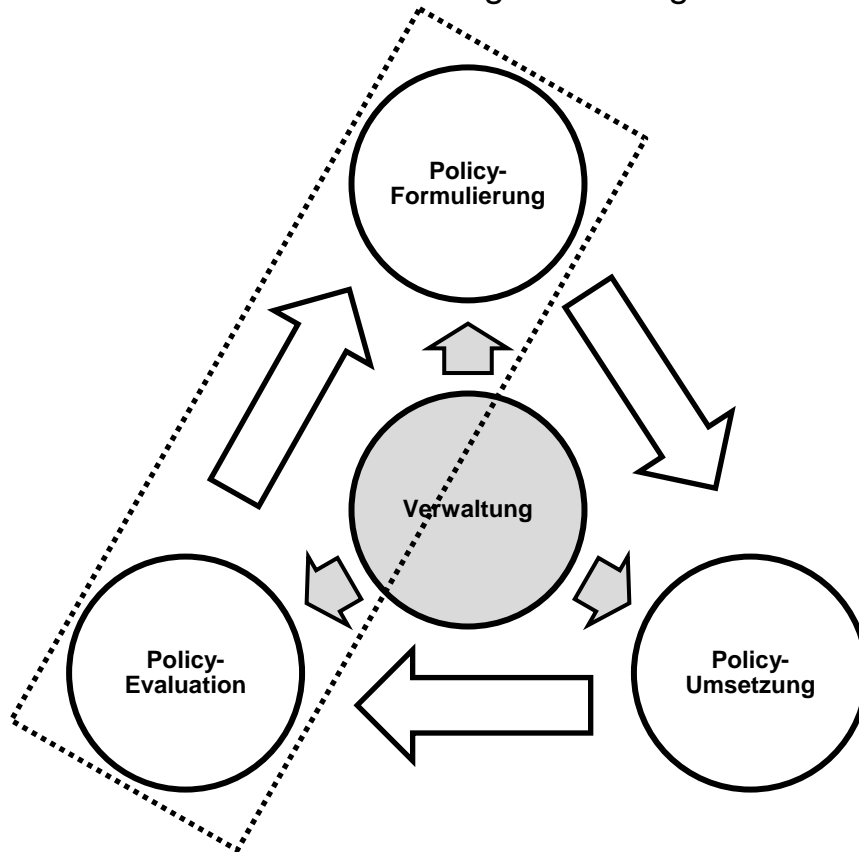
Zwar ist das Konzept der Evidenzbasierung mit der Übertragung von der Medizin auf den Bereich der Politik methodisch offener geworden (Hansen & Rieper 2010), doch bleibt es umstritten (Donaldson et al. 2008; Nutley et al. 2010). Der vorliegende Beitrag geht von der Prämisse aus, dass es verschiedene Arten politikrelevanten Wissens gibt, die es voneinander zu unterscheiden gilt. Es wird eine Typologie entwickelt, die neben Evidenz auch Expertise als Wissensart berücksichtigt und zusätzlich nach den Politikdimensionen *Policy*, *Politics* und *Polity* unterscheidet. Theoriegeleitet wird anschliessend die Rolle der Verwaltung und der von ihr beigezogenen externen Experten hinsichtlich des Einbezugs der verschiedenen Wissensarten bei der Formulierung von Gesetzen untersucht. Dabei wird die frühe Phase des vorparlamentarischen Verfahrens, die bisher nie eingehend erforscht wurde, anhand einer Analyse der vertraulichen Verwaltungsarchive zu vier Gesetzgebungsprozessen durchleuchtet. Es zeigt sich, dass die Verwaltung selber generell nur über wenig Expertise verfügt, sie diese Wissenslücke aber bis zu einem gewissen Grade über extern beigezogene Expertise sowie über Evidenz, die sie selber generiert oder durch Dritte generieren lässt, füllt. Systematische, nachvollziehbare Evidenz dient der Verwaltung bis zu einem gewissen Grade als Bollwerk gegen eine übermässige Abhängigkeit von externem, interessen gebundenem Wissen.

Der zweite Beitrag befasst sich mit der Frage, unter welchen Bedingungen Evidenz aus dem Stadium der Politik-Evaluation in die Reformulierung einer Policy einfließt (vgl. Abbildung 1-3). Er ist im Jahr 2012 unter dem Titel „Exploring the Necessary Conditions for Evaluation Use in Program Change“ im *American Journal of Evaluation* erschienen.

Die Forschung über die Nutzung von Evaluationen ist äusserst üppig und hat eine Vielzahl potenziell relevanter Faktoren identifiziert, doch sind die Erkenntnisse bisher insgesamt wenig schlüssig und zuweilen widersprüchlich. Im vorliegenden Artikel wird argumentiert, dass dies daran liegt, dass der Komplexität des Phänomens zu wenig Beachtung geschenkt wurde; weil es verschiedene Verwendungsarten gibt und diese jeweils von unzähligen Faktoren abhängen, lassen sich kaum allgemein gültige Ergebnisse generieren. Deshalb muss die Art der betrachteten

Nutzung klar eingegrenzt und der jeweilige Nutzungskontext berücksichtigt werden.

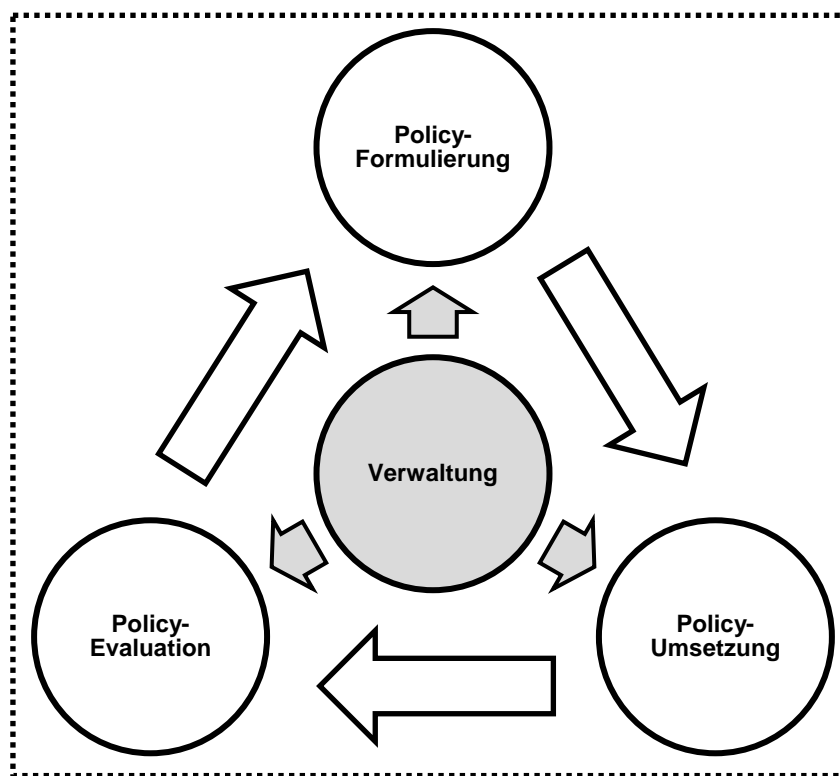
**Abbildung 1-3:** Verortung des Artikels „Exploring the Necessary Conditions for Evaluation Use in Program Change“



Im Beitrag wird die direkte, instrumentelle Nutzung von Evaluationen bei der Reformulierung von politischen Programmen als zu erklärendes Phänomen definiert und anschliessend nach dem Vorbild der realistischen Evaluationstheorie (Pawson & Tilley 1997) mittels kontextspezifischer Veränderungsmechanismen, welche Evaluationen auslösen können, modelliert. Zudem wird angesichts der Komplexität der Erklärungsanspruch zurückgenommen, indem es lediglich darum geht, kontextspezifische notwendige Bedingungen zu identifizieren. Empirisch wird das vorgeschlagene Modell anhand der Nutzung von elf Programmevaluationen der Direktion für Entwicklung und Zusammenarbeit – einer evaluationserfahrenen Einheit der Schweizer Bundesverwaltung – überprüft. Der Fallvergleich wird mittels „Qualitative Comparative Analysis“ (QCA) durchgeführt, wobei im Gegensatz zum üblichen Einsatz dieser Metho-

de, bei dem es um hinreichende Bedingungen geht, eine abgeänderte Form zur Analyse notwendiger Bedingungen zur Anwendung kommt. Der gewählte theoretische und methodische Ansatz ermöglicht neue Erkenntnisse. So zeigt sich beispielsweise, dass bestimmte in der Nutzungsforschung oft erwähnte Mechanismen wie der „Truth Test“ (Weiss & Bucuvalas 1980) nur unter bestimmten Voraussetzungen ausgelöst werden. Die Bedingungen, unter welchen ein Programm evaluiert wird, bestimmen somit, wie die Evidenz, welche die Evaluation hervorbringt, genutzt werden kann.

**Abbildung 1-4:** Verortung des Artikels „Problem erkannt, aber nicht gebannt – Der Nutzen einer Verknüpfung von Konzept- und Umsetzungs-evaluation am Beispiel der Strategie ‚Migration und Gesundheit‘“



Der dritte Fachartikel, der zusammen mit Fritz Sager verfasst wurde, liefert schliesslich eine Antwort auf die in der EBP-Literatur bisher weitgehend vernachlässigte Gretchenfrage, ob die Evidenzbasierung einer Policy für ihre Wirksamkeit eine Rolle spielt. Um diese Frage zu beantworten, ist es notwendig, den gesamten Politikzyklus zu betrachten: Es muss festgestellt werden, inwiefern Evidenz in die Politikformulierung

---

eingeflossen ist, ob die Policy planmässig umgesetzt wurde und inwieweit sie Wirkungen entfaltet hat (vgl. Abbildung 1-4). Dies wird im 2009 in der *Zeitschrift für Evaluation* publizierten Beitrag „Problem erkannt, aber nicht gebannt – Der Nutzen einer Verknüpfung von Konzept- und Umsetzungsevaluation am Beispiel der Strategie ‚Migration und Gesundheit‘“ geleistet.

Um die drei Stadien des Politikzyklus systematisch miteinander in Verbindung zu stellen, wurde in Anlehnung an die Literatur zur Policy-Analyse (Knoepfel et al. 1997, 2011: 157ff.) ein umfassendes Untersuchungsrastrer entwickelt. Dessen empirische Anwendung anhand der Strategie „Migration und Gesundheit“ des Bundes zeigt exemplarisch die Grenzen von EBP auf: Angesichts mangelnden Wissens zur Ausgangslage formulierte die Bundesverwaltung nur sehr allgemeine strategische Vorgaben, die angesichts des von Anfang an sehr beschränkten Wirkungspotentials der Strategie nicht zu erreichen waren. Zudem wurde die spärlich vorhandene Evidenz bei der Strategieformulierung in einzelnen Punkten nicht berücksichtigt, was sich negativ auf die Wirksamkeit der entsprechenden Massnahmen auswirkte. Die bescheidene Wirkungsbilanz der Strategie kann jedoch nur zum Teil auf Defizite bei der Evidenzbasierung zurückgeführt werden; zusätzlich waren auch deutliche Vollzugsmängel feststellbar. EBP allein kann somit die Wirksamkeit von Politik nicht sicherstellen.

### **1.3            Fazit**

Die Dissertation unterstreicht die Rolle der Bundesverwaltung als zentrale Akteurin in Bezug auf ein EBP in der Schweiz. Die Verwaltung ist nicht nur für die Evidenzbasierung auf der Ebene von Programmen und Projekten ausschlaggebend, sondern – wie namentlich der erste Beitrag zeigt – auch bei der Formulierung von Gesetzen, indem sie einerseits selber Evidenz beiträgt und andererseits den Evidenzbeitrag externer Akteure steuert. Als faktisch formulierende Instanz ist die Verwaltung dafür verantwortlich, dass die vorhandene Evidenz in der Policy auch tatsächlich ihren Niederschlag findet. Dies geschieht in unterschiedlichem



Ausmass, je nach fallspezifischem Kontext. Über die drei Beiträge hinweg als wichtig erweisen sich die Art des vorhandenen Wissens, die politikanalytische Kapazität der zuständigen Verwaltungsakteure sowie die Konfliktivität der Policy. Wie auch bei Frey (2012) stellen dagegen die institutionellen Eigenheiten des politischen Systems, die in der Schweizer Literatur prominent behandelt werden, keine klar erkennbaren Hinderungs- oder Förderfaktoren für EBP dar. Um deren Bedeutung systematisch abzuschätzen, ist allerdings ein Ländervergleich notwendig. Es wäre deshalb zu begrüessen, wenn künftige Forschungsvorhaben hier ansetzen würden.

Des Weiteren verdeutlicht gerade der dritte Beitrag, dass so genannte „Black Box“-Studien, welche zwar die Wirkungen von Massnahmen messen, deren Umsetzung dagegen ausser Acht lassen, ungeeignet sind, um Aussagen zur Bedeutung von EBP und letztlich zur Wirksamkeit von Politik zu machen. Denn während das klinische Umfeld der evidenzbasierten Medizin eine gleichmässige Anwendung der Intervention und eine Randomisierung der Subjekte ermöglicht, ist das Umfeld, in dem Politik formuliert und umgesetzt wird, viel komplexer und weniger kontrollierbar. Dies wird in allen drei Beiträgen deutlich. Entsprechend müssen nicht nur die theoretischen Ansätze und Methoden, mittels welcher Aussagen über die Nutzung von Evidenz im Politikprozess gemacht werden, der Komplexität gerecht werden – worauf im Rahmen dieser Dissertation geachtet wurde –, sondern auch die Theorien und Methoden, mittels welcher überhaupt erst Evidenz über die Wirksamkeit von Politik generiert wird, müssen diesen Umstand berücksichtigen. Gültige Evidenz kann deshalb nicht alleine durch RCTs generiert werden; vielmehr kann solche Evidenz wohl nur in Ausnahmefällen durch RCTs generiert werden. In den meisten Fällen sind dagegen Forschungsdesigns zu wählen, die den Unwägbarkeiten der Umsetzung von politischen Massnahmen Rechnung tragen. In diesem Sinne unterstreicht die Dissertation die Notwendigkeit einer theoretischen und methodischen Offenheit des EBP-Konzepts.

Schliesslich setzt die Komplexität des Politikprozesses, die in allen drei Beiträgen offensichtlich wird, dem EBP-Ansatz Grenzen. Politik kann

---

immer nur bis zu einem gewissen Grade evidenzbasiert sein; es bleibt jeweils eine Vielzahl von Unsicherheitsfaktoren bestehen. Die von gewissen EBP-Kritikern befürchtete rein technokratische Politikgestaltung (z.B. Parsons 2004; Sanderson 2002) ist deshalb faktisch gar nicht möglich. Auch wenn die Stärke der Evidenzbasierung variiert, so scheint die Realität zumindest in der Schweiz generell weit vom EBP-Ideal einer rein wissens- statt ideologiegeleiteten Politikgestaltung entfernt.

#### 1.4 Bibliographie

- Balthasar, A. (2007). *Institutionelle Verankerung und Verwendung von Evaluationen*. Chur/Zürich: Rüegger.
- Beywl, W., & Widmer, T. (2009). Evaluation in Expansion: Ausgangslage für den intersektoralen Dreiländer-Vergleich. In T. Widmer, W. Beywl, & C. Fabian (Hg.), *Evaluation: ein systematisches Handbuch* (pp. 13–23). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Bussmann, W. (1997). Evaluationen in der Schweiz. In W. Bussmann, U. Klöti, & P. Knoepfel (Hg.), *Einführung in die Politikevaluation* (pp. 13–35). Basel, Frankfurt a.M.: Helbing & Lichtenhahn.
- Bussmann, W. (2008). The Emergence of Evaluation in Switzerland. *Evaluation*, 14(4), 499–506.
- Cabinet Office (1999). *Modernising Government. Whitepaper*. London.
- Donaldson, S. I., Christie, C. A., & Mark, M. M. (2008). *What Counts as Credible Evidence in Applied Research and Evaluation Practice?* Newbury Park, CA: Sage.
- Frey, K. (2010). Revising Road Safety Policy: The Role of Systematic Evidence in Switzerland. *Governance*, 23(4), 667–690.
- Frey, K. (2012). *Evidenzbasierte Politikformulierung in der Schweiz: Gesetzesrevisionen im Vergleich*. Baden-Baden: Nomos.
- Frey, K., & Ledermann, S. (2010). Evidence-Based Policy: A Concept in Geographical and Substantive Expansion. *German Policy Studies*, 6(2), 1–15.

- Frey, K., & Widmer, T. (2011). Revising Swiss Policies: The Influence of Efficiency Analyses. *American Journal of Evaluation*, 32(4), 494–517.
- Germann, R. E. (Hg.) (1985). *Experts et commissions de la confédération*. Lausanne: Presses polytechniques romandes.
- Hansen, H. F., & Rieper, O. (2010). The Politics of Evidence-Based Policy-Making: The Case of Denmark. *German Policy Studies*, 6(2), 87–112.
- Head, B. W. (2010). Reconsidering Evidence-Based Policy: Key Issues and Challenges. *Policy and Society*, 29(2), 77–94.
- Heinrich, C. J. (2007). Evidence-Based Policy and Performance Management: Challenges and Prospects in Two Parallel Movements. *American Review of Public Administration*, 37(3), 255–277.
- Howlett, M. (2009). Policy Analytical Capacity and Evidence-Based Policy-Making: Lessons from Canada. *Canadian Public Administration*, 52(2), 153–175.
- Klöti, U., & Schneider, G. (1989). *Informationsbeschaffung des Gesetzgebers*. Grösch: Rüegger.
- Knoepfel, P., Larrue, C., Varone, F., & Hill, M. (2011). *Public Policy Analysis*. Bristol: The Policy Press.
- Knoepfel, P., Varone, F., Bussmann, W., & Mader, L. (1997). Evaluationsgegenstände und Evaluationskriterien. In W. Bussmann, U. Klöti, & P. Knoepfel (Hg.), *Einführung in die Politikevaluation* (pp. 78–118). Basel, Frankfurt a.M.: Helbing & Lichtenhahn.
- Lüthi, R. (2009). Im Spannungsfeld zwischen Verwaltung und Politik: Erfahrungen aus der Erarbeitung von Erlassentwürfen durch parlamentarische Kommissionen. *LeGes*, 2009(3), 363–374.
- Nideröst, B. (2002). Erfolgsbedingungen für Evaluationen. *LeGes*, 13(1), 39–55.
- Nutley, S., Morton, S., Jung, T., & Boaz, A. (2010). Evidence and Policy in Six European Countries: Diverse Approaches and Common Challenges. *Evidence & Policy*, 6(2), 131–145.

- 
- Nutley, S., Walter, I., & Davies, H. T. O. (2007). *Using Evidence: How Research Can Inform Public Services*. Bristol: The Policy Press.
- OECD (2006). *Regulatory Reform in Switzerland: Enhancing Market Openness through Regulatory Reform*. Paris.
- Parsons, W. (2004). Not Just Steering But Weaving: Relevant Knowledge and the Craft of Building Policy Capacity and Coherence. *Australian Journal of Public Administration*, 63(March), 43–57.
- Pawson, R., & Tilley, N. (1997). *Realistic Evaluation*. London, Thousand Oaks, New Delhi: Sage.
- Pawson, R., Wong, G., & Owen, L. (2011). Known Knowns, Known Unknowns, Unknown Unknowns: The Predicament of Evidence-Based Policy. *American Journal of Evaluation*, 32(4), 518–546.
- PVK (2005). *Die drei „KMU-Tests“ des Bundes: bekannt? genutzt? wirkungsvoll?* Bern: Parlamentarische Verwaltungskontrollstelle (PVK).
- Radaelli, C. M., & Meuwese, A. C. M. (2009). Better Regulation in Europe: Between Public Management and Regulatory Reform. *Public Administration*, 87(3), 639–654.
- Rissi, C., & Sager, F. (2012). Types of knowledge utilization of regulatory impact assessments: Evidence from Swiss policymaking. *Regulation & Governance*. “Early View”-Publikation. doi:10.1111/rego.12002.
- Sager, F. (2007a). Making transport policy work: polity, policy, politics and systematic review. *Policy & Politics*, 35(2), 269–288.
- Sager, F. (2007b). Habermas’ Models of Decisionism, Technocracy, and Pragmatism in Times of Governance: The Relationship of Public Administration, Politics, and Science in the Alcohol Prevention Policies of the Swiss Member States. *Public Administration*, 85(2), 429–447.
- Sager, F., & Pawson, R. (2008). The Politics of Evidence-Based Policy-Making: Workshop Description. *ECPR Joint Sessions of Workshops*. Rennes.

- Sager, F., & Rissi, C. (2011). The limited scope of policy appraisal in the context of referendum democracy: The case of regulatory impact assessment in Switzerland. *Evaluation*, 17(2), 151–163.
- Sanderson, I. (2002). Evaluation, Policy Learning and Evidence-Based Policy Making. *Public Administration*, 80(1), 1–22.
- Sanderson, I. (2004). Getting Evidence into Practice: Perspectives on Rationality. *Evaluation*, 10(3), 366–379.
- Sciarini, P. (2006). Le processus législatif. In U. Klöti, P. Knoepfel, H. Kriesi, W. Linder, & Y. Papadopoulos (Hg.), *Handbuch der Schweizer Politik* (4. Aufl., pp. 491–525). Zürich: Neue Zürcher Zeitung.
- Solesbury, W. (2001). Evidence Based Policy: Whence it Came and Where it's Going. London.
- Spinatsch, M. (2002). Evaluation in Switzerland: Moving Toward a Decentralized System. In J.-E. Furubo, R. C. Rist, & R. Sandahl (Hg.), *International Atlas of Evaluation* (pp. 375–391). New Brunswick, London: Transaction Publishers.
- Spinatsch, M. (2006). Kompetent, hartnäckig, konstruktiv: Die Hauptabteilung Staats- und Verwaltungsrecht des Bundesamtes für Justiz im Urteil ihrer Partner in der Bundesverwaltung. Bern: Dr. Markus Spinatsch, Beratung für Politik und Verwaltung.
- Strydom, W. F., Funke, N., Nienaber, S., Nortje, K., & Steyn, M. (2010). Evidence-based policymaking: A review. *South African Journal of Science*, 106(5/6), 1–8.
- Van Dooren, W. (2008). Nothing New Under the Sun? Change and Continuity in the Twentieth-Century Performance Movements. In W. Van Dooren & S. Van de Walle (Hg.), *Performance Information in the Public Sector: How it is Used* (pp. 11–23). Basingstoke: Palgrave Macmillan.
- Varone, F. (2007). Développer les capacités d'évaluation. In S. Jacob, F. Varone, & G. J.-L. (Hg.), *L'évaluation des politiques au niveau régional* (pp. 157–177). Bruxelles: Lang.

- Weiss, C. H., & Bucuvalas, M. J. (1980). Truth Tests and Utility Tests: Decision-Makers' Frames of Reference for Social Science Research. *American Sociological Review*, 45(2), 302–313.
- Widmer, T., & Beywl, W. (2009). Divergente Entwicklungspfade der Evaluation. In T. Widmer, W. Beywl, & C. Fabian (Hg.), *Evaluation: ein systematisches Handbuch* (pp. 509–527). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Widmer, T., & Frey, K. (2006). Evaluation von Mehrebenen-Netzwerkstrategien. *Zeitschrift für Evaluation*, 2006(2), 287–316.
- Widmer, T., & Neuenschwander, P. (2004). Embedding Evaluation in the Swiss Federal Administration: Purpose, Institutional Design and Utilization. *Evaluation*, 10(4), 388–409.



---

## 2 EVIDENZ UND EXPERTISE IM VORPARLAMENTARISCHEN PROZESS: DIE ROLLE VON VERWALTUNG UND EXTERNEN EXPERTEN<sup>1</sup>

SIMONE LEDERMANN

*Unter dem Schlagwort „Evidence-Based Policy“ hat die Frage, wie Wissen in die Politikformulierung einfließt, an Aktualität gewonnen. Allerdings ist der Evidenzbegriff diffus geblieben und stösst der Ansatz einer evidenzbasierten Politik angesichts der Komplexität von Politikprozessen an Grenzen. In diesem Artikel wird deshalb eine Typologie politikrelevanten Wissens entwickelt, die neben Evidenz auch Expertise berücksichtigt sowie nach den Politikdimensionen Inhalt, Prozess und Recht unterscheidet. Damit lässt sich der Wissensbeitrag der Akteure zur Politikformulierung differenziert analysieren. Die empirische Untersuchung von vier vorparlamentarischen Gesetzgebungsverfahren in der Schweiz zeigt, dass die federführenden Verwaltungsakteure über beschränkte Expertise verfügen, diesen Mangel jedoch je nach Politikdimension verwaltungsintern oder durch externe Experten sowie durch Evidenz kompensieren. Der Wissensbeitrag der Akteure wird durch den fallspezifischen politischen und institutionellen Kontext geprägt. Schliesslich gibt es Hinweise, dass der Einbezug externer Experten zwar funktional angebracht ist, jedoch ein potentielles demokratisches Risiko darstellt.*

Die Politikanalyse strebt seit ihren Anfängen in den 1950er Jahren danach, Wissen im und für den Politikprozess bereitzustellen (Parsons 2004). Diese Bestrebungen erlebten unter dem Label „Evidence-Based Policy“ (EBP) im vergangenen Jahrzehnt einen Aufschwung (Sanderson 2004; Solesbury 2001; Van Dooren 2008). Ursprünglich von Blairs Labour-Regierung gefördert, steht EBP mit dem Slogan „What matters is

---

<sup>1</sup> Dieser Beitrag entstand im Rahmen einer Untersuchung, die vom Schweizerischen Nationalfonds unterstützt wurde (Projekt Nr. 113420, « La conception des lois. La démarche légistique entre théorie et pratique »). Es handelt sich um die Vorversion des folgenden Artikels: Ledermann, S. (2014). Evidenz und Expertise im vorparlamentarischen Gesetzgebungsprozess: Die Rolle von Verwaltung und externen Experten. *Swiss Political Science Review*, doi: 10.1111/spsr.12114.



what works“ für einen Staat, der seine Politik nicht nach ideologischen Gesichtspunkten gestaltet, sondern danach, was erwiesenermassen funktioniert (Cabinet Office 1999a; Nutley et al. 2007: 127; Solesbury 2001). Seither hat EBP in verschiedenen Politikbereichen und Ländern Verbreitung gefunden (Head 2010a; Nutley et al. 2010).

Was als Evidenz zählt, ist indessen unklar (Frey & Ledermann 2010). Die Labour-Regierung hat Evidenz ursprünglich 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 1999b: 33)

Häufig werden nur wissenschaftliche Erkenntnisse als Evidenz gezählt, wobei umstritten ist, welche Forschungsmethoden verlässliche Evidenz hervorbringen (Davies et al. 2000: ch. 1; Donaldson et al. 2008).

Nicht nur ist der Evidenzbegriff uneinheitlich, der EBP-Ansatz stösst ausserdem angesichts der Komplexität von Politik an Grenzen. Einer rationalen Logik folgend, sollten verschiedene Alternativen anhand von möglichst viel und guter Evidenz überprüft werden, damit schliesslich jene Massnahme ausgewählt wird, die nachweislich die besten Ergebnisse hervorbringt. Doch bereits Simon (1947) zeigte auf, dass die rationale Logik angesichts einer Vielzahl von Einflussfaktoren scheitert, so dass die Folgen staatlichen Handelns bis zu einem gewissen Grade unvorhersehbar bleiben. Die Komplexität beeinträchtigt zudem die Mess- und Vorhersehbarkeit der Wirkungen staatlichen Handelns, was die Aussagekraft von wissenschaftlichen Erkenntnissen schmälert (Sanderson 2006; Schneider 2008; Weiss et al. 2008). Häufig ist bereits das Ausgangsproblem nicht eindeutig abgegrenzt (Parsons 2004: 50), womit unklar bleibt, welche Massnahmen für die Problemlösung überhaupt in Betracht kommen und somit einer evidenzbasierten Überprüfung unterzogen werden sollten. Gleichzeitig sind aufgrund der Komplexität und Kontextabhängigkeit zu jeder einzelnen politischen Massnahme so

---

vielfältige wissenschaftliche Abklärungen mit entsprechend grossem Aufwand notwendig (vgl. das Fallbeispiel von Pawson et al. 2011), dass nur eine sehr begrenzte Anzahl von Massnahmen überprüft werden kann. Ferner muss Evidenz aufbereitet werden, damit sie von politischen Akteuren zur Kenntnis genommen wird (Frey 2012; Tenbenschel 2004). EBP stösst somit aufgrund der Komplexität an praktische Grenzen. Politik kann deshalb nie bis ins letzte evidenzbasiert sein, sondern es handelt sich um ein Kontinuum (Monaghan 2008), weshalb in der Literatur auch von „evidence-informed“, „evidence-inspired“ oder „evidence-aware“ gesprochen wird (Nutley et al. 2003, 2010).

Um der Begrenztheit des EBP-Ansatzes zu begegnen, wurde verschiedentlich vorgeschlagen, neben wissenschaftlichen Studien andere Arten von Wissen wie namentlich Praxis- und Erfahrungswissen als Teil politikrelevanter Evidenz zu berücksichtigen (Head 2008; Nutley et al. 2003, 2007; Parsons 2004; Sanderson 2009, 2010; Strydom et al. 2010; Vreugdenhil & Ker Rault 2010). Diese weiteren Arten von Wissen wurden allerdings nicht klar abgegrenzt. Zudem erscheint eine Ausweitung des Evidenzbegriffs nicht wünschenswert, da angenommen werden muss, dass verschiedene Wissensarten eine unterschiedliche Rolle bei der Politikgestaltung spielen. In diesem Artikel werden stattdessen verschiedene Arten politikrelevanten Wissens unterschieden. Neben wissenschaftlicher Evidenz wird Expertise als personengebundenes Wissen, das durch die langwierige Auseinandersetzung mit einem Thema erworben wird, berücksichtigt. Evidenz und Expertise werden ausserdem danach typisiert, auf welche Politikdimension sie sich beziehen.

Wissen kann durch unterschiedliche Akteure in die Politikformulierung eingebracht werden. EBP fokussierte bisher stark auf die Rolle der Verwaltung. In vielen Ländern der westlichen Welt ist es zudem gang und gäbe, externe Experten<sup>2</sup> beizuziehen. Im Theorieteil werden Hypothesen dazu formuliert, welchen Beitrag Verwaltung und Experten zu den verschiedenen Arten von Evidenz und Expertise leisten. Diese Hypothesen werden anhand von vier Gesetzgebungsprozessen im schweizerischen

---

<sup>2</sup> In diesem Beitrag wird generell die männliche Form verwendet, wobei weibliche Personen immer mitgemeint sind.

Bundesstaat überprüft. Betrachtet wird die erste Phase des vorparlamentarischen Prozesses (bis zum Vernehmlassungsentwurf), während welcher der stärkste Einbezug von Wissen zu erwarten ist und die bisher noch nicht vertieft untersucht worden ist. Der gewählte Forschungskontext zeichnet sich durch eine hohe Komplexität aus: Es wird die Formulierung neuer Gesetze bzw. die Totalrevision von Gesetzen untersucht, deren Gegenstand zu Beginn noch nicht festgelegt ist, womit offen ist, welches Wissen überhaupt relevant ist. Zudem enthalten Gesetze üblicherweise eine Vielzahl von unterschiedlich ausgerichteten Massnahmen (Bussmann 2010). Gleichzeitig werden aufgrund der in der Schweiz starken Kompromissorientierung, die sich aus den direkt-demokratischen Instrumenten heraus entwickelt hat (Neidhart 1970), zahlreiche Akteure an der Politikformulierung beteiligt. Gemäss dem allgemeinen Tenor der Literatur setzt das Kompromisserfordernis einem evidenzbasierten Ansatz klar Grenzen (Bussmann 1997, 2008; OECD 2006; PVK 2005; Sager & Rissi 2011; Sager 2007a; Spinatsch 2002, 2006; Widmer & Neuenschwander 2004). Es stellt sich jedoch die Frage, inwiefern die verschiedenen Akteure Expertise einbringen, welche eine allenfalls mangelnde Evidenz kompensiert.

Im Diskussionsteil werden die empirischen Resultate vergleichend mit Bezug auf die Hypothesen reflektiert und die Determinanten des Einbezugs der verschiedenen Wissensarten diskutiert. Daraus werden die Schlussfolgerungen gezogen, wobei der Bezug externer Experten auch aus demokratietheoretischer Sicht kritisch betrachtet wird.

## **2.1 Typologie politikrelevanten Wissens**

Evidenz wird in der Literatur von sehr eng (z.B. Coalition for Evidence-Based Policy 2012) bis zu sehr breit (z.B. Nutley et al. 2007) definiert. Das Problem einer engen Begriffsdefinition besteht darin, dass viel für die Politikformulierung relevantes Wissen unbeachtet bleibt, während ein breiter Evidenzbegriff Gefahr läuft, dass durch die Subsummierung verschiedener Wissensarten, die durch unterschiedliche Akteure eingebracht werden und durch unterschiedliche Determinanten bestimmt wer-

---

den, Kausalzusammenhänge nicht mehr erkannt werden. In der Folge wird deshalb eine analytische Trennung verschiedener politikrelevanter Wissensarten entwickelt. Dabei wird zunächst eine Unterscheidung zwischen Evidenz und Expertise eingeführt. Diese beiden Wissensarten werden anschliessend anhand von drei Politikdimensionen weiter differenziert.

Der vorliegende Beitrag wählt bezüglich *Evidenz* einen Begriff mittlerer Breite: Evidenz wird verstanden als forschungsbasierte Information, ohne dass dabei eine bestimmte Forschungsmethode vorausgesetzt würde. Evidenz muss den beiden Kernmerkmalen von Wissenschaftlichkeit entsprechen (Sager & Ledermann 2008): Sie muss einerseits mittels eines *systematischen Verfahrens* generiert und andererseits in *intersubjektiv nachvollziehbarer* Form festgehalten worden sein. Gefordert wird somit nicht nur ein systematisches (Nutley et al. 2003: 128) oder strenges (Head 2010b: 77: „rigorous“) Vorgehen, sondern zusätzlich, was Frey die Transparenz der Ergebnisse nennt:

„Das Kriterium des *systematischen* Verfahrens verlangt, dass die Informationen nach den Regeln der qualitativen und/oder quantitativen Methoden erhoben werden. Die Darstellung und Interpretation der erhobenen Informationen muss *transparent* erfolgen, damit die Angemessenheit bzw. die Validität und Reliabilität der Erkenntnisse beurteilt werden kann.“ (Frey 2012: 22, Hervorhebungen im Original)

Evidenz muss somit unabhängig beobachtet und überprüft werden können (Rycroft-Malone et al. 2004). Dafür müssen die Informationen objektgebunden sein, d.h. sie müssen physisch auf Papier oder elektronisch in einem Computer bzw. auf einem Server schriftlich festgehalten und zugänglich sein. Die Urheberschaft von Evidenz wird bewusst offen gelassen; es kann sich um wissenschaftliche Studien von Hochschulen oder Auftragsarbeiten von Privaten handeln, aber auch um Informationen, welche die Verwaltung selber zusammenträgt, sofern sie den obigen Anforderungen genügen. Evidenz kann sowohl spezifische staatliche Interventionen als auch generelle gesellschaftliche Phänomene und Entwicklungen betreffen (Head 2008: 4).

Durch die Objektivierung wird Evidenz statisch, d.h. sie ist nicht so schnell veränderbar. Dies unterscheidet die objektgebundene Evidenz von personengebundener *Expertise*. Als Experte wird im allgemeinen Sprachgebrauch eine Person bezeichnet, die auf einem bestimmten Gebiet überdurchschnittlich viel weiss, sei dies, weil sie es beforscht hat, viel Erfahrung darin hat oder sich intensiv damit beschäftigt hat. *Expertise* geht gemäss der Kognitionspsychologie allerdings über reines Wissen hinaus und umfasst auch die Fähigkeit, Kenntnisse situationsgerecht flexibel abzurufen und anzuwenden (Dreyfus & Dreyfus 2005; Gigerenzer 2008). Experten zeichnen sich dadurch aus, dass sie auch komplexe Situationen schnell erfassen und deuten können (Chi et al. 1988). Um diese Fähigkeit zu erlangen, ist eine *langjährige* und *bewusste Beschäftigung* mit einem Thema notwendig. Gemeinhin werden in der Literatur zehn Jahre als notwendig erachtet, damit jemand *Expertise* erlangen kann (Ericsson 2006).

Durch Expertenbefragungen kann *Expertise* systematisch erhoben und zu Evidenz werden. Dabei geht die Flexibilität und situative Anwendung, die *Expertise* gegenüber Evidenz auszeichnet, allerdings bis zu einem gewissen Grade verloren. Zudem umfasst *Expertise* neben explizitem auch implizites Wissen (zu den verschiedenen Wissensarten vgl. Nutley et al. 2003), das auch mithilfe bestimmter Techniken (vgl. Bekker et al. 2010; Rycroft-Malone et al. 2004; Tourmen 2009) nur zum Teil explizierbar ist. Um implizites Wissen zu berücksichtigen, wie dies verschiedentlich in der EBP-Literatur gefordert wurde (Hansen & Rieper 2010; Parsons 2004; Sanderson 2010; Tenbenschel 2004; Weiss et al. 2008), ist der direkte Einbezug von Experten in den Politikprozess deshalb unabdingbar.

Dem EBP-Slogan „What matters is what works“ folgend, fokussierte EBP bisher auf Wissen über die Zielerreichung von politischen Massnahmen. In der Literatur wurde verschiedentlich darauf hingewiesen, dass für eine erfolgreiche wissensbasierte Politikformulierung weiteres Wissen relevant ist (vgl. z.B. Head 2008, 2010a; Nutley et al. 2003; Pawson et al. 2011). Dieses politikrelevante Wissen wird in der Folge in Anlehnung an die drei Politikdimensionen *Policy*, *Politics* und *Polity* (Schubert &

Bandelow 2003) in Wissen zum Inhalt der Politik, Wissen zum politischen Prozess und Wissen zum Recht als institutioneller Rahmen gegliedert. Alle drei Kategorien von Wissen können sowohl in Form von Evidenz als auch von Expertise vorliegen (vgl. Abbildung 2-1).

**Abbildung 2-1:** Typologie politikrelevanten Wissens



(1) *Wissen zum Inhalt der Politik*: Die inhaltliche Dimension von Politik wird im Englischen als *Policy* bezeichnet. Das für die Politikformulierung notwendige substantielle Wissen kann, mit Rückgriff auf die Wissenstypologie von Nutley, Walter und Davies (2003: 128), in drei Bestandteile eingeteilt werden: (a) *Know-what works*: Die Kenntnis dessen, welche Interventionen die erwünschten Wirkungen zeitigen. Dies entspricht dem ursprünglichen Kern von EBP. (b) *Know-about problems*: Wissen über das Problem, das durch die Intervention gelöst werden soll, und seine Ursachen. Welche Gruppen sind vom Problem betroffen? Wie stark sind die Beeinträchtigungen? Was sind die Auslöser des Problems und wie entwickelt es sich voraussichtlich in Zukunft? Um eine zielgerichtete politische Massnahme zu formulieren, müssen diese Fragen beantwortet werden (vgl. auch Pawson et al. 2011: 520, Box 2, Frage 1). (c) *Know-how to put into practice*: Wissen, wie die Massnahmen umgesetzt werden können, wird in der Literatur als sehr wichtig erachtet (Ledermann & Sager 2009; Parsons 2004; Sanderson 2010; Tenbensen 2004; Weiss et al. 2008). Dieses praktische Vollzugswissen (practical implementation knowledge bei Head 2008) umfasst insbesondere das Wissen der Fachpersonen, die mit den Adressaten der Massnahmen in Kontakt stehen. Unter Umständen verfügen auch die Adressaten („Kunden“) der

Massnahmen über dieses Wissen (Head 2010a). Letztlich geht es darum, dass die formulierten Gesetze auch tatsächlich umsetzbar sind (vgl. auch Pawson et al. 2011: 520, Box 2, Frage 4)

(2) *Wissen zum Prozess*: Hierbei handelt es sich um das Wissen zur Prozessdimension von Politik, der mit dem englischen *Politics* erfasst wird. Es geht um Wissen zur Gestaltung von Politikformulierungsprozessen, also dazu, *wie* man Gesetze macht. Es umfasst die Kenntnis der verschiedenen Phasen eines Politikformulierungs- bzw. Gesetzgebungsprozesses. Welche Grundlagen sind zu welchem Zeitpunkt notwendig, welche Konsultationen werden wann durchgeführt usw. (vgl. z.B. Bundesamt für Justiz 2007). Dazu gehört auch, was Head (2008) politisches Wissen (*political knowledge*) nennt: die Kenntnis von Strategien, um Themen auf die politische Agenda zu setzen, sie zu einer Priorität zu machen, und die Hauptbotschaften für die politisch relevanten Akteure verständlich zu kommunizieren. Schliesslich gehört dazu Wissen, wie Verhandlungen geführt und Kompromisse geschmiedet werden und wie gegebenenfalls die Unterstützung der Bevölkerung sicherzustellen ist (vgl. auch Pawson et al. 2011: 520, Box 2, Fragen 2 und 3).

(3) *Rechtliches Wissen*: Das rechtliche Wissen entspricht der institutionellen Dimension von Politik (*Polity*). Im Gesetzesstaat, wie er in Kontinentaleuropa einschliesslich der Schweiz existiert (Kickert 2008) und in welchem – im Unterschied zur angelsächsischen Welt (Christensen et al. 2010) – sämtliches staatliches Handeln an ein formales, vom Parlament legitimiertes Gesetz gebunden sein muss (Pötzsch 2009: 28–31), kommt dem rechtlichen Wissen ein hoher Stellenwert zu. Gesetze sind generell und abstrakt (Busmann 2010: 279): Generell bedeutet, dass Gesetze für eine Vielzahl von natürlichen und juristischen Personen, einschliesslich Behörden, gelten – häufig für die ganze Bevölkerung und das gesamte staatliche Territorium. Abstrakt meint, dass sie eine Vielzahl von Sachverhalten regeln. Das Formulieren solch generell-abstrakter Rechtsnormen, die für die Anwendung im Einzelfall gleichzeitig klar und unmissverständlich sein müssen (Pawson et al. 2011: 520), ist anspruchsvoll. Es erfordert *juristisches Spezialwissen* im jeweiligen

Sachgebiet, einschliesslich Wissen zur Rechtsprechung und zum internationalen Recht.

Aus der Differenzierung zwischen Evidenz und Expertise einerseits und den drei Politikdimensionen andererseits ergibt sich eine Typologie von sechs Arten politikrelevanten Wissens. Nun stellt sich die Frage, welchen Beitrag verschiedene Akteure zu diesen Wissensarten leisten.

## **2.2 Hypothesen**

Moderne Staaten verfügen über ein System der Politikberatung, das sowohl staatliche als auch private Akteure umfasst (Howlett 2011). Im Folgenden werden aus der Literatur die theoretischen Annahmen zum Beitrag von Verwaltung und externen Experten zu den verschiedenen Arten von Evidenz und Expertise in der Politikformulierung abgeleitet.

### **2.2.1 Rolle der Verwaltung**

Während Max Weber die Verwaltung noch als von der Politik getrennte, ausführende Gewalt verstand (Sager & Rosser 2009), ist mittlerweile anerkannt, dass die Verwaltung nicht nur Politik umsetzt, sondern auch an der Formulierung von Politik wesentlich beteiligt ist (Nicholson-Crotty 2005; Page 2003; Sager 2007b). Besonders für die Verwertung von Wissen spielt die Verwaltung bei der Politikformulierung gemäss der Literatur eine wichtige Rolle (Freeman et al. 2011; Frey 2012; Howlett 2011; Sager & Anderegg 2012; Sager 2007b). Bereits Max Weber betrachtete die Verwaltung als Hort des Wissens:

„Die bürokratische Verwaltung bedeutet: Herrschaft kraft Wissen: dies ist ihr spezifisch rationaler Grundcharakter.“ (Weber 1947)

Inwiefern Webers Idealtypus den herrschenden Gegebenheiten entspricht, ist jedoch fraglich. So setzt laut Howlett (2009) der Einbezug von *Evidenz* politikanalytische Kapazitäten (*policy analytical capacity*) voraus, über welche die Verwaltung nur begrenzt verfügt:



„It [policy analytical capacity] refers to the amount of basic research a government can conduct or access, its ability to apply statistical methods, applied research methods, and advanced modelling techniques to this data and employ analytical techniques such as environmental scanning, trends analysis, and forecasting methods in order to gauge broad public opinion and attitudes, as well as those of interest groups and other major policy players, and to anticipate future policy impacts” (Howlett 2009: 162–163).

Politikanalytische Kapazitäten sind gemäss Howlett erforderlich, damit die Verwaltung Evidenz generieren, aber auch damit sie vorhandene Evidenz verwerten kann. Gemäss der Literatur erzeugt die Verwaltung allemal kaum selber Evidenz. Nur ein kleiner Anteil der in der Politikformulierung tätigen Verwaltungsangestellten in Kanada widmet sich beispielsweise Analyseaufgaben (Howlett & Walker 2012), und sogar spezialisierte Politikanalyseeinheiten generieren selber wenig Evidenz, sondern geben Studien zumeist extern in Auftrag (Howlett 2009: 250). Howletts Definition politikanalytischer Kapazitäten bezieht sich auf Evidenz zu Inhalt und Prozess. Diese Kapazitäten sind laut Howlett in Westeuropa mit der traditionell starken rechtlichen und finanziellen Kontrolle generell schwächer ausgeprägt als in der angelsächsischen Welt. Dies dürfte auch auf die Schweiz zutreffen, gibt es doch kaum spezialisierte Politikanalyseeinheiten in der Verwaltung. Frey (2012) stellte nur in einem von vier untersuchten Politikbereichen verwaltungsinterne politikanalytische Kapazitäten fest.

Gleichzeitig wurde die Annahme geäussert, die Schweizer Bundesverwaltung sei im letzten Jahrhundert nicht nur kräftig gewachsen, sondern auch professioneller geworden (Sciarini 2006). Gemäss anderen Autoren verfügt die Verwaltung jedoch vorwiegend über juristische, hingegen weniger über inhaltliche *Expertise* (Klöti & Schneider 1989; Spinatsch 2006). Frey (2012) differenziert, dass die inhaltliche *Expertise* der Bundesverwaltung insbesondere in Regelungsbereichen, die neu sind oder für deren Umsetzung sie nicht selber zuständig ist, begrenzt ist. Dass die für die Politikformulierung zuständigen Verwaltungsangestellten kaum über fachliche *Expertise* verfügen, zeigten Page und Jenkins (2005) für England auf. Dort liegt die Politikformulierung nicht in den Händen von

Juristen oder Fachexperten, sondern von Prozessexperten. Allerdings ist der institutionelle Kontext in England ein anderer, weil die Politikformulierung normalerweise im Regierungsstab geschieht, während die Gesetzesformulierung in der Schweiz angesichts der bescheidenen Stäbe der Bundesräte (Varone 2006) in den Verwaltungseinheiten stattfindet. Inwiefern die einzelnen Verwaltungseinheiten über Prozessexpertise verfügen, ist dabei fraglich, zumal sie nicht tagtäglich mit der Formulierung von Gesetzen beschäftigt sein dürften.

Aus diesen Ausführungen lassen sich verschiedene Hypothesen zum Wissensbeitrag der Verwaltung ableiten, die in Tabelle 2-1 zusammengefasst sind.

**Tabelle 2-1:** Hypothesen zum Wissensbeitrag der Verwaltung

	<i>Evidenz</i>	<i>Expertise</i>
<i>Inhalt</i>	gering	gering
<i>Prozess</i>	gering	?
<i>Recht</i>	?	gross

Es gilt die Annahme, dass die Bundesverwaltung über beschränkte politikanalytische Kapazitäten verfügt und ihr Beitrag zu *Evidenz* bezüglich Inhalt und Prozess deshalb beschränkt ist. In Bezug auf rechtliche Evidenz lässt sich aus der Literatur keine Hypothese ableiten. Weiter ist zu vermuten, dass die Verwaltung aufgrund der legalistischen Tradition vor allem juristische *Expertise* in die Politikformulierung einbringt, während ihre Rolle in Bezug auf Expertise zum Inhalt begrenzt sein dürfte und sie hinsichtlich der Prozessexpertise ungewiss ist.

### 2.2.2 Rolle der externen Experten

Zwischen 1995 und 1999 wurde etwa ein Fünftel der im Schweizer Parlament beratenen Gesetzesentwürfe durch eine formell eingesetzte Expertenkommission vorbereitet (Sciarini 2006). Daneben existieren weitere Formen des Expertenbeizugs. So werden beispielsweise häufig einzelne Experten zu spezifischen Fragen mit Gutachten beauftragt oder angehört. In der Schweiz ist der Expertenbeizug damit traditionell sehr

ausgeprägt (Sciarini 2004), doch sind externe Experten beispielsweise auch in Deutschland (Döhler 2012), den Niederlanden (Bekker et al. 2010), Island (Guðmundsson et al. 2010), dem Vereinigten Königreich (Monaghan 2010), Griechenland (Ladi 2005), den USA (Lavertu & Weimer 2010) und Neuseeland (Tenbensen 2004) regelmässig an der Politikformulierung beteiligt. Als „Experten“ werden oft alle extern beigezogenen Akteure bezeichnet, wobei nicht gesichert ist, dass ihnen in jedem Fall ein Expertenstatus im Sinne eines vertieften Wissens bezüglich des Politikfelds zukommt.

Gemäss empirischen Ergebnissen werden Experten häufig in erster Linie aufgrund eines Bedarfs an *Expertise* in die Politikformulierung eingebunden, nicht allein zur Überbrückung von Personalengpässen (Döhler 2012; Lavertu & Weimer 2010). Diese Art der Wissensbeschaffung ist funktional angebracht (Sager & Anderegg 2012), insbesondere wenn eine dauerhafte Anstellung von qualifiziertem Personal sich nicht lohnt oder nicht möglich ist (z.B. Personalstopp). Gerade in komplexen, neuen und sich schnell ändernden Politikbereichen bietet sich der Rückgriff auf Experten für die Verwaltung an:

„Administrative agencies in the United States face a challenge in maintaining adequate in-house expertise to help them perform their regulatory functions in domains involving rapidly changing technology or knowledge. Civil service laws make it difficult to hire and fire, federal agencies often are unable to compete with the private sector for expert personnel, and the ad hoc nature of the expertise agencies require may not justify hiring additional staff. Consequently, agencies routinely turn to advisory committees for supplemental expertise.” (Lavertu & Weimer 2010: 1)

Der Beizug von Externen führt aber auch zu einem Legitimitätsproblem, handelt es sich bei der Gesetzgebung doch um das Mittel des Staates schlechthin (Döhler 2012), das ihn von anderen Akteuren unterscheidet (Leresche 2001). Über das Gesetz soll der Rechtsstaat gewährleistet werden, weshalb an den Gesetzgebungsprozess hohe Anforderungen gestellt werden. Ein Beizug externer Experten macht nur Sinn, wenn auf ihren Rat gehört wird. Gleichzeitig stellt sich in diesem Fall die demokratietheoretische Frage, ob gewisse Partikulärinteressen in nicht legitimier-

---

ter Weise Gewicht erhalten (Lavertu & Weimer 2010). Erschwert wird das Dilemma dadurch, dass eine gewisse Nähe der externen Experten zur Politik in der Literatur als notwendig erachtet wird, damit deren Expertise dem politisch Machbaren entspricht und verwendet werden kann (Bekker et al. 2010; Döhler 2012; Lavertu & Weimer 2010; Sager & Anderegg 2012).

In der Schweiz stellt sich die demokratietheoretische Frage besonders prominent, da externen Experten bewusst eine Repräsentationsfunktion zugeschrieben wird. Die Konsultation externer Personen während dem vorparlamentarischen Prozess (via Vernehmlassungen, Expertenkommissionen, Anhörungen) entwickelte sich aus dem systembedingtem Kompromisserfordernis heraus. Auf diese Weise sollen die verschiedenen Interessen möglichst frühzeitig eingebunden werden (Bussmann 2008; Neidhart 1970; Sciarini 2006). Gleichzeitig erhoffte man sich davon zusätzliches Fachwissen und Sachverstand (Germann & Frutiger 1979; Germann 1981, 1985; Sciarini 2006).<sup>3</sup>

Bezüglich der verschiedenen Politikdimensionen können externe Experten vor allem fachliche Expertise sowie, wenn es sich um Juristen mit entsprechender Spezialisierung handelt, rechtliche Expertise beitragen. Hingegen ist von Experten als Aussenstehende keine vertiefte Prozesskenntnis zu erwarten.

Ein Beitrag zur *Evidenzbasierung* von Gesetzen ist vor allem von Experten aus der Wissenschaft zu erwarten. Allerdings sind Kantone bzw. Gemeinden als Vollzugsträger sowie Interessenorganisationen in Expertenkommissionen gemäss den verfügbaren Zahlen, die allerdings bereits älteren Datums sind (Germann & Frutiger 1979; Germann 1981, 1985), stärker vertreten als Personen aus der Wissenschaft. Hinzu kommt, dass die Generierung, aber auch die systematische Aufbereitung von Evidenz aufwendig und kostspielig ist, weshalb von externen Experten in der Regel nur ein Evidenzbeitrag zu erwarten, wenn sie von der Verwaltung einen bezahlten Auftrag erhalten. Dass externe Akteure von sich aus Evi-

---

<sup>3</sup> Die doppelte Funktion von Interessenvertretung und Expertise zeigt sich z. B. auch in Griechenland, wo Experten häufig enge Parteibindungen aufweisen (Ladi 2005).

denz generieren oder aufbereiten (vgl. Beispiele bei Frey 2010), dürfte aufgrund der damit verbundenen Kosten nur in Ausnahmefällen vorkommen. Indes ist es durchaus üblich, dass die Verwaltung entsprechende Aufträge an externe Experten vergibt (Howlett 2011).

Der Evidenzbeitrag von externen Experten kann sich grundsätzlich sowohl auf den Inhalt (z.B. Evaluation bestehender Massnahmen), den Prozess (z.B. Befragung der Interessengruppen zur Unterstützung möglicher Massnahmen) als auch das Recht (z.B. internationaler Rechtsvergleich) beziehen – je nach Auftrag der Verwaltung und Spezialisierung der externen Experten.

**Tabelle 2-2:** Hypothesen zum Wissensbeitrag externer Experten

	<i>Evidenz</i>	<i>Expertise</i>
<i>Inhalt</i>	gering, ausser starker Rolle von	gering
<i>Prozess</i>	Wissenschaftlern und explizitem	unbedeutend
<i>Recht</i>	Auftrag	gering

Tabelle 2-2 stellt die Hypothesen, die sich aus der Literatur ergeben, in Kurzform dar. Einen wesentlichen *Evidenzbeitrag* dürften externe Experten nur leisten, wenn Personen aus der Wissenschaft eine wichtige Rolle spielen, vor allem wenn sie von der Verwaltung einen entsprechenden Auftrag erhalten. Die Evidenz kann sich dabei auf alle drei Dimensionen beziehen. Zudem ist anzunehmen, dass die durch externe Experten eingebrachte *Expertise* beschränkt bleibt, weil ihre Repräsentationsfunktion im Vordergrund steht. Von externen Experten ist vor allem inhaltliche und rechtliche Expertise, weniger dagegen Prozessexpertise zu erwarten, weil letztere eine gute Kenntnis der verwaltungsinternen Abläufe voraussetzt.

### 2.3 Empirisches Vorgehen

Die aufgestellten Hypothesen sollen in der Folge empirisch überprüft werden. Dazu wird ein vergleichendes Fallstudiendesign angewandt. Fallstudien eignen sich besonders gut, um kausal komplexe Phänomene

---

wie den Einbezug von Wissen in die Gesetzgebung zu untersuchen (Freeman et al. 2011).

Die Auswahl der Fälle erfolgte aufgrund von theoretischen Überlegungen (Blatter et al. 2007: 149–157). Um die Vergleichbarkeit zu gewährleisten, wurde ein homogener Kontext gewählt (Berg-Schlosser & De Meur 2009). Den institutionellen Rahmen bildet die schweizerische Bundesverwaltung. Sie spielt bei der Gesetzesformulierung eine zentrale Rolle, weil der Regierungsstab des Bundesrates klein ist (Varone 2006) und die Verwaltung auch gegenüber dem Milizparlament eine starke Stellung einnimmt (Lüthi 2009). Für die Gesetzesformulierung wird jeweils eine federführende Verwaltungseinheit bestimmt, die den Prozess (z.B. Bezug von verwaltungsinternen und -externen Akteuren) und die Inhalte der Vorlage im Rahmen allfälliger Vorgaben von Bundesrat und Departementvorsteher gestalten kann. In der Regel werden Gesetzesentwürfe anschliessend bei Kantonen und Interessenverbänden in die Vernehmlassung gegeben, wobei dies im Entscheid des Bundesrates liegt.

Als Grundgesamtheit werden komplexe Politikformulierungsprozesse betrachtet: umfassende Gesetze, die neu erarbeitet oder komplett überarbeitet (Totalrevision) werden. Als Fall definiert wurde der Gesetzgebungsprozess ab der Erteilung des Mandats an die Verwaltungseinheit bis zur Eröffnung der Vernehmlassung. Dass bezüglich des vorparlamentarischen Prozesses Forschungslücken bestehen (Sciarini 2006; Varone 2006), gilt für diese frühe Phase, für welche die Arbeitsweise der Verwaltung und ihre Zusammenarbeit mit externen Experten noch nie vertieft analysiert wurde (Frey 2012), ganz besonders. Prozessanalysen weisen darauf hin, dass Wissen in dieser frühen Phase, wenn der Inhalt von Gesetzen noch nicht festgelegt ist, am besten einfließen kann (Widmer & Frey 2008; Widmer 2009).

Um den Kontext möglichst konstant zu halten, wurde eine zeitliche Abgrenzung eingeführt. Gewählt wurden Fälle, bei welchen während der 46. Legislatur (1999-2003) die Vernehmlassung eröffnet wurde und bei denen mindestens in den vier Jahren zuvor kein Wechsel an der Departementsspitze stattgefunden hat, so dass sich der politische Rahmen über eine gewisse Stabilität auszeichnete. Die Fälle mussten zum Zeit-

punkt, als die Arbeiten am Forschungsprojekt starteten (Oktober 2006), aus Verwaltungssicht abgeschlossen sein, d.h. die Botschaft des Bundesrates an das Parlament musste vorliegen, damit der Zugang zu den Dokumenten gewährt werden konnte.<sup>4</sup>

Da sowohl die Rolle der Verwaltung als auch jene der externen Experten hinsichtlich ihres Wissensbeitrags beobachtet werden soll, wurden Fälle mit unterschiedlichen Erarbeitungsformen gewählt, namentlich vorwiegend verwaltungsintern und vorwiegend extern erarbeitete Vernehmlassungsentwürfe. Zudem wurden Fälle aus verschiedenen Verwaltungseinheiten gewählt, in der Annahme, dass deren politikanalytische Kapazität unterschiedlich ist (Balthasar 2007; Frey 2012). Schliesslich zeigt die Literatur einhellig, dass Konflikte einem Einbezug von Wissen in die Politik grundsätzlich hinderlich sind (Alkin 1985; Chelimsky 1995; Frey 2012; Ledermann 2012; Monaghan 2010; Rissi & Sager 2012; Sager & Ledermann 2008). Im Sinne einer Kontrolle dieses Faktors wurden bewusst keine sehr umstrittenen Gesetzgebungsprozesse betrachtet, sondern nur solche, die sich – zumindest während der untersuchten Phase – durch einen tiefen oder mittleren Konfliktgrad auszeichneten.

Die Fälle sind in Tabelle 2-3 aufgeführt. In Anhang I finden sich zudem die Eckdaten zu jedem Fall.

**Tabelle 2-3:** Fallwahl (federführende Verwaltungseinheit in Klammer)

		<i>politische Konfliktualität</i>	
		<i>mittel</i>	<i>tief</i>
<i>Einbezug</i> <i>Externer</i>	<i>schwach</i>	Sicherheitskontrollgesetz (Generalsekretariat UVEK)	Transplantationsgesetz (Bundesamt für Gesundheit)
	<i>stark</i>	Sprachengesetz (Bundesamt für Kultur)	Vormundschaftsrecht (Bundesamt für Justiz)

Methodisch steht die Dokumentenanalyse im Zentrum. Sämtliche in den Archiven der federführenden Verwaltungseinheit vorhandenen Doku-

<sup>4</sup> Das Öffentlichkeitsgesetz, das den Zugang zu Dokumenten der Bundesverwaltung gewährleistet, trat erst am 1. Juni 2006 in Kraft und gilt nicht rückwirkend. Gewisse Verwaltungsstellen lehnten die im Rahmen des Forschungsprojekts gestellten Gesuche um Dokumenteneinsicht denn auch ab.

mente wurden gesichtet und im Hinblick auf die interessierenden Dimensionen anhand eines Fallstudienprotokolls in Form eines Rasters analysiert (Yin 2003). Erfasst wurden die beteiligten Personen, die Art ihrer Beteiligung sowie ihr institutioneller und fachlicher Hintergrund, Hinweise bezüglich Konflikten zwischen den beteiligten Personen, der Inhalt der Erörterungen (worüber wurde gesprochen, was wurde abgeklärt bezüglich Inhalt, Prozess, Recht), archivierte wissenschaftliche Studien, Arbeitspapiere und Grundlagen sowie Verweise auf Studien.

Zur Validierung wurden mit ein bis zwei Personen pro Fall Interviews durchgeführt. Dabei handelte es sich um die bei den Verwaltungseinheiten für die Projektleitung zuständigen Personen. In einem Fall (Vormundschaftsrecht) wurde zusätzlich mit einem externen Experten gesprochen, weil der Prozess zu Beginn fast vollständig externalisiert worden war. Da die Gesetzgebungsprojekte für die Personen meist schon mehrere Jahre zurücklagen, war ihr Erinnerungsvermögen beschränkt (vgl. Diekmann 1998). Bei den Gesprächen ging es deshalb vor allem darum sicherzustellen, dass kein wichtiger Aspekt des Falles übersehen wurde. Weiter dienten die Interviews dazu, zusätzliche Informationen über die beteiligten Personen einzuholen. Diese Informationen wurden anschliessend mittels einer Internetrecherche ergänzt.

Hinsichtlich des Einbezugs von Expertise und Evidenz wurde eine skalierende Inhaltsanalyse (Mayring 2003) der Fallraster vorgenommen. Die angewandte Skala ist in Tabelle 2-4 aufgeführt.



**Tabelle 2-4:** Skalen zum Wissensseinbezug im Gesetzgebungsprozess

<i>Ausprägung</i>	<i>Juristische Expertise</i>	<i>Fachexpertise</i>	<i>Prozessexpertise</i>	<i>Evidenz (zu jeweiliger Dimension)</i>
<i>unbe- deutend</i>	Keine oder marginale Beteiligung von Personen mit juristischer Expertise.	Keine oder marginale Beteiligung von Personen mit Fachexpertise.	Keine oder marginale Beteiligung von Personen mit Prozessexpertise.	Keine Rezeption wissenschaftlicher Studien UND keine Generierung von Evidenz.
<i>gering</i>	Effektive Beteiligung von mindestens einer Person mit generaler, aber nicht spezialisierter juristischer Expertise.	Effektive Beteiligung von mindestens einer Person mit beschränkter Fachexpertise (z.B. erworben durch Erfahrung, nicht Ausbildung).	Effektive Beteiligung von mindestens einer Person mit beschränkter Prozessexpertise (beschränkte Gesetzgebungserfahrung bzw. Kenntnis des Gesetzgebungsleitfadens usw.)	Einzelner Hinweis auf eine bis zwei wissenschaftliche Studien UND/ODER Generierung von Evidenz, deren Systematik etwas fragwürdig ist.
<i>moderat</i>	Punktuelle effektive Beteiligung von einer oder zwei Personen mit spezialisierter juristischer Expertise.	Punktuelle effektive Beteiligung von einer oder zwei Personen mit ausgewiesener Fachexpertise.	Punktuelle effektive Beteiligung von einer oder zwei Personen mit Erfahrung in Gesetzgebungsprojektleitung.	Wiederholte Hinweise auf eine bis zwei wissenschaftliche Studien UND/ODER Generierung von systematischer Evidenz zu einem bis zwei Themen.
<i>eher viel</i>	Punktuelle effektive Beteiligung von drei oder mehr Personen mit spezialisierter juristischer Expertise.	Punktuelle effektive Beteiligung von drei oder mehr Personen mit ausgewiesener Fachexpertise.	Punktuelle effektive Beteiligung von drei oder mehr Personen mit Erfahrung in Gesetzgebungsprojektleitung.	Hinweise auf drei oder mehr wissenschaftliche Studien ODER Generierung von Evidenz zu drei oder mehr Themen.
<i>viel</i>	Während gesamtem Prozess effektive Beteiligung von jeweils mindestens zwei Personen mit spezialisierter juristischer Expertise.	Während gesamtem Prozess effektive Beteiligung von jeweils mindestens zwei Personen mit ausgewiesener Fachexpertise.	Während gesamtem Prozess effektive Beteiligung von jeweils mindestens zwei Personen mit Erfahrung in Gesetzgebungsprojektleitung.	Hinweise auf drei oder mehr wissenschaftliche Studien UND Generierung von Evidenz zu drei oder mehr Themen.

---

## 2.4 Resultate

In diesem Kapitel werden die Ergebnisse der Fallstudien strukturiert dargestellt, indem zunächst kurz der Kontext beschrieben wird (Thematik des Gesetzgebungsprozesses, federführende Verwaltungsakteure, Phasen des Erarbeitungsprozesses, Konfliktualität des Prozesses), bevor die einbezogene Evidenz und Expertise dargestellt und der Wissensbeitrag der verschiedenen Akteure tabellarisch zusammengefasst wird.<sup>5</sup>

### 2.4.1 Sicherheitskontrollgesetz

*Thematik:* Auslöser des Gesetzgebungsprozesses war ein Seilbahnunglück auf der Riederalp im Dezember 1996, bei dem sich die Frage nach der Rolle des Bundes als Aufsichtsbehörde stellte. Eine Überprüfung des zuständigen Departements UVEK zeigte, dass die Kontrolle der technischen Sicherheit von Anlagen, Fahrzeugen, Geräten und Komponenten, für welche die Ämter des UVEK zuständig sind, sehr unterschiedlich wahrgenommen wird. Der Departementsvorsteher beauftragte daraufhin sein Generalsekretariat, ein neues Gesetz zu erarbeiten, um die Verfahren zu vereinheitlichen und so zu regeln, dass der Staat die Kontrolle nur noch wo notwendig selber wahrnimmt und sie ansonsten an Private auslagert.

*Federführung:* Generalsekretariat UVEK (GS-UVEK). Als Stabsstelle des Departementsvorstehers ist das GS-UVEK normalerweise für die Koordination der Geschäfte der Ämter, hingegen nicht für Gesetzgebungsprozesse zuständig. In diesem Fall wird dem GS-UVEK die Federführung übertragen, weil rund zehn Ämter des UVEK betroffen und diese dem Gesetzgebungsprojekt gegenüber kritisch eingestellt sind (vgl. Konfliktualität).

---

<sup>5</sup> Wie im Rest des Beitrags wird auch bei der Darstellung der Fälle ausschliesslich die männliche Form verwendet, auch wenn es sich um weibliche Personen handelte.

*Erarbeitungsprozess:* Es lassen sich zwei Phasen unterscheiden:

1. Februar 1998 bis Januar 2000: Konzeptphase. Eine 3-köpfige GS-interne Arbeitsgruppe erstellt zwei Zwischenberichte zu den möglichen Stossrichtungen des Gesetzes, die dem Departementsvorsteher als Entscheidungsgrundlage unterbreitet werden.
2. Februar 2000 bis September 2001: Formulierungsphase. Die 3-köpfige GS-interne Arbeitsgruppe leitet weiterhin die Arbeiten. Zusätzlich werden grössere Arbeitsgruppen gebildet, in denen die betroffenen Verwaltungseinheiten des UVEK vertreten sind und welche die Gesetzesbestimmungen für die jeweiligen Sachgebiete formulieren sollen. Der Bund sucht den Kontakt zu den Kantonen, welche zwei Vertreter delegieren. Zudem finden Aussprachen mit einzelnen Verbänden statt.

*Konfliktualität:* Mittel. Die Ämter des UVEK lehnen das neue Gesetz ab, weil sie die technische Sicherheit als gewährleistet erachten und eine Einschränkung ihres Handlungsspielraums befürchten. Im Bereich der Motorfahrzeugkontrolle wehren sich die Kantone gegen eine Auslagerung an Private, weil sie zum Teil eigene Infrastrukturen erstellt haben. Verschiedene Verbände schalten sich ein, weil sie einen Mehraufwand und Mehrkosten befürchten. Eine öffentliche Debatte findet in der betrachteten Phase nicht statt.

*Evidenz:* Die GS-interne Arbeitsgruppe erhebt mittels einer schriftlichen Befragung mit anschliessender Validierungsrunde systematisch die rechtlichen Grundlagen sowie die Umsetzung der sicherheitsrelevanten Aufgaben des Departements. Zudem konsultiert sie sechs Mal die betroffenen Ämter und wertet deren Stellungnahmen jeweils systematisch aus, was ihr ein abgestütztes Bild der Akzeptanz des Gesetzgebungsvorhabens ermöglicht. Die verwaltungsintern beigezogenen Akteure weisen die federführende Stelle nur in einem Fall auf eine wissenschaftliche Publikation zu risikobasiertem Recht hin. Eines der Ämter führt auf Anstoss der internen Arbeitsgruppe eine Befragung zur Art der Risikobewertung in den Ämtern durch und generiert damit fachliche Evidenz. Die beigezogenen Kantonsvertreter und konsultierten Verbände tragen weder zur Rezeption noch zur Generierung von Evidenz bei.

*Expertise:* Die Federführung obliegt drei Juristen im GS-UVEK, die weder über rechtliches noch fachliches Wissen auf dem Gebiet der technischen Sicherheit, noch über Gesetzgebungserfahrung, sondern lediglich über theoretische Kenntnisse der Gesetzgebung verfügen. In der zweiten Erarbeitungsphase werden Juristen aus den einzelnen Ämtern einbezogen, die eine gute Kenntnis der rechtlichen Bestimmungen zur technischen Sicherheit in ihren jeweiligen Sachgebieten, jedoch nur beschränktes Fachwissen haben. Zudem wird eine Person mit Erfahrung in Gesetzgebung aus dem Bundesamt für Justiz beigezogen. Die zwei von den Kantonen delegierten Vertreter verfügen über Fachexpertise zur Motorfahrzeugkontrolle, jedoch ist diese weitgehend beschränkt auf ihre jeweiligen Kantone. Zudem bringen sie keine Kenntnisse zur technischen Sicherheit in anderen Bereichen, für welche die Kantone zuständig sind, mit. Aufgrund der ablehnenden Haltung gegenüber dem neuen Gesetz zeigen sich sowohl die beigezogenen Personen aus den UVEK-Ämtern als auch die kantonalen Vertreter und die Verbände eher zurückhaltend bei der Weitergabe ihrer Expertise.

**Tabelle 2-5:** Wissensbeitrag der Akteure im Fall Sicherheitskontrollgesetz

<i>Wissensart</i>	<i>Dimension</i>	<i>Federführende Verwaltungs- akteure</i>	<i>Beigezogene Verwaltungs- akteure</i>	<i>Beigezogene Externe</i>
<i>Evidenz</i>	<i>Inhalt</i>	moderat	moderat	unbedeutend
	<i>Prozess</i>	eher viel	unbedeutend	unbedeutend
	<i>Recht</i>	moderat	gering	unbedeutend
<i>Expertise</i>	<i>Inhalt</i>	gering	moderat	gering
	<i>Prozess</i>	gering	moderat	unbedeutend
	<i>Recht</i>	gering	eher viel	unbedeutend

## 2.4.2 Sprachengesetz

*Thematik:* Im März 1996 stimmen Volk und Stände der Revision des Sprachenartikels in der Verfassung zu. Das Rätoromanische wird zur Landes- und Teilamtsprache des Bundes erhoben, die bestehenden Massnahmen des Bundes zur Förderung des Rätoromanischen und Italienischen werden auf Verfassungsstufe verankert, und Bund und Kanto-

ne erhalten den Auftrag, die Verständigung zwischen den Sprachgemeinschaften zu fördern. Mit der Totalrevision der Verfassung wird 1999 zusätzlich die Unterstützung der mehrsprachigen Kantone durch den Bund beschlossen. Durch ein neues Sprachengesetz soll der Verfassungsauftrag umgesetzt werden.

*Federführung:* Bundesamt für Kultur (BAK). Auf Entscheid des Bundesrates erhält nicht die Bundeskanzlei, sondern das Eidg. Departement des Innern die Federführung. Sie wird im Bundesamt für Kultur an eine Sektion delegiert, die bis dahin nur am Rande mit der Thematik zu tun und auch keine Gesetzgebungsprozesse geleitet hatte. Die Leitung der Sektion wechselt während des Prozesses mehrmals; der zuständige Mitarbeiter bleibt dagegen konstant.

*Erarbeitungsprozess:* Es lassen sich drei Phasen unterscheiden:

1. Januar 1997 bis Juni 1998: Geplant sind zwei separate Gesetze – ein Amtssprachengesetz und ein Verständigungsgesetz. Das BAK setzt je eine interdepartementale Arbeitsgruppe ein. Die Arbeitsgruppe zum Amtssprachengesetz formuliert einen Vorentwurf, den sie in die Ämterkonsultation schickt. Aufgrund der Stellungnahmen wird entschieden, ein umfassendes Sprachengesetz statt zweier separater Gesetze zu erarbeiten.
2. November 1998 bis September 1999: Zwei Arbeitsgruppen von Bund und Kantonen erörtern die Umsetzung des Sprachenartikels: Die Arbeitsgruppe „EDI/EDK“ befasst sich mit der Umsetzung des Verständigungsauftrags, die Arbeitsgruppe „Mehrsprachige Kantone“ mit der Unterstützung der mehrsprachigen Kantone. Ergebnis ist ein Bericht zuhanden der Departementsleitung.
3. Februar 2000 bis Oktober 2001: Die „Paritätische Arbeitsgruppe Sprachengesetz“, bestehend zu gleichen Teilen aus Vertretern von Bund und Kantonen sowie einzelnen unabhängigen Experten, formuliert einen Vernehmlassungsentwurf.

*Konfliktualität:* Mittel. Die Kantone wollen nicht, dass der Bund eigene Massnahmen zur Förderung der Verständigung unternimmt. Der Kompetenzkonflikt dominiert die Diskussion in sämtlichen gemischten Arbeits-

gruppen. Bezüglich der Amtssprachenregelung sind in der Bundesverwaltung die Meinungen, wie weit die Gleichbehandlung aller Sprachen gehen soll, geteilt. Eine öffentliche Debatte zur Sprachenfrage findet während der betrachteten Phase nicht statt.

*Evidenz:* Zu Beginn des Prozesses führt die federführende Verwaltungsstelle mit einem breiten Kreis von Adressaten und Betroffenen Aussprachen über die Erwartungen an ein Sprachengesetz durch und wertet diese aus, wobei die Systematik etwas fragwürdig ist. Ansonsten bringt die federführende Verwaltungsstelle keine Evidenz in den Prozess ein. Verwaltungsintern trägt das Bundesamt für Justiz mit einem Gutachten zu den Verfassungskompetenzen juristische Evidenz bei. Die Kantone geben hierzu ihrerseits zwei Gutachten in Auftrag und bringen diese in die Paritätische Arbeitsgruppe ein. Daneben liefern die in dieser Arbeitsgruppe vertretenen Sprachwissenschaftler vereinzelt Hinweise auf wissenschaftliche Erkenntnisse.

*Expertise:* Den federführenden Verwaltungsakteuren fehlt weitgehend relevante Expertise. Die Fachexpertise ist sehr beschränkt, Erfahrung mit Gesetzgebungsprozessen fehlt gänzlich, und juristische Expertise stellt das BAK nur in marginalem Umfang zur Verfügung. Hingegen bringen andere Verwaltungsstellen juristische sowie vor allem fachliche Expertise in Sprachenfragen in die interdepartementale Arbeitsgruppe Amtssprachengesetz ein. Mit dem Einbezug einer erfahrenen Person aus dem Bundesamt für Justiz gelangt zudem eine gewisse juristische und Prozessexpertise in die Paritätische Arbeitsgruppe. In dieser Arbeitsgruppe sind auch drei Sprachwissenschaftler und ein Rechtsprofessor mit fachlicher bzw. juristischer Expertise vertreten, doch wird die Diskussion so stark durch die Kompetenzfrage dominiert, dass sie ihre Expertise nur begrenzt einbringen können.

**Tabelle 2-6:** Wissensbeitrag der Akteure im Fall Sprachengesetz

<i>Wissensart</i>	<i>Dimension</i>	<i>Federführende Verwaltungs- akteure</i>	<i>Beigezogene Verwaltungs- akteure</i>	<i>Beigezogene Externe</i>
<i>Evidenz</i>	<i>Inhalt</i>	unbedeutend	unbedeutend	gering
	<i>Prozess</i>	gering	unbedeutend	unbedeutend
	<i>Recht</i>	unbedeutend	moderat	moderat
<i>Expertise</i>	<i>Inhalt</i>	gering	eher viel	moderat
	<i>Prozess</i>	unbedeutend	moderat	unbedeutend
	<i>Recht</i>	unbedeutend	moderat	gering

### 2.4.3 Transplantationsgesetz

*Thematik:* Die Transplantationsmedizin lag ursprünglich im Kompetenzbereich der Kantone, wobei Art und der Umfang der vorhandenen kantonalen Bestimmungen stark variierten. Angesichts der steigenden Zahl von kostspieligen Transplantationen und einiger HIV-Ansteckungen fordern breite Kreise eine Bundesregelung. Das Parlament erlässt im März 1996 eine Übergangsbestimmung. Im Februar 1999 heissen Volk und Stände die notwendige Verfassungsbestimmung mit grossem Mehr gut.

*Federführung:* Bundesamt für Gesundheit (BAG). Die Federführung wird zunächst bei einer Person in der Abteilung Recht angesiedelt. Im Laufe des Frühjahrs 1996 stellt das BAG drei zusätzliche Personen (davon zwei Juristen) für den Transplantationsbereich an und schafft eine eigene Sektion. Die vier Personen bilden die Arbeitsgruppe, welche das Transplantationsgesetz formuliert.

*Erarbeitungsprozess:* Die 4-köpfige BAG-interne Arbeitsgruppe zieht von Beginn weg einen Rechtsprofessor als Berater bei. Bis Ende 1997 arbeitet sie hauptsächlich an der Verfassungsbestimmung. Ab Anfang 1998 wendet sie sich dem Gesetzesentwurf zu.

*Konfliktualität:* In der betrachteten Phase gering (vgl. auch Engeli & Varone 2011). Die Kantone wünschen ausdrücklich eine Regelung der Transplantationsmedizin auf Bundesebene. Zwar geht es beim Gesetz um ethisch heikle Fragen, doch sind sich die Mitglieder der Arbeitsgruppe weitgehend einig und werden ihre Vorschläge von der Departements-

leitung und vom Bundesrat unterstützt. Der Vernehmlassungsentwurf wird abseits der Öffentlichkeit erarbeitet.

*Evidenz:* Die Arbeitsgruppe berücksichtigt zahlreiche juristische und medizinische Studien. Auch wird ein systematischer Vergleich der Regelungen auf internationaler Ebene sowie ein Rechtsvergleich mit bestimmten Ländern unternommen. Die Arbeitsgruppe trägt zudem diverse Zahlen zur Zahl der verschiedenen Transplantationen im In- und Ausland zusammen (fachliche Evidenz). Während von den verwaltungsintern beigezogenen Akteuren kein Evidenzbeitrag auszumachen ist, liefert ein Gutachten zur Definition des Todes, das ein spezialisierter Rechtsprofessor im Auftrag der Arbeitsgruppe erstellt, zusätzliche juristische Evidenz.

*Expertise:* Die federführenden Verwaltungsakteure verfügen kaum über relevante Expertise. Da die Mitglieder der Arbeitsgruppe konstant bleiben, können sie sich aber mit der Zeit in beschränktem Rahmen rechtliches und fachliches Spezialwissen aneignen. Prozessexpertise, die über die Kenntnis des Gesetzgebungslaufes hinausgehen würde, fehlt den zuständigen Personen. Einmal wird eine in Gesetzgebung erfahrene Person aus dem Bundesamt für Justiz beigezogen. Zudem lässt sich die Arbeitsgruppe durch einen Rechtsprofessor mit einer guten Kenntnis des Transplantationsbereichs beraten und zieht im Rahmen von Hearings und bilateralen Gesprächen weiteres spezialisiertes juristisches und fachliches Wissen zur Transplantationsmedizin bei, wobei die verfügbare Expertise insgesamt beschränkt ist. Die Juristen aus der Bundesverwaltung, welche die Arbeitsgruppe konsultiert, verfügen dagegen nicht über Spezialwissen zur Transplantationsmedizin.



**Tabelle 2-7:** Wissensbeitrag der Akteure im Fall Transplantationsgesetz

<i>Wissensart</i>	<i>Dimension</i>	<i>Federführende Verwaltungs- akteure</i>	<i>Beigezogene Verwaltungs- akteure</i>	<i>Beigezogene Externe</i>
<i>Evidenz</i>	<i>Inhalt</i>	viel	unbedeutend	unbedeutend
	<i>Prozess</i>	unbedeutend	unbedeutend	unbedeutend
	<i>Recht</i>	eher viel	unbedeutend	moderat
<i>Expertise</i>	<i>Inhalt</i>	gering	unbedeutend	eher viel
	<i>Prozess</i>	gering	gering	unbedeutend
	<i>Recht</i>	gering	gering	eher viel

#### 2.4.4 Vormundschaftsrecht

*Thematik:* Als einziger Teil des Zivilgesetzbuches wurde das Vormundschaftsrecht seit 1912 praktisch nicht angepasst. Kritisiert wird, dass die Terminologie (z.B. Vormund, Mündel) nicht mehr zeitgemäss sei, dass die im Gesetz vorgesehenen Massnahmen es nicht erlauben, auf die individuelle Situation der Betroffenen einzugehen, und dass die Organisation des Vormundschaftswesens zu wenig professionell sei. Mit der Totalrevision soll das gesamte Vormundschaftsrecht überholt werden.

*Federführung:* Bundesamt für Justiz (BJ). Die Federführung liegt beim Direktionsbereich Privatrecht. Dieser ist zu Beginn stark belastet mit den Revisionen der übrigen Teile des Zivilgesetzbuches, so dass die Arbeiten zu einem grossen Teil ausgelagert werden.

*Erarbeitungsprozess:* Es lassen sich drei Phasen unterscheiden:

1. September 1993 bis Juli 1995: Drei externe Experten, allesamt Hochschulprofessoren, erarbeiten im Auftrag des BJ einen Grundlagenbericht für die Totalrevision des Vormundschaftsrechts. Der Bericht wird an einer Fachtagung vorgestellt und die Fachkreise erhalten Gelegenheit zur schriftlichen Stellungnahme.
2. November 1996 bis Juni 1998: Die drei Hochschulprofessoren erarbeiten in regelmässiger Absprache mit Vertretern des BJ einen Gesetzesvorentwurf.

- 
3. Februar 1999 bis Juni 2003: Eine breit abgestützte Expertenkommission verfasst auf der Basis des bestehenden Vorentwurfs einen Vernehmlassungsentwurf.

*Konfliktualität:* Tief. Die Notwendigkeit und Stossrichtungen der Totalrevision sind breit anerkannt. Im Detail gibt es in der dritten Phase unter den Mitgliedern der Expertenkommission Meinungsverschiedenheiten (z.B. Organisation des Vormundschaftswesens, Zwangsbehandlung), die jedoch konsensual gelöst werden. Eine öffentliche Debatte findet nicht statt.

*Evidenz:* Die drei externen Experten rezipieren vor allem in der ersten Phase unzählige juristische und fachliche Publikationen und verarbeiten diese in Form von Arbeitspapieren. Im Auftrag des BJ erstellen zudem zwei Rechtsinstitute systematische internationale bzw. interkantonale Rechtsvergleiche zu drei Themenbereichen, womit reichlich juristische Evidenz von extern beigetragen wird. In der dritten Phase der Expertenkommission macht das BJ juristische Literatur zugänglich und bereitet diese in Arbeitspapieren auf. Nicht-juristische Literatur rezipiert das BJ dagegen kaum. Prozessevidenz generieren die federführenden Akteure insofern, als sie die schriftlichen Stellungnahmen zum ersten Bericht der Experten systematisch auswerten, um die Akzeptanz der vorgeschlagenen Massnahmen abzuschätzen. Ein BJ-intern erstelltes Gutachten zu Zwangsbehandlungen liefert von Verwaltungsseite zusätzliche rechtliche Evidenz.

*Expertise:* Indem die federführenden Verwaltungsakteure bereits die übrigen Teile des Zivilgesetzbuchs revidiert haben, verfügen sie über grosse Prozess Erfahrung. Hingegen ist ihre juristische und fachliche Expertise beschränkt, weil der Bund im Vormundschaftswesen keine Aufsichts- oder Vollzugstätigkeiten erfüllt. Durch die langjährige Beschäftigung mit der Revision erlangen die zuständigen Mitarbeiter bis zu einem beschränkten Grade jedoch fachliches und rechtliches Spezialwissen. Verwaltungsintern, insbesondere auch BJ-intern, wird zusätzliche juristische Expertise beigezogen, während fachliche Expertise nur in einem Fall von einem anderen Amt her eingeflossen ist. Die drei Hochschulprofessoren sowie etliche Mitglieder aus der Expertenkommission verfügen

über ein vertieftes juristisches und fachliches Wissen im Vormundschafsbereich. Ausserdem wird in der ersten und dritten Phase über Hearings zusätzliche Rechts- und Fachexpertise zu spezifischen Fragen eingeholt. Die extern eingebrachte Prozessexpertise ist dagegen gering; sie beschränkt sich bei den Hochschulprofessoren auf die frühere Mitgliedschaft in Expertenkommissionen.

**Tabelle 2-8:** Wissensbeitrag der Akteure im Fall Vormundschaftsrecht

<i>Wissensart</i>	<i>Dimension</i>	<i>Federführende Verwaltungs- akteure</i>	<i>Beigezogene Verwaltungs- akteure</i>	<i>Beigezogene Externe</i>
<i>Evidenz</i>	<i>Inhalt</i>	gering	unbedeutend	eher viel
	<i>Prozess</i>	moderat	unbedeutend	unbedeutend
	<i>Recht</i>	eher viel	moderat	viel
<i>Expertise</i>	<i>Inhalt</i>	gering	gering	viel
	<i>Prozess</i>	viel	unbedeutend	gering
	<i>Recht</i>	gering	eher viel	viel

## 2.5 Diskussion

Die Hypothesen zur Rolle der Verwaltung und der externen Experten beim Einbezug der verschiedenen Arten von Evidenz und Expertise werden in der Folge anhand eines Fallvergleichs erörtert und die Erkenntnisse daraus synthetisiert. Zum Schluss werden die Faktoren diskutiert, welche den Einbezug der verschiedenen Wissensarten beeinflussen.

### 2.5.1 Rolle der Verwaltung

Bezüglich der Rolle der Verwaltung beim Einbezug von *Evidenz* wurde ausgehend von der Literatur die These aufgestellt, dass deren Beitrag insbesondere bezüglich systematischem Inhalts- und Prozesswissen wegen schwacher politikanalytischer Kapazitäten gering ist. Tatsächlich hat die Verwaltung nur in einem der vier Fälle (Transplantationsgesetz) vorhandene inhaltliche Evidenz aufbereitet und verwertet. Jedoch hat sie entgegen der Annahme in sämtlichen Fällen selber Evidenz generiert. So wurde beim Sicherheitskontrollgesetz bei den Ämtern des UVEK so-

---

wohl inhaltliche Evidenz als auch Prozessevidenz (Akzeptanz der Massnahmen) erhoben. Beim Vormundschaftsrecht und in beschränktem Masse beim Sprachengesetz wurde durch Adressatenbefragungen ebenfalls Prozessevidenz generiert. Beim Transplantationsgesetz trug die Verwaltungsstelle fachliche Evidenz in Form von in- und ausländischen Vergleichszahlen zusammen. Es handelte sich in allen Fällen um ein systematisches Zusammentragen und Auswerten von Fakten bzw. Einschätzungen, ohne dass dabei besondere methodische Verfahren, wie sie in Howletts (2009) Definition politikanalytischer Kapazitäten zentral sind, angewandt worden wären. Angesichts des juristischen Hintergrunds der meisten beteiligten Verwaltungsakteure gilt die Vermutung, dass die politikanalytischen Kapazitäten in diesem Sinne gefehlt hätten.

Der Evidenzbeitrag der Verwaltung fiel bezüglich der Dimension des Rechts, zu der keine Hypothese formuliert wurde, am grössten aus: Sowohl beim Vormundschaftsrecht als auch beim Transplantationsgesetz haben die federführenden Akteure die juristische Literatur systematisch aufbereitet. Beim Sicherheitskontrollgesetz haben sie die rechtlichen Grundlagen der technischen Sicherheit aller Ämter gezielt zusammengetragen und verglichen. Beim Sprachengesetz hat zwar nicht die federführende Stelle selber, jedoch das Bundesamt für Justiz rechtliche Fragen systematisch abgeklärt. Die Bundesverwaltung hat somit in erster Linie juristische Evidenz in die Vernehmlassungsentwürfe eingebracht.

Tatsächlich scheint die Gesetzesarbeit, wie Klöti und Schneider (1989) seinerzeit behaupteten, in gewissem Sinne Juristensache zu sein: Mit Ausnahme des Sprachengesetzes waren die federführenden Verwaltungsakteure fast ausschliesslich Juristen. Diese verfügten jedoch im fraglichen Politikbereich nicht über juristisches Spezialwissen, sondern nur über generelles rechtliches Wissen. Damit erwies sich die Vermutung, die Bundesverwaltung bringe aufgrund der kontinentaleuropäischen legalistischen Tradition selber juristische *Expertise* in den Gesetzgebungsprozess ein, als weniger zutreffend. Kompensiert wurde die beschränkte juristische Expertise der federführenden Akteure in drei von vier Fällen dadurch, dass zu spezifischen Fragen verwaltungsinterne Rechtsexpertise beigezogen wurde.

Der Annahme entsprechend wiesen die federführenden Akteure dagegen nur ein beschränktes Fachwissen auf. In neuen Regelungsbereichen (Transplantationsgesetz) oder in Politikfeldern, in welchen die Bundesverwaltung keine Aufsichts- oder Vollzugskompetenzen besitzt (Vormundschaftsrecht), fehlt in der Verwaltung eine kompetente Stelle. In den anderen Fällen wurde die Federführung wohl aus politischen Überlegungen nicht der fachkundigsten Einheit zugewiesen. Im Falle des Sprachengesetzes wurde das Bundesamt für Kultur als zuständig erklärt, obwohl die Bundeskanzlei gerade im Bereich der Amtssprachen über deutlich mehr Expertise verfügt hätte. Beim Sicherheitskontrollgesetz wurde die Federführung in sehr unüblicher Weise dem Generalsekretariat zugewiesen, weil von den eigentlich kompetenten Ämtern Widerstand erwartet wurde. Fachliche Expertise wurde weniger stark verwaltungsintern kompensiert als juristische Expertise (z.B. wurden beim Sicherheitskontrollgesetz nur beschränkt verwaltungsinterne Fachpersonen beigezogen).

Schliesslich können sich die federführenden Verwaltungsakteure in der Schweiz im Gegensatz zu ihren englischen Kollegen nicht mit besonderer Prozessexpertise brüsten (Page 2003). Immerhin wurde in drei von vier Fällen die Beratung des Bundesamtes für Justiz in Anspruch genommen, womit die mangelnde Prozessexpertise bis zu einem gewissen Grade wettgemacht werden konnte. Dass die Federführung gleich dem Bundesamt für Justiz übertragen würde, steht in der stark departementalisierten Verwaltungsstruktur (vgl. Varone 2006), in der jeder Departementsvorsteher seine Kompetenzen verteidigt, hingegen wohl ausser Frage. Aus Sicht des längerfristigen Wissensmanagements macht dies auch wenig Sinn; wie die Fallstudien zeigen, bieten die langwierigen Gesetzgebungsprozesse für die Verwaltung die Gelegenheit, juristische und fachliche Expertise aufzubauen, die dann in der Phase der Umsetzung genutzt werden kann.

Die Ergebnisse des Fallvergleichs zur Rolle der Verwaltung beim Einbezug von Wissen in Gesetzgebungsprozessen sind in Tabelle 2-9 synthetisiert.

**Tabelle 2-9:** Erkenntnisse zum Wissensbeitrag der Verwaltung

	<i>Evidenz</i>	<i>Expertise</i>
<i>Inhalt</i>	moderat: z.T. systematisches Zusammentragen von Fakten	gering
<i>Prozess</i>	moderat: z.T. systematische Erhebung von Einschätzungen	Federführung gering, aber oft verwaltungsintern kompensiert
<i>Recht</i>	moderat – gross	Federführung gering, aber oft verwaltungsintern kompensiert

### 2.5.2 Rolle der externen Experten

Von externer Seite, so wurde vermutet, sei ein *Evidenzbeitrag* vor allem von Personen aus der Wissenschaft zu erwarten, und zwar aufgrund des mit der Evidenzgenerierung und -aufbereitung verbundenen Aufwands vor allem dann, wenn diese einen vergüteten Auftrag von der Verwaltung erhielten, wobei sich das systematische Wissen auf alle drei Politikdimensionen beziehen könne. Diese Annahmen werden weitgehend bestätigt. Tatsächlich wurde in jenem Fall, bei dem niemand aus der Wissenschaft beigezogen wurde (Sicherheitskontrollgesetz), auch kein externer Evidenzbeitrag geleistet. Gleichzeitig machten in den Expertenkommissionen (Sprachengesetz, Vormundschaftsrecht) vorwiegend die Wissenschaftler auf Studien aufmerksam. Externe Experten trugen dabei ausschliesslich Evidenz zu Inhalt und Recht, nicht jedoch zum Prozess bei.

Eine über die Aufbereitung bestehender Evidenz hinaus gehende *Generierung* wissenschaftlicher Erkenntnisse fand nur im Rahmen von Aufträgen statt, wobei diese durchwegs auf juristische Evidenz ausgerichtet waren: Beim Transplantationsgesetz ging es um das rechtliche Konzept des Hirntods, beim Vormundschaftsrecht um systematische Rechtsvergleiche und beim Sprachengesetz um die Definition der Kompetenzen von Bund und Kantonen. Die Tatsache, dass die Verwaltung vor allem bereit ist, Geld für juristische Abklärungen auszugeben, kann als Hinweis für die Privilegierung dieser Wissensart aufgrund der legalistischen Verwaltungstradition gewertet werden. Der Fall des Sprachengesetzes zeigt allerdings, dass nicht nur die Verwaltung bereit ist, Studien zu finanzieren, sondern auch andere Akteure, die ein grosses Interesse an Evidenz

aufweisen (vgl. auch Frey & Widmer 2011). In diesem Fall waren es die Kantone, die zwei Rechtsgutachten in Auftrag gaben und diese dann in den Prozess einbrachten.

Die Hypothesen, dass externe Experten bezüglich *Expertise* vor allem inhaltliches und rechtliches Wissen, hingegen als Aussenstehende kaum Prozessexpertise beitragen, werden durch die betrachteten Fälle klar bestätigt. Dass der Expertisebeitrag externer Experten hingegen grundsätzlich gering ist, weil Experten im Schweizer System vor allem als Repräsentanten von Interessen einbezogen werden, stützt der Fallvergleich dagegen nur zum Teil. In allen Fällen war der Beizug externer Personen zumindest teilweise dadurch motiviert, zusätzliches Wissen einzubeziehen, was bisherige Studien bestätigt (Döhler 2012; Germann 1981; Lavertu & Weimer 2010). Offensichtlich im Vordergrund stand der Wissensbeitrag beim Beizug von Wissenschaftlern wie den drei Hochschulprofessoren beim Vormundschaftsrecht oder den Sprachwissenschaftlern beim Sprachengesetz. Aber auch vom Einbezug von Kantonsvertretern erhoffte sich die Verwaltung zusätzliche Expertise: beim Sprachengesetz zu den staatlichen Möglichkeiten einer Förderung der Verständigung sowie zur Problematik einer mehrsprachigen Verwaltung; beim Vormundschaftsrecht zu den Vor- und Nachteilen verschiedener Vollzugsarrangements und beim Sicherheitskontrollgesetz zu den in den Kantonen angewandten Verfahren der Sicherheitsüberprüfung.

Inwiefern die extern beigezogenen Personen tatsächlich Expertise lieferten, hing wesentlich vom Konfliktgrad ab. So konnten sich die Sprachwissenschaftler als Fachexperten beim Sprachengesetz nur begrenzt in der Paritätischen Arbeitsgruppe einbringen, weil die Vertreter von Bund und Kantonen die Diskussion stark mit der Kompetenzfrage dominierten. Beim Sicherheitskontrollgesetz waren die Kantonsvertreter primär darauf bedacht, ihre Kompetenzen im Bereich der Motorfahrzeugkontrolle zu verteidigen, ohne dass sie ihre (ohnehin beschränkte) Expertise einzubringen versuchten. Umgekehrt brachten sich die einzelnen Personen in der Expertenkommission zum Vormundschaftsrecht trotz der Vielfalt der vertretenen Interessen mit ihrem Wissen in die Diskussion ein. Auch die Hearings, die beim Vormundschaftsrecht und Transplantationsgesetz mit

diversen Stakeholdern durchgeführt wurden, ermöglichten der Verwaltung den Einbezug von zusätzlicher Expertise (z.B. Wissen zur Organzuteilung bei Transplantationen), während die Aussprachen mit Verbänden beim konfliktuelleren Sicherheitskontrollgesetz keinen erkennbaren Wissensmehrwert lieferten. Bereits moderate Konflikte führten somit dazu, dass die Verteidigung von Positionen einen sachlichen Austausch verhinderte.

In einem konfliktuellen Umfeld wurden die beigezogenen Personen zudem nicht durch die federführende Verwaltungsstelle ausgewählt, sondern von übergeordneter Stelle delegiert, wobei die Expertise nicht das primäre Auswahlkriterium darstellte. Den beiden beim Sicherheitskontrollgesetz delegierten Kantonsvertretern kam beispielsweise nicht der erhoffte Expertenstatus bezüglich sicherheitsrelevanter kantonaler Aufgaben zu. Auch einige Kantons- und Bundesvertreter in der Paritätischen Arbeitsgruppe Sprachengesetz wiesen nur geringe Kenntnisse der Materie auf. In beiden Fällen wurde die Auswahl der Personen den Kantonen bzw. den betroffenen Bundesämtern überlassen. Bei der Expertenkommission zum Vormundschaftsrecht erfolgten die Ernennungen dagegen *ad personam* durch die federführende Verwaltungsstelle. Obwohl auch hier die Vertretung der verschiedenen Interessengruppen wichtig war, gelang es der Verwaltung mit allenfalls einer Ausnahme, Experten im Sinne von Fachpersonen mit einem vertieften Wissen zur Mitarbeit zu bewegen.

Tabelle 2-10 gibt die Erkenntnisse zur Rolle der externen Experten aus dem Fallvergleich gerafft wieder.

**Tabelle 2-10:** Erkenntnisse zum Wissensbeitrag der externen Experten

	<i>Evidenz</i>	<i>Expertise</i>
<i>Inhalt</i>	gering: nur Wissenschaftler und Aufbereitung bestehender Evidenz	unterschiedlich: abhängig von Auswahl der Personen und Konfliktualität
<i>Prozess</i>	unbedeutend	unbedeutend
<i>Recht</i>	moderat: bei Auftrag auch Generierung neuer Evidenz	unterschiedlich: abhängig von Auswahl der Personen und Konfliktualität



### **2.5.3 Determinanten des Einbezugs der verschiedenen Wissensarten**

Die vier Fälle liefern Hinweise, dass der Einbezug von Evidenz und Expertise nicht in technokratischer Art und Weise geschieht, sondern durch den politisch-institutionellen Kontext geprägt wird. Die institutionelle Kompetenzordnung scheint erstens die Verfügbarkeit von fachlicher und juristischer Expertise zu bestimmen: Wo der Bund nur über beschränkte Kompetenzen verfügt, ist die verwaltungsintern vorhandene politikbereichsspezifische Expertise bescheiden. Auch gibt es Anzeichen, dass der Einbezug externer Expertise in diesem Fall erschwert ist, weil es bei neuen Regelungsbereichen kaum Experten gibt bzw. weil es bei bisher kantonal geregelten Bereichen schwierig ist, Experten zu finden, die einen Überblick haben.

Zudem scheint bereits ein mittlerer Konfliktgrad den Einbezug von Wissen zu hemmen, und zwar gilt dies nicht nur für Evidenz, wie in der Literatur dokumentiert, sondern auch für den Einbezug von Expertise. Im Falle des Sicherheitskontrollgesetzes waren Ämter, Kantone und Verbände aufgrund ihrer ablehnenden Haltung allesamt zurückhaltend bei der Weitergabe ihrer Expertise. Auch beim Sprachengesetz verhinderten die Kompetenzstreitigkeiten zwischen Bund und Kantonen eine sachliche Auseinandersetzung, in die sich namentlich die Wissenschaftler mit Evidenz und Expertise hätten einbringen können. In einem konfliktuellen Umfeld ist zudem die Prozessgestaltungsmacht der federführenden Verwaltungsstelle eingeschränkt, was insbesondere zur Folge hat, dass sie die Auswahl der einzubeziehenden Personen beschränkt beeinflussen kann. Die delegierenden Einheiten achten weniger auf die Expertise der Personen als darauf, dass diese ihre Positionen gut verteidigen.

In zwei Fällen bemühten sich die federführenden Verwaltungsakteure, die begrenzten Möglichkeiten, Expertise einzubeziehen, durch eine verstärkte Abstützung auf Evidenz wettzumachen. Im Falle des Sicherheitskontrollgesetzes wich die federführende Stelle angesichts des Widerstrebens der Ämter bei der Mitarbeit auf systematische Befragungen sowohl zu den rechtlichen Rahmenbedingungen als auch zum Inhalt sicherheitsrelevanter Aufgaben wie auch zur Akzeptanz der vorgeschla-

---

genen Massnahmen (Prozessdimension) aus. Hier führten Konflikte somit zum Einbezug von weniger Expertise, aber mehr Evidenz. Beim Transplantationsgesetz konsultierten die Zuständigen aufgrund des sowohl verwaltungsinternen als auch -externen Mangels an Experten die vorhandene internationale wissenschaftliche Rechts- und Fachliteratur und trugen in Form von Rechtsvergleichen und der Erhebung von Fallzahlen selber neue Evidenz bei. Evidenz ist in diesen Fällen dort zum Einsatz gekommen ist, wo die Expertise geendet hat.

Schwierigkeiten beim Beizug von Expertise können somit als Push-Faktoren für den Einbezug von Evidenz dienen. Dies erklärt das Ausmass an beigezogener Evidenz aber nicht in allen Fällen. Denn auch beim Sprachengesetz war der Einbezug von Expertise durch Konflikte mittleren Grades erschwert, dennoch wurde nicht auf Evidenz ausgewichen, während beim Vormundschaftsrecht ein beachtlicher Einbezug von Expertise stattfand und trotzdem zusätzlich Evidenz beigezogen wurde. Ein weiterer Erklärungsfaktor sind die politikanalytischen Kapazitäten der federführenden Akteure. Allerdings deuten die Fallstudien darauf hin, dass unter politikanalytischer Kapazität nicht das Beherrschen ausgeklügelter sozialwissenschaftlicher Methoden zu verstehen ist, wie sie bei Howletts (2009) Definition im Zentrum stehen. Vielmehr reicht ein grundsätzlicher wissenschaftlicher Zugang. Dieser zeigt sich an einem Interesse an vorhandenen wissenschaftlichen Erkenntnissen und an der Fähigkeit, systematische Daten zusammenzutragen. Beim Sprachengesetz fehlten politikanalytische Kapazitäten im Sinne eines solchen wissenschaftlichen Zugangs weitestgehend.

Der Fallvergleich bestätigt somit die Erkenntnisse von Frey (2012), dass der Einbezug von Wissen wesentlich von fallspezifischen Faktoren, namentlich der Konfliktivität des Gegenstands und der politikanalytischen Kapazität der Verwaltungseinheit – allerdings im soeben präzisierten Sinne – abhängt. Darüber hinaus dürften allerdings auch Systemmerkmale beeinflussen, inwiefern die verschiedenen Wissensarten Eingang in die Gesetzgebung finden. So kann die Tatsache, dass bei den betrachteten Fällen ausschliesslich bezahlte Studien zur Bereitstellung juristischer Evidenz vergeben wurden, als Hinweis für das legalistische Verwal-

tungsverständnis in der Schweiz genommen werden (vgl. auch Bekke & van der Meer 2000).

## 2.6 Schlussfolgerungen

Der „Evidence-Based Policy“-Ansatz hat international Reformen hin zu einer stärker wissensbasierten Politikgestaltung ausgelöst. Eine klare Definition, was als Evidenz zählt, fehlt aber nach wie vor. Gemeinhin werden darunter forschungsbasierte Resultate verstanden, doch in Anbetracht der Grenzen, auf welche wissenschaftliche Erkenntnisse angesichts der Komplexität von Politik stossen, wurde in letzter Zeit vermehrt eine Ausweitung des Begriffs verlangt. Der EBP-Ansatz läuft dabei Gefahr, einerseits beliebig und andererseits blind zu werden für Unterschiede bezüglich des Einbezugs verschiedener Wissensarten in den Politikprozess. In diesem Beitrag wurde stattdessen eine Typologie politikrelevanter Wissensarten eingeführt, die einerseits zwischen wissenschaftlicher – d.h. systematischer, intersubjektiv nachvollziehbarer – Evidenz und erfahrungsbasierter Expertise unterscheidet, und andererseits sämtliche drei Politikdimensionen – *Policy* als Inhalt, *Politics* als Prozess und *Polity* als rechtlicher Rahmen – berücksichtigt. In der Folge wird erörtert, welche Erkenntnisse diese Typologie ermöglicht hat, bevor abschliessend die demokratietheoretischen Implikationen dieser Erkenntnisse reflektiert werden.

Wie die Anwendung anhand von vier vorparlamentarischen Gesetzgebungsprozessen im Schweizer Bundesstaat gezeigt hat, bewährt sich die vorgeschlagene Typologie, um das für die Politikgestaltung relevante Wissen differenziert zu erfassen und die Rolle der verschiedenen Akteure beim Einbezug von Wissen zu analysieren. Die vertieften Archivstudien, ergänzt durch Interviews und Internetrecherchen, weisen darauf hin, dass die federführenden Stellen ihr Manko an Expertise je nach Politikdimension auf verschiedene Art kompensieren. Prozessexpertise wurde, der Hypothese entsprechend, ausschliesslich verwaltungsintern beigeleitet; Aussenstehenden fehlt das entsprechende Insider-Wissen. Juristische Expertise wurde ebenfalls vorwiegend innerhalb der Verwal-

tung beigezogen, wobei bei fehlendem internem Rechtswissen auch auf externe Experten zurückgegriffen wurde. Fachliche Expertise wurde dagegen etwa in gleichem Masse über den Beizug verwaltungsinterner als auch -externer Personen in den Prozess eingebracht, letzteres vor allem, wenn es in der Verwaltung zum betreffenden Thema keine fachlich kompetenten Personen gab.

Mit einer Ausnahme kam neben Expertise auch Evidenz zum Tragen. So wurde in zwei Fällen die vorhandene Rechts- und Fachliteratur relativ breit aufbereitet, was wesentlich zur Evidenzbasierung beitrug. Zudem wurde zu allen drei Politikdimensionen Evidenz generiert, doch beschränkte sie sich auf klar abgrenzbare Problemstellungen, wie beispielsweise in inhaltlicher Hinsicht auf die Zusammenstellung von Fallzahlen, in rechtlicher Hinsicht auf einen internationalen Rechtsvergleich oder die vertiefte Abklärung bestimmter Rechtsbegriffe oder in Bezug auf den Prozess auf das systematische Einholen und Auswerten von Stellungnahmen gewisser Anspruchsgruppen. Dies bestätigt einerseits die einleitend erwähnten Grenzen von systematischer und nachprüfbarer Evidenz, die in der Produktion aufwendig ist und deshalb immer nur einen kleinen Ausschnitt aus der komplexen Realität abbilden kann. Andererseits stellt die Tatsache, dass es bei der berücksichtigten Evidenz um einfache systematische Zusammenstellungen von Fakten und Einstellungen handelte, die Definition politikanalytischer Kapazitäten von Howlett (2009) mit ihrem Fokus auf die Kenntnis spezialisierter methodischer Verfahren in Frage. Gerade angesichts der Komplexität von Politikprozessen scheint weniger ein Bedarf an Studien mit komplizierten Modellen als ein Bedürfnis nach grundlegendem Wissen zu bestehen, bei dem sich die Frage nach den Grenzen der Aussagekraft weniger eminent stellt. Um solche Evidenz einzubeziehen, ist ein grundsätzliches wissenschaftliches Interesse und ein systematisches, nachvollziehbares Vorgehen, nicht jedoch ein wissenschaftlich-methodisches Spezialwissen erforderlich. Die Fallstudien weisen somit darauf hin, dass das Konzept der politikanalytischen Kapazitäten allenfalls angepasst werden müsste.

Daneben deuten die Fälle darauf hin, dass der Einbezug der verschiedenen Wissensarten primär von fallspezifischen politisch-institutionellen

Faktoren, namentlich der bestehenden Kompetenzordnung im fraglichen Politikbereich und dem Konfliktgrad abhängt. Die federführenden Verwaltungsstellen weichen insbesondere dann auf Evidenz aus, wenn ihre Möglichkeiten, Expertise einzubeziehen, beschränkt sind. Evidenz und Expertise scheinen somit bis zu einem gewissen Grade austauschbar zu sein, sind sie doch auf das gleiche Ziel ausgerichtet: Gesetze wissenschaftsorientiert zu gestalten, damit sie die beabsichtigten Wirkungen entfalten.

Die Typologie politikrelevanter Wissensarten ermöglicht ferner, systembedingte Einflussfaktoren zu untersuchen. Die Kompromissorientierung des schweizerischen Systems hemmt den Einbezug von Expertise gemäss den Fallstudien nicht grundsätzlich, sondern nur bei Konflikten, wobei sich bereits moderate Spannungen auswirken. Diese führen dazu, dass die Interessenvertretung zum einzigen Kriterium bei der Auswahl der einbezogenen Personen wird und deren Wissensbeitrag unbeachtet bleibt. Ferner scheint im legalistisch geprägten Verwaltungskontext der Schweiz eine Präferenz für juristische gegenüber anderer Evidenz zu bestehen. Im Unterschied zur englischen Verwaltung dürfte schliesslich eine Leitung von Gesetzgebungsverfahren durch Prozessexperten im stark departementalisierten schweizerischen System ausser Frage stehen, weil damit das Bundesamt für Justiz, das primär über Prozessexpertise verfügt, viel Macht erhielte, was von den übrigen Departementen nicht akzeptiert würde. Gesetzgebungsprozesse dauern in der Schweiz im Vergleich zu England allerdings um ein Vielfaches länger. Wie die empirische Analyse gezeigt hat, bieten sie deshalb für die Verwaltungseinheiten die Gelegenheit, fachliche und juristische Expertise aufzubauen, die für die spätere Umsetzung von Nutzen sein kann. Diese Chance würde durch die Übertragung der Federführung an Prozessexperten verpasst. Solche systembedingten Unterschiede könnten künftig anhand von internationalen Vergleichsstudien, in welchen die dargestellten Wissensarten systematisch unterschieden werden, vertieft analysiert werden.

---

### **2.6.1 Demokratiethoretische Implikationen des Expertenbeizugs**

Wie die Fälle gezeigt haben, erwies sich die Expertise der federführenden Verwaltungsakteure in Bezug auf alle drei Politikdimensionen als gering. Umso dringlicher stellt sich die demokratiethoretisch relevante Frage, ob die Verwaltung zum Spielball von Interessen wird. Erhalten gewisse Partikulärinteressen angesichts der fehlenden Expertise der Verwaltung in nicht legitimierter Weise Gewicht? Die Fälle weisen darauf hin, dass in der Schweiz diesbezüglich ein potentiell Risiko besteht, doch fehlen konkrete Anhaltspunkte für eine problematische Einflussnahme. Einerseits hat die federführende Verwaltungsstelle den Mangel an Expertise zumindest in Bezug auf den Prozess und das Recht oft verwaltungsintern sowie zum Teil durch die Aufbereitung und Generierung von Evidenz kompensiert. Andererseits spricht die konkrete Ausgestaltung des Expertenbeizugs gegen einen übermässigen Einfluss einzelner Interessen. So scheint die breit abgestützte, 20-köpfige Expertenkommission beim Vormundschaftsrecht unbedenklich. Auf den ersten Blick kritischer ist dagegen, dass beim Sprachengesetz die Kantone von Beginn weg sehr eng in die Ausarbeitung des Gesetzes einbezogen wurden, während sämtliche übrigen Interessen aussen vor blieben. Jedoch gibt es keine Hinweise, dass es den Kantonsvertretern gelungen wäre, die Vorlage stark zu beeinflussen. So setzte der Bund durch, dass verschiedene Massnahmen in den Vernehmlassungsentwurf aufgenommen wurden, gegen die sich eine Mehrheit der Kantonsvertreter ausgesprochen hatte. Die Meinungsbildung wurde somit nicht von vornherein eingeschränkt. Demokratiethoretisch grundsätzlich heikel ist die Beratung durch eine Einzelperson, wie sie beim Transplantationsgesetz vorgekommen ist, wo die Verwaltung einen Rechtsprofessor zur Begleitung des Prozesses beigezogen hat. Hier ist es wichtig, dass es sich um eine politisch unabhängige Person handelt und sich die Verwaltung nicht allein auf ihren Rat verlässt, sondern selber zusätzliche Informationen beschafft, was im konkreten Fall in ausgeprägtem Masse geschehen ist. Auch beim Vormundschaftsrecht hat sich die Verwaltung nicht allein auf den Rat der drei Hochschulprofessoren verlassen, sondern deren Bericht

an einer Fachtagung zur Diskussion gestellt und deren Gesetzesvorlage in einer breit abgestützten Expertenkommission weiterentwickelt.

Die auf der Kompromissorientierung basierende Praxis des Einbezugs von Akteuren dürfte somit durchaus zu grundsätzlich demokratisch heiklen Konstellationen führen, in welchen gewissen Akteuren ein grosser Einfluss zukommen könnte. In allen untersuchten sensiblen Fällen hat die Verwaltung jedoch gewisse „Checks“ durchgeführt. Konkrete Hinweise für Abhängigkeiten des Staates von einzelnen externen Meinungsmachern fehlen somit.

Gleichzeitig ist der Beizug von externen Experten in Gesetzgebungsprozessen durchaus funktional angebracht. Es handelt sich um eine vorübergehende Aufgabe der Verwaltung, die spezifisches Wissen erfordert, das genauso gut extern aufbereitet oder generiert werden kann, so dass der verwaltungsinterne Mehrbedarf an Ressourcen reduziert werden kann. Häufig liegt das relevante Wissen allemal bei externen Akteuren mit je eigenen Interessen. Deren Expertise darf von der Verwaltung nicht ohne weiteres als bare Münze genommen werden. Sofern jedoch eine Überprüfung stattfindet, stellt auch der Einbezug interessengebundener Expertise einen Erfolgsfaktor für Gesetzgebungsprozesse dar, weil sie das politisch Machbare widerspiegelt. Wissen ist eben letztlich nie gänzlich objektivierbar. Dies gilt für Evidenz und noch viel deutlicher für Expertise. Das Ziel des EBP-Ansatzes einer ideologiefreien Politikgestaltung, die sich allein nach dem vorhandenen Wissen richtet, bleibt deshalb allemal unerreichbar. Wir müssen uns wohl oder übel mit einer „wissensbewussten Politik“, die bewusst mit den Risiken, die mit den verschiedenen Wissensarten verbunden sind, umgeht, zufrieden geben.

## **2.7 Bibliographie**

Alkin, M. C. (1985). *A Guide for Evaluation Decision Makers*. Beverly Hills, CA: Sage.

Balthasar, A. (2007). *Institutionelle Verankerung und Verwendung von Evaluationen*. Chur/Zürich: Rüegger.

- 
- Bekke, H. A. G. M., & Van der Meer, F. M. (Hg.) (2000). *Civil Service Systems in Western Europe*. Cheltenham: Edward Elgar.
- Bekker, M., Egmond, S. Van, Wehrens, R., Putters, K., & Bal, R. (2010). Linking Research and Policy in Dutch Healthcare: Infrastructure, Innovations and Impacts. *Evidence & Policy*, 6(2), 237–254.
- Berg-Schlosser, D., & De Meur, G. (2009). Comparative Research Design. Case and Variable Selection. In B. Rihoux & C. C. Ragin (Hg.), *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques* (pp. 19–32). Los Angeles, London, New Delhi, Singapore: Sage.
- Blatter, J. K., Janning, F., & Wagemann, C. (2007). *Qualitative Politikanalyse: Eine Einführung in Forschungsansätze und Methoden*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Bundesamt für Justiz. (2007). *Gesetzgebungsleitfaden: Leitfaden für die Ausarbeitung von Erlassen des Bundes* (3. Aufl.). Bern: Bundesamt für Justiz (BJ).
- Bussmann, W. (1997). Evaluationen in der Schweiz. In W. Bussmann, U. Klöti, & P. Knoepfel (Hg.), *Einführung in die Politikevaluation* (pp. 13–35). Basel, Frankfurt a.M.: Helbing & Lichtenhahn.
- Bussmann, W. (2008). The Emergence of Evaluation in Switzerland. *Evaluation*, 14(4), 499–506.
- Bussmann, W. (2010). Evaluation of Legislation: Skating on Thin Ice. *Evaluation*, 16(3), 279–293.
- Cabinet Office (1999a). *Modernising Government. Whitepaper*. London.
- Cabinet Office (1999b). *Professional Policy Making for the Twenty First Century*. London.
- Chelimsky, E. (1995). The Political Environment of Evaluation and What It Means for the Development of the Field. *Evaluation Practice*, 16(3), 215–225.
- Chi, M. T. H., Glaser, R., & Farr, M. J. (Hg.) (1988). *The Nature of Expertise*. Hillsdale NJ: Lawrence Erlbaum Associates.
- Christensen, R. K., Goerdel, H. T., & Nicholson-Crotty, S. (2010). Management, Law, and the Pursuit of the Public Good in Public



- Administration. *Journal of Public Administration Research and Theory*, 21, i125–i140.
- Coalition for Evidence-Based Policy (2012). What Works in Social Policy? Findings From Well-Conducted Randomized Controlled Trials. Retrieved from [www.evidencebasedprograms.org](http://www.evidencebasedprograms.org).
- Davies, H. T. O., Nutley, S., & Smith, P. C. (Hg.) (2000). *What Works? Evidence-Based Policy and Practice in Public Services*. Bristol: The Policy Press.
- Diekmann, A. (1998). *Empirische Sozialforschung: Grundlagen, Methoden, Anwendungen* (4. Aufl., p. 635). Reinbek bei Hamburg: Rowohlt.
- Döhler, M. (2012). Gesetzgebung auf Honorarbasis – Politik, Ministerialverwaltung und das Problem externer Beteiligung an Rechtsetzungsprozessen. *Politische Vierteljahresschrift*, 53(2), 181–210.
- Donaldson, S. I., Christie, C. A., & Mark, M. M. (2008). *What Counts as Credible Evidence in Applied Research and Evaluation Practice?* Newbury Park, CA: Sage.
- Dreyfus, H. L., & Dreyfus, S. E. (2005). Peripheral Vision: Expertise in Real World Contexts. *Organization Studies*, 26(5), 779–792.
- Engeli, I., & Varone, F. (2011). Governing Morality Issues through Procedural Policies. *Swiss Political Science Review*, 17(3), 239–258.
- Ericsson, K. A. (2006). The Influence of Experience and Deliberate Practice on the Development of Superior Expert Performance. In K. A. Ericsson, N. Charness, R. R. Hoffman, & P. J. Feltovich (Hg.), *Cambridge Handbook of Expertise and Expert Performance* (pp. 685–706). Cambridge: Cambridge University Press.
- Freeman, R., Griggs, S., & Boaz, A. (2011). The Practice of Policy Making. *Evidence & Policy*, 7(2), 127–136.
- Frey, K. (2010). Revising Road Safety Policy: The Role of Systematic Evidence in Switzerland. *Governance*, 23(4), 667–690.
- Frey, K. (2012). *Evidenzbasierte Politikformulierung in der Schweiz: Gesetzesrevisionen im Vergleich*. Baden-Baden: Nomos.

- 
- Frey, K., & Ledermann, S. (2010). Evidence-Based Policy: A Concept in Geographical and Substantive Expansion. *German Policy Studies*, 6(2), 1–15.
- Frey, K., & Widmer, T. (2011). Revising Swiss Policies: The Influence of Efficiency Analyses. *American Journal of Evaluation*, 32(4), 494–517.
- Germann, R. E. (1981). *Ausserparlamentarische Kommissionen: Die Milizverwaltung des Bundes*. Bern, Stuttgart: Paul Haupt.
- Germann, R. E. (1985). Les hommes de science dans les commissions. In R. E. Germann (Hg.), *Experts et commissions de la Confédération*. Lausanne: Presses polytechniques romandes.
- Germann, R. E., & Frutiger, A. (1979). *Les commissions extraparlimentaires créées de 1970 à 1977: Tableaux statistiques*. Genève: Université de Genève.
- Gigerenzer, G. (2008). *Rationality for Mortals: How People Cope with Uncertainty*. Oxford: Oxford University Press.
- Guðmundsson, H. S., Jónsdóttir, S., & Júlíusdóttir, S. (2010). Evidence-Based Policy and Practice in Social Services in Iceland. *Evidence & Policy*, 6(2), 195–211.
- Hansen, H. F., & Rieper, O. (2010). The Politics of Evidence-Based Policy-Making: The Case of Denmark. *German Policy Studies*, 6(2), 87–112.
- Head, B. W. (2008). Three Lenses of Evidence-Based Policy. *Australian Journal of Public Administration*, 67(1), 1–11.
- Head, B. W. (2010a). Reconsidering Evidence-Based Policy: Key Issues and Challenges. *Policy and Society*, 29(2), 77–94.
- Head, B. W. (2010b). Water Policy—Evidence, Learning and the Governance of Uncertainty. *Policy and Society*, 29(2), 171–180.
- Howlett, M. (2009). Policy Analytical Capacity and Evidence-Based Policy-Making: Lessons from Canada. *Canadian Public Administration*, 52(2), 153–175.
- Howlett, M. (2011). Public Managers as the Missing Variable in Policy Studies: An Empirical Investigation Using Canadian Data. *Review of Policy Research*, 28(3), 247–263.

- Howlett, M., & Walker, R. M. (2012). Public Managers in the Policy Process: More Evidence on the Missing Variable? *Policy Studies Journal*, 40(2), 211–233.
- Kickert, W. J. M. (2008). The Study of Public Management in Europe and the United States. *der moderne staat*, 2008(1), 221–236.
- Klöti, U., & Schneider, G. (1989). *Informationsbeschaffung des Gesetzgebers*. Grösch: Rüegger.
- Ladi, S. (2005). The Role of Experts in the Reform Process in Greece. *West European Politics*, 28(2), 279–296.
- Lavertu, S., & Weimer, D. L. (2010). Federal Advisory Committees, Policy Expertise, and the Approval of Drugs and Medical Devices at the FDA. *Journal of Public Administration Research and Theory*, 21(2): 211-237.
- Ledermann, S. (2012). Exploring the Necessary Conditions for Evaluation Use in Program Change. *American Journal of Evaluation*, 33(2), 159-178.
- Ledermann, S., & Sager, F. (2009). Problem erkannt, aber nicht gebannt – Der Nutzen einer Verknüpfung von Konzept- und Umsetzungsevaluation am Beispiel der Strategie “Migration und Gesundheit”. *Zeitschrift für Evaluation*, 8(1), 7–25.
- Leresche, J.-P. (2001). *Gouvernance et coordination des politiques publiques* (pp. 31–65). Paris: Pedone.
- Lüthi, R. (2009). Im Spannungsfeld zwischen Verwaltung und Politik: Erfahrungen aus der Erarbeitung von Erlassentwürfen durch parlamentarische Kommissionen. *LeGes*, 2009(3), 363–374.
- Mayring, P. (2003). *Qualitative Inhaltsanalyse: Grundlagen und Techniken* (8. Aufl.). Weinheim, Basel: Beltz.
- Monaghan, M. P. (2008). Appreciating Cannabis: The Paradox of “Evidence” in Evidence-Based Policy-Making. *Evidence & Policy*, 4(2), 209–231.
- Monaghan, M. P. (2010). Adversarial Policies and Evidence Utilization: Modeling the Changing Evidence and Policy Connection. *German Policy Studies*, 6(2), 17–52.

- 
- Neidhart, L. (1970). *Plebiszit und pluralitäre Demokratie: Eine Analyse der Funktionen des schweizerischen Gesetzesreferendums*. Bern: Francke.
- Nicholson-Crotty, S. (2005). Bureaucratic Competition in the Policy Process. *Policy Studies Journal*, 33(3), 341–361.
- Nutley, S., Morton, S., Jung, T., & Boaz, A. (2010). Evidence and Policy in Six European Countries: Diverse Approaches and Common Challenges. *Evidence & Policy*, 6(2), 131–145.
- Nutley, S., Walter, I., & Davies, H. T. O. (2003). From Knowing to Doing: A Framework for Understanding the Evidence-into-Practice Agenda. *Evaluation*, 9(2), 125–148.
- Nutley, S., Walter, I., & Davies, H. T. O. (2007). *Using Evidence: How Research Can Inform Public Services*. Bristol: The Policy Press.
- OECD (2006). *Regulatory Reform in Switzerland: Enhancing Market Openness through Regulatory Reform*. Paris.
- Page, E. C. (2003). The Civil Servant as Legislator: Law Making in British Administration. *Public Administration*, 81(4), 651–679.
- Page, E. C., & Jenkins, B. (2005). *Policy Bureaucracy: Government with a Cast of Thousands*. New York: Oxford University Press.
- Parsons, W. (2004). Not Just Steering But Weaving: Relevant Knowledge and the Craft of Building Policy Capacity and Coherence. *Australian Journal of Public Administration*, 63(March), 43–57.
- Pawson, R., Wong, G., & Owen, L. (2011). Known Knowns, Known Unknowns, Unknown Unknowns: The Predicament of Evidence-Based Policy. *American Journal of Evaluation*, 32(4), 518–546.
- Pöttsch, H. (2009). *Die Deutsche Demokratie* (5. Aufl.). Bonn: Bundeszentrale für politische Bildung.
- PVK (2005). *Die drei „KMU-Tests“ des Bundes: bekannt? genutzt? wirkungsvoll?* Bern: Parlamentarische Verwaltungskontrollstelle (PVK).
- Rissi, C., & Sager, F. (2012). Types of knowledge utilization of regulatory impact assessments: Evidence from Swiss policymaking. *Regulation & Governance*. “Early View”-Publikation. doi:10.1111/rego.12002.

- Rycroft-Malone, J., Seers, K., Titchen, A., Harvey, G., Kitson, A., & McCormack, B. (2004). What Counts as Evidence in Evidence-Based Practice? *Journal of Advanced Nursing*, 47(1), 81–90.
- Sager, F. (2007a). Making Transport Policy Work: Polity, Policy, Politics and Systematic Review. *Policy & Politics*, 35(2), 269–288.
- Sager, F. (2007b). Habermas' Models of Decisionism, Technocracy, and Pragmatism in Times of Governance: The Relationship of Public Administration, Politics, and Science in the Alcohol Prevention Policies of the Swiss Member States. *Public Administration*, 85(2), 429–447.
- Sager, F., & Anderegg, C. (2012). Dealing With Complex Causality in Realist Synthesis: The Promise of Qualitative Comparative Analysis. *American Journal of Evaluation*, 33(1), 60–78.
- Sager, F., & Ledermann, S. (2008). Valorisierung von Politikberatung. In S. Bröchler & R. Schützeichel (Hg.), *Politikberatung* (pp. 310–325). Stuttgart: Lucius & Lucius.
- Sager, F., & Rissi, C. (2011). The Limited Scope of Policy Appraisal in the Context of Referendum Democracy: The Case of Regulatory Impact Assessment in Switzerland. *Evaluation*, 17(2), 151–163.
- Sager, F., & Rosser, C. (2009). Weber, Wilson, and Hegel: Theories of Modern Bureaucracy. *Public Administration Review*, 69(6), 1136–1147.
- Sanderson, I. (2004). Getting Evidence into Practice: Perspectives on Rationality. *Evaluation*, 10(3), 366–379.
- Sanderson, I. (2006). Complexity, “Practical Rationality” and Evidence-Based Policy Making. *Policy and Politics*, 34(1), 115–132.
- Sanderson, I. (2009). Intelligent Policy Making for a Complex World: Pragmatism, Evidence and Learning. *Political Studies*, 57(4), 699–719.
- Sanderson, I. (2010). Evidence, Learning and Intelligent Government: Reflections on Developments in Scotland. *German Policy Studies*, 6(2), 53–85.
- Schneider, V. (2008). Komplexität, politische Steuerung, und evidenzbasiertes Policy-Making. In F. Janning & K. Toens (Hg.), *Die Zukunft*

- 
- der Policy-Forschung* (pp. 55–70). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Schubert, K., & Bandelow, N. C. (2003). Politikdimensionen und Fragestellungen der Politikfeldanalyse. In K. Schubert & N. C. Bandelow (Hg.), *Lehrbuch der Politikfeldanalyse* (pp. 1–21). München, Wien: Oldenbourg.
- Sciarini, P. (2004). The Decision-Making Process. In U. Klöti, P. Knoepfel, H. Kriesi, W. Linder, & Y. Papadopoulos (Hg.), *Handbook of Swiss Politics* (pp. 509–562). Zurich: Neue Zürcher Zeitung.
- Sciarini, P. (2006). Le processus législatif. In U. Klöti, P. Knoepfel, H. Kriesi, W. Linder, & Y. Papadopoulos (Hg.), *Handbuch der Schweizer Politik* (4. Aufl., pp. 491–525). Zürich: Neue Zürcher Zeitung.
- Simon, H. A. (1947). *Administrative Behavior*. New York: Free Press.
- Solesbury, W. (2001). Evidence Based Policy: Whence it Came and Where it's Going. London.
- Spinatsch, M. (2002). Evaluation in Switzerland: Moving Toward a Decentralized System. In J.-E. Furubo, R. C. Rist, & R. Sandahl (Hg.), *International Atlas of Evaluation* (pp. 375–391). New Brunswick, London: Transaction Publishers.
- Spinatsch, M. (2006). Kompetent, hartnäckig, konstruktiv: Die Hauptabteilung Staats- und Verwaltungsrecht des Bundesamtes für Justiz im Urteil ihrer Partner in der Bundesverwaltung. Bern: Dr. Markus Spinatsch, Beratung für Politik und Verwaltung.
- Strydom, W. F., Funke, N., Nienaber, S., Nortje, K., & Steyn, M. (2010). Evidence-based policymaking: A review. *South African Journal of Science*, 106(5/6), 1–8.
- Tenbenschel, T. (2004). Does more evidence lead to better policy? *Policy Studies*, 25(3), 189–207.
- Tourmen, C. (2009). Evaluators' Decision Making: The Relationship Between Theory, Practice, and Experience. *American Journal of Evaluation*, 30(1), 7–30.
- Van Dooren, W. (2008). Nothing New Under the Sun? Change and Continuity in the Twentieth-Century Performance Movements. In

- W. Van Dooren & S. Van de Walle (Hg.), *Performance Information in the Public Sector: How it is Used* (pp. 11–23). Basingstoke: Palgrave Macmillan.
- Varone, F. (2006). L'administration fédérale. In U. Klöti, P. Knoepfel, H. Kriesi, W. Linder, Y. Papadopoulos, & P. Sciarini (Hg.), *Handbuch der Schweizer Politik* (4. Aufl., pp. 289–316). Zürich: Verlag NZZ.
- Vreugdenhil, H., & Ker Rault, P. (2010). Pilot Projects for Evidence-Based Policy-Making: Three Pilot Projects in the Rhine Basin. *German Policy Studies*, 6(2), 115–151.
- Weber, M. (1947). *Wirtschaft und Gesellschaft: Grundriss der verstehenden Soziologie* (5. Aufl.). Tübingen: Mohr Siebeck.
- Weiss, C. H., Murphy-Graham, E., Petrosino, A., & Gandhi, A. G. (2008). The Fairy Godmother – and Her Warts: Making the Dream of Evidence-Based Policy Come True. *American Journal of Evaluation*, 29(1), 29–47.
- Widmer, T. (2009). The Contribution of Evidence-Based Policy to the Output-Oriented Legitimacy of the State. *Evidence & Policy*, 5(4), 351–372.
- Widmer, T., & Frey, K. (2008). *Erfolgsfaktoren der Programmentwicklung beim Bundesamt für Gesundheit: Synthesepapier*. Bern: Bundesamt für Gesundheit.
- Widmer, T., & Neuenschwander, P. (2004). Embedding Evaluation in the Swiss Federal Administration: Purpose, Institutional Design and Utilization. *Evaluation*, 10(4), 388–409.
- Yin, R. K. (2003). *Case Study Research: Design and Methods* (3. Aufl.). Thousand Oaks, London, New Delhi: Sage.

## 2.8

## Anhang

Tabelle: Eckdaten der Fälle

<i>Fall</i>	<i>Sicherheitskontrollgesetz</i>	<i>Sprachengesetz</i>	<i>Transplantationsgesetz</i>	<i>Vormundschaftsrecht</i>
<i>Offizieller Titel</i>	Bundesgesetz über die Prüfung und Kontrolle der technischen Sicherheit	Bundesgesetz über die Landessprachen und die Verständigung zwischen den Sprachgemeinschaften	Bundesgesetz über die Transplantation von Organen, Geweben und Zellen	Totalrevision des Vormundschaftsrechts
<i>Federführung</i>	Generalsekretariat UVEK (GS-UVEK), Eidg. Departement für Umwelt, Verkehr, Energie und Kommunikation (UVEK)	Bundesamt für Kultur (BAK), Eidg. Departement des Innern (EDI)	Bundesamt für Gesundheit (BAG), Eidg. Departement des Innern (EDI)	Bundesamt für Justiz (BJ), Eidg. Justiz- und Polizeidepartement (EJPD)
<i>Mandat</i>	Februar 1998	Januar 1997	Juli 1996	September 1993
<i>Vernehmlassung</i>	September 2001	Oktober 2001	Dezember 1999	Juni 2003
<i>Botschaft</i>	Juni 2006	--- (April 2004: Ablehnung Bundesrat; September 2006: Bericht zu parlamentarischer Initiative 04.429)	September 2001	Juni 2006
<i>Parlamentsbeschluss</i>	--- (September 2009: Nichteintretensentscheid)	Oktober 2007	Oktober 2004	Dezember 2008
<i>Inkraftsetzung</i>	---	Januar 2010	Juli 2007	Januar 2013

Legende: Untersuchungszeitraum: Mandat bis Vernehmlassung (Angabe entspricht Eröffnungsdatum).





---

### 3 EXPLORING THE NECESSARY CONDITIONS FOR EVALUATION USE IN PROGRAM CHANGE<sup>6</sup>

SIMONE LEDERMANN

*Research has identified a wide range of factors that affect evaluation use but continues to be inconclusive as to their relative importance. This article addresses the complex phenomenon of evaluation use in three ways: first, it draws on recent conceptual developments to delimitate the examined form of use; second, it aims at identifying conditions that are necessary but not necessarily sufficient for evaluation use; third, it combines mechanisms of evaluation use, context conditions, and actor perceptions. The study reported here examines the use of 11 program and project evaluations by the Swiss Agency for Development and Cooperation (SDC). The article makes use of qualitative comparative analysis (QCA), a method that is well suited to the study of context-bound necessity. It is concluded that the analysis of conditions that are necessary to trigger mechanisms of evaluation use in certain contexts is challenging, but promising to face the complexity of the phenomenon.*

The questions of whether and how evaluations are used are nearly as old as the evaluation business itself. A first round of empirical research about evaluation use took place in the mid-70s, when evaluators realized that their results were not implemented as they had expected (Alkin et al. 1979; Weiss 1972). However, empirical research on evaluation use has remained disappointingly inconclusive (Frey & Ledermann 2010). In the 1980s, a rich body of mostly quantitative empirical studies identified many determinants of evaluation use, but their relative importance remained contested (Alkin 1985; Cousins & Leithwood 1986). Recent reviews of the empirical literature on evaluation use also identify many potentially relevant factors, but find it impossible to state which ones are

---

<sup>6</sup> From *American Journal of Evaluation* (2012), 33(2): 159-178, doi: 10.1177/1098214011411573. An earlier version of this article has been presented at the ECPR Joint Sessions in Rennes, France, April 11–16, 2008.

most related to increasing evaluation use (Johnson et al. 2009). What has become evident through all the studies is that the phenomenon of “use” is multifaceted, and there have been numerous attempts to conceptualize it more clearly.

The study reported in this article follows the calls in recent years that it is time to abandon the ambition of finding “the important” characteristic for use and to adopt a focus on context-bound mechanisms of use instead (Henry & Mark 2003; Mark & Henry 2004). It draws on these contributions in order to conceptualize and explain use. This article remains modest as to the possible scope of explanation. Given the complexity of the phenomenon, we will never be able to fully account for any form of evaluation use. The “spirit of humility” advocated by Weiss, Murphy-Graham, and Birkeland (2005) is thus shared. Yet, this article aims to contribute to a better understanding of the mechanisms at work and to identify necessary—though not sufficient—conditions to trigger these mechanisms in specific contexts. I argue that in the face of complex phenomena like evaluation use, research on necessary conditions constitutes a good answer. The article presents a method for the identification of necessary conditions, which draws on qualitative comparative analysis (QCA; cf. De Meur & Rihoux 2002; Ragin 1987, 2000, 2008; Rihoux & Ragin 2009; Schneider & Wagemann 2007).

This method will be applied to evaluations commissioned by the Swiss Agency for Development and Cooperation (SDC). The SDC has a long evaluation tradition and belongs to the offices with the most evaluations in the Swiss Government (Balthasar 2007; Spinatsch 2002). There are different types of evaluations within SDC. The present study examines so-called external evaluations of programs and large-scale projects funded by SDC. These evaluations are commissioned by the desk manager who is responsible for strategic management of the program or project at SDC head office in Switzerland. External evaluations are carried out by external consultants and are usually scheduled toward the end of a program or project cycle in order to inform decision making by the desk manager about whether and how to continue.

The article begins with recent conceptual developments that will be drawn upon to clarify the focus of this study on decisions to substantially change the evaluated program or project. Then, hypotheses will be developed consisting of mechanisms through which an evaluation can lead to a decision of change and of conditions that are necessary to trigger these mechanisms.

These hypotheses will subsequently be tested in a qualitative comparison of the use of 11 external evaluations by SDC. It will be shown that depending on context, different mechanisms bring about a change decision and different actor perceptions are necessary to trigger these change mechanisms. The article concludes that a context-bound focus on mechanisms and on necessary conditions is expedient in the face of the complexity of evaluation use.

### **3.1 Conceptualization of Evaluation Use**

The interest in use arises from the fact that evaluation draws its legitimacy in part from practical use, that is, its ability to help improve policy, programs, or projects (Patton 1997). There is a normative expectation for evaluations to be used and SDC, for instance, has established principles to ensure successful use (SDC 2004). From the 1970s onwards, three categories of effects of evaluations were distinguished (Sager & Ledermann 2008): (a) Instrumental use: Evaluation recommendations and findings inform decision making and lead to change in the object of evaluation. (b) Conceptual use: Evaluation results lead to a better understanding or a change in the conception of the object of evaluation. (c) Symbolic use: Evaluation results are used to justify or legitimize a preexisting position, without actually changing it. Later, Patton (1997) extended the typology to include process use: Participation in an evaluation bringing about change, regardless of the evaluation results.

These categories of evaluation use have been widely applied in the relevant literature. In the last decade, however, the lack of a coherent definition has been criticized and several alternatives have been proposed. Mark and Henry (2004) distinguish between effects of evaluations at the

cognitive level (thoughts and feelings) and at the behavioral (action) level. Conceptual use pertains to the first kind and instrumental use to the second, while process and symbolic use can be both. Kirkhart (2000) proposes three dimensions according to which the effects of evaluations can be described: source for change, intention, and time. The source for change can be either the evaluation process or the evaluation results. This is what distinguishes process use from the other three types of use, which refer to the results of evaluations (cf. also Mark & Henry 2004; Weiss et al. 2005). Intention is what characterizes symbolic use. In contrast, conceptual and instrumental use can be both intended or not. As to time, Kirkhart distinguishes between immediate, end-of-cycle and long-term effects of evaluations. Furthermore, Kirkhart (2000) maintains that “utilization” and “use” are inappropriate expressions, because they suggest purposeful action, whereas in reality, the effects of evaluations are more diffuse. As an alternative, she prefers the term “evaluation influence,” denoting “the capacity or power of persons or things to produce effects on others by intangible or indirect means” (Kirkhart, 2000: 7).

In spite of these conceptual developments, most recent studies work with the conventional types of evaluation use, even though many acknowledge their weaknesses and have welcomed conceptual contributions (e.g., Balthasar 2006; Fleischer & Christie 2009; Johnson et al. 2009; Weiss et al. 2005). As Weiss and her colleagues (2005) put it: “the three constructs of instrumental, conceptual, and political [here: symbolic] use appear to capture much of the experience in the empirical literature and practical experience.”

This article combines old and new. It focuses on evaluation-based decisions to change the object of evaluation as a specific type of instrumental use and draws on the conceptual developments in order to delineate this type of use. Instrumental use is the most straightforward effect of evaluations (cf. Weiss 1998; Weiss et al. 2005). In Mark and Henry’s terms (2004), it is the behavioral outcome of “program continuation, cessation, or change” that is of interest. Change in existing structures—and cessation is just a form of radical change—is likely to lead to opposition. Under what conditions are SDC desk managers prepared to take a decision to

change a program or project based on an external evaluation even though they might have to face opposition? This is the central question of the present article. The study concentrates on decisions with some bearing on the program or project (e.g., change of location or partner organization, modification of strategic orientation, termination of the program or of part of it, etc.), because it is assumed that in these cases, the preconditions for evaluation-based decision making emerge more clearly than in the case of decisions to continue with the status quo. Furthermore, the focus on evaluation-based change decisions corresponds to the purpose of SDC's external evaluations. The SDC's principles on external evaluation underline the importance of preparing for change throughout the evaluation process (SDC 2004). The SDC has showed great interest in increasing this kind of use, also with the help of the present study.

With respect to the time dimension, I examine the end-of-cycle effects of evaluations that are looked at, while the more diffuse long-term effects are ignored. The study has covered the period of roughly half a year after the end of the evaluations. The change decisions had to have been formally taken, and there had to be a written "proof" for a decision to be counted as such. No attempt has been made to find out whether the decisions were actually carried out in the end. Neither did I distinguish between the process and the results as possible sources of change. The information from the evaluation certainly had to have some influence, or in other words "leverage" (Cronbach 1982), on the desk manager's decision for it to be considered an evaluation-based decision. But I did not care whether the decision had been taken because of the process of evaluation or because of the findings. Given the situation that SDC desk managers are usually at a great geographical distance from the program or project site, the results of the evaluation are likely to be more important than in other contexts.

### **3.2 A Context-Mechanism-Actor Model of Evaluation Use**

In Weiss' (1998) words: "Use is about change. Any theory of evaluation use has to be a theory of change." I propose a theoretical model of eval-

uation-based change constructed according to a “realistic” scheme of theory-building (Pawson & Tilley 1997) consisting of different mechanisms of evaluation use and different contexts.

Mark and Henry propose a “theory of evaluation influence” with an abundance of interdependent mechanisms (Henry & Mark 2003; Mark & Henry 2004). While their list is comprehensive, the high number of mechanisms and possible interdependencies is likely to lead to pathways for evaluation use that look different for each case (Weiss et al. 2005). In order to prevent an “individualization” of each case, where explanatory patterns are not visible (Berg-Schlosser & De Meur 2009), I only investigated a small number of mechanisms with a sound empirical foundation.

In a well-known review of research on evaluation use, Cousins and Leithwood (1986) propose a conceptual framework with 12 factors that are assigned to two dimensions—evaluation implementation and decision/policy setting. I selected two factors from each dimension. But instead of examining their separate influence on evaluation use, as it is usually done, I adopted a conjunctural approach (Amenta & Poulsen 1994; Yamasaki & Rihoux 2009), where it is combinations of factors that only together produce an outcome. This conjunctural approach is applied both in theory-building as well as in the empirical analysis (see next section).

In the dimension of evaluation implementation, the “truth test” described by Weiss and Bucuvalas (1980) singles out evaluation quality and the nature of the findings as two factors that are by now widely accepted to be important. They found that a study is perceived as useful, either if it confirms users’ preconceptions or if it is considered good-quality research. The factors are interdependent: If a study challenges users’ preexisting beliefs and reveals something new to them, it must be judged high quality to be considered useful, whereas quality is less important if the study confirms users’ expectations. This means that the perceived quality of a study must be higher, the higher its “novelty value.”<sup>7</sup> The

---

<sup>7</sup> Weiss and Bucuvalas were surprised to find that in addition to a study’s consistency with users’ knowledge and values (truth test), consistency with institutional norms (utility test) emerged as a separate dimension in factor analysis. In the present study, the two dimensions are taken together, be-

novelty value and quality of evaluations are assumed to affect users' judgments about the usefulness of an evaluation and, as a consequence, are likely to be relevant to whether or not an evaluation is actually used as a basis for decisions about change in the program or project.

The added value of this article is that the truth test is put into context and further differentiated. This is done based on an empirically derived model by Valovirta (2002) that illustrates the significance of two specific characteristics of the decision and policy-setting dimension. According to Valovirta, evaluation use is an argumentative process in which evaluations provide arguments that correspond more or less to users' beliefs and expectations and which they can draw on or reject. The use of the arguments provided by an evaluation will depend on how actors perceive the study. This perception, according to Valovirta, depends on context. Based on empirical research about the use of evaluations in the Finnish Government, he singles out two major context factors: the level of conflict among the stakeholders and the amount of pressure for change. These two context conditions together result in four mechanisms in which an evaluation can bring about change as shown in Figure 3-1.

**Figure 3-1:** Mechanisms of Evaluations Use in Different Contexts

		<i>Level of conflict</i>	
		<i>Low</i>	<i>High</i>
<i>Pressure for change</i>	<i>Low</i>	(1) Awakener	(4) Conciliator
	<i>High</i>	(2) Trigger	(3) Referee

*Source:* Valovirta (2002, Figure 2) with own adaptations.

The cells of the figure can be explained as follows: (a) In a situation of low pressure for change and a low level of conflict, an evaluation can reveal unknown problems and act as an awakener. (b) In an environment of high pressure and low conflict, an evaluation can be a trigger for changes that are broadly accepted as necessary. (c) Under conditions of

---

cause they could not empirically be isolated. The resulting "novelty value" factor shows whether the evaluation revealed something new or unexpected to the users that challenged either their personal beliefs or their institutional norms or both.



high pressure and high conflict, the evaluation is likely to function as a referee, deciding what ought to be done, even though the solution might not satisfy all the stakeholders. (d) In a situation of low pressure and high conflict, the evaluation is likely to be exploited to defend one's positions and criticize others, without necessarily being used as a basis for change. Exceptionally, however, an evaluation in such a situation can act as a conciliator between the conflicting parties, thereby enabling change to happen.

Will the actors perform a sober truth test on the evaluation in all of these situations before making use of it? According to the underlying hypothesis of this article, this is not the case. Depending on the context, actors' perceptions of the novelty value and evaluation quality are likely to be more or less relevant for whether an evaluation is taken as a basis for change decisions.

Table 3-1 summarizes the necessary actor perceptions for an evaluation to bring about change in the four specified context and through the four specified mechanisms.

**Table 3-1:** Hypothesized Necessary Actor Perceptions for Use in Different Contexts

<i>Context conditions</i>		<i>Mechanism</i>	<i>Assumed necessary actor conditions</i>	
<i>Pressure for change</i>	<i>Level of conflict</i>		<i>Novelty value</i>	<i>Evaluation quality</i>
low	low	Awakener	high	high
high	low	Trigger	irrelevant	high
high	high	Referee	irrelevant	irrelevant
low	high	Conciliator	high	high

It can be hypothesized that:

- To function as an awakener, the evaluation must first of all reveal something new to the users. Moreover, given the neutral context, actors are likely to perform a truth test, so the quality of an evaluation also tends to be relevant for the evaluation to be considered trustworthy.

- To act as a trigger for change in a consensual situation, where people are aware of problems, an evaluation is likely to be used even without revealing much new. The absence of conflict allows for a sober consideration of the evaluation, so its quality is assumed to be important.
- In a conflict-laden situation with a strong pressure for change, an evaluation is likely to be used as a referee, no matter its quality or novelty value, because something has to be done. Different groups are likely to draw on different parts of the evidence and interpret it differently (Jewell & Bero 2007).
- Where there is a lack of problem awareness among stakeholders that are in conflict with one another, the chances for an evaluation to be used for change decisions are small. To act as a conciliator, it is to be assumed that an evaluation must be of high quality and must show some new ways out of the situation.

To sum up, this article examines the interplay between the policy and decision setting on the one hand and specific characteristics of evaluation implementation on the other hand. I investigate how the truth test, which is based on the actor perceptions about the novelty value and quality of an evaluation, works under varying levels of conflict and pressure for change. The question is whether these factors contribute to an explanation of why some evaluations are used as a basis for decisions to change the evaluated program or project while others are not.

It is not contended here that the context and actor-related factors fully account for the occurrence of any evaluation-based change decision. I will not try to identify all the ingredients needed for an evaluation to cause change. As past research has shown, evaluation use is a complex phenomenon in which so many aspects are potentially relevant, that this would appear to be an excessive claim.

The assertion is a more modest one: To find conditions that are necessary for evaluations to cause change in certain contexts. In other words, I am interested in a specific set of change mechanisms, each operating in a specific context. I attempted to find some preconditions for an evaluation to be able to act as an awakener, trigger, referee, or conciliator un-

der the above-mentioned contexts characterized by the presence or absence of conflict or pressure for change.

### **3.3 Method**

The hypotheses in Table 3-1 are tested with a comparative case study design. This section describes the research context and case selection, data collection, and data processing. In particular, it introduces the method of QCA.

#### **3.3.1 Research Context and Case Selection**

The theoretical framework described above has been applied to a set of external evaluations of the SDC. Because organization matters for evaluation use (e.g., Weiss 1998) and because organizational factors are not part of the research model, the focus on just one type of evaluation within only one organization is a crucial control. The SDC has internal guidelines about how to carry out external evaluations that guarantee a certain degree of similarity of the process. The research design thus corresponds to a “comparative-cases strategy” (Lijphart 1975), in which cases are selected according to the “most similar” design (Przeworski & Teune 1970) that maximizes the variance of the explanatory factors and minimizes the variance of the control conditions. There are about 30 planned external evaluations per year (cf. SDC 2002 and following years) that provide enough variance. As it was not possible to ensure the “intimacy” with each of the 30 cases—something that is crucial for the application of QCA (Berg-Schlosser & De Meur 2009; Ragin 1994; Rihoux & Lobe 2009)—11 cases were selected from among the population.

External evaluations are scheduled toward the end of a funding cycle in order to inform decision making about whether and how to continue with a specific program or project. They are carried out by one to three external consultants. The terms of reference with the evaluation questions are set by the desk manager at the head office in Switzerland responsible for the program or project, usually in collaboration with the local SDC field

office. Field office staff frequently assist evaluators in organizing visits to program and project sites. At the end of the visit, a debriefing between the evaluators and the local SDC staff is common, sometimes together with program and project staff. However, the main intended users of the external evaluations are the desk managers. They receive the evaluation report and in most cases, a second debriefing with the evaluators takes place at SDC headquarters in Switzerland to discuss the report. Finally, the desk managers have to take the decisions about what to do with the programs and projects based on the evaluation.

In order to assure a certain breadth, the cases were selected according to the structure of the organization (cf. Merkens 2003). The basis of selection was the SDC evaluation program (SDC 2002). Out of 30 external evaluations listed in the program, 24 had actually taken place and form the universe from which the cases were selected. As a first step, one external evaluation was drawn at random from each of the five organizational units (called departments) within SDC that had carried out at least one external evaluation. As a second step, seven more cases were selected at random to achieve a proportional representation of each department with respect to the total number of external evaluations. The case from the so-called Thematic Department had to be dropped in the course of data analysis due to a lack of information on its use by the desk manager. The final sample listed in Table 3-2 consists of 11 cases and covers 4 departments. The cases cover evaluations of SDC funded program and projects all around the world. At SDC's request for anonymity, the cases are just referred to by random uppercase characters.

**Table 3-2: Cases by Department**

<i>SDC Department</i>	<i>Cases</i>
Eastern Europe and CIS	A, B, C, D, E
Bilateral Cooperation	F, G, H
Humanitarian Aid	I, J
Multilateral Cooperation	K

*Note:* SDC = Swiss Agency for Development and Cooperation.

### **3.3.2 Data Collection**

Given the desk managers' central role as commissioners and main users of external evaluations, I used interviews with the desk managers responsible for the programs and projects as the main source of information for the case studies. Due to SDC internal staff rotation, half of the desk managers had already left their position at the time of the interview and were often abroad. In some cases, it was necessary to interview more than one desk manager on the same project or program evaluation. In total, 14 desk managers were interviewed, half of them face-to-face, four on the phone, and three in written form by e-mail. The interview guides consisted of questions on the purpose and context of the evaluation, the selection of the evaluators and the process of evaluation, the quality of the evaluation, and its use for the desk managers and the organization. For each topic, there was a mix of closed questions where desk managers were asked to give their assessment on a 4-point scale and open questions where they were asked to elaborate on their view. Face-to-face interviews took about one and a half hours; telephone interviews were a bit shorter. In the case of the written interviews, there were at least two rounds in which desk managers were asked to give more detailed information on certain aspects. In addition to the interviews with desk managers, six telephone and four e-mail interviews were conducted with the external evaluators as a complementary source of information. The guideline consisted of open questions, especially on the context of the evaluation and the follow-up process. The telephone interviews took about 40 min.

Document analysis was used first to prepare the interview guides and second to cross-check the interview information. The documents included the evaluation reports, the terms of reference and the evaluation contracts, as well as program and project proposals for the following funding cycle. Empirical work was carried out mainly in 2004 roughly half a year after the end of the evaluations. At this point, the interviewees were still able to recall the evaluation process and the desk managers had already taken their decisions about how to continue.

---

### 3.3.3 QCA and the Concept of Necessity

Earlier it was noted that “intimacy” with each case is important for an accurate comparative case analysis. Therefore as a first step, each case has been regarded as a unique narrative of evaluation-based decision making and has been analyzed on its own. The interplay between actor perceptions, mechanisms, and context has been traced in the form of a “thick description” (Geertz 1973) of how the evaluation has been conducted, in what context, how it has been perceived and used for decision making by the desk manager. The case narrative corresponds to the abovementioned conjunctural approach (Amenta & Poulsen 1994; Yamasaki & Rihoux 2009) in that it traces the interplay between actor perceptions and context that in their combination lead to a change decision (or the lack of it).

As a second step, the information from the case studies was systematized in order to allow for the subsequent comparison of the cases by the method of QCA (Ragin 1987). QCA uses set relations and formal logic to find commonalities between various cases with the same outcome. Contrary to statistical methods, which attempt to measure the “net effect” of single, independent variables on an outcome, QCA tries to explain outcomes through combinations of interdependent conditions or in other words through “configurations” (Ragin 2008). This configurational thinking about causally relevant conditions that only together lead to an outcome is common in qualitative research in general. What is special about QCA is the degree of systematization that allows for the comparison of a greater number of cases.

Each case was systematically classified as to how well it fulfills each of the conditions (pressure for change, level of conflict, novelty value of evaluation, evaluation quality) and whether the outcome (evaluation-based change decision) is present or not. Once all the cases were classified, they were compared with each other by formal logic in order to find the configurations that are necessary or sufficient for the outcome. Given the interest of the present study in the interplay between contexts and actor perceptions, QCA as a configurational method was well-suited for the endeavor (Befani et al. 2007).

There have recently been important methodological advances (e.g., Brady & Collier 2004; Mahoney & Goertz 2006). The concepts of necessity and sufficiency have been proven to be helpful to deal with the causal complexity we are confronted with in the real-world (Goertz 2006a, 2006b; Goertz & Starr 2003; Ragin 2000). As the present article is interested in finding the conditions that are necessary for evaluation-based change decisions, it concentrates on the concept of necessity, leaving aside sufficiency.

The claim that a condition is necessary for an outcome is a strong one, because it means that this condition always has to be present if an outcome is to occur. Often, however, there are different causal paths to the same outcome, and it is only very trivial conditions that must always be present. Contrary to sufficiency, however, necessity does not imply that the condition will really always lead to the outcome. It is just a precondition for the outcome. In addition, this study will investigate context-bound claims of necessity, so the claim is a weaker one. It will try to identify necessary, but probably insufficient conditions for an evaluation to inform program/project change decisions under particular circumstances. The result is what George and Bennett (2005) call “contingent generalizations”.

The analysis of necessary conditions is usually considered to be just the first step of a QCA analysis, before turning to the analysis of sufficiency (Schneider & Wagemann 2007). But with regard to complex social phenomena, such as evaluation use, where so many factors might be relevant, context-bound necessity claims seem more adequate than sufficiency claims. This is why necessity actually deserves an analysis in its own right.

### **3.3.4 Data Dichotomization**

Several QCA techniques have recently been developed (cf. Rihoux 2006; Rihoux & Ragin 2009; Schneider & Wagemann 2007). The present article applies the basic technique of crisp-set QCA (cf. Grofman & Schneider 2009 for a short general introduction). Crisp-set QCA requires

a dichotomization. For every case, it has been decided whether the outcome (evaluation-based change decision) has occurred or not and whether each of the four conditions listed is high or low.

**Table 3-3:** Specification of Conditions and Outcome

<i>Element</i>	<i>Fulfilled (= 1), if:</i>
Outcome: Evaluation-based change decision	Desk manager mentions that based on the evaluation, she or he has decided to make significant changes in the program or project, such as: unplanned termination of (part of) the program or project change in partner organization important strategic change in the program/project with major consequences at the operational level The change decision has to be documented in some way (e.g. project proposal for next funding cycle, letter to partner organization, etc.). It is not necessary that the change decision has been implemented.
Context condition 1: Pressure for change	Desk manager mentions problems with the program or project that required a change and that he or she was already aware of before receiving any evaluation information. Interview information was cross-checked with evaluation purpose statement in the terms of reference.
Context condition 2: Level of conflict	Desk manager or evaluator mentions specific conflicts in the project or program that affected the desk manager in his or her function.
Actor condition 1: Novelty value	Desk manager mentions specific points she or he learnt from the evaluation or was surprised to hear from the evaluator.
Actor condition 2: Evaluation quality	Desk manager gave positive ratings for the evaluation on the majority of the closed questions on the following accuracy standards : <ul style="list-style-type: none"> <li>• precise description of procedures</li> <li>• appropriate application of research methods</li> <li>• trustworthy sources of information</li> <li>• substantiated conclusions</li> <li>• neutral reporting</li> </ul>

Mayring's (2003) method of qualitative content analysis was used to produce the dichotomized data matrix. For this systematic, rule-guided qualitative text coding, the categories are deduced from the theoretical context. In this case, it was the four conditions and the outcome of evaluation-based change decision that served as categories. After coding a



subset of the interview transcripts and documents, the categories have been refined and then applied to the rest of the corpus.

The codings for each category have been summarized for each case, and based on these summaries, the cases have been compared and the categories have been rated high or low. It is crucial that this dichotomization is performed on the basis of the intimate knowledge of the cases. Table 3-3 specifies the categories for the outcome and the four conditions as applied.

In the rare instances of contradicting information on the same condition (contradicting interview statements or document information), the condition was only coded “high” if the majority of the sources pointed in this direction, otherwise it was assessed to be low. The dichotomization led to the data matrix in Table 3-4, which lists all the cases and codings.

**Table 3-4:** Data Matrix

<i>Case</i>	<i>Context conditions</i>		<i>Actor conditions</i>		<i>Outcome</i>
	<i>Pressure for change</i>	<i>Level of conflict</i>	<i>Novelty value</i>	<i>Evaluation quality</i>	<i>Change decision</i>
A	0	1	1	1	0
B	1	1	1	1	1
C	0	0	1	1	1
D	1	1	0	0	1
E	0	0	1	1	1
F	1	0	0	1	1
G	0	0	0	1	0
H	1	1	0	1	1
I	0	1	0	0	0
J	0	0	0	0	0
K	1	0	0	0	0

*Note:* 0 = condition low / outcome absent; 1 = condition high / outcome present.

In the following paragraphs, I have selected particular cases to illustrate the dichotomization of the conditions and the outcome. More detail on each case can be found below in the “findings” section of the article.

In all 6 out of the 11 external evaluations informed a decision to substantially change the evaluated program or project. In case C, for instance, the substantial change consisted in an additional component for the existing emergency medicine program (delivery of consumable material to local hospitals). In case E, the desk manager decided to terminate the project based on the evaluation, although this decision has finally not been implemented for political reasons. Given the focus of this study on decision making by the desk manager, the outcome is nonetheless rated positive. Five external evaluations did not lead to any important change decisions. For example, evaluation J, which concerned a building project, did not inform the desk manager’s decision because he could not find an answer to his strategic questions in the evaluation, so the next funding cycle was started without any major changes. In case A, the decision to terminate the program had already been taken when the evaluation delivered its results. The lessons learnt from this evaluation had not informed any decisions about other programs by the time of the present study.

As to the context conditions, the pressure for change was graded high in case D, for example, because the desk manager had encountered major problems with an important partner all along. In three out of five cases, where the level of conflict was high, this was due to the fact that the program or project was very controversial within SDC or between SDC and a second funding body in the federal administration (cases A, B, and H). In the other two cases, there were severe conflicts between SDC and the local implementing agencies. In case F, there were conflicts between local stakeholders the desk manager did not know about and which did not affect him; the level of conflict is, therefore, considered low.

Concerning the desk managers’ perceptions of the evaluations, the novelty value was rated high in four cases. For instance, the lack of consumable material in local hospitals, which evaluation C revealed, came as a great surprise to the desk manager. In all 7 out of the 11 evaluations

were considered of high quality by the desk managers. The cases B, F, and H, for example, were rated positively on all five accuracy standards, whereas evaluation I was assessed negatively throughout. In case K, the description of the procedure, neutrality, and the soundness of the conclusions were negative.

### **3.3.5 Steps of a QCA Analysis**

The data matrix in Table 3-4 served as a basis for the QCA analysis. The analysis of necessity focuses on the cases where the outcome has occurred, that is, the evaluations that have actually led to a change decision (cf. Schneider & Wagemann 2007, for the steps of a necessity analysis). It is analyzed whether these cases show a common condition or a common absence of a certain condition. If they do, this is an indication that this condition or its absence, respectively, might be necessary for the outcome to occur. The claim that a condition is necessary must be substantiated in qualitative terms based on the knowledge of the cases (cf. Rihoux & Lobe 2009). It must be argued why the presence of the necessary condition was in fact causally relevant for the occurrence of the outcome.

To further consolidate the necessity claim, the analysis can turn to the cases where the asserted necessary condition is absent. It follows logically that the outcome in these cases did not occur. Based on the case knowledge, it can be checked whether the absence of the necessary condition was causally relevant for the absence of the outcome. If it is, the necessity claim is substantiated further. However, there might be other reasons why the outcome did not occur. Even though the necessary condition is present, the outcome might be absent, but this does not challenge the necessity claim. For instance, an evaluation might not have been used for decision making because the information was not available in time. This, however, would not contradict the claim that good evaluation quality is necessary for evaluation-based decision making.

In the present study, the analysis of necessity is context-bound. This means that for each combination of the two context conditions “pressure

for change” and “level of conflict,” the analysis of necessity is carried out separately. First, we look at the cases with low pressure for change and a low level of conflict and examine whether the evaluations that were used for decision making under these context conditions show any common actor conditions that can be considered necessary. Then, we turn to the cases with low pressure for change, but high conflict, and check for necessity and so on and so forth.

In brief, data analysis with QCA involves a three-step procedure: First, each case is analyzed separately in order to trace the mechanisms that have led to a change decision or, on the contrary, have failed to do so. Second, based on the dichotomous data matrix, the cases in each context constellation are systematically compared with each other. Third, the results of this comparison are interpreted based on the intimate knowledge of each case. The findings from this analysis will be presented in the next section.

### 3.4 Findings of the Comparative Case Analysis

The data matrix in Table 3-4 shows that the cases with a positive outcome (desk manager has taken an evaluation-based change decision) do not share any common condition, nor do the cases with a negative outcome. This is to say that no single condition alone is necessary for the occurrence or nonoccurrence of an evaluation-based change decision. This complex structure in the data hints at the complexity of the underlying causal links and suggests that a context-bound view is more adequate.

**Figure 3-2:** Distribution of Cases by Context

		<i>Level of conflict</i>	
		<i>Low</i>	<i>High</i>
<i>Pressure for change</i>	<i>Low</i>	(1) Awakener <i>C, E, G, J</i>	(4) Conciliator <i>A, I</i>
	<i>High</i>	(2) Trigger <i>F, K</i>	(3) Referee <i>B, D, H</i>

*Note:* Characters in italics denote cases where the outcome is present.

Analogous to Figure 3-1, Figure 3-2 shows the distribution of the cases along the two context dimensions and indicates the outcome, that is, whether in the respective cases an evaluation-based change decision has been taken or not.

The 11 cases are scattered among the four contexts. In the “awakener” and “trigger” contexts, there are both cases with a positive and a negative outcome. In the “referee” context, all the cases show a positive outcome, which means that all three external evaluations were used as a basis for a change decision. In contrast, in the “conciliator” context, none of the evaluations has been used in that way.

For each of the four contexts in Figure 3-2, I examine whether the claimed change mechanisms have been present and whether the necessity claims are substantiated or not.

### **3.4.1 Evaluation as an Awakener (Low Pressure, Low Conflict)**

In the context of low pressure and low conflict, it had been claimed that an evaluation can cause change by awakening people, provided that it reveals something new and that it is of good quality. There are two cases, C and E, where the desk managers have actually taken a change decision, and two cases, G and J, where they have not.

Case C fulfills the two actor conditions that have been claimed necessary: The desk manager was convinced of the good quality of the evaluation, which had been carried out by two highly competent logistics specialists. Furthermore, the evaluation revealed something new, as it disclosed a severe shortcoming in the program (lack of consumables in emergency medicine) in one of the towns that had been considered the model project site. As a consequence, discussions were held with the state health ministry and the desk manager decided to integrate the missing component (consumable supply) in the program. The evaluation did clearly work as an awakener.

Case E also fulfills the two supposedly necessary conditions of novelty and good quality. The evaluation was assessed as good by the desk

manager and gave her a lot of new information, given that she did not know much about the project before. In fact, the evaluation detected several deficiencies and suggested to terminate the project, which the desk manager decided to do. However, for political reasons, the SDC hierarchy intervened and decided that despite of the clearly negative evidence, the project was to continue. So the desk manager had taken a change decision based on the evaluation (positive outcome), but in the end it has not been implemented.

To further substantiate the claim that the two conditions of novelty and high quality are necessary for an evaluation to trigger change in a consensual situation with a low pressure for change, it is useful to consider the negative cases G and J, where one or both of the necessary conditions were absent. Evaluation G was considered good quality but did not disclose anything unknown. Instead, it confirmed the project strategy, which had been adopted before the evaluation was made. The lack of novelty was a crucial reason why the evaluation did not result in a change decision.

Evaluation J was estimated poor quality by the desk manager. The report consisted of only seven pages of unstructured text, describing the project site, which the desk manager knew from his own visits to the place. Many evaluation questions remained unanswered and the evaluation did not provide any new information. In this evaluation, bad quality and a lack of novelty go together.

### **3.4.2 Evaluation as a Trigger (High Pressure, Low Conflict)**

In a consensual environment where stakeholders are aware of problems that must be solved, evaluations are assumed to trigger change only if they are of good quality. They do not need to show something new. It is sufficient if they confirm a strategy of action the stakeholders have thought of already.

Case F corroborates the account of how an evaluation can act as a trigger for change. There were several signs that the program needed modification (cancellation of the cooperation contract by the partner universi-

ty, high staff fluctuation, lack of impact on government). The evaluation was carried out in a top-down manner, without much participation of the local stakeholders. The SDC's general strategy for the state in question was to foster decentralization to allow for a more balanced development of the country as a whole. So, when the evaluation recommended decentralizing the activities, the suggestion was taken up and the program was adapted accordingly. The evaluation triggered strategic program modifications SDC had already considered.

In case K, there were also signs for a need for change, such as high overhead costs of the organization in charge, but the evaluation was not used as a basis for a change decision in the end.

One of SDC's main evaluation questions was whether to continue the program or not and with what organization. However, the evaluation failed to address these questions. Rather, it restricted itself to how the program could be improved, without challenging the existing arrangement in a more fundamental way. The desk manager considered the evaluation as biased and of bad quality, which is why it failed to act as a trigger for a change.

On the whole, the evidence supports the assumption that an evaluation must be conceived good quality in order to trigger change in a consensual context with high pressures for change and that novelty is less an issue.

### **3.4.3 Evaluation as a Referee (High Pressure, High Conflict)**

In a conflict-laden, high-pressure environment, it has been assumed that neither novelty nor quality is necessary for evaluations to be used as a referee to decide what to change. However, the conclusions of the evaluations are likely to be accepted only by one part of the stakeholders.

The three cases B, D, and H, which fall in this context, all led to decisions for substantial change in the evaluated programs. They differ with respect to their quality and novelty value, so the assumption that these actor perceptions are less important than in other contexts is confirmed.

At a closer look, however, it appears that the quality is important if the evaluations challenge existing beliefs.

In case H, the evaluation showed that the development fund, which was the object of evaluation, had to change its strategy, because there were too few requests for financial assistance. The evaluation suggested that the fund start to develop its own projects instead of waiting for demands for financial assistance from other organizations. The evaluation confirmed the desk manager's opinion, but challenged the position of another involved Swiss federal agency, with which the SDC desk manager had been in conflict for a long time. The evaluation results also contradicted some of the local stakeholders, all of whom were part of the decision committee of the fund, which had to agree to the strategic change. So, even though the results were not new to the desk manager who took them as a basis to advocate a strategic change (positive outcome), they were new to some of the other decision makers. The latter were finally convinced by the evaluation not least thanks to its very good quality. The evaluation provided a good description of the procedure and the criteria.

Together with case H, evaluation B is the most sophisticated in the sample. The desk manager was totally convinced of the program, but there had been severe internal conflicts about its appropriateness for a development agency like SDC. The evaluation showed that the program was effective, and given good evaluation quality, internal criticisms toward the program ebbed down. The desk manager was supported in her decision to expand the program. In one respect, however, the desk manager's position was challenged as the evaluation showed that one of the project organizations was very expensive compared to others. The desk manager considered the evidence for this point to be sound and, as a consequence, drove down cooperation with this organization. Here again, good quality was necessary for the desk manager to take the evaluation as a basis for the change decision.

Evaluation D confirmed the desk manager's impression that the local project organization failed to implement an important part of the planned activities, which had been a cause of conflict throughout program implementation. As a consequence of the evaluation, the desk manager re-



placed the organization. The desk manager assessed evaluation quality rather negatively in the closed questions, but emphasized several times that she did not much care about it. It seems that evaluation quality did not matter to the desk manager, because the results were as she had expected.

The discussion of these three cases shows that in the context of high conflict and high pressure for change, evaluation users decided according to the truth test (Weiss & Bucuvalas 1980): High evaluation quality is only necessary if an evaluation challenges preexisting beliefs of decision makers. Users were only willing to revise their opinion if they were convinced by the quality of the results.

Conversely, if evaluation results confirm decision makers' opinions, quality is less an issue. None of the two conditions seems, however, always necessary.

#### **3.4.4 Evaluation as a Conciliator (Low Pressure, High Conflict)**

According to the hypothesis, in a situation of conflict, where stakeholders are not much aware of a need for change, substantial change decisions are only taken if an evaluation is regarded good quality and shows new ways out of disagreement. Situations where an evaluation acts as a conciliator are deemed to be rare.

Both evaluations in the sample, undertaken in a conflict-laden environment without pressure for change (cases A and I), were not used for a change decision, so strictly speaking it is not possible to assess the necessity claims.

In case I, there was a conflict between two sections within SDC about the aims of the evaluation. The role of the evaluators was not clear from the start. After several attempts to rewrite the intermediate evaluation report in order to suit both sections, the evaluation was stopped midway. Conflict was all-pervasive and determined desk managers' perceptions of evaluation quality and its novelty value; the involved desk managers were unable to give a differentiated judgment, so that the measurement

of these two conditions is questionable. The desk managers were totally unwilling to take decisions upon the results of the evaluation in the intermediate report. Overall, the evaluation acted as a trigger for open conflict rather than as a conciliator.

In case A, there was clear mistrust between SDC and the project organization. Against the desk manager's expectations, SDC rather than the project organization was criticized by the evaluation. The desk manager considered the evaluation of good quality and accepted the critique, but did not act on it, because the decision to terminate the project had already been taken. The aim of the evaluation was merely to draw the lessons learnt from the project, but this did not have an impact on the decisions by the desk manager or SDC in general (at least in the period of study). The conditions that have been claimed necessary for an evaluation to act as a conciliator were fulfilled, but the evaluation was not used to for decision making because the decision had already been taken. Otherwise, it might have been used.

In the absence of cases with a positive outcome, it is not really possible to test the assumption that an evaluation must show something new and be considered of good quality to motivate a change decision in a conflict-laden environment, where the pressure for change is low. But case A points into this direction.

### **3.5 Discussion**

The inherent complexity of the phenomenon of use is an important reason why research results on evaluation use have largely remained inconclusive so far. In order to address this complexity, I adopted three strategies: first, I built on recent conceptual contributions to confine the analyzed type of "use," namely evaluation-informed decisions to substantially change the evaluated program or project as one specific type of instrumental use. Change decisions imply that much is at stake, so it can be assumed that the ingredients that are necessary for such an important evaluation-based decision appear more clearly than in an analysis of minor decisions.

Second, I did not try to predict such change decisions. Rather, I examined how the decision makers (in this case, the desk managers in charge of the program or project) must perceive the evaluation so that they might take such a decision. In other words, I analyzed necessary conditions. Third, I presumed that the necessary conditions depend on the context, because depending on the context it is other mechanisms that bring about a change decision (cf. Astbury & Leeuw 2010). Because of context dependency, this study has adopted a conjunctural approach to both hypothesis formulation and testing. According to this approach, it is combinations of factors instead of single conditions that cause an outcome.

The empirical part of this article is exploratory in character. A total of 11 external evaluations on projects or programs funded by SDC have been used for a comparative case analysis. Decision making by the desk managers who commissioned the evaluations has largely followed the assumed patterns. Depending on the level of conflict and on existing pressures for change, evaluation information has been used differently, and it is different conditions that have been necessary to trigger the different mechanisms of use.

**Table 3-5:** Results about Necessary Actor Perceptions for Evaluation Use in Different Contexts

<i>Context conditions</i>		<i>Mechanism</i>	<i>Necessary actor conditions</i>	
<i>Pressure for change</i>	<i>Level of conflict</i>		<i>Novelty value</i>	<i>Evaluation quality</i>
low	low	<i>Awakener</i>	high	high
high	low	<i>Trigger</i>	irrelevant	high
high	high	<i>Endorser</i>	<i>low</i>	<i>irrelevant</i>
		<i>Reviser</i>	<i>high</i>	<i>high</i>
low	high	<i>Conciliator</i>	?	?

*Note:* Italics indicates refinement of hypothesis; ? indicates necessity test not possible.

Table 3-5 summarizes the detailed results of the hypothesis test for the four context constellations that have been examined. For the first two context constellations, the cases confirm the assumptions that have been presented in Table 3-1: In a consensual environment with low pressure for change, evaluations can act as awakeners, if they disclose something unknown and are considered good quality. In high-pressure consensual situations, evaluations can function as a trigger for change decisions that are accepted as indispensable, but only provided that evaluation quality is deemed to be good. In contrast, novelty is not a precondition.

For the third context of high conflict and high pressure for change, the hypotheses can be refined based on the case studies. The evaluation does not simply act as a referee between different positions; rather the mechanism appears to depend on the novelty value. If the evaluation confirms the decision maker's expectations and does not provide much new information, it can be used to endorse one's decision as to how to change the evaluated program or project. There was one case in the sample which suggests that the quality of the evaluation does not matter much in this situation. In contrast, if the evaluation does not confirm the decision maker's expectations, it might still be used as a basis for a change decision, if the decision maker is persuaded that she or he had been wrong. In order for decision makers to revise their opinion, it is necessary that they consider the evaluation of high quality. This means that in this situation of high conflict and high pressure for change, decision makers actually performed a "truth test" in the way as described by Weiss and Bucuvalas (1980), where novelty and evaluation quality are interdependent, and none of the two alone is necessary.

Finally, the extent to which novelty and evaluation quality are necessary for an evaluation to act as a conciliator in a situation of conflict and low pressure could not be tested because the two cases in the sample that fall into this context were not used as a basis for a change decision. Given that, according to Valovirta (2002), it is rare that evaluations are actually used in such contentious situation where change is not pressing, this lack of an empirical case does not come as a surprise.

Based on the results of this article, certain recent findings about evaluation use can be somewhat differentiated. In a recent survey among American evaluators (Fleischer & Christie 2009), beliefs and values of key stakeholders have been considered one of the main barriers to evaluation use. According to the above results, this is mostly true: if beliefs and values of decision makers are challenged by new evaluation information, the hurdle for its use for decision making is higher, as the evaluation needs to be considered of good quality, too. However, there are contexts where the novelty value of an evaluation does not seem to be a relevant barrier to change decisions, namely where the pressure for change is high and stakeholders get along well with each other.

According to this study, evaluation quality appears to matter in most contexts, whereas in a recent review of empirical studies about use (Johnson et al. 2009) and in the above-cited survey among American evaluators (Fleischer & Christie 2009), high standards of methodological rigor figured among the less important factors for use. Evaluation quality in the present article has been measured as perceived by the decision makers (the desk managers). The same external evaluations have been part of a meta-analysis that was made in parallel to the present study and where evaluation quality was assessed by professional evaluators according to evaluation standards similar to those of the Joint Committee (1994). In several cases, the desk managers' quality ratings were much higher than those of the meta-analysis, where many of the evaluations were considered of rather poor quality. The evaluations suffered from several shortcomings that according to Thomas (2010) are common in development evaluation. In fact, quality as rated by the meta-analysis is not a necessary condition for evaluation-based change decisions in any of the contexts. This underlines that, as Patton (1997) puts it, use depends on evaluation quality as "matters of subjective user judgment" rather than on preset standards.

---

### 3.6 Limitations of the Study

At a conceptual level, a first limitation follows from the focus on evaluation-based change decisions, which does not cover all forms of instrumental use. Conditions that appear necessary for the examined outcome of evaluation-based decisions to substantially change the program or project might be less so for minor change decisions or decisions to continue the program or project, because the hurdle is lower, but there is no reason to believe that results would be completely different. As an additional conceptual limitation, this contribution examines only four conditions, leaving aside many others. It is well possible that in certain contexts, other conditions that have been neglected due to a focus on a set of specific hypotheses are more relevant to evaluation-based change decisions than the ones examined. Even so, the selected factors are theoretically grounded and certainly not trivial necessary conditions that can always be taken for granted; the cases provide evidence that the perceived novelty value and quality of evaluations as well as the decision setting vary considerably. In tracing the mechanisms behind the change decisions, I am quite confident that the examined conditions are in fact relevant necessary preconditions for use, but they are indeed unlikely to be sufficient. So other conditions must be fulfilled too. One aspect that has repeatedly been claimed to be (most) central is stakeholder involvement (cf. recent studies like Fleischer & Christie 2009; Johnson et al. 2009; Skolits et al. 2009). The key stakeholders in the present study are the desk managers. Their participation in the evaluation has been measured indirectly: it can be observed that the more they have been involved in the evaluations, the better their assessment of evaluation quality thanks to a better understanding of the evaluation process. Local stakeholder participation in the evaluation at the program or project site has not been included in the analysis, because it did not seem to affect decision making by the desk manager at the head office in Switzerland.

This study is based on theoretical claims that are context-bound, so that it can only lead to “contingent generalizations” (George & Bennett 2005) that will remain restricted in scope. Given the explorative nature of the empirical part of the present study, there are further important limitations, in particular with respect to external validity. First, SDC has a particular

evaluation tradition and a particular decision-making setting which differs from others. For internal validity, the fact that all cases have been chosen from the same organizational setting was important, given that organizational conditions were not included in the analysis. Furthermore, the SDC decision-making setting where responsibility for decision making relies on one person, the desk manager, has facilitated data collection. At the same time, this means that results cannot be simply generalized to other contexts.

Second, the study is based on 11 cases. The author was lucky that there were cases for each of the four context constellations of interest, even though cases with a positive outcome were missing for one of the constellations, so that the respective hypotheses could not really be tested. In the other three contexts, the number of positive cases lies between just one and three. The causal relationships that have been claimed are based on an analysis of the mechanisms underlying the processes that led to evaluation-based change decisions. So it is not just correlations based on a small number of cases. However, even in a qualitative case comparison, a higher number of cases is desirable because it helps to refine our understanding of the mechanisms that produce the outcome and of the conditions that are necessary to trigger these mechanisms. At the same time, it is indispensable to have an intimate knowledge of each of the cases to reveal the mechanisms in the specific context; mechanisms are hard to identify because they are usually hidden and context dependent (Astbury & Leeuw 2010). So there is always a trade-off between a higher number of cases and a better knowledge of each case. With 11 cases, the present article lies well beyond the mean of 3.4 cases that are usually analyzed in studies about evaluation use based on case study method (Brandon & Singh 2009).

As to internal validity, the study relies mainly on interview information from the desk managers that commissioned the analyzed external evaluations and were supposed to make use of them. There is a risk that desk managers overstated the quality of the evaluations they commissioned. I tried to deal with this in checking different dimensions of evaluation quality and in asking for the reasons of their quality ratings. The desk

managers were interviewed about half a year after the end of the evaluation. Certain measures were taken to reduce the risk that they could not remember the evaluation. Before the interview, desk managers were prompted to provide certain documents, in order to get them to deal with the evaluation once again. Furthermore, interview information, namely information about change decisions, was cross-checked with document information. Interviews with evaluators provided supplementary information. In most cases, the data were rich, but in two or three cases where there had been a change in the desk manager during the evaluation or just after its completion, there remained certain gaps in the process reconstruction. However, these gaps are likely to exist not just with respect to this analysis but also in reality in the evaluation process and in the process of evaluation use.

### **3.7 Conclusions**

Awareness for the complexity of the phenomenon of evaluation use has been growing in parallel with the body of research. This article has applied multiple strategies to handle this complexity: the form of use examined is clearly delineated; theoretical claims and empirical analysis are contextbound; a conjunctural approach is adopted in hypothesis formulation and in the QCA analysis; the study merely wants to find the necessary preconditions for change decisions in specific contexts and does not try to fully explain their occurrence or nonoccurrence. The explorative empirical analysis suggests that this combination of strategies is likely to pay off. The results suggest that even the understanding of well-known mechanisms can be refined: the “truth test,” for instance, could only be observed under one specific circumstance.

Depending on the context, the perceived novelty value and quality of an evaluation seem to matter more or less. This is likely to be true for most of the factors that are related to evaluation use and has implications for research on this subject. It is high time that we do not just control for contextual factors but make context explicit. We need adequate theoretical models that depict interdependencies between causal mechanisms (or



factors) and contexts. The realist methodology by Pawson and Tilley (1997) gives some guidance. The disadvantage is that even if we only examine a small number of factors (for example, four conditions as in the present study), theoretical models can get quite complicated. However, we might have to accept this higher degree of theoretical complexity if we strive for a better understanding of our complex reality.

Furthermore, we need adequate methods to test the complex theoretical models. QCA is one possibility. Recent developments of the method, namely “fuzzy-set QCA” (Ragin 2000, 2005, 2008; Schneider & Wagemann 2007), avoid dichotomization and accommodate more easily a higher number of cases, even though this goes to some extent at the cost of losing the direct link between the data and the cases, which is one of the strengths of the crisp-set application of QCA presented in this article. Another possibility would be a mixed-method design (e.g., Bergman 2008), for instance, combining a statistical analysis of survey data with in-depth case studies. It is, however, important to pay attention to the “compatibility” of theoretical claims and method; in contrast to QCA, statistical methods do not allow for a test of necessity claims. Necessity claims are, however, very valuable to advance research on phenomena like evaluation use that depend on multiple and often unpredictable elements, which can never be fully accounted for. A possible way forward for research on evaluation use is a focus on mechanisms that lead to well-circumscribed forms of use and on the necessary ingredients to trigger these mechanisms in different contexts. Such knowledge about context-specific necessary conditions for different kinds of evaluation use is also valuable for evaluation practitioners seeking advice about how to promote the utilization of their findings.

---

### 3.8 Bibliography

- Alkin, M. C. (1985). *A Guide for Evaluation Decision Makers*. Beverly Hills, CA: Sage.
- Alkin, M. C., Daillak, R., & White, P. (1979). *Using Evaluations. Does Evaluation Make a Difference?* Beverly Hills, CA: Sage.
- Astbury, B., & Leeuw, F. L. (2010). Unpacking Black Boxes: Mechanisms and Theory Building in Evaluation. *American Journal of Evaluation*, 31(3), 363–381.
- Amenta, E., & Poulsen, J. D. (1994). Where to Begin: A Survey of Five Approaches to Selecting Independent Variables for Qualitative Comparative Analysis. *Sociological Methods & Research*, 23(1), 22–53.
- Balthasar, A. (2006). The Effects of the Institutional Design on the Utilitation of Evaluation. *Evaluation*, 12(3), 353–371.
- Balthasar, A. (2007). *Institutionelle Verankerung und Verwendung von Evaluationen*. Chur/Zürich: Rüegger.
- Befani, B., Ledermann, S., & Sager, F. (2007). Realistic Evaluation and QCA: Conceptual Parallels and an Empirical Application. *Evaluation*, 13(2), 171–192.
- Berg-Schlosser, D., & De Meur, G. (2009). Comparative Research Design. Case and Variable Selection. In B. Rihoux & C. C. Ragin (Eds.), *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques* (pp. 19–32). Los Angeles, London, New Delhi, Singapore: Sage.
- Bergman, M. M. (Ed.) (2008). *Advances in Mixed Methods Research*. London, Thousand Oaks, CA, New Delhi, and Singapore: Sage.
- Brady, H. E., & Collier, D. (Eds.) (2004). *Rethinking Social Inquiry*. Lanham, MD: Rowman & Littlefield.
- Brandon, P. R., & Singh, J. M. (2009). The Strength of the Methodological Warrants for the Findings of Research on Program Evaluation Use. *American Journal of Evaluation*, 30(2), 123–157.

- Cousins, J. B., & Leithwood, K. A. (1986). Current Empirical Research on Evaluation Utilization. *Review of Educational Research, 56*(3), 331–364.
- Cronbach, L. J. (1982). *Designing Evaluations of Educational and Social Programs*. San Francisco, CA: Jossey-Bass.
- De Meur, G., & Rihoux, B. (2002). *L'analyse quali-quantitative comparée (AQQC-QCA): Approche, techniques et applications en sciences humaines*. Louvain-la-Neuve: Bruylant-Academia.
- Fleischer, D. N., & Christie, C. A. (2009). Evaluation Use: Results From a Survey of U.S. American Evaluation Association Members. *American Journal of Evaluation, 30*(2), 158–175.
- Frey, K., & Ledermann, S. (2010). Evidence-Based Policy: A Concept in Geographical and Substantive Expansion. *German Policy Studies, 6*(2), 1–15.
- Geertz, C. (1973). *The Interpretation of Cultures: Selected Essays*. New York, NY: Basic Books.
- George, A. L., & Bennett, A. (2005). *Case Studies and Theory Development in the Social Sciences*. Cambridge: MIT Press.
- Goertz, G. (2006a). Assessing the Trivialness, Relevance, and Relative Importance of Necessary or Sufficient Conditions in Social Science. *Studies in Comparative International Development, 41*(2), 88–109.
- Goertz, G. (2006b). *Social Science Concepts: A User's Guide*. Princeton, NJ and Oxford: Princeton University Press.
- Goertz, G., & Starr, H. (Eds.) (2003). *Necessary Conditions: Theory, Methodology, and Applications*. Lanham, MD: Rowman & Littlefield.
- Grofman, B., & Schneider, C. Q. (2009). An Introduction to Crisp Set QCA, with a Comparison to Binary Logistic Regression. *Political Research Quarterly, 62*(4), 662–672.
- Henry, G. T., & Mark, M. M. (2003). Beyond Use: Understanding Evaluation's Influence on Attitudes and Actions. *American Journal of Evaluation, 24*(3), 293–314.
- Jewell, C., & Bero, L. (2007). Public Participation and Claimsmaking: Evidence Utilization and Divergent Policy Frames in California's

- Ergonomics Rulemaking. *Journal of Public Administration Research and Theory*, 17(4), 625–650.
- Johnson, K., Greenseid, L. O., Toal, S. A., King, J. A., Lawrenz, F., & Volkov, B. (2009). Research on Evaluation Use: A Review of the Empirical Literature from 1986 to 2005. *American Journal of Evaluation*, 30(3), 377–410.
- Joint Committee. (1994). *The Program Evaluation Standards*. Thousand Oaks, CA: Sage.
- Kirkhart, K. E. (2000). Reconceptualizing Evaluation Use: An Integrated Theory of Influence. *New Directions for Evaluation*, 2000(88), 5–23.
- Lijphart, A. (1975). The Comparable-Cases Strategy in Comparative Research. *Comparative Political Studies*, 8(2), 158–177.
- Mahoney, J., & Goertz, G. (2006). A Tale of Two Cultures: Contrasting Quantitative and Qualitative research. *Political Analysis*, 14(3), 227–249.
- Mark, M. M., & Henry, G. T. (2004). The Mechanisms and Outcomes of Evaluation Influence. *Evaluation*, 10(1), 35–57.
- Mayring, P. (2003). *Qualitative Inhaltsanalyse: Grundlagen und Techniken* (8th ed.). Weinheim and Basel: Beltz.
- Merkens, H. (2003). Auswahlverfahren, Sampling, Fallkonstruktion. In U. Flick, E. von Kardorff, & I. Steinke (Eds.), *Qualitative Forschung: Ein Handbuch* (2nd ed., pp. 286–299). Reinbeck bei Hamburg: Rowohlt.
- Patton, M. Q. (1997). *Utilization-focused Evaluation: The New Century Text* (3rd ed.). Thousand Oaks, CA, London, and New Delhi: Sage.
- Pawson, R., & Tilley, N. (1997). *Realistic Evaluation*. London, Thousand Oaks, CA, and New Delhi: Sage.
- Przeworski, A., & Teune, H. (1970). *The Logic of Comparative Social Inquiry*. New York, NY: Wiley-Interscience.
- Ragin, C. C. (1987). *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies*. Berkeley, CA, Los Angeles, CA, and London: University of California Press.
- Ragin, C. C. (1994). *Constructing Social Research: The Unity and Diversity of Method*. Thousand Oaks, CA: Pine Forge Press.

- Ragin, C. C. (2000). *Fuzzy-Set Social Science*. Chicago, IL and London: The University of Chicago Press.
- Ragin, C. C. (2005). From Fuzzy Sets to Crisp Truth Tables. Compass Working Paper No. 28.
- Ragin, C. C. (2008). *Redesigning Social Inquiry: Fuzzy Sets and Beyond*. Chicago, IL and London: Chicago University Press.
- Rihoux, B. (2006). Qualitative Comparative Analysis (QCA) and Related Systematic Comparative Methods: Recent Advances and Remaining Challenges for Social Science Research. *International Sociology*, 21(5), 679–706.
- Rihoux, B., & Lobe, B. (2009). The Case for Qualitative Comparative Analysis (QCA): Adding Leverage for Thick Cross-Case Comparison. In D. Byrne, & C. C. Ragin (Eds.), *The SAGE Handbook of Case-based Methods* (pp. 222–242). London, Thousand Oaks, CA, New Delhi, and Singapore: Sage.
- Rihoux, B., & Ragin, C. C. (2009). *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques*. Los Angeles, CA, London, New Delhi, and Singapore: Sage.
- Sager, F., & Ledermann, S. (2008). Valorisierung von Politikberatung. In S. Bröchler, & R. Schützeichel (Eds.), *Politikberatung* (pp. 310–325). Stuttgart: UTB.
- Schneider, C. Q., & Wagemann, C. (2007). *Qualitative Comparative Analysis (QCA) und Fuzzy Sets: Ein Lehrbuch für Anwender und jene, die es werden wollen*. Opladen and Farmington Hills, MI: Barbara Budrich.
- SDC. (2002). *Ongoing Evaluation Programme for 2002–2003 of SDC*. SDC: Berne.
- SDC. (2004). *10 Principles to Ensure Successful Use of Evaluations*. SDC: Berne. Retrieved from [www.deza.admin.ch](http://www.deza.admin.ch).
- Skolits, G. J., Morrow, J. A., & Burr, E. M. (2009). Reconceptualizing Evaluator Roles. *American Journal of Evaluation*, 30(3), 275–295.
- Spinatsch, M. (2002). Evaluation in Switzerland: Moving Toward a Decentralized System. In J. E. Furubo, R. C. Rist, & R. Sandahl

- 
- (Eds.), *International Atlas of Evaluation* (pp. 375–391). New Brunswick, NJ and London: Transaction.
- Thomas, V. (2010). Evaluation Systems, Ethics, and Development Evaluation. *American Journal of Evaluation*, 31(4), 540–548.
- Valovirta, V. (2002). Evaluation Utilization as Argumentation. *Evaluation*, 8(1), 60–80.
- Weiss, C. H. (1972). Utilization of Evaluation: Toward Comparative Study. In C. H. Weiss (Ed.), *Evaluating Action Programs: Readings in Social Action and Education*. Boston, MA: Allyn & Bacon.
- Weiss, C. H. (1998). Have We Learned Anything New About the Use of Evaluation? *American Journal of Evaluation*, 19(1), 21–33.
- Weiss, C. H., & Bucuvalas, M. J. (1980). Truth Tests and Utility Tests: Decision-Makers' Frames of Reference for Social Science Research. *American Sociological Review*, 45(2), 302–313.
- Weiss, C. H., Murphy-Graham, E., & Birkeland, S. (2005). An Alternate Route to Policy Influence: How Evaluations Affect D.A.R.E. *American Journal of Evaluation*, 26(1), 12–30.
- Yamasaki, S., & Rihoux, B. (2009). A Commented Review of Applications. In B. Rihoux & C. C. Ragin (Eds.), *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques* (pp. 123–145). Los Angeles, London, New Delhi, Singapore: Sage.



#### **4 PROBLEM ERKANNT, ABER NICHT GEBANNT – DER NUTZEN EINER VERKNÜPFUNG VON KONZEPT- UND UMSETZUNGSEVALUATION AM BEISPIEL DER STRATEGIE „MIGRATION UND GESUNDHEIT“<sup>8</sup>**

SIMONE LEDERMANN, PROF. DR. FRITZ SAGER

*Der vorliegende Beitrag befasst sich mit dem Zusammenhang zwischen Konzeptevaluation und der Bewertung der Umsetzung von Programmen. Am Beispiel der Evaluation der schweizerischen Strategie „Migration und Gesundheit“ wird aufgezeigt, welchen Nutzen die Umsetzungsevaluation aus der Bewertung des Programmkonzepts zieht. Die Unterscheidung von Konzept und Umsetzung erlaubt nicht nur die Zuordnung von Wirkungsdefiziten auf „Policy Failure“ und „Implementation Failure“, sondern ist auch notwendige Bedingung, um aus der gekoppelten Betrachtungsweise der Erkenntnisse zielführende und realistische Lehren für die weitere Programmentwicklung zu ziehen. Der Artikel unterstreicht damit den Bedarf an Globalevaluationen.*

*Causarum enim cognitio cognitionem eventorum facit.*

[Die Kenntnis der Ursachen bewirkt die Erkenntnis der Ergebnisse.]

Cicero, Topica 67

Der vorliegende Beitrag befasst sich mit dem Zusammenhang zwischen Konzeptevaluation und der Bewertung der Umsetzung von Programmen. Die Konzeptevaluation befasst sich mit dem Design eines Programms ungeachtet dessen Umsetzung. Das Programmkonzept wird in der Literatur unterschiedlich benannt: logisches Modell, Programmtheorie, Programmmodell oder Kausalkette sind einige der Begriffe, die verwendet werden, um zu bezeichnen, welche Wirkungsüberlegungen einem Pro-

---

<sup>8</sup> Aus *Zeitschrift für Evaluation* (2009) 8(1): 7-25. Der Beitrag stützt sich auf die Evaluation der Strategie M+G (Ledermann et al. 2006).



programm zugrunde liegen. Im Rahmen von Bestrebungen zu stärker evidenzbasierter Politik erlebt die Konzeptanalyse eine neue Blüte (vgl. Elkeles 2006; Hanney et al. 2003). Es geht darum, die Eignung von Programmen und Massnahmen im Voraus, also ex-ante zu evaluieren. Aber auch bei Ex-post-Evaluationen nimmt die Konzeptanalyse einen wichtigen Stellenwert ein. Gemäss Rossi, Lipsey und Freeman (2004) stellt sie die erste Stufe der Evaluationshierarchie dar. Erst auf dieser Basis macht es gemäss diesen Autoren Sinn, die Umsetzung eines Programms zu bewerten.

Für ein erfolgreiches Programm braucht es beides, ein gutes Konzept und eine gute Umsetzung. Dieser Umstand ist schon vor langem erkannt worden und hat mit der Unterscheidung von *Policy Failure* (oder *Program Failure*) und *Implementation Failure* in die Literatur Eingang gefunden (vgl. Linder & Peters 1987; Robertson 1984). Während bei der *Policy Failure* der Nicht-Erfolg eines politischen Programms auf das Konzept zurückzuführen ist, liegt das Problem bei der *Implementation Failure* nicht an der Programmanlage, sondern an seiner Realisierung. Nur die Kombination von korrektem Konzept und adäquater Umsetzung führt zum Programmerfolg.

Dieser Beitrag legt den Nutzen einer gekoppelten Evaluation des Konzepts und der Umsetzung dar. Empirische Basis der Analyse ist die schweizerische Strategie „Migration und Gesundheit“ (kurz: Strategie M+G). Bereits 1991 lancierte das Bundesamt für Gesundheit (BAG) ein Pilotprojekt zum Thema Migration und Gesundheit, das dann mehrmals verlängert wurde und schliesslich in das Grundsatzpapier „Migration und Gesundheit: Strategische Ausrichtung des Bundes 2002–2006“ (BAG 2002) mündete. Die Strategie M+G wurde 2002 von der Regierung verabschiedet und mit deren Beschluss bis 2007 verlängert. Sie lässt sich als politisches Programm verstehen, das heisst als Bündel von Massnahmen, die auf ein gemeinsames Ziel ausgerichtet sind.

In der nachfolgenden Analyse der Strategie M+G wird zunächst der analytische Zugang zum Programmkonzept als Evaluationsgegenstand dargelegt und der Zusammenhang zur Umsetzung hergestellt. Zudem wird das empirische Vorgehen erläutert. Danach wird das Programmkonzept

der Strategie M+G deskriptiv dargestellt, bevor es im anschließenden Abschnitt bewertet wird. Es wird gezeigt, dass das Programmkonzept insgesamt empirisch fundiert und in sich stimmig ist, wobei der Konkretisierungsgrad der Vorgaben gering bleibt.

Nach der Betrachtung des Strategiekonzepts wendet sich der Beitrag der Umsetzung zu. Die Bilanz fiel gemischt aus, was sich durch die Art der Bearbeitung der zahlreichen Berührungspunkte zu anderen Politikbereichen erklären lässt. Die Evaluation kommt zu dem Ergebnis, dass sich bei den vertikalen Schnittstellen eine *Policy Failure* und eine *Implementation Failure* kumuliert haben, während die Umsetzung bei den horizontalen Schnittstellen mit einer Verbesserung der Interventionsstrategie einherging. In den Schlussfolgerungen wird der Nutzen aus der Verbindung von Konzept- und Umsetzungsevaluation verdeutlicht. Daraus lassen sich konzeptionelle Schlüsse zu den Anforderungen an eine Strategie im Bereich Migration und Gesundheit sowie in weiteren Querschnittsbereichen föderal strukturierter Politik ziehen.

#### **4.1 Programmkonzept und Umsetzung als Evaluationsgegenstände**

In der Programmevaluation werden verschiedene Evaluationsgegenstände betrachtet. Grob kann zwischen dem Programmkonzept, der Programmumsetzung und den Programmwirkungen unterschieden werden. Im Programmkonzept wird festgelegt, was das Problem ist, was mit dem Programm diesbezüglich verbessert werden soll, mit welchen Massnahmen die Verbesserung erreicht werden soll und wie die Massnahmen umgesetzt werden sollen.<sup>9</sup> Für die vorliegende Untersuchung wird das Konzept in folgende vier Bestandteile zerlegt:<sup>10</sup>

---

<sup>9</sup> Im Detail variiert die Definition der exakten Gegenstände. In Bezug auf das Programmkonzept sprechen Rossi, Lipsey und Freeman (2004) einerseits von Bedarfsanalyse (*Assessment of the Need for a Program*) und andererseits von Programmtheorie (*Program Theory*). Knoepfel (vgl. Knoepfel/Bussmann 1997; Knoepfel et al. 2001) unterscheidet zunächst zwischen Politikkonzept und Policy-Design, später zwischen öffentlicher Problemdefinition (*définition politique du problème public*) und politisch-administrativem Programm (*programme politico-administratif*). Hier

- *Problemdefinition:* Annahmen zur Art des gesellschaftlichen Problems und zu seinen Ursachen (Kausalhypothesen). Was das Programm als problematisch definiert, entspricht nur einem Ausschnitt der gesellschaftlichen Wirklichkeit.
- *Zielvorgaben:* Vorgaben dazu, wie der Zustand im Problemfeld zu verändern ist. Sie legen fest, was das Programm zur Lösung des Problems beitragen soll.
- *Operative Vorgaben:* Vorgaben zu den Aktivitäten, die zur Erreichung der Zielvorgaben und somit zur Behebung des Problems bzw. seiner Ursachen unternommen werden sollen.
- *Organisatorische Vorgaben:* Strukturelle Vorgaben, die festlegen, wer die Aktivitäten durchführen soll (Zuständigkeiten) und welche Ressourcen dafür bereitzustellen sind, sowie prozedurale Vorgaben zu den Verfahren, die bei der Umsetzung zu beachten sind.

Es handelt sich bei den vier Elementen um Vorgaben, also um Plan- oder Soll-Angaben. Zwischen den vier Bestandteilen wird eine hierarchische Beziehung angenommen. Die *Planung* eines Programms folgt idealtypisch der obigen Reihenfolge. Probleme bestehen nicht per se, sondern sie müssen zuerst als solche definiert werden. Jeder Policy-Vorschlag, jedes Programm enthält eine explizite oder implizite Diagnose des zu lösenden ‚Problems‘, eine Problemdefinition. In den Annahmen zur Art des Problems widerspiegeln sich auch bestimmte Werte und Prioritäten (vgl. Bacchi 1999). Daraufhin wird überlegt, inwiefern das Problem gelöst werden soll, d.h. welche Ziele gesetzt werden sollen. Die Ziele bestimmen dann, welche Massnahmen (operative Vorgaben) getroffen werden müssen, um sie zu erreichen. Je nach geplanten Massnahmen ist schliesslich eine andere Organisationsform vorzusehen.

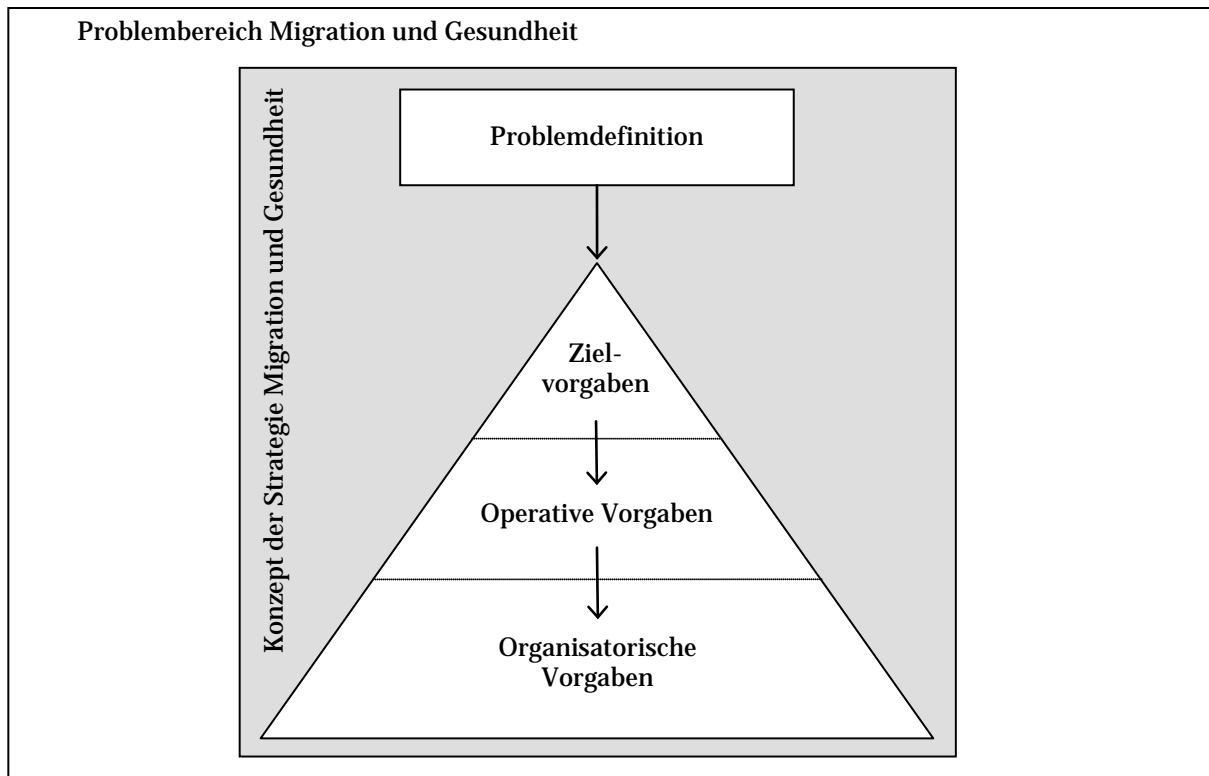
---

werden diese Gegenstände aufgrund ihres gemeinsamen konzeptionellen Charakters zusammengefasst.

<sup>10</sup> Die Definition der folgenden Bestandteile lehnt sich an das Kern-Schale-Modell von Knoepfel und Weidner (1982; vgl. auch Knoepfel et al. 2001: 175ff.; 1997: 84ff.) an. Deren Modell wird allerdings vereinfacht, indem auf eine Unterscheidung zwischen Zielsetzungen und evaluativen Elementen (Indikatoren zur Messung der Zielerreichung) verzichtet und die strukturellen und prozeduralen Elemente zu den organisatorischen Vorgaben zusammengefasst werden. Zudem wird das Modell um die Problemdefinition erweitert, womit der Zusammenhang zwischen den Annahmen zum Problem (Kausalhypothesen) und den Annahmen, wie dieses gelöst werden kann (Interventionshypothesen, vgl. Knoepfel et al. 1997), hervorgehoben wird.

Diese idealtypische hierarchische Beziehung zwischen den konstitutiven Elementen eines Programmkonzepts wird in Abbildung 4-1 am Beispiel der Strategie verdeutlicht.

**Abbildung 4-1:** Konzeptbestandteile der Strategie M+G



*Quelle:* eigene Darstellung.

Bei der *Umsetzung* des Programms sieht das idealtypische Vorgehen genau umgekehrt aus: Zuerst muss eine Organisation auf die Beine gestellt werden, damit die Massnahmen überhaupt umgesetzt werden können. Die Wirkungen der Massnahmen sollten wenn möglich den Zielvorgaben entsprechen und die ursprüngliche Problemlage verändern. Kurz: Bei der Programmierung klettert man im Modell in Abbildung 4-1 von oben nach unten, bei der Umsetzung von unten nach oben.

In Wirklichkeit wird dieses idealtypische Vorgehen weitgehend weder bei der Programmierung noch bei der Umsetzung in dieser Weise eingehalten. Dennoch ist das Modell hilfreich, da es zum Ausdruck bringt, dass die verschiedenen Elemente und Umsetzungsschritte eines Programms aufeinander abgestimmt sein müssen.

## **4.2 Empirisches Vorgehen und Bewertungskriterien**

Die verschiedenen Bestandteile eines Programms lassen sich nicht auf den ersten Blick erkennen, sondern müssen durch eine Analyse erst rekonstruiert werden. Im nachfolgenden Abschnitt werden die Ergebnisse des Rekonstruktionsprozesses dargestellt. Daraufhin wird untersucht, wie ‚gut‘ dieses Konzept ist. Als Kriterien werden einerseits die wissenschaftliche Abstützung der zugrunde liegenden Problemdefinition sowie andererseits die innere und äussere Kohärenz der Vorgaben zur Interventionsstrategie herangezogen.

### **4.2.1 Empirische Evidenz**

Die Problemrekonstruktion ist das Resultat einer qualitativen Inhaltsanalyse nach Mayring (2003), bei welcher in Anlehnung an Knoepfel, Varone und Larrue (2001: 151) drei Problemdimensionen als Kategorien verwendet werden: Betrachtet wird, worin das Problem laut der Strategie M+G besteht (Intensität), welche Gruppen davon betroffen sind (Verbreitung), und welches die angenommenen Ursachen sind (Kausalhypothesen). Die Problemdefinition wird auf ihre empirische Evidenz hin beurteilt. Dazu wurde ein Screening der im Strategiepapier angegebenen Quellen sowie ergänzender wissenschaftlicher Literatur durchgeführt, es wurden die vorhandenen Daten im Gesundheitsbereich gesichtet und Expertengespräche mit Personen aus der Wissenschaft geführt.

### **4.2.2 Innere Kohärenz**

Das Kriterium der inneren Kohärenz bezieht sich auf die Vorgaben zur Interventionsstrategie, also darauf, was der Bund gegen die identifizierten Probleme unternehmen will. Die gewählte Interventionsstrategie umfasst die Zielvorgaben, die operativen Vorgaben (Massnahmen) sowie die organisatorischen Vorgaben. Diese Vorgaben sollten einerseits zum definierten Problem passen und andererseits in sich selbst stimmig sein. Ein diffuses, lückenhaftes oder widersprüchliches Programmkonzept kann die Umsetzung und Wirkung erheblich beeinträchtigen und zu ei-

nem unnötigen Ressourcenverbrauch führen. Deshalb wird bei den verschiedenen Bestandteilen des Programmkonzepts der Strategie M+G deren Klarheit und Vollständigkeit untersucht und deren logische Kohärenz analysiert.

### **4.2.3 Äussere Kohärenz**

Weil die Massnahmen eines Programms die Erreichung von Zielen anderer Programme fördern oder beeinträchtigen können, wird in der Folge die Abstimmung zwischen der Strategie M+G und ihrem Kontext auf horizontaler Ebene untersucht. Dazu wurden erstens jene Tätigkeitsbereiche des Bundes identifiziert, in welche die Strategie M+G hineinwirkt. Zweitens wurden die Bereiche der Gesundheits- und der Migrationspolitik des Bundes erfasst, die sich auf die Strategie M+G auswirken könnten. Die relevanten Politikprogramme wurden einzeln auf ihre Kohärenz überprüft.

## **4.3 Strategiekonzept**

In diesem Abschnitt werden die einzelnen Konzeptbestandteile, wie sie im Strategiepapier (BAG 2002) dargestellt sind, beschrieben.

### **4.3.1 Problemdefinition**

Das Programmkonzept der Strategie geht von zwei Hauptproblemen aus: fehlende gesundheitliche Chancengleichheit und Kostensteigerungen im Gesundheitswesen. Die Migrationsbevölkerung in der Schweiz ist gemäss Problemdefinition gesundheitlich benachteiligt. Sie weist „gesundheitliche Defizite“ auf, und die Wahrscheinlichkeit von Gesundheitsproblemen ist gemäss Strategiepapier grösser, weil die Migrationsbevölkerung „erhöhten“ oder „kumulierten Gesundheitsrisiken“ ausgesetzt sei. Zur Migrationsbevölkerung zählt die Strategie M+G alle Personen – unabhängig von ihrem Aufenthaltsstatus und ihrer Nationalität –, die entweder selbst aus einem anderen Land in die Schweiz gekommen sind oder direkte Nachkommen solcher Personen sind. „Gesundheit“ wird im

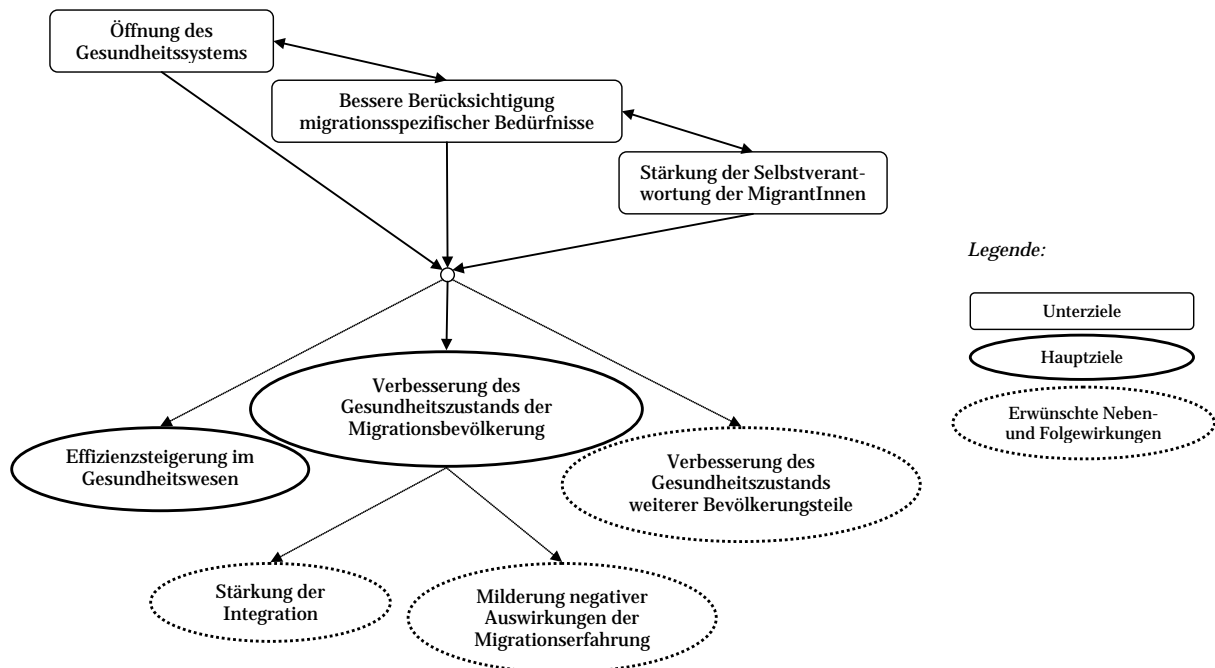
Programmkonzept nicht als Abwesenheit von Krankheit, sondern in einem ganzheitlichen Sinne als Lebensqualität verstanden. Die Gesundheitsprobleme betreffen nur einen „Teil der Migrationsbevölkerung“. Frauen seien in vielerlei Hinsicht stärker betroffen als Männer, und es bestünden Unterschiede zwischen Asylsuchenden und Arbeitsmigrantinnen und -migranten. In Bezug auf die Kostensteigerungen im Gesundheitswesen geht das Strategiepapier implizit davon aus, dass ein Teil der höheren Kosten migrationsbedingt sei.

Die Ursachen der Probleme werden im Programmkonzept umfassend erörtert. Ausgehend von der Migrationspolitik einerseits und der Migrationserfahrung andererseits führt die Kausalkette über verschiedene Zwischenschritte zu den Hauptproblemen, den gesundheitlichen Defiziten bei der Migrationsbevölkerung und der migrationsbedingten Zunahme der Gesundheitskosten. Wichtige Ursachen sind dabei ungünstige Rahmenbedingungen für die Integration, die sozioökonomische Benachteiligung sowie Verständigungs- und Zugangsprobleme im Gesundheitssystem (vgl. auch Abbildung 4-3).

### **4.3.2 Zielvorgaben**

Die langfristige *Vision*, auf der das Programm aufbaut, ist die Verwirklichung der gesundheitlichen Chancengleichheit für die Migrationsbevölkerung. Dieser Vision will die Strategie durch das *Hauptziel* der Verbesserung des Gesundheitszustands der Migrationsbevölkerung näher kommen. Als zweites Hauptziel soll ausserdem eine Effizienzsteigerung im Gesundheitswesen erreicht werden (vgl. Abbildung 4-2).

Ferner werden drei Prinzipien genannt, die als *Unterziele* aufgefasst werden können: (1) Öffnung des Gesundheitssystems (Prinzip der Chancengleichheit); (2) Verbesserung der Berücksichtigung migrationspezifischer Bedürfnisse bei der Versorgung (Prinzip der adäquaten Leistungserbringung); (3) Stärkung der Selbstverantwortung und Problemlösungskompetenz von Migrantinnen und Migranten (Prinzip der Selbstverantwortung und des Empowerments).

**Abbildung 4-2: Zielsystem der Strategie**

Quelle: eigene Darstellung.

Schliesslich postuliert die Strategie M+G zusätzliche *Neben- und Folgewirkungen* wie die Verbesserung des Gesundheitszustands von weiteren Bevölkerungsteilen oder die Verbesserung der Integration der Migrationsbevölkerung.

### 4.3.3 Operative Vorgaben

In Bezug auf die Aktivitäten, die im Rahmen der Strategie M+G unternommen werden sollen, legt das Strategiepapier fünf Interventionsbereiche in klarer Prioritätenordnung mit je zwei bis drei Massnahmen fest (vgl. Tabelle 4-1). Insgesamt enthält das Konzept 13 Massnahmen.



**Tabelle 4-1: Operative Vorgaben**

<i>Interventionsbereich/ Massnahmen</i>	<i>Leistungsziele der Massnahmen</i>
<b>1. Interventionsbereich Bildung (Aus-, Fort- und Weiterbildung)</b>	
Aus- und Fortbildung interkulturelle Übersetzung	In der Schweiz existiert eine professionelle Aus- und Fortbildung für interkulturelle Übersetzung im Gesundheitsbereich.
Aus- und Fortbildung interkulturelle Vermittlung	In der Schweiz existieren eine professionelle Aus- und Fortbildung sowie eine anerkannte Zertifizierung für interkulturelle Vermittlung.
Fort- und Weiterbildung Gesundheitsberufe	Es existieren für alle Gesundheitsberufe regelmässige Fort- und/oder Weiterbildungsangebote im Bereich M+G.
<b>2. Interventionsbereich Information, Prävention und Gesundheitsförderung</b>	
Prävention und Gesundheitsförderung	Bei den verschiedenen Gruppen in der Migrationsbevölkerung finden systematisch gezielte präventive Interventionen statt.
Sensibilisierung Leistungserbringer	Bei den wichtigsten leistungserbringenden Diensten in der Schweiz wurden zum Thema M+G Sensibilisierungsbemühungen unternommen.
<b>3. Interventionsbereich Gesundheitsversorgung</b>	
Leicht zugängliche Angebote	In allen grossen Agglomerationen existieren leicht zugängliche Angebote, die unbürokratisch Hilfe bei Gesundheitsproblemen anbieten und gezielt Dienste vermitteln.
Vermittlungs- und Koordinationsdienste	In allen Regionen der Schweiz existieren Vermittlungsdienste für den Gesundheits- und wenn möglich auch für den Sozialbereich.
Einsatz von interkultureller Übersetzung	In allen Regionen der Schweiz werden professionelle interkulturelle Übersetzungsdienste angeboten, deren Finanzierung geklärt ist.
<b>4. Interventionsbereich Therapieangebote für Traumatisierte im Asylbereich</b>	
Differenzierung bestehendes Angebot	Verbesserung des bestehenden Angebots, insbesondere für Traumatisierte mit längerfristiger Aufenthaltsperspektive.
Aufbau niederschwelliger Angebote	Aufbau leicht zugänglicher, dezentralisierter Angebote, insbesondere für Traumatisierte mit ungeklärter Aufenthaltsperspektive.
<b>5. Interventionsbereich Forschung (Grundlagen, Evaluation und Monitoring)</b>	
Gesundheitsmonitoring	In der Schweiz existiert ein Gesundheitsmonitoring in Bezug auf die Migrationsbevölkerung.
Problemorientierte Grundlagenforschung	Es werden Forschungsprojekte im Bereich M+G durchgeführt.
Evaluation	Die Strategie M+G und ihre wichtigsten Massnahmen werden evaluiert.

*Quelle:* eigene Darstellung, in Anlehnung an BAG 2002: 48 (Tabelle 5).

#### **4.3.4 Organisatorische Vorgaben**

In organisatorischer Hinsicht teilt das Strategiepapier die Federführung dem BAG zu. Es ist für die Gesamtkoordination und vier Interventionsbereiche zuständig, während für den Interventionsbereich der Therapieangebote für Traumatisierte im Asylbereich das Bundesamt für Migration (BFM) als zuständig erklärt wird. Das BAG soll gemäss Konzept mit dem BFM sowie mit der Eidgenössischen Ausländerkommission (EKA), die für die Integrationsförderung des Bundes zuständig ist, zusammenarbeiten. Sowohl BFM als auch EKA waren bereits an der Erarbeitung des Strategiepapiers beteiligt gewesen.

Der Einbezug weiterer Akteure sei aufgrund des interinstitutionellen Charakters des Bereichs Migration und Gesundheit zudem unabdingbar. Um die Ziele der Strategie zu erreichen, sei es nötig, dass verschiedene Bundesstellen, die Kantone, die Gemeinden sowie auch private Organisationen zusammenarbeiten würden. „Der Bund versteht sich für alle Aktivitäten als Promotor und Koordinator und setzt selber keine Massnahmen um“ (BAG 2002: 39). Mit der Umsetzung seien vielmehr externe feld- und problemnahe Partner zu beauftragen. Im Strategiepapier war vorgesehen, alle beteiligten Akteure in einer sogenannten „Projektgruppe“, welche für die Steuerung des gesamten Programms zuständig gewesen wäre, zusammenzufassen. Bei einer späteren Konkretisierung des Programmkonzepts in einem sogenannten Projektantrag wurden die Aufgaben dagegen auf verschiedene Gremien verteilt.

#### **4.4 Bewertung des Konzepts**

Die Problemdefinition des Programms wurde bezüglich ihrer empirischen Evidenz beurteilt, während bei den übrigen Vorgaben die innere und äussere Kohärenz bewertet wurde.

##### **4.4.1 Empirische Evidenz**

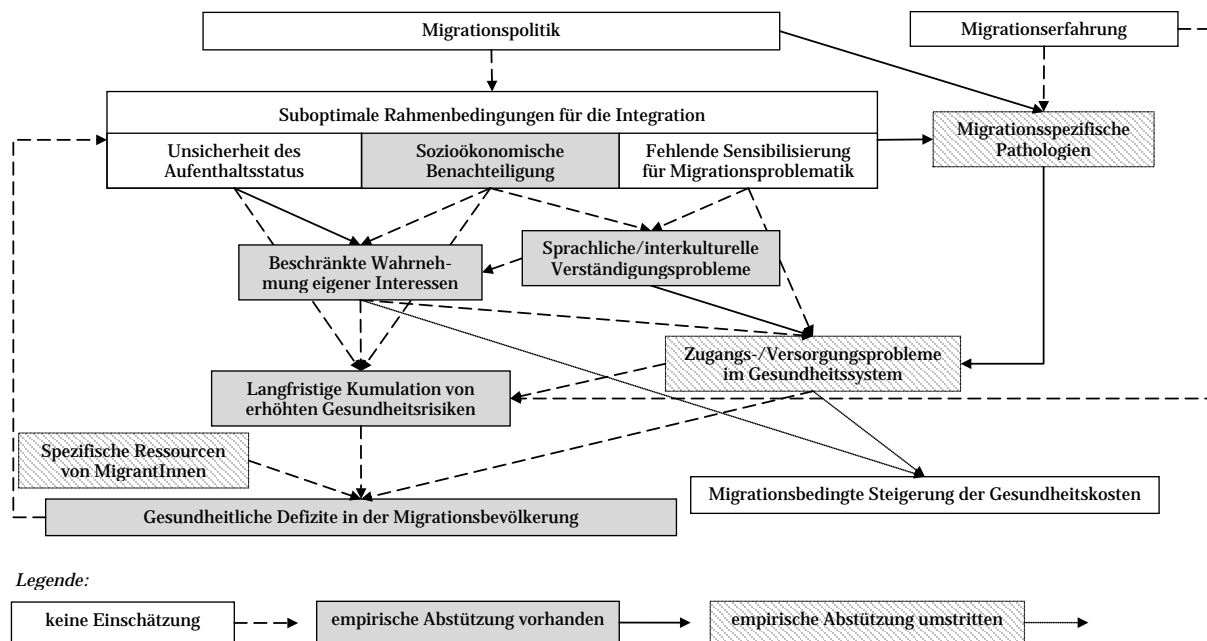
Das Strategiepapier verweist nur punktuell auf Studien oder Daten, welche die Aussagen zum Problem stützen würden. Die Evaluation stellte

beim Vergleich der Problemannahmen mit weiteren Quellen jedoch fest, dass die Annahmen zum Problem im Grossen und Ganzen gestützt werden.

Die im Programmkonzept postulierten gesundheitlichen Defizite der Migrationsbevölkerung lassen sich mit epidemiologischen Daten belegen: Die Auswertungen der Schweizerischen Gesundheitsbefragungen zeigen im Vergleich zur schweizerischen Bevölkerung sowohl ein schlechteres subjektives Wohlbefinden als auch ein häufigeres Auftreten verschiedener körperlicher Beschwerden (vgl. BFS 2000: 125; BFS 2004: 120; Vranjes et al. 1996). Die Daten aus der Gesundheitsbefragung weisen auch auf erhebliche Unterschiede innerhalb der Migrationsbevölkerung hin, lassen jedoch kaum Rückschlüsse auf einzelne Gruppen zu (vgl. Vranjes et al. 1996). Immerhin zeigen sie, dass Migrantinnen sowohl nach subjektiven als auch nach objektiven Indikatoren einen vergleichsweise schlechteren Gesundheitszustand aufweisen, womit die im Programmkonzept angenommene stärkere Betroffenheit von Frauen belegt ist.

Expertengespräche mit Personen aus der Wissenschaft ergaben, dass die Annahmen zu den Ursachen der gesundheitlichen Defizite insgesamt ebenfalls plausibel sind. Abbildung 4-3 stellt die Kausalannahmen im Strategiekonzept in verdichteter Form dar und zeigt gleichzeitig, inwiefern die einzelnen Annahmen empirisch belegt werden können. Dabei konnte im Rahmen des Literaturscreenings nur ein Teil der Kausalketten überprüft werden.

Einige Problemannahmen werden durch die vorhandene Literatur klar belegt. Dazu gehört, dass sich die sozioökonomische Lage sehr stark auf die Gesundheit auswirkt. Die gesundheitlichen Defizite bei Migrantinnen und Migranten lassen sich aber nicht allein auf die tendenziell tiefere soziale Schicht zurückführen (vgl. Vranjes et al. 1996). Vielmehr spielen daneben auch migrationsspezifische Gründe eine Rolle.

**Abbildung 4-3: Problemanalyse und empirische Abstützung**

*Quelle:* eigene Darstellung.

Zu gewissen Annahmen liegen auch widersprüchliche wissenschaftliche Erkenntnisse vor. In Bezug auf migrationspezifische Pathologien gibt es beispielsweise empirische Hinweise, dass seit den 1980er Jahren eine Pathologisierung der Migrationserfahrung mithilfe der Trauma-Diagnose stattgefunden hat (vgl. Salis Gross 2004), was jedoch im Strategiekonzept nicht angesprochen wird. Die Darstellung des Problems fällt in diesem Punkt deshalb zu wenig differenziert aus. Beschränkt ist die empirische Evidenz in Bezug auf die Zugangs- und Versorgungsprobleme. Zwar existieren hierzu international vergleichende Arbeiten (vgl. z.B. Bollini 1992; Bollini/Siem 1995), eine tatsächliche Unterversorgung konnte für die Schweiz jedoch nur bei Vorsorgeleistungen bei Migrantinnen belegt werden.

Hinsichtlich der Kostensteigerungen im Gesundheitssystem fehlen empirische Angaben dazu, welcher Anteil der Kosten auf Migrantinnen und Migranten entfällt. Die Annahme einer migrationsbedingten Kostenzunahme wird in der Literatur allerdings eher bezweifelt.

Insgesamt lässt sich feststellen, dass die Problemdefinition in Bezug auf die gesundheitlichen Defizite bei der Migrationsbevölkerung durch wissenschaftliche Erkenntnisse weitgehend gestützt wird. Nicht empirisch belegt ist hingegen die migrationsbedingte Kostensteigerung.

#### 4.4.2 Innere Kohärenz

Die einzelnen Konzeptbestandteile wurden nach ihrer Klarheit, Vollständigkeit und ihrer Kohärenz zu den übergeordneten Bestandteilen bewertet. Insgesamt ergab sich dabei ein leicht positives Bild (vgl. Tabelle 4-2).

**Tabelle 4-2:** Innere Kohärenz

<i>Konzeptbestandteil</i>	<i>Klarheit</i>	<i>Vollständigkeit</i>	<i>Hierarchische Kohärenz</i>
Problemdefinition	-	0	XX
Zielvorgaben	-	0	+
Operative Vorgaben	0	0	+
Organisatorische Vorgaben	0	+	+

*Legende:* + = hoch; 0 = mittel; - = tief; XX = nicht vorgesehen.

*Quelle:* eigene Darstellung

Im Strategiepapier sind viele Annahmen zu den zugrunde liegenden Problemen nur implizit vorhanden. Die Problemanalyse ist umfassend, doch fehlt eine Konkretisierung, was insbesondere auch mit den fehlenden empirischen Grundlagen zusammenhängen dürfte.

Die Zielvorgaben sind im Programmkonzept unklar, weil quer durch das Strategiepapier verstreut. Ausserdem sind sie sehr allgemein formuliert und es fehlt eine Konkretisierung. Die vorhandenen Ziele sind jedoch gut auf die Annahmen zur Art und zu den Ursachen der Probleme abgestimmt. Die Hauptziele sind auf die beiden Hauptprobleme der gesundheitlichen Defizite in der Migrationsbevölkerung sowie der Kostensteigerungen ausgerichtet.

Auf operativer Ebene sind die Interventionsbereiche klar definiert und enthalten je zwei bis drei thematisch passende Massnahmen. Der Bezug

---

der Massnahmen zur Problemdefinition ist klar gegeben. Sie sind in erster Linie auf Verbesserungen beim Gesundheitszustand der Migrationsbevölkerung und erst in zweiter Linie auf Effizienzsteigerungen im Gesundheitswesen ausgerichtet. Zwischen den Interventionsbereichen und Massnahmen lassen sich zudem Synergien erkennen, was für eine hohe Kohärenz spricht. Das Strategiepapier enthält jedoch eine Fülle von Vorgaben zu den einzelnen Massnahmen, deren Bedeutung unklar bleibt, während wichtige Angaben wie jene zu den Instrumenten und Adressaten fehlen. Die operativen Vorgaben sind deshalb etwas undeutlich und unvollständig.

Die organisatorischen Vorgaben sind im Strategiepapier vage, wurden dann aber im Projektantrag verdeutlicht. Die Angaben dort sind relativ detailliert und vollständig.

#### **4.4.3 Äussere Kohärenz**

Die Strategie M+G ist ein Querschnittsprogramm und weist, wie bereits aus ihrem Namen hervorgeht, Verknüpfungen zur Gesundheits- und zur Migrationspolitik auf. Daneben berühren die fünf Interventionsbereiche weitere Politikfelder, so etwa die Berufsbildungspolitik, die verschiedenen Präventions- und Gesundheitsförderungsinitiativen, die Krankenversicherungspolitik und die Forschungspolitik. In Bezug auf die Ziele der meisten Politiken, zu welchen die Strategie konzeptionelle Berührungspunkte aufweist, besteht eine positive Beziehung. Insbesondere die verschiedenen Gesundheitsförderungsbemühungen zielen in dieselbe Richtung. In einem ambivalenten Verhältnis steht das Strategiekonzept vor allem zur Asyl- und Zulassungspolitik. Dies ergibt sich aus dem Umstand, dass die Asyl- und Zulassungspolitik die Integration von Teilen der Migrationsbevölkerung erschweren kann, indem eine problematische sozioökonomische Struktur miterzeugt und gesundheitsbeeinträchtigende Faktoren wie die Unsicherheit des Aufenthalts verstärkt werden. Im Übrigen ist festzustellen, dass rund die Hälfte der Massnahmen kantonale Kompetenzen bei der Gesundheitsversorgung tangiert.

#### 4.4.4 Gesamtbewertung des Programmkonzepts

Insgesamt zeigt die Evaluation des Programmkonzepts, dass die Strategie M+G auf einem komplexen Ursachengefüge basiert, das weitgehend empirisch abgestützt ist. Die Probleme wurden somit erkannt. Gerade bei der Strategie M+G, bei der eine Vielzahl von Zielsetzungen und fünf Interventionsbereiche mit je mehreren Massnahmen vorhanden sind, die durch eine Vielzahl von Akteuren umgesetzt werden sollen, ist die Frage nach der Kohärenz wichtig. Es lässt sich denn auch feststellen, dass die verschiedenen Bestandteile insgesamt in sich stimmig und kohärent zu den jeweils übergeordneten Vorgaben sind, jedoch blieb die notwendige Konkretisierung aus. Durch ihre Positionierung in zwei Politikfeldern sowie die Tatsache, dass die fünf Interventionsbereiche über den Gesundheitsbereich hinausreichen, weist die Strategie eine Vielzahl von Schnittstellen mit anderen Politikfeldern auf. Die Bearbeitung dieser Schnittstellen erwies sich bei der Umsetzung als kritischer Erfolgsfaktor.

#### 4.5 Umsetzung der Strategie

Die Konzeptevaluation diene zunächst dazu, die Soll-Vorgaben für die Umsetzung der Strategie M+G systematisch zu erheben. Verglichen an diesen Vorgaben fällt die Leistungsbilanz der Strategie auf den ersten Blick eher bescheiden aus: Von den im Strategiepapier vorgesehenen 13 Massnahmen konnten lediglich vier vollständig oder weitgehend umgesetzt werden. Bei vier Massnahmen konnten die operativen Vorgaben teilweise, bei zwei weiteren Massnahmen nur sehr bedingt erreicht werden. Die übrigen drei Massnahmen wurden nicht umgesetzt. Damit wurden die operativen Vorgaben in allen Interventionsbereichen nur teilweise erfüllt (vgl. Tabelle 4-3).

Ein wichtiger Grund für diese durchgezogene Bilanz liegt darin, dass die Vorgaben selbst zu wenig konkret waren, wie bei der Konzeptevaluation festgestellt wurde. So allgemein, wie die Vorgaben formuliert waren, konnten sie nicht erfüllt werden. Bei den festgestellten Mängeln bei der Umsetzung handelt es sich somit zuallererst um eine *Policy Failure*. Es muss allerdings anerkannt werden, dass es zum Zeitpunkt der Strategie-

entwicklung nur sehr begrenzt möglich war, für die Konkretisierung der Vorgaben auf wissenschaftliche Erkenntnisse zurückzugreifen.

**Tabelle 4-3:** Erfüllung der Leistungs- und Wirkungsziele

<i>Interventionsbereich/ Massnahmen</i>	<i>Erfüllung der Leistungsziele</i>	<i>Erfüllung der Wirkungsziele</i>
<b>1. Interventionsbereich Bildung (Aus-, Fort- und Weiterbildung)</b>		
Aus- und Fortbildung interkulturelle Übersetzung	+	(+)
Aus- und Fortbildung interkulturelle Vermittlung	–	–
Fort- und Weiterbildung Gesundheitsberufe	(–)	(+)
<b>2. Interventionsbereich Information, Prävention und Gesundheitsförderung</b>		
Prävention und Gesundheitsförderung	(+)	(+)
Sensibilisierung Leistungserbringer	(+)	–
<b>3. Interventionsbereich Gesundheitsversorgung</b>		
Leicht zugängliche Angebote	–	(+)
Vermittlungs- und Koordinationsdienste	–	(+)
Einsatz von interkultureller Übersetzung	(+)	(+)
<b>4. Interventionsbereich Therapieangebote für Traumatisierte im Asylbereich</b>		
Differenzierung bestehendes Angebot	+	(+)
Aufbau niederschwelliger Angebote	(–)	(–)
<b>5. Interventionsbereich Forschung (Grundlagen, Evaluation und Monitoring)</b>		
Gesundheitsmonitoring	(+)	(–)
Problemorientierte Grundlagenforschung	+	–
Evaluation	+	(+)

*Legende:* + = ja; (+) = eher ja; (–) = eher nein; – = nein.

*Quelle:* eigene Darstellung.

Die mittelmässige Bilanz muss ausserdem mit Blick auf die Wirkungsziele relativiert werden, wobei vorauszuschicken ist, dass die Einschätzungen der Wirkungen mit einer gewissen Unsicherheit behaftet bleiben, da es im Rahmen der Evaluation nur sehr beschränkt möglich war, Wirkungen bei den Adressaten und Begünstigten selbst zu erheben. Auf der Basis der vorhandenen Daten kommt die Evaluation zu dem Schluss, dass die Strategie M+G bei acht der 13 Massnahmen einen Beitrag zu den Wirkungszielen geleistet hat. Diese Wirkungsbeiträge wurden allerdings zuweilen über andere als die ursprünglich geplanten Massnahmen



erreicht. Dies erklärt, weshalb die Wirkungseinschätzung teilweise positiv ausgefallen ist, obwohl die Massnahmen selbst nicht oder nur teilweise umgesetzt wurden. Auch zu den übergeordneten Zielen (vgl. Abbildung 4-2) konnte die Strategie M+G punktuell einen Beitrag leisten.

Die Gesamtbilanz der Umsetzung der Strategie M+G lässt sich zu einem wesentlichen Teil dadurch erklären, wie die vorhandenen Schnittstellen bearbeitet wurden. So konnten die Beiträge zu den Wirkungszielen insbesondere dank der horizontalen Zusammenarbeit mit Akteuren auf Bundesebene erreicht werden, während Schwierigkeiten in der Kooperation mit kantonalen Akteuren bei der vertikalen Koordination die Umsetzung der Strategie M+G behinderten. Die Bearbeitung der Schnittstellen soll deshalb näher erläutert werden.

#### **4.5.1 Bearbeitung der horizontalen Schnittstellen**

Gemäss der vorhandenen empirischen Evidenz ist die soziale Benachteiligung der Hauptgrund für die gesundheitlichen Defizite der Migrationsbevölkerung. Um eine Verbesserung zu erreichen, müssten somit die Rahmenbedingungen für die Integration und damit die Asyl- und Zulassungspolitik verändert werden, zu welcher die Strategie M+G Widersprüche aufweist (vgl. oben). Im Rahmen der Umsetzung der Strategie M+G fand zwar eine Koordination mit dem BFM als zuständigem Amt statt, sie beschränkte sich aber auf den Interventionsbereich Therapieangebote für Traumatisierte im Asylbereich und war insgesamt eher lose. Eine Veränderung der politischen Voraussetzungen stand dagegen nicht zur Diskussion, denn die Asyl- und Zulassungspolitik werden stark durch gesetzliche Bestimmungen geprägt, die durch einen parlamentarischen Entscheid oder gar eine Volksabstimmung legitimiert sind. Die Strategie M+G kann als Verwaltungsprogramm nicht auf solche Politikfelder Einfluss nehmen, weshalb hier weder eine *Policy Failure* noch eine *Implementation Failure* vorliegt.

In anderen Bereichen konnten bei der Umsetzung des Programms Synergien realisiert werden. Enger als mit dem BFM war die Koordination mit der Integrationsförderung des Bundes, die über das Sekretariat der

---

EKA abgewickelt wird. Während bis dahin sowohl die EKA als auch das BAG Massnahmen zur Ausbildung von interkulturellen Übersetzenden unterstützt hatten, beschlossen die beiden Stellen, dass künftig das BAG für die Ausbildung sorgen und neben dem Gesundheits- auch den Sozial- und Bildungsbereich einbeziehen würde. Im Gegenzug erklärte sich die EKA bereit, Vermittlungsstellen für interkulturelle Übersetzung zu fördern – eine Massnahme, die im Strategiepapier M+G ebenfalls vorgesehen war.

Die EKA setzte sich bei ihren eigenen Massnahmen klar für einen integrativen Ansatz ein: Anstatt spezielle Angebote für die Migrationsbevölkerung bereitzustellen, sollen öffentliche Dienste ihre Leistungserbringung so anpassen, dass sie für die gesamte Bevölkerung einschliesslich der Migrantinnen und Migranten geeignet ist. Dieser integrative Ansatz gewann im Zuge der Zusammenarbeit mit der EKA auch bei der Umsetzung der Strategie M+G immer grössere Bedeutung. Gewisse Massnahmen wie die Schaffung von speziellen, leicht zugänglichen Angeboten für Migrantinnen und Migranten wurden bewusst gar nicht umgesetzt. Zuweilen wurde auch die Ausgestaltung bestehender Massnahmen im Sinne des integrativen Ansatzes angepasst.

Mit der stärkeren Ausrichtung auf einen integrativen Ansatz trat auch das „Migration Mainstreaming“ stärker in den Vordergrund. In Anlehnung an das „Gender Mainstreaming“ (BAG 2004) wird darunter eine Strategie zur Beseitigung bestehender Diskriminierungen von Migrantinnen und Migranten verstanden, indem ihre Bedürfnisse bei der Ausgestaltung von Massnahmen berücksichtigt werden sollen. Voraussetzung für das „Migration Mainstreaming“ ist die Sensibilisierung der im Bereich tätigen Institutionen. Damit wurde im Rahmen der Umsetzung der Strategie M+G die Vernetzung immer wichtiger. Vor allem BAG-intern wurde eine Vernetzung angestrebt. Am erfolgreichsten war die Zusammenarbeit mit der Sektion *Aids*. Daneben fand auch ein Migration Mainstreaming in Bezug auf die Aktivitäten des Fachbereichs *Gender Health* sowie der Sektionen *Drogen, Alkohol, Tabak* und *Grundlagen* statt. Andere Sektionen und Bereiche wurden zwar kontaktiert, doch konnte noch keine konkrete Koordination erreicht werden. Dies lag einerseits an den be-

schränkten personellen Ressourcen, welche für die Umsetzung der Strategie M+G zur Verfügung standen, zum Teil aber auch an der fehlenden Unterstützung durch die Amtshierarchie.

Daneben wurde über die sogenannte Interinstitutionelle Begleitgruppe der Strategie M+G die Vernetzung mit national tätigen Institutionen aus dem Gesundheits- und Migrationsbereich gesucht (Bundesstellen, Stiftung Gesundheitsförderung Schweiz, Berufs- und Interessenverbände aus dem Gesundheitsbereich, interkantonale Konferenzen). Bis zu einem gewissen Grade hat die Evaluation eine Sensibilisierung der in diesem Gremium vertretenen Organisationen für die Migrationsproblematik festgestellt, doch kam nur punktuell eine konkrete Zusammenarbeit zustande. Mehrheitlich blieb die Vernetzung unverbindlich. Es zeigte sich, dass die unterschiedlichen Organisationskulturen und -ziele ein Hindernis darstellen.

Zusammenfassend lässt sich festhalten, dass bei der Strategie M+G sinnvollerweise auf die Bearbeitung von Schnittstellen mit gesetzlich legitimierten Politikbereichen verzichtet wurde. Die Strategie M+G wirkte jedoch darauf hin, dass andere Verwaltungsprogramme und Projekte des BAG im Sinne des integrativen Ansatzes an die Bedürfnisse der Migrationsbevölkerung angepasst wurden, womit zusätzliche Wirkungen erzielt werden konnten. Durch den integrativen Ansatz fand im Rahmen der Umsetzung gewissermassen ein *Policy Upgrade* statt. Es verbleibt aber in Bezug auf Aktivitäten weiterer nationaler Gesundheits- und Migrationsorganisationen ein ungenutztes Synergiepotenzial.

#### **4.5.2 Bearbeitung der vertikalen Schnittstellen**

Die durchgezogene Umsetzungsbilanz der Strategie M+G lässt sich nicht allein damit erklären, dass auf Massnahmen, die dem integrativen Ansatz nicht entsprachen, verzichtet wurde, sondern auch damit, dass die Strategie M+G an Kompetenzgrenzen im föderalen Staatsaufbau stiess. Während im Migrationsbereich die Kompetenzen zwar weitgehend beim Bund liegen, die Politik aber auf Gesetzesebene gemacht wird und damit ausserhalb des Einflussbereichs der Strategie M+G liegt, sind im Ge-

---

sundheitsbereich die Kantone die primären Akteure (vgl. Sager & Rüefli 2005). Der Bund ist lediglich für die gesetzliche Regelung der Krankenversicherung zuständig. Die Massnahmen in den Interventionsbereichen Gesundheitsversorgung und Angebote für Traumatisierte im Asylbereich sowie die Massnahmen zur Sensibilisierung der Leistungserbringer betrafen dagegen eindeutig kantonale Zuständigkeiten. Wie bei der Beschreibung des Programmkonzepts erwähnt, war der Einbezug der Kantone in die Umsetzung geplant, jedoch wird ihnen neben den weiteren Akteuren keine Sonderstellung zugewiesen. Die gültige föderale Kompetenzordnung wurde also vernachlässigt, was als grundsätzliche *Policy Failure* bezeichnet werden muss.

Die Beteiligung der Kantone fand über die bereits erwähnte Interinstitutionelle Begleitgruppe statt. Durch sie sollten die kantonalen Stellen sensibilisiert werden, um sie für eine Mitwirkung zu gewinnen und damit eine Dezentralisierung der Strategie M+G zu ermöglichen. Faktisch waren die Kantone aber nur über eine interkantonale Konferenz, die Schweizerische Konferenz der kantonalen Gesundheitsdirektorinnen und -direktoren, in der Interinstitutionellen Begleitgruppe vertreten. Zudem sass ein kantonaler Integrationsdelegierter als Vertreter der Schweizerischen Konferenz der kommunalen, regionalen und kantonalen Integrationsdelegierten in diesem Gremium.

Die Einbindung über die Gesundheitsdirektorenkonferenz erwies sich als schwierig, da die Konferenz selbst heterogene Interessen vertreten musste und über begrenzte Kapazitäten zur Bearbeitung von Themen verfügte. Aufgrund der tiefen Priorität des Themas Migration und Gesundheit gelang es nicht, Anliegen der Strategie über die Direktion des BAG an einem Treffen der kantonalen Gesundheitsdirektorinnen und -direktoren zur Sprache zu bringen. Eine regionale und kantonale Verankerung hat deshalb über die Gesundheitsdirektorenkonferenz höchstens am Rande stattgefunden. Dies muss als *Implementation Failure* verstanden werden.

Die Konferenz der Integrationsdelegierten unterstützte die Strategie M+G dagegen engagiert und trug so zu einer gewissen Sensibilisierung in den Kantonen und Städten bei, allerdings in sehr unterschiedlichem Aus-

mass. Durch ihr Engagement konnten gewisse Kantone für eine Mitfinanzierung von Ausbildungsgängen für interkulturelle Übersetzende gewonnen werden. Auch spielten die Integrationsdelegierten beim Aufbau von Vermittlungsstellen für interkulturelle Übersetzende durch die EKA eine wichtige Rolle.

Von Seiten des BAG wurde aufgrund von Mängeln bei der Programmkommunikation nur wenig Lobby- und Sensibilisierungsarbeit bei den Kantonen betrieben. Diese weitere *Implementation Failure* mag zu einem gewissen Teil erklären, weshalb die Anliegen einzelner Projekte bei den kantonalen Entscheidungsträgern auf wenig Gehör stiessen. So konnte beispielsweise eine vom BAG unterstützte Fachstelle zu Gesundheit und Integration die Auflage einer finanziellen Beteiligung der Kantone trotz grosser Bemühungen nur zum Teil erfüllen.

Trotz einzelner Versuche, die Kantone an der Strategie M+G zu beteiligen, ist dies insgesamt betrachtet somit nur sehr beschränkt gelungen. Die Interinstitutionelle Begleitgruppe erwies sich dazu als ungeeignet, weil ihr Teilnehmerkreis zu heterogen war. Die mangelnde Bearbeitung von vertikalen Schnittstellen lässt sich aber auch durch Ressourcenbeschränkungen beim Programmmanagement des BAG, durch Probleme bei der Programmkommunikation sowie durch das Fehlen eines sichtbaren Engagements der Spitzen der beteiligten Bundesämter für die Thematik Migration und Gesundheit erklären. Als Folge konnten einige Massnahmen nicht oder nur teilweise umgesetzt werden. Dass die Strategie M+G nicht, wie ursprünglich vorgesehen, in die Kantone hinausgetragen wurde, ist also darauf zurückzuführen, dass die *Policy Failure* durch verschiedene *Implementation Failures* verstärkt worden ist.

## **4.6 Schlussfolgerungen**

Dieser Beitrag hatte zum Ziel, den Nutzen aus der Bewertung der beiden Evaluationsgegenstände Programmkonzept und Umsetzung sowie insbesondere den Erkenntnisgewinn aus ihrer Zusammenführung für die Evaluation von Programmen aufzuzeigen. Zu diesem Zweck wurden zunächst die konstitutiven Elemente der Strategie Migration und Gesund-

---

heit 2002 bis 2007 definiert. Dabei wurde die Problemdefinition, welche der Strategie M+G zugrunde liegt, herausgearbeitet und die wissenschaftliche Fundierung diskutiert. Anschliessend wurde betrachtet, welche Zielvorgaben, welche operativen und welche organisatorischen Vorgaben gemacht wurden und inwiefern diese in sich stimmig und auf die zugrunde liegende Problemdefinition abgestimmt sind. Der Beitrag kommt zu dem Schluss, dass das Programmkonzept zwar in sich schlüssig erscheint, aber in vielerlei Hinsicht viel zu allgemein bleibt.

Umgesetzt wurden die Vorgaben des Programmkonzepts nur zum Teil. Dies liegt einerseits im Sinne einer *Policy Failure* daran, dass die Vorgaben selbst zu wenig spezifisch waren, als dass sie hätten erfüllt werden können. Zudem wurden bei der Durchführung sowohl Grenzen des Programms als auch gewisse Möglichkeiten der Erweiterung deutlich. Die gekoppelte Analyse von Konzept und Umsetzung führte im Falle der horizontalen Schnittstellen die Kompensation der *Policy Failure* oder gar ein *Policy Upgrade* durch innovative inhaltliche Anpassungen in der Durchführungsphase vor Augen, während im Falle der vertikalen Schnittstellen die *Policy Failure* durch *Implementation Failures* noch verstärkt wurde.

So zeigte sich im Falle der horizontalen Berührungspunkte auf Bundesebene, dass zwar Zielkonflikte mit der Asyl- und Ausländerpolitik bestehen, diese jedoch nicht aus dem Weg geräumt werden können, weil sie auf gesetzlicher Ebene legitimiert sind, während es sich bei der Strategie um ein Verwaltungsprogramm handelt. Durch die gewählte Interventionsstrategie werden diese Beschränkungen allerdings berücksichtigt.

Neue Möglichkeiten eröffnete der Übergang zu einem stärker integrativen Ansatz, der durch die Zusammenarbeit mit der EKA begünstigt wurde. Er zielt darauf ab, bestehende Angebote im Gesundheitsbereich „migrationskompatibel“ zu machen anstatt spezielle Strukturen für diesen Bevölkerungsteil zu schaffen. Auf Massnahmen, die diesem integrativen Ansatz nicht entsprachen, wurde verzichtet, dafür nutzte das BAG Schnittstellen zu bestehenden Programmen, um den integrativen Ansatz zu fördern. Vor allem BAG-intern konnten im Präventionsbereich Synergiepotenziale genutzt werden. Diese BAG-internen Anstrengungen stärk-

ten auch die Glaubwürdigkeit gegenüber anderen Organisationen und Institutionen aus dem Migrations- und Gesundheitsbereich. Die vorhandenen Organisationsstrukturen boten die Gelegenheit, Kontakte zu weiteren national tätigen Akteuren zu knüpfen, wobei erst wenige konkrete Ergebnisse erzielt worden sind. Insgesamt ist es aber gelungen, einen Beitrag zu den Wirkungszielen der Strategie M+G zu leisten – dies in manchen Fällen jedoch über andere als die ursprünglich vorgesehenen Massnahmen.

Nicht aufgewogen worden, sind im zweiten Fall dagegen jene Defizite, die sich auf vertikale Schnittstellen mit kantonalen Zuständigkeiten im Gesundheitsbereich beziehen. Das vorhandene Gremium erwies sich für den Einbezug der Kantone als ungeeignet. Ausserdem behinderten gewisse *Implementation Failures* wie Probleme bei der Programmkommunikation und ein fehlendes sichtbares Engagement der höheren hierarchischen Ebenen eine stärkere Sensibilisierung bei den Kantonen. Seine Promotorenfunktion hat der Bund im föderalen Aufgabengefüge nur sehr beschränkt wahrgenommen. Gewisse Massnahmen konnten daher auch nicht umgesetzt werden.

Die vorliegende Studie unterstreicht die Notwendigkeit von umfassenden Globalevaluationen. Die Unterscheidung von Konzept und Umsetzung erlaubt nicht nur die Zuordnung von Wirkungsdefiziten auf *Policy Failure* und *Implementation Failure*, sondern ist auch notwendige Bedingung für die Möglichkeit, aus der gekoppelten Betrachtungsweise der Erkenntnisse realistische und zielführende Lehren für die weitere Programmentwicklung zu ziehen.

Die Erkenntnisse aus der Untersuchung der Strategie M+G lassen den Schluss zu, dass die während der Umsetzung erfolgte konzeptionelle Ausrichtung auf einen integrativen Ansatz richtig war. Dem integrativen Ansatz zufolge ist auf die Anpassung bestehender Angebote hinzuwirken. Liegen diese aber in kantonaler Kompetenz, mag der Bund im besten Fall über ein zeitlich beschränktes Programm kurzfristige Veränderungen bewirken. Auf die Dauer ist jedoch die kantonale Verankerung zentral. Ein Bundesprogramm, das in kantonale Zuständigkeitsbereiche hineingreift, muss deshalb auf allen Ebenen einen Einbezug kantonaler

Akteure vorsehen (vgl. auch Sager & Rüefli 2005). Die Art des Einbezugs muss dabei auf die Akteure abgestimmt sein; politisch Verantwortliche sind über andere Wege anzusprechen als Vertreter einer kantonalen Fachbehörde.

Die Zusammenarbeit mit den Kantonen erfordert von Seiten des Bundes zunächst einen beträchtlichen Ressourcenaufwand, ist bei Massnahmen in kantonalen Zuständigkeitsbereichen aus Gründen der Nachhaltigkeit jedoch unerlässlich. Um die Kantone für ein Mitwirken zu gewinnen, ist ein sichtbares Bundesengagement Voraussetzung. Fehlt dieses Engagement, ist ein Verwaltungsprogramm, das Kompetenzen der Kantone tangiert, zumindest längerfristig zum Scheitern verurteilt.

#### **4.7 Bibliographie**

- Bacchi, C. L. (1999). *Women, Policy and Politics: The Construction of Policy Problems*. London: Sage.
- BAG (2002). *Migration und Gesundheit: Strategische Ausrichtung des Bundes 2002–2006*. Bern: Bundesamt für Gesundheit (BAG).
- BAG (2004). *Faktenblatt: Gender Mainstreaming in Public Health*. Bern: Bundesamt für Gesundheit (BAG).
- BFS (2000). *Migration und ausländische Bevölkerung in der Schweiz*. Neuchâtel: Bundesamt für Statistik (BFS).
- BFS (2004). *Ausländerinnen und Ausländer in der Schweiz: Bericht 2004*. Neuchâtel: Bundesamt für Statistik (BFS).
- Bollini, P. (1992). Health Policies for Immigrant Populations in the 1990s: A Comparative Study in Seven Receiving Countries. *Migration XXX* (Special Issue 'Migration and Health in the 1990s'). Genf: International Organization for Migration.
- Bollini, P., & Siem, H. (1995). No Real Progress Towards Equity: Health of Migrants and Ethnic Minorities on the Eve of the Year 2000. *Social Science & Medicine*, 41(6), 819–828.



- Elkeles, T. (2006). Evaluation von Gesundheitsförderung und die Forderung nach Evidenzbasierung: Fünf Thesen zur Anwendbarkeit auf Gesundheit. *Zeitschrift für Evaluation*, 2006(1), 39–70.
- Hanney, S. R., Gonzalez-Block, M. A., Buxton, M. J., & Kogan, M. (2003). The Utilisation of Health Research in Policy-Making: Concepts, Examples and Methods of Assessment. *Health Research Policy and Systems*, 1(2), 1–28.
- Knoepfel, P., & Weidner, H. (1982). Formulation and Implementation of Air Quality Control Programmes: Patterns of Interest Consideration. *Policy and Politics*, 10(1), 85–110.
- Knoepfel, P., & Bussmann, W. (1997). Die öffentliche Politik als Evaluationsobjekt. In W. Bussmann, U. Klöti, & P. Knoepfel (Hg.), *Einführung in die Politikevaluation* (pp. 58–77). Basel, Frankfurt a.M.: Helbing & Lichtenhahn, 58–77.
- Knoepfel, P., Larrue, C., & Varone, F. (2001). *Analyse et pilotage des politiques publiques*. Genf, Basel, München: Helbing & Lichtenhahn.
- Knoepfel, P., Varone, F., Bussmann, W., & Mader, L. (1997). Evaluationsgegenstände und Evaluationskriterien. In W. Bussmann, U. Klöti, & P. Knoepfel (Hg.), *Einführung in die Politikevaluation* (pp. 78–118). Basel, Frankfurt a.M.: Helbing & Lichtenhahn.
- Ledermann, S., Hammer, S., Sager, F., Dubas, D., Rüefli, C., Schmidt, N., Trageser, J., Vettori, A., & Zeyen Bernasconi, P. (2006). *Evaluation der Strategie „Migration und Gesundheit 2002–2006“: Schlussbericht und Beilagenbände 1–3*. Bern: Bundesamt für Gesundheit.
- Linder, S. H., & Peters, B. G. (1987). A Design Perspective on Policy Implementation: The Fallacies of Misplaced Prescription. *Policy Studies Review*, 6(3), 459–475.
- Mayring, P. (2003). *Qualitative Inhaltsanalyse: Grundlagen und Techniken*. Weinheim, Basel: Beltz.
- Robertson, D. B. (1984). Program Implementation versus Program Design: Which Accounts for Policy ‘Failure’? *Policy Studies Review*, 3(3–4), 391–405.

- Rossi, P. H., Lipsey M. W., & Freeman, H. E. (2004). *Evaluation: A Systematic Approach*. Thousand Oaks, London, New Delhi: Sage.
- Sager, F., & Rüefli, C. (2005). Die Evaluation öffentlicher Politiken mit föderalistischen Vollzugsarrangements: Eine konzeptionelle Erweiterung des Stufenmodells und eine praktische Anwendung. *Swiss Political Science Review*, 11(2), 101–129.
- Salis Gross, C. (2004). Struggling with Imaginaries of Trauma and Trust: The Refugee Experience in Switzerland. *Culture, Medicine and Psychiatry*, 28(2), 151–167.
- Vranjes, N., Bisig, B., & Gutzwiller, F. (1996). *Gesundheit der Ausländer in der Schweiz*. Bern: Bundesamt für Gesundheit (BAG), Facheinheit Sucht und Aids.